DEVELOPMENT OF A SUPERVISED MACHINE LEARNING MODEL TO ENHANCE URBAN WATER SYSTEM MANAGEMENT: A CASE STUDY OF STELLENBOSCH MUNICIPALITY

by
REJOICE VAN DER WALT

Dissertation presented for the degree of
Doctor of Philosophy in Military Science in the Faculty of
Military Science at Stellenbosch University

Supervisor: Prof. K. I. Theletsane
Co-supervisor: Prof. A. Taigbenu

December 2023

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained herein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe upon any third-party rights, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2023

ABSTRACT

Globally, the challenges of conserving freshwater resources are becoming increasingly complex. Among the reasons cited by several researchers are the continuing growth of the world’s population, urbanisation, and the adverse effects of climate change on rainfall amounts and cycles. The complexity stems from the fact that human and natural systems are inextricably linked and interdependent. This makes managing urban water systems a major challenge that requires an integrated management approach capable of addressing the increasing variables that are interdependent and interrelated in an urban water system. To this end, tools continue to be developed to assist water resources managers to improve their management strategies, while data-driven methods are currently gaining popularity.

Researchers have consistently emphasised the importance of accurately predicting the water demands of an urban water system as a prerequisite for effective freshwater management. However, the increasingly interconnected and interdependent variables that result from the interactions between human and natural systems pose a significant challenge to accurately predicting water demand. Consequently, traditional modelling tools are also increasingly becoming inadequate. The impacts of climate change, which lead to uncertainties in precipitation cycles, and rapid urbanisation are the main causes of the inadequacy of traditional modelling tools, as they cannot accurately quantify the uncertainties that arise in the system. As a result, data-driven machine learning techniques are becoming more common and are currently widely used in the Global North. In contrast, their use in the Global South is currently very limited, which is also true in South Africa.

Another challenge posed by climate change is the changes in evapotranspiration and precipitation that limit terrestrial water storage and necessitate the search for alternative water sources. Among several options for alternative water sources, the case study area (Stellenbosch Municipality) has considered the reuse of municipal wastewater. However, to date, Stellenbosch Municipality has not developed this resource to any significant extent. It is therefore imperative to investigate the barriers to the development of this resource in the Stellenbosch Municipality. The main goal of this study was to use technology to develop a strategy for the sustainable management of Stellenbosch Municipality’s urban water system.

The transdisciplinary research approach was the overarching research methodology used in this study because it provided the researcher with the flexibility to choose methods from different research traditions. Other research methods used in the transdisciplinary approach included a critical systematic literature review, interactive management, simulation, a standard cross-industry data-mining research process, and a case study. The mixed-methods exploratory sequential research design, characterised by two phases, was applied to the Stellenbosch Municipality as the case study, where the unit of analysis was urban water demand. The first phase consisted of collecting qualitative data through a soft management systems interactive research method from a purposively selected focus group of municipal wastewater specialists and community representatives. The collected qualitative data were modelled using Concept Star decision-making tools. The second phase consisted of quantitative data collection and simulation guided by standard cross-industry processes for data-mining research. Both traditional time series models and supervised machine learning models were developed for forecasting and predicting run-of-river abstraction for the Stellenbosch Municipality.

Qualitative studies conducted on the factors that hinder the implementation of municipal wastewater reuse as an alternative water source in the Stellenbosch Municipality found that social issues were the main cause, followed by deficiencies in water laws, policies, and guidelines for the implementation of municipal wastewater reuse projects. The four principles of human-centred design were identified as an appropriate methodology for desirable implementation of wastewater reuse projects in the Stellenbosch Municipality. Quantitative studies that predicted urban water demand in the Stellenbosch Municipality showed nonlinearity between total water consumption and population/household growth, which should be the norm. From the exploratory data analysis (EDA), the variable run-of-river abstraction was set as the dependent variable for the modelling processes. The following models were developed: traditional Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average models and supervised machine learning models; thus AdaBoost, Gradient Boosting, Stochastic Gradient Boosting, Random Forest, and Artificial Neural Networks. The model with the best performance was Random Forest, followed by Stochastic Gradient Boosting, both of which the researcher saved and recommended for production.

The study's application of the transdisciplinary research methodology is a unique contribution to urban water management research. In addition, this study helps to highlight the importance of a human-centred design approach and the use of data-driven supervised machine learning techniques in the management of urban water systems, which the researcher considers a human-centred data-driven technological triad for the management of urban water systems. It is an effective framework for deploying novel approaches to water management in an urban setting that can be applied to other communities.

OPSOMMING

Die bewaring van varswaterbronne en die herontdekking van nuwe bronne is wêreldwyd ’n wetenskaplike probleem van groot belang. Daarbenewens het die bestuur van veral stedelike afloopwater in ingewikkeldheid toegeneem. Navorsers wat die probleem bestudeer het, het tot die gevolgtrekking gekom dat daar ’n oorsaaklike verband bestaan tussen die voortdurende toename in die wêreldbevolking, verstedeliking, die negatiewe uitwerking van klimaatsverandering en die sikliese verband met reënval, en die afname en besoedeling van stedelike waterbronne. Die ingewikkeldheid van die probleem word verder vermeerder deur die wisselwerking tussen mens en natuur. Dit is duidelik dat die wisselwerking tussen bogemelde faktore die bestuur van stedelike waterbronne ’n uitdagende taak maak. Dit is voor die hand liggend dat ’n oorhoofse en geïntegreerde bestuursbenadering noodsaaklik is wat sowel die onderlinge interafhanklikheid, asook die oorhoofse wisselwerking, kan aanspreek. Nuwe bestuursmetodes word voortdurend ondersoek om die probleem van stedelike water en afloopwater maksimaal na te vors.

Faktore wat die ingewikkeldheid van die probleem verder beïnvloed is die feit dat die vraag na die beskikbaarheid van en die omvang van bestaande waterbronne akkuraat voorspel moet word. Akkurate voorspelling het sy eie probleme deurdat historiese data nie geredelik beskikbaar is nie. Die historiese akkuraatheid van waterdata is ook nie betroubaar nie. ’n Verdere probleem is dat die impak van klimaatsverandering verdamping en reënval uiters nadelig beïnvloed.

Die tekortkominge van huidige metodes en tegnieke het aanleiding gegee dat data-gedrewe tegnieke en simulasie toenemend gebruik word. Die “machine learning model” is die metode wat huidiglik toenemend gebruik word.

Die hoof doelwit van hierdie studie was om ’n masjiengedrewe simulasie te ontwerp, die implementering daarvan te toets, gebruikers te leer hoe dit werk, en sodanig waterbestuursprobleme te kan hanteer en oplossings te bied om wetgewing, regulasies, en beleidsvoorskrifte in werking te stel.

Die transdissiplinêre metodologiese benadering is as die oorhoofse navorsingsmetodologie gebruik omdat dit die navorser die ruimte gebied het om verskillende dissiplines se wetenskapsbenaderings bymekaar te bring. Die metodiek is aangevul deur ’n kritiese literatuuroorsig, interaktiewe bestuursimulasie, en ’n

gevallestudie. Die gemengde-metodes ondersoekende opvolgende navorsingsontwerp is in twee fases aangewend. Dit is eerstens toegepas op Stellenbosch Plaaslike Munisipaliteit as die gevallestudie waar stedelike wateraanvraag as die eenheid van ontleding gebruik is. Die eerste fase het bestaan uit die versameling van data deur ’n “soft management systems” interaktiewe proses met voorafgekeurde spesialiste en gemeenskapsverteenwoordigers. Die kwalitatiewe data was versamel en verwerk deur gebruik te maak van Concept Star se besluitnemingsinstrumente. Die tweede fase het bestaan uit tradisionele kwalitatiewe dataversameling. Beide die tydreeks- en die “machine learning” prosesse was ontwerp vir die voorspelling van afloopwater van Stellenbosch Munisipale Werke.

Kwalitatiewe studies van die faktore wat inhiberend inwerk op die hersirkulering van stedelike afloopwater het getoon dat sosiale faktore negatief inwerk op die hergebruik van afloopwater. Ander faktore wat aangedui was, was gebrekkige wetgewing en ’n gebrek aan beleidsvoorskrifte en standaarde. Die inhiberende faktore word veroorsaak deur menslike persepsies rakende die implementering van werkbare alternatiewe waterbronne. Verdere studies het aangetoon dat ’n nie-lineêre tendens waarneembaar is tussen totale water verbruik en die toename in bevolkingsgetalle en huishoudings. Hierdie bevinding is teenstrydig met die algemene verwagting en normatiewe gebruik. Vir die voorlopige data-ontleding van die onttrekking van rivierwater was “run-of-river abstraction” gestel as die afhanklike veranderlike in die moduleringsproses. Die volgende modelle was ontwikkel: die tradisionele Autoregressive Integrated Moving Average en die Seasonal Autoregressive Integrated Moving Average modelle, asook die “machine learning model”; dus AdaBoost, Gradient Boosting en Stochastic Boosting, Random Forest, en Artificial Neural Networks. Die model met die beste resultate was Random Forest, gevolg deur Stochastic Gradient Boosting. Die navorser beveel beide hierdie modelle aan.

Die toepassing van die transdissiplinêre navorsingsmetodologie is ’n unieke kombinasie en toevoeging tot waterbestuursnavorsing. Voorts help die studie om die belangrikheid van ’n mensgedrewe ontwerpbenadering en die gebruik van data-gesentreerde “machine learning” tegnieke in waterbestuursnavorsing as die eerste opsie te oorweeg. Die oorhoofse transdissiplinêre metodiek, ’n mensgesentreerde ontwerp, en die “machine learning” werktuig is volgens die navorser die beste kombinasie van ’n wetenskapsgeoriënteerde “gereedskapskis” om waterbestuursprobleme te ondersoek.

DEDICATION

This dissertation is dedicated to many people without whose inspiration and support I would not have succeeded in completing my studies, and to whom I owe a great gratitude.

To my late father, Victor Malisa, who inspired and encouraged me throughout my life, as he greatly emphasised the importance of education. He provided a solid educational foundation by enrolling me in good primary and secondary schools. He dreamed of me getting a doctorate in my field of study, which I was very passionate about. On his deathbed, I promised him that with God's grace, I would be able to do that. To my grandparents, Phillip and Violet Malisa, whose love of education inspired me and the rest of the Malisa clan. They have waited decades for this moment and have always believed that I could achieve this level of research. Most importantly, my brother Clifton Malisa, who went to be with God on 15 September 2018. I dedicate this dissertation to you. I will always remember you and hear your voice telling me with pride how capable I am.

To my husband, André van der Walt, who tirelessly supported me throughout my studies and believed in me to the extent that I could not afford to give up, even when things were very difficult. André, you are one in a million husbands, a man of integrity and humanity. Thank you for being part of my academic and overall life journey. To my two sons, Patson Malisa and Anesu Malisa, who have always supported me through all the valleys and mountains and always believed that their mother is the greatest of all mothers.

Professor Eugene Cloete, without your unfailing support in every form, it would have been impossible for me to even enrol in studies at Stellenbosch University. I thank you for giving me a father figure that is the equivalent of my biological father. Your mentorship has provided me with leadership skills that I believe will greatly benefit my future career. You believed that I could achieve great things and that a PhD was a task I could accomplish as a stepping stone to greater responsibility, while most people around you did not.

ACKNOWLEDGMENTS

First and foremost, I am grateful to God, who made it possible for me to complete my PhD at Stellenbosch University against all odds. My God, you gave me supernatural tenacity to persevere when all hell broke loose. You miraculously provided me with angels in the form of my two promoters who desperately wanted me to get my PhD. In the midst of a multitude of challenges, you gave me wisdom on how to strategically approach and overcome the challenges I faced. You gave me the courage to stand up to anyone who ruthlessly tried to diminish my dignity and undermine my intellectual abilities because you so wonderfully shaped me into a black woman. My God, you proved once again that when you said yes, no one can say no. Thank you, my good shepherd.

To my promoter, Prof. K.I. Theletsane, who opened the door for me when it was closed at Stellenbosch University and offered to be my promoter and gave me hope in my darkest hour. I want to say that you were appointed to the Faculty of Military Sciences at that exact time. Working with you opened my eyes and showed me the power of believing in someone and how that can change a person’s career and life. You have taught me the practical meaning of the words “service” and “leadership” that I will carry into my future career. Throughout my research, your knowledge, understanding, and support have contributed significantly to my doctoral research experience. I appreciate the extensive knowledge and research skills you have provided me. Your structured way of mentoring and timely presence in the midst of your hectic schedule kept me happy.

To my co-promoter, Prof. A. Taigbenu, whom God miraculously sought and appointed as he is omniscient. I want to thank you for answering the call to make my dream of a PhD come true. Your unwavering commitment to my work strengthened my research morale when I was at wit’s end. Your fatherly care and extensive knowledge in water resources management contributed significantly to the success of my doctoral dissertation.

To Dr F. Babi, whom I consulted on machine learning modelling techniques, thank you for making yourself available amid your hectic schedule and keeping my hopes alive that I would develop meaningful models with the little data that I was provided by the respective institutions.

To Ms Maria Basson, the faculty liaison, who professionally guided me through the administrative processes. She always took care of me in a timely manner and helped me in situations when I was stuck.

Finally, I would like to thank the Dean of the Faculty of Military Sciences, Prof. Samuel Tshehla, for giving me the opportunity to conduct my research at the faculty even though I am a civilian. I enjoyed learning about military business practices, and being in a military environment was a pleasure.

April 2023

Rejoice van der Walt

TABLE OF CONTENTS

CHAPTER 2: THE HISTORICAL CONTEXT OF THE CASE STUDY

CHAPTER 3: INTERNATIONAL PERSPECTIVE ON MUNICIPAL WASTEWATER REUSE FOR AGRICULTURAL PURPOSES

CHAPTER 4: THE APPLICATION OF MACHINE LEARNING TECHNIQUES IN URBAN WATER SYSTEM MANAGEMENT

CHAPTER 5: RESEARCH METHODOLOGY

CHAPTER 6: QUALITATIVE RESEARCH FINDINGS

CHAPTER 7: MODEL DEVELOPMENT

CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

¥	Chinese yen
ACF	Auto correlation function
AdaBoost	Adaptive Boosting
AFS	Adaptive Fourier Series
AI	Artificial intelligence
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BOD	Biochemical oxygen demand
CART	Classification and regression tree
CDPH	California Department of Public Health
CEC	Contaminant of emerging concern
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
cfu	Colony forming unit
CG	Conjugate gradient
CMA	Catchment management agency
CMC	Catchment management committee
CMF	Catchment management forum
COD	Chemical oxygen demand
COD_{Cr}	Mean chemical oxygen demand
CONAGUA	Comisión Nacional del Agua
CRISP-DM	Cross-Industry Standard Process for Data Mining
CSV	Comma-separated values
DE	Differential evolution
DO	Dissolved oxygen
DWA	Department of Water Affairs
DWAF	Department of Water Affairs and Forestry
DWS	Department of Water and Sanitation
EC	Electrical conductivity
EDA	Exploratory data analysis
EPA	Environmental Protection Agency
ES	Exponential smoothing
EU	European Union
FAO	Food and Agriculture Organization
FC	Faecal coliform

GBM	Gradient Boosting Model
GDP	Gross domestic product
HC-T-DD	Human-centred data-driven technological [triad/framework]
IDP	Integrated Development Plan
ISO	International Organization for Standardization
IUWM	Integrated urban water management
IWRM	Integrated water resources management
LAS	Sodium dodecylbenzene sulfonate
LASSO	Least Absolute Shrinkage and Selection Operator
MAPE	Mean absolute percentage error
MICE	Multiple Imputation by Chained Equations
MLR	Multiple linear regression
MPN/L	Most probable number per litre
MSE	Mean squared error
mtmin	Monthly minimum temperature
mtmax	Maximum temperature
ND	Not detected
NGO	Non-governmental organisation
NTU	Nephelometric turbidity units
NWP	National Water Policy
NWRS	National Water Resource Strategy
OMEGA	Outil Mèthodologique de Gestion Intègrèe des Eaux Urbaines
Ova	Nematode and cestode eggs
PACF	Partial auto correlation function
pH	Potential hydrogen
RMSE	Root mean square error
RoRabs	Run-of-river abstraction
RSA	Republic of South Africa
RWU	Regional water utility
SARIMA	Seasonal Autoregression Integrated Moving Average
SAWS	South African Weather Service
SGB	Stochastic Gradient Boosting
spre	Sum precipitation
SS	Suspended solids
Stats SA	Statistics South Africa
SVM	Support Vector Machine

SVR	Support Vector Regression
SWRCB	State Water Resources Control Board [California]
TDS	Total dissolved solids
T-N	Total nitrogen
T-P	Total phosphorus
TSAMA	Transdisciplinary, Sustainability, Analysis, Modelling and Assessment
TSS	Total suspended solids
UN	United Nations
USA	United States of America
VAR	Vector Autoregression
WA	Wavelet
WFD	Water Framework Directive
WHO	World Health Organization
WRFP	Water Recycling Funding Program
WRP	Water reclamation plant
WSA	Water services authority
WSC	Water services committee
WSI	Water services institution
WSP	Water services provider
WTP	Water treatment plant
WTWs	Water treatment works
WUA	Water user association
WWTP	Wastewater treatment plant
XGBoost	Extreme Gradient Boosting

LIST OF UNITS

BCM	Billion cubic metre(s)
ha	Hectare(s)
hm^3	Cubic hectometre(s)
kl/m	Kilolitre(s) per month
km^2	Square kilometre(s)
m^3	Cubic metre(s)
m^3/d	Cubic metre(s) per day
mg/L	Milligram(s) per litre
ml	Millilitre(s)

ML/d	Megalitres per day
mm	Millimetre(s)
mm^3/yr	Cubic millimetre(s) per year
uS/cm	Microsiemens per centimetre

LIST OF CHEMICAL SYMBOLS

Al	Aluminium
As	Arsenic
B	Boron
Be	Beryllium
Cd	Cadmium
CN	Cyanide
Co	Cobalt
Cr	Chromium
Cu	Copper
F	Fluoride
Fe	Iron
Hg	Mercury
Li	Lithium
Mb	Molybdenum
Mn	Manganese
NH_3-N	Ammoniacal nitrogen
Ni	Nickel
Pb	Lead
Se	Selenium
Sn	Tin
Ti	Thallium
V	Vanadium
Zn	Zinc

CHAPTER 1:
INTRODUCTION

1.1 BACKGROUND TO THE RESEARCH

Freshwater is an essential resource for humanity and the existence of all living organisms. The challenge of conserving this resource is ever increasing. The challenge has led scholars such as Pradhan (2017), Orlando (2015), and others to postulate the possibility of global peace and security deficits due to water scarcity, which could culminate in a third world war as nations fiercely compete for water resources. Kofi Annan, former secretary-general of the United Nations (UN), once stated during his tenure that "fierce competition for fresh water may well become a source of conflict and war in the future". Among the reasons for projected freshwater scarcity are the continued growth of the world's population, urbanisation, and the negative impacts of climate change on freshwater availability (Eftelioglu *et al*., 2017).

According to Akhtar *et al.* (2021), the complexity of freshwater resource management arises from both anthropogenic activities and changes in natural systems. This is because human and natural systems are inextricably linked and interdependent. For example, population growth and urbanisation increase water-intensive socioeconomic development activities and food demand, which increase competition for water resources in an urban setting (Flörke *et al.,* 2018). Conversely, food production is typically associated with using large amounts of fertilisers and pesticides, which in turn have a high potential to pollute natural waters and coastal ecosystems and negatively impact freshwater availability (Lovarelli *et al.*, 2018). In addition, the negative impacts of climate change instigate increases in average temperature and uneven shifts in precipitation patterns that lead to extreme events such as heat waves, droughts, and floods, which pose significant challenges to forecasting and predicting urban water demand and supply (Koutroulis *et al.*, 2018). Research has shown that water agencies can sustainably manage the urban water system by accurately forecasting their water demand across all horizons.

Sustainable management of urban water systems is only possible if current and future water demands are known to allow adequate short- and long-term planning of key management actions. For this reason, extensive research is being conducted on forecasting techniques for water demand and supply. Efforts are being made to address the challenges posed by the rapidly changing environment, which require responsive political and economic policies. Water agencies are expected to sustainably manage the urban water system despite various challenges. To this end, there has been an increased search for decision support tools capable of managing the complexity caused by the variability and uncertainty in the urban water system and providing accurate forecasts and projections of urban water demand (Healy *et al.*, 2015). In this process, machine learning modelling techniques have emerged and have proven to be more robust in creating and training models that can accurately predict and forecast the water demand and supply of an urban water system. Their versatility stems from the fact that the developed models provide a better understanding and interpretation of the interconnectedness and interdependence of natural and human systems. The ability to quantify uncertainty makes machine learning techniques superior to traditional data-driven techniques (Tiwari & Adamowski, 2017).

However, Hadjimichael *et al*. (2016) pointed out that artificial intelligence (AI) applications, such as machine learning, in the water sector have not been fully explored in functional decision support systems. There is a disconnect between the water engineering and computer engineering fields. Although continuous research is being conducted to develop machine learning models for water management, the practical utility of these models is limited. Currently, machine learning techniques are practically widely used in the Global North, while they have little application in the Global South, including South Africa. In this study, data-driven supervised machine learning techniques were used to capture the relationship between several variables that affect water demand and supply in an urban water system in Stellenbosch Municipality, a municipality in South Africa (Oyebode *et al*., 2014).

1.2 RATIONALE FOR DEPLOYING SUPERVISED MACHINE LEARNING TECHNIQUES IN URBAN WATER SYSTEM MANAGEMENT

Because of the above challenges in managing an urban water system, it is becoming increasingly important to equip water agencies and managers with tools that will enable them to satisfy their consumers efficiently. This is because consumer satisfaction can only be achieved through the efficient and continuous provision of water in sufficient quantity and of good quality at an acceptable pressure and price while maintaining a reliable water distribution network. To meet these consumer expectations, comprehensive and accurate planning is required in conjunction with good decision-making processes. To this end, Oyebode and Ighravwe (2019) recommended short-term water demand forecasts (hours, days, weeks), which primarily help in planning and optimising urban water systems and allow water agencies to accurately plan and manage water demands, make better-informed water budget management decisions, accurately plan maintenance, and conduct financial planning (Rinaudo, 2015). Other researchers have examined medium-term forecasts (one to ten years), which help water agencies to make accurate water demand projections that are primarily influenced by changes in population size and demographics (Smolak *et al.*, 2020). Accordingly, long-term forecasting is considered a window for predicting water demand over decades. It is considered a basis for planning and designing future infrastructure development. These long-term forecasts can comprehensively address planning for the size and operation of reservoirs, pump stations, and pipeline capacity, as well as water pricing policies or water restrictions (Brentan *et al*., 2017).

However, due to increasing uncertainty in the urban water supply system, traditional, steady-state decision-making tools to predict future demands, which require simple static designs and upgrades, are no longer practical to meet consumer needs (Brown *et al*., 2010). This is because a decision that does not account for uncertainty could have disastrous consequences. There is currently a search for decision-making methods that can integrate various tools to take uncertainty into account and improve decision making. These new tools are believed to enable water agencies to better understand and predict urban water system behaviour (House-Peters & Chang, 2011). Considering this challenge, Herrfahrdt-Pähle (2013) indicated that the management of urban water systems needs to shift to a more adaptive management paradigm. Machine learning has emerged and has proven successful in building and training models that enable water managers and policy makers to understand and interpret variability in the urban water system. The most crucial feature of machine learning as a scientific method is the reduction of risks to the water system and reducing uncertainties that result from the negative impacts of climate challenges and rapid urbanisation (Tiwari & Adamowski, 2017).

Machine learning techniques have been and are being widely used in the Global North but are still in their infancy in the Global South. A significant drawback in the Global South is the lack of large datasets required to use machine learning techniques.

However, due to increasing uncertainties in the urban water system, the use of machine learning techniques is no longer an option in the Global South. This is because water demand models developed using traditional techniques are becoming increasingly inaccurate. The consequences are dire; considering that overestimation can lead to the construction of oversized facilities, and underestimation to service constraints to consumers and water scarcity. Researchers have reported that actual water demand is overestimated by up to 100% in some cases (Shabani *et al*., 2016). Accurate water demand prediction is essential for water supply agencies to meet consumer expectations, and using machine learning techniques to predict urban water demand and supply is no longer an option but a necessity.

1.3 INTEGRATED URBAN WATER MANAGEMENT (IUWM)

In addition to predicting water demand and supply, other elements, such as management approaches, must be considered for the sustainable management of urban water systems. This is because an urban water system typically includes a natural freshwater ecosystem with infrastructure to supply freshwater to urban centres and surrounding areas. In addition, there is a separate system for collecting, discharging, and treating wastewater from urban centres. There is global consensus on how to manage these two systems sustainably. Principles of urban water management, such as IUWM (Furlong *et al.*, 2015) have emerged. The main goal of the IUWM principle is to improve urban water resources management through resource diversification, efficient water use, and conservation.

Closas *et al.* (2012) and Kirshen *et al.* (2018) described IUWM as a principle that aims to manage and coordinate all water services, sources, and actors in an urban water system in a sustainable manner. Bahri (2012) considered IUWM as a mindset rather than a method and emphasised that there is no one-size-fits-all solution, but a mix of good water management principles adapted to local sociocultural and economic conditions. Furthermore, the World Bank (2016) defined IUWM as a flexible, participatory, and iterative process that integrates the elements of an urban water system, such as water supply, sanitation, stormwater management, and waste management, and incorporating both urban development and river basin management to maximise economic, social, and environmental benefits in an equitable manner. The researcher adopted the description of IUWM by Closas *et al.* (2012) and Kirshen *et al.* (2018) without rejecting the other views presented.

Koop and Van Leeuwen (2017), who advocated the principles of IUWM, outlined the following benefits of IUWM: (i) improved environmental protection, (ii) improved quality of life for poor urban dwellers through the health benefits of a clean environment emanating from improved sanitation and efficient drainage systems, and (iii) improved inclusive urban planning that brings social, environmental, and economic benefits to all. Accordingly, there are several vital principles of IUWM, including the recognition of the value of alternative water sources, which drives the use of "purposeful" water use and efficient management of water storage, distribution, treatment, reuse, and disposal as a full cycle. Bahri (2012) summarised these fundamental principles and highlighted their role in promoting a sustainable relationship between water resources, land use, and energy while ensuring economic efficiency, social equity, environmental sustainability, and consumer satisfaction.

To support the importance of the IUWM approach, Garcia *et al.* (2016) described a general urban water system, as illustrated in Figure 1.1, which shows the interconnectedness and interdependence of the elements that make up the urban water system. It shows how managing these elements in a fragmented approach would lead to uncertainty about urban water systems.

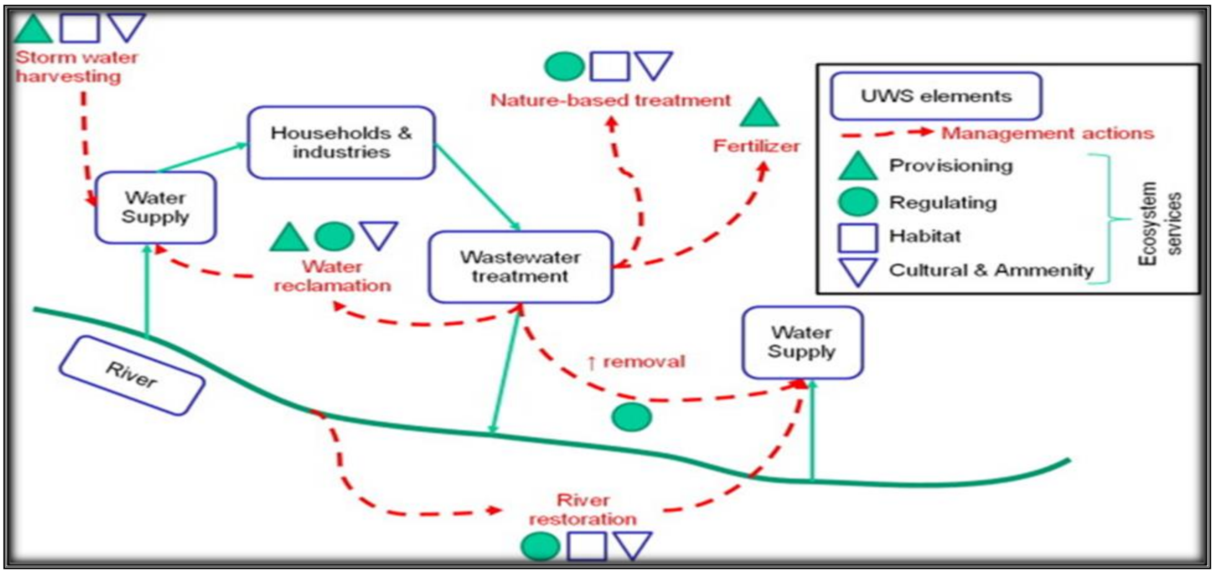


Figure 1.1: Urban water system; with the most common urban water system elements, implemented management actions, and affected ecosystem services

Source: Garcia *et al.* (2016)

As well researched as IUWM approaches have been, and continue to be, their application has been widespread only in the Global North. Several reasons are cited for the limited application of the approach in the Global South, which includes South

Africa. Among the challenges described are underdeveloped infrastructure and governance structures (Bahri *et al*., 2016). Other reasons are fragmented institutional arrangements, strict regulatory frameworks, and inappropriate economic and financial models that do not promote implementing IUWM principles (Jacobsen *et al.,* 2012; Tsegaye *et al.*, 2012).

Among the key IUWM principles cited are the efficient management of water storage, distribution, treatment, recycling, and disposal as a complete cycle. The benefits of such a paradigm for managing urban water systems are twofold. Firstly, it alleviates water stress, and secondly, it underscores the goal of collecting and treating urban wastewater and transforming it into high-quality effluent for reuse. With the looming global water crisis, applying IUWM principles is extremely important. A large portion of this study was devoted to alternative water sources in order to render the water supply of the urban water system sustainable. Several alternative water sources are being considered worldwide, including seawater desalination, groundwater abstraction, and the reuse of treated municipal wastewater. Since the reuse of treated municipal wastewater is a favourable option for use in the case study but has not been applied to date, the researcher investigated the reasons for its lack of use and the impact it would have if used to improve the water supply of the case study area.

1.4 TRANSDISCIPLINARY RESEARCH IN URBAN WATER SYSTEM MANAGEMENT

In addition to the accurate prediction of water demand and supply and an appropriate approach to urban water management, the current water management landscape requires a new approach to water management research. Due to the increasing complexity of water management, monodisciplinary research methods are becoming increasingly inadequate. The current water management landscape requires cross-disciplinary research to find holistic solutions for water management. A research approach that facilitates knowledge exchange between disciplines is critical. It underscores the fact that “soft” scientists are familiar with water research that focuses on people, but they lack technical knowledge, whereas “hard” scientists generally have a technical understanding of water management but are unable to link their technological research findings to socioeconomic challenges.

Research shows that in natural resource management studies, including water management, knowledge sharing between disciplines and non-disciplines is essential to provide a deeper understanding of research problems to enable researchers to find appropriate solutions (Chan *et al.*, 2021; Cvitanovic *et al.*, 2015). They can bridge the gap described by Westley *et al.* (2011) between the ever-growing global challenges and the effort to find appropriate solutions on time. According to Jacobs and Nienaber (2011), the ongoing water-related challenges require a transdisciplinary approach to research. Researchers and practitioners must work together to develop robust research solutions to address the ongoing water challenges. Pahl-Wostl (2007) and Reyers *et al.* (2010) also emphasised the need for a shift towards a transdisciplinary research approach to address the increasing challenges of water management holistically and effectively.

A transdisciplinary research methodology was used in this study to gather knowledge from different disciplines and to provide a holistic approach to answering the research questions. The researcher chose the transdisciplinary research methodology because it creates a conducive environment for solving complex, real-world problems such as water management. Furthermore, the transdisciplinary research methodology seeks to bridge the knowledge gap between multidisciplinary and interdisciplinary approaches by co-producing knowledge with society, such as water professionals from different water disciplines and the community (McGregor, 2011). Such a practice has been described by Scholz *et al.* (2006) as science with society rather than for society. Mobjörk (2010) considered this as a structured process of mutual learning between society and scientists from different disciplines to reduce the "resourcefulness gap" identified by Westley *et al.* (2011). Hadorn *et al.* (2008) recommended the use of transdisciplinary research methods when knowledge about a societal problem is uncertain and contested, and if the problem may lead to serious consequences if not resolved in a timely manner. The challenges of water management fit this problem description.

This study thus employed the transdisciplinary research methodology to identify the systems knowledge, target knowledge, and transformation knowledge of the research questions. Table 1.1 summarises these forms of knowledge and the respective research questions they sought to answer during the research process.

Table 1.1: Three forms of knowledge the research questions sought to address

Form of knowledge	Research questions
Systems knowledge	Questions about genesis and possible further development of a problem and about interpretations of the problem in the real world.
Target knowledge	Questions related to determining and explaining the need for change, desired goals, and better practices.
Transformation knowledge	Questions about technical, social, legal, cultural, and other possible means of acting that aim to transform existing practices and introduce desired ones.

Source: Pohl and Hardon (2007)

1.5 DEPLOYMENT OF SUPERVISED MACHINE LEARNING TECHNIQUES IN URBAN WATER SYSTEM MANAGEMENT

Since the main objective of this study was to apply machine learning techniques to predict the water demand of an urban water system, an overview of machine learning is provided in this section. The definition of machine learning is the art and science of allowing computers the ability to make decisions from data without being explicitly programmed. The two main categories of machine learning are supervised machine learning and unsupervised machine learning. Both techniques have been used in urban water management, with supervised machine learning techniques being used in this study. The reason is that they are widely used in water management because historical data from government agencies are becoming more readily available. In addition, there are several supervised learning algorithms, among which are Artificial Neural Networks (ANNs) (Suh *et al*., 2015), regression-based algorithms (Schleich & Hillenbrand, 2009), and time series algorithms (Arandia *et al*., 2015). These techniques have been used to develop models that attempt to capture the relationship between an outcome variable of interest and a set of explanatory or predictive variables.

1.6 PROBLEM STATEMENT

Sustainable urban water management, which includes water supply and demand management, is critical in any urban environment. Researchers have cited accurate forecasting and projections of water demand as prerequisites for the sustainable management of an urban water system. However, the ever-increasing interconnected and interdependent variables that result from the interactions between human and natural systems pose a significant challenge to water demand forecasting. The conventional modelling tools that are formally used are becoming increasingly inadequate. The impacts of climate change, which create uncertainties in precipitation cycles, exacerbate the inadequacy of traditional modelling tools because they cannot accurately quantify the uncertainties that arise in the system. As a result, the use of data-driven machine learning techniques is increasing, and they are currently widely used in countries in the Global North. However, the application is minimal in the Global South, which is also true for South Africa. Since machine learning modelling techniques are still in their infancy in managing urban water systems in South Africa, their introduction would provide water authorities, policy makers, and decision makers with tools to sustainably manage urban water systems and match water demand with supply in the face of increasing complexity and uncertainty in the systems.

As climate change alters evapotranspiration and precipitation rates and limits terrestrial water storage, the search for alternative water sources continues to generate significant interest. Among the options being considered, the reuse of municipal wastewater is emerging as a favourable option. However, South Africa has not tapped this resource to any significant extent. It is therefore necessary to investigate the barriers to the development of this resource in South Africa.

This study explored solutions for the sustainable management of the urban water system amid increasing complex challenges arising from the following:

- How variables in the system are becoming increasingly interconnected and interdependent, which is due to the interaction between natural and human systems or climate variability.
- The need for the water agency to capture and interpret the variability in the system and quantify the uncertainties introduced into the system by population growth, urbanisation, increased economic activity, and climate change.
- The urgency with which water management agencies must accurately plan, forecast, and predict a city’s short-, medium-, and long-term water demand and supply to prevent demand from exceeding supply.

Currently, water resources management research is conducted from a monodisciplinary perspective. However, this approach has been criticised for inadequately addressing the research problems facing the water sector. Accordingly, this study adopted a transdisciplinary research methodology to holistically understand the research questions and to find solutions to the problem the research aimed to address.

1.6.1 Hypotheses

This study sought to test the following hypotheses:

Null hypothesis (H_0): Supervised machine learning models can accurately predict and forecast urban water demand compared to conventional models.

Alternative hypothesis (H_A): Supervised machine learning do not accurately predict and forecast urban water demand compared to conventional models.

1.6.2 Research goal

The main goal of this study was to use technology to develop a strategy for the sustainable management of the Stellenbosch Municipality's urban water system.

1.7 OBJECTIVES OF THE STUDY

In order to achieve the above research goal and solve the research problem, the following objectives formed the basis of this study:

i. To review urban water management approaches globally and in South Africa.
ii. To investigate impediments to municipal wastewater reuse in irrigated agriculture globally and in South Africa.
iii. To explore challenges in implementing municipal wastewater reuse in Stellenbosch Municipality.
iv. To develop, train and deploy a highly accurate water demand and supply prediction and forecasting model for Stellenbosch Municipality

1.8 RESEARCH STRATEGY AND SCOPE OF THE STUDY

This study followed a transdisciplinary research approach that allowed the researcher to combine multiple disciplines, including non-disciplines, to holistically address the research questions and find solutions to a prevailing social problem. The research strategy followed is illustrated in Figure 1.2. A literature review on the development of South African water laws, policies, and governance was conducted to provide the researcher a deeper understanding of the research problem in the delineated study area. In order to achieve the main goal of the study, both the supply and demand management of the urban water system were investigated. To this end, the study introduced the philosophy of alternative water sources to improve the system's water supply, and the reuse of treated municipal wastewater was selected as a favourable option for the case study. Accordingly, literature on international perspectives of the reuse of treated municipal wastewater was reviewed to measure the effectiveness of the practice. In addition, the study conducted a literature review on the use of machine learning techniques in urban water management to identify trends and gaps in their use.

Both qualitative and quantitative research methods were used for data collection and processing. The qualitative component was used to inform quantitative aspects, including data collection and modelling. The application of machine learning modelling techniques to manage urban water systems was limited to Stellenbosch Municipality in the Western Cape province of South Africa. The objective was to investigate the impact of alternative water sources and the application of machine learning techniques in predicting water demand and supply in a municipality, thereby improving its water system management. The study recognises that there are several alternative sources of water; however, only treated municipal wastewater was considered in this study. In addition, several variables affect the management of an urban water system that were not considered in this study. Instead, water balance data, population distribution data, and weather statistics were considered. Research shows that daily, weekly, and monthly forecasts should be conducted when forecasting water demand in the short term. However, in this study, only monthly and annual forecasts could be conducted because daily and weekly water supply and demand data were unavailable for the case study. For the modelling strategy, several widely used water demand forecasting algorithms were selected and their performances were compared to determine the best algorithm that water agencies could use.

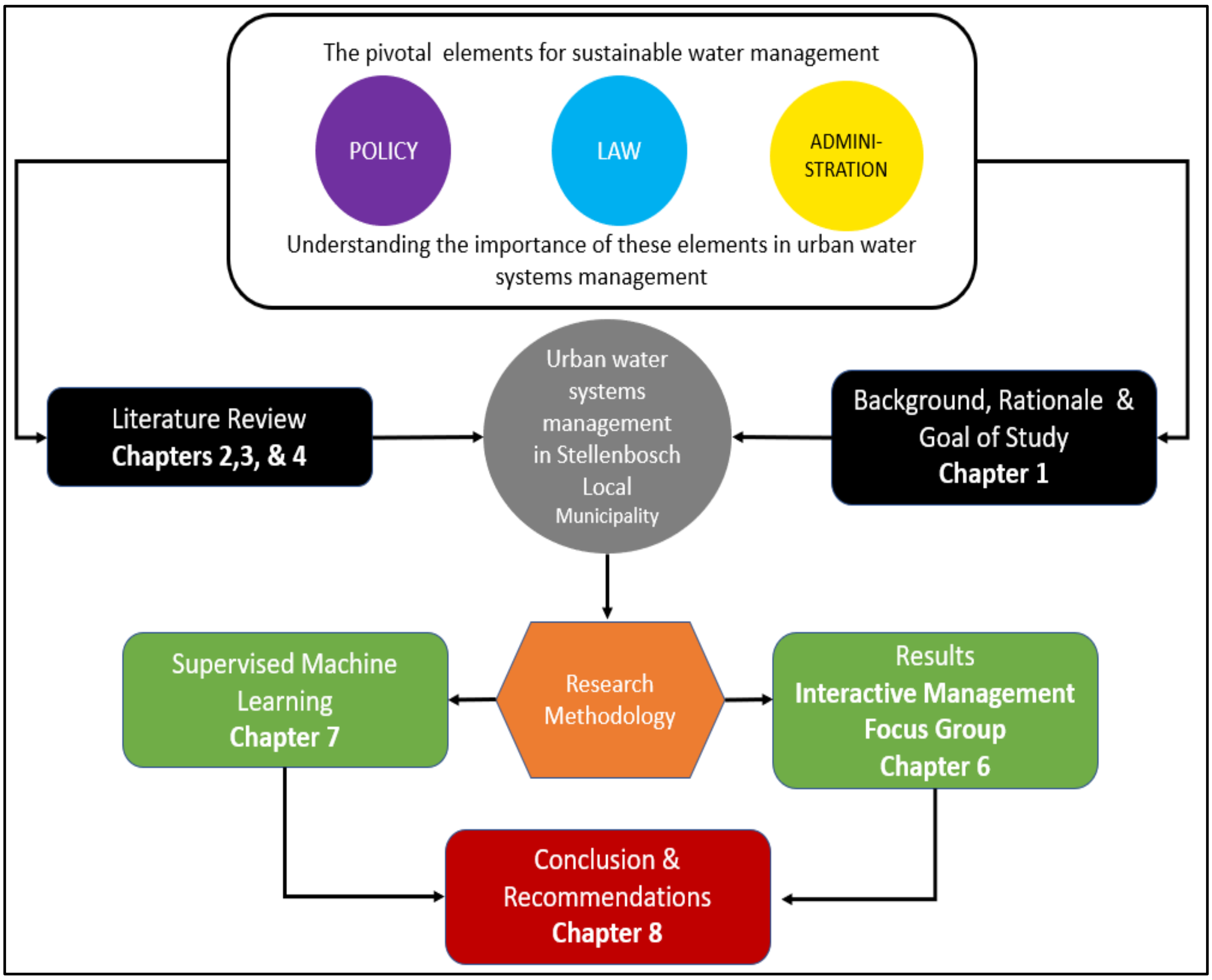


Figure 1.2: General overview of the research strategy

1.9 LAYOUT OF THE DISSERTATION

This dissertation contains eight chapters, which are structured as follows:

CHAPTER 1
- Presents the background of the study, as well as the research problem, hypothesis, main goal, objectives, scope, and the outline of the study.

CHAPTER 2
- Commences by delineating the historical background of the case study, followed by a literature review on the following aspects: evolution of South African water laws, policies, and administration, and global water management principles. A synopsis on the state of treated municipal wastewater reuse in South Africa is presented.

CHAPTER 3
- Literature review on the international perspective of treated municipal wastewater reuse in irrigated agriculture around laws, policies, and guidelines. The following regions were selected as case studies: California in the United States of America (USA), Spain in Europe, Mexico in Central America, China in Asia, and Egypt in North Africa.

CHAPTER 4
- Application of supervised machine learning techniques in urban water system management is presented, which details the fundamentals of the algorithms deployed.

CHAPTER 5
- Provides the overarching research methodology and research methods employed. Interactive management conducted to qualitatively understand the research problem in the context of the case study, along with quantitative supervised machine learning method used to develop, train, and deploy the water demand forecasting and prediction model.

CHAPTER 6
- The intearactive management workshop results and discussions on impediments in the application of treated municipal wastewater reuse in the case study are presented.

CHAPTER 7
- The results of the short- and intermediate-term water demand and supply models developed are presented and discussed.

CHAPTER 8
- Summary of research findings, conclusion, and recommendations are delineated.

Figure 1.3: General content of dissertation chapters

CHAPTER 2:
THE HISTORICAL CONTEXT OF THE CASE STUDY

2.1 OVERVIEW

The case study of this research was the Stellenbosch Municipality, which is named after the main town in its jurisdiction. This chapter presents the background of the Stellenbosch Municipality and the town of Stellenbosch. The historical perspective and socio-political context of the town of Stellenbosch strongly influence the development of Stellenbosch Municipality's urban water system. In order to provide context, the researcher deemed it appropriate to devote some consideration to these two aspects of the town. The management of urban water systems in South Africa has been significantly influenced by the political environment during the different political eras. An additional account of how South African water laws have evolved over the course of the various political dispensations is presented. This is followed by a detailed account of the period in which a striking change took place in 1994, which led to the adoption of a new, inclusive constitution from which the country's current laws and policies draw inspiration. The water laws and policies are followed by changes in the institutional arrangements of South African water management. In this context, the current institutional structure of the Stellenbosch Municipality is described, which ensures equitable water supply by the municipality. The chapter concludes with an overview of global approaches to water management and their evolution, along with the principle of IUWM. The researcher explores the application of these water management principles by Stellenbosch Municipality in Chapter 6.

2.2 BACKGROUND OF STELLENBOSCH MUNICIPALITY: THE CASE STUDY

Stellenbosch Municipality is estimated to cover an area of 900 km^2. Figure 2.1 shows the area. Its jurisdiction includes two major towns – Stellenbosch and Franschhoek – and several hamlets, including Wemmershoek, Klapmuts, and Jamestown, and several informal settlements. The combination of enormous environmental resources and high scenic beauty gives businesses in Stellenbosch Municipality a competitive advantage over surrounding towns. As a result, Stellenbosch is home to a disproportionate number of chief executive officers (CEOs) and managing directors compared to other cities in the country. As a result, the town can maintain a comparatively high level of economic activity and consumer services, regardless of its geographic location and population size. A detailed report on the historical prospects of the town of Stellenbosch is presented in the following subsection.

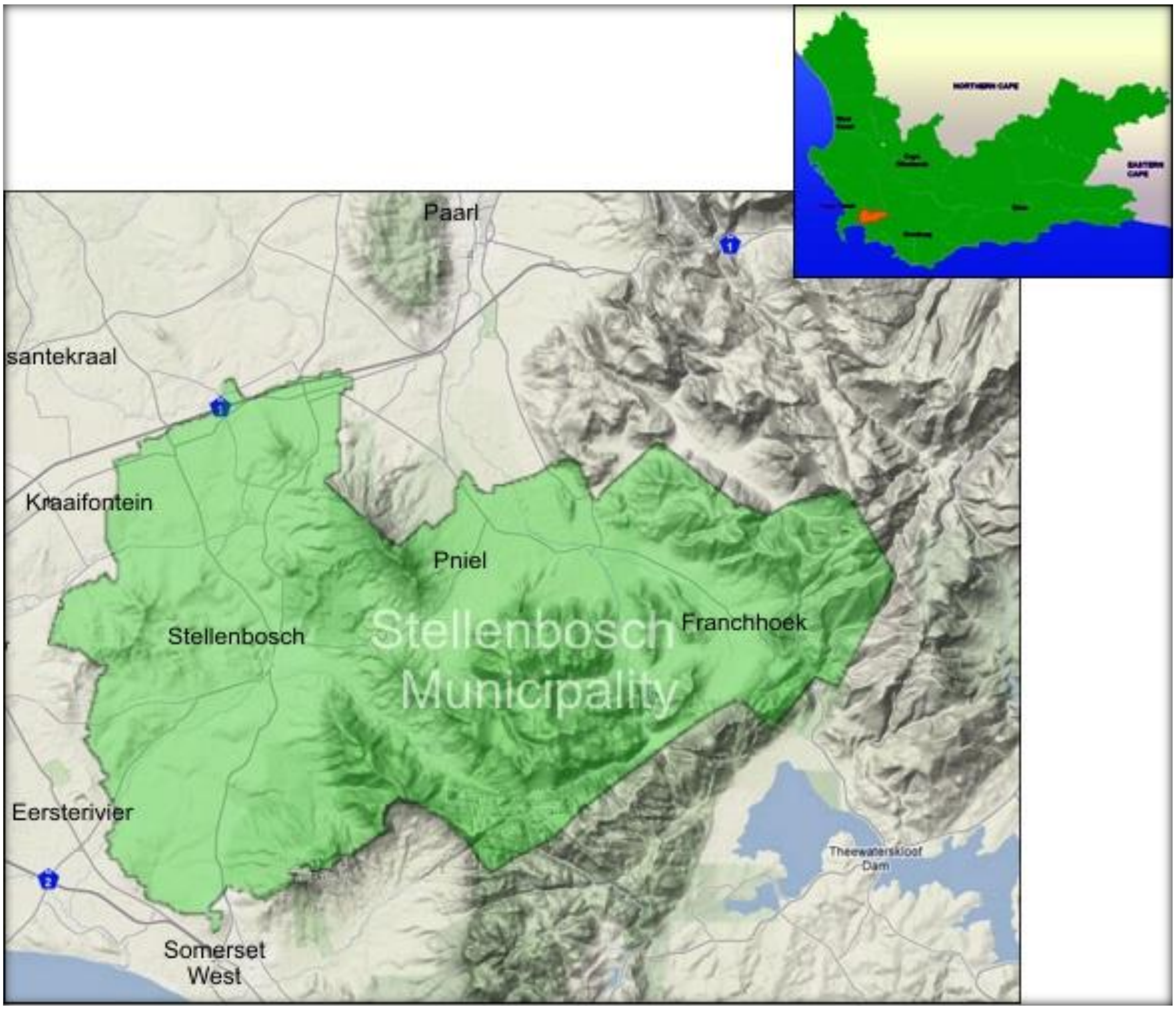


Figure 2.1: Map of Stellenbosch Municipality

Source: Western Cape Government (2017)

2.2.1 Historical perspective of Stellenbosch

Stellenbosch was founded in October 1679 by Simon van der Stel, then governor of the Cape of Good Hope, as an outpost of the Dutch East India Company. The outpost supplied Dutch merchant ships that were rounding the southern tip of the African continent. Van der Stel encouraged colonists to settle on the banks of the Eerste River (Hattingh, 1983).

Fairbridge (1922) summarises Simon van der Stel's vision on the morning of 29 October 1679 with tender empathy:

> It is two hundred and forty years since Simon van der Stel rode into the smiling valley surrounded by mountains and watered by the Eerste on its way to the sea from the slopes above Jan de Junkers Hoek. The land must have been white as Ornithogalum and pink as heather and monsonia that spring morning. The river must have rippled as merrily over the round stones and clumps of reeds as it does today. But there were no white houses, no great oaks, no vineyards, no peach blossoms, and no theological seminary. It was a wild landscape he looked upon when, filled with love for the land and a desire for its expansion – a desire the Society did not share – he founded the town of Stellenbosch in his first year in office in 1679 and named it after himself.

It is almost a century since Fairbridge (1922) so eloquently described the sensation of Simon van der Stel 335 years ago. Stellenbosch has evolved from humble beginnings into a town known for its scenic beauty, university, architecture, and wineries. Between 1684 and 1750, large tracts of land in Stellenbosch were granted to "vryburgers". These were farmers who were in the service of the Dutch East India Company but had the freedom to farm for themselves. They settled and farmed the land to produce goods for the company, which supplied the ever-growing fleet of ships heading east and back west. These farms are still part of the agricultural activities around Stellenbosch and Franschhoek today. The farms were worked in part by 8 500 slaves from West Africa, East Africa, Batavia, Suriname, Java, and India (Hattingh, 1983).

2.2.2 The socio-political context of Stellenbosch

In 1948, the National Party introduced the apartheid policy at a meeting in Stellenbosch, which secured the town's place in the political and social injustice of the past. The city is known as "the cradle of the apartheid doctrine", which was based on racial discrimination and the segregation of the four main population groups: white, Indian, coloured, and black people. This distinction between the four groups, as enshrined in the apartheid constitution, encompassed all aspects of life and made it punishable to live with and marry someone not of one's own race. The segregation line determined where to do business, what line to wait in, where to swim, how to conduct business, and where to reside legally. The entire public sector was a social mechanism for administering the apartheid doctrine. Before 1948, Stellenbosch had a racially integrated social structure. After 1948, Stellenbosch residents were forcibly relocated to segregated residential areas. The two laws that the researcher considers inhumane in the context of urban racial segregation were the Natives (Urban Areas) Act (No. 21 of 1923), which created a conducive environment for urban racial segregation, although it was not compulsory, and the Group Areas Act (No. 41 of 1950), which was brutal as many communities were segregated by race and lost their properties (Erasmus, 2010). A major concern of the above practices was the development of basic service infrastructure. The apartheid government prioritised the development of basic service infrastructure in white areas and neglected development in black areas.

The apartheid system was abolished in 1994. A new constitution was enacted based on universal suffrage and liberal values. Despite the new political circumstances, Stellenbosch, like most other cities and towns, remains racially divided. White residential areas are individually well served, while most black residents live on the outskirts of the city, where basic municipal services are non-existent or rudimentary (Seekings, 2008). Despite a constitutional amendment at the dawn of democracy that required the government to ensure that every citizen has access to clean water and adequate sanitation, as well as a safe environment (Republic of South Africa [RSA], 1996), racial segregation remains deeply entrenched in Stellenbosch. In its draft of the Stellenbosch Spatial Development Framework, the Sustainability Institute divided Stellenbosch into the following areas:

- Stellenbosch North: Welgevonden Estate, Cloetesville, Kayamandi, and Plankenburg;
- Stellenbosch East: Idas Valley, Simonswyk, Mostertsdrift, Rozendal, and Karindal;
- Stellenbosch Central: Central Stellenbosch, Dorp Street, and Stellenbosch University;
- Stellenbosch West: Onder Papegaaiberg, Devon Valley, and Die Boord; and
- Stellenbosch South: Krigeville, Dalsig, Brandwacht, Paradyskloof, Technopark, De Zalze, and Jamestown (Stellenbosch Municipality, 2012).

The above-mentioned division of the town resembles a profound urban racial segregation that has been practised since the apartheid era. Of concern is that geographic demarcation and population composition influence decision making regarding infrastructure development and basic service delivery. The northern part of the town consists of black neighbourhoods with informal and formal mixed housing. The spatial development of the population and the town itself significantly impact the management of urban water systems, including initiatives to introduce new approaches to urban water management, as population distribution influences the development of water and wastewater infrastructure. The size and growth rate of a population determine whether the town can provide adequate clean water and wastewater services within its jurisdiction. Despite projections showing that the black population is growing at the fastest rate compared to other races, the City of Stellenbosch, to date, does not have a basic service policy designed to effectively address the skew in population distribution along racial lines.

2.3 THE EVOLUTION OF SOUTH AFRICAN WATER LAWS

Since the 20th century, South African water laws have been written and adapted according to policy strategies and priorities. The water laws presented in this study shed light on how water resources have been developed and managed. Significantly, the four main water laws discussed in this thesis clearly reflect the political influence of the period in which they were written. This is evident in the shift in priorities that correlates with the change in the political landscape from 1912 to 1998. In addressing current and future urban water management challenges, current decision-making processes should consider South Africa’s political history in order to address the consequences of abuses committed by various administrations over the past century.

2.3.1 The Water Act (No. 8 of 1912)

The year 1910 marked the establishment of the Union of South Africa and the time when parliament enacted the Water Act (No. 8 of 1912) (Thompson, 2006). The Act divided water into a public category (*res communis*) and a private category (*res privatae*). The central aspects of the legislation were irrigation activities, riparian water rights, and the conservation of water before it entered the receiving natural waters (Tempelhoff, 2017). The responsibilities of the Department of Irrigation at the time were to manage large volumes of water and to develop large irrigation systems.

Minimal, if any, attention was paid to urban wastewater management. The urban population served was relatively small and the water supply was easily managed. The government conducted water management using technocratic methods; water infrastructure construction projects were therefore highly prioritised (Kidd, 2009; Thompson *et al*., 2001).

Rapid and massive technological inventions characterised the period after the end of World War II in 1945. During this time, water was considered an infinite natural resource. Changes in attitude and technology affected the South African landscape. They necessitated the repeal of the Water Act of 1912, which was subsequently replaced by the new Water Act (No. 54 of 1956).

2.3.2 The Water Act (No. 54 of 1956)

The Water Act of 1956 was greatly influenced by an increase in industrial and agricultural activities associated with urbanisation. These changes in the socioeconomic environment necessitated a shift from riparian rights to government control of water resources. The Act was enacted eight years after the beginning of the apartheid era. Among its fundamental principles was enforcing riparian rights while the state exercised pseudo-control over all water resources. There were strict control measures on industrial and groundwater use. The Act's reinforcement of partial riparian rights resulted in unequal access to water and jeopardised the black community's access to water (Kidd, 2009). During this time, the focus of the Department of Irrigation shifted to sustainable socioeconomic development that was in line with global trends. Priority was given to developing water infrastructure for industrial activities and to meet household needs due to rapid urbanisation. Control of water pollution was a high priority, and the need for effective management was emphasised. Complementary laws were enacted to control, conserve, and use water for domestic, agricultural, and industrial purposes. In addition, water authorities were established to manage water for urban and industrial uses and regional wastewater systems within their jurisdiction. However, the Water Act of 1956 contained inconsistencies regarding water pollution issues that were differentiated according to water use activities (Glazewski, 1999). Although the 1956 Water Act was intended to strengthen government control over water (*dominus fluminis*), the government's powers were not used extensively to weaken riparian rights (Kidd, 2009). The Water

Act provided adequate riparian rights to privileged white farmers. The majority of black farmers were excluded from access to water rights.

As urbanisation progressed, the government introduced formal subsidies to support infrastructural development of wastewater treatment plants (WWTPs) and water treatment facilities in urban areas (Tempelhoff, 2017). These schemes excluded black residential areas until 1977 when black municipal councils were introduced, and efforts were made at the local and regional government levels to improve water and sanitation in South African townships (Tempelhoff, 2004). However, these efforts were peripheral and insignificant. This compelled the founders of the 1993 Interim Constitution to prioritise the elimination of the previous regime's historical disregard for Africans' human right to adequate access to clean water and sanitation. Accordingly, South Africa revised and drafted new water laws at the onset of its democracy to address the abuses of the previous oppressive regime.

2.4 LEGISLATIVE FRAMEWORK FOR THE WATER SECTOR IN THE POST-APARTHEID ERA

2.4.1 The Constitution of the Republic of South Africa (1996)

In 1994, a new democratic order began in South Africa, during which a new constitution was drafted. The Constitution of the Republic of South Africa (RSA, 1996) is currently the supreme law of the land. To redress past grievances, the Bill of Rights (Chapter 2) of the Constitution guarantees every citizen the right to water by stating that "everyone has the right to have access to adequate food and water". In the democratic order, access to water is considered a human right. Accordingly, the Bill of Rights also provides that access to water should be fair and equitable for every citizen. It obligates the state to take appropriate legislative and other measures, within its means, to achieve the "progressive realisation" of everyone's right to have access to sufficient water (Article 27(2)). In addition, other constitutional rights emphasise water security as a core element in implementing these rights.

From a water management perspective, the Constitution addresses environmental issues in Article 24 of the Bill of Rights, which states that everyone has the right to:

"(a) to an environment not detrimental to his health and well-being; and (b) to the protection of the environment for the benefit of present and future generations through appropriate legislative acts and measures that:

i. prevent pollution and environmental degradation,
ii. promote conservation; and
iii. ensure environmentally sustainable development and use of natural resources while promoting defensible economic and social development."

Following the adoption of the new Constitution, the White Paper on a National Water Policy (Department of Water Affairs and Forestry [DWAF], 1997) was developed through a two-year consultation and participation process. The main objective of the consultation process was to ensure an all-inclusive process for all South African citizens in the development of the new National Water Act, which aimed to restore a semblance of equitable access to water.

2.4.2 The National Water Policy (NWP)

In 1994, when the democratically elected South African government sought to eliminate the evils of apartheid through projects that would provide equitable and sustainable social and economic development, existing laws prohibited them. Consequently, new policies were designed and adopted. Among these policies was the NWP of 1997 (DWAF, 1997), which redefined the ownership and allocation of the country's water resources. The main objective of the NWP was to support the realisation of the democratically elected government's aspirations regarding water resources management and development projects. The central principle established by consensus was that all water is a common resource and that control over all water resources should rest with the state (*res fluminis*). The statutory principles of riparian rights and water ownership were abolished. Before the NWP, 28 basic principles and objectives were developed for the new water law. In this section, the researcher refers to Principle 7, which emphasises the efficient management of the quality, quantity, and reliability of the nation's water resources to achieve optimal, long-term environmental sustainability and social and economic benefits for all South Africans. The main objective of the NWP was to fulfil the provisions of the Bill of Rights of the South African Constitution of 1996 through equitable, sustainable, efficient, and effective water use for optimal social and economic benefit (DWAF, 2004; 1997). In addition, several water-related policies and laws have been drafted and adopted to be administered by various ministries and government sectors involved in water-related activities.

By adopting a clear water policy after the apartheid era, the new government demonstrated its commitment to redressing the grievances of the previous regimes in this sector. A clear policy is a cornerstone for achieving a government's development aspirations. Without clear policies, it is complicated to muster the political will to address society's problems. A well-developed policy framework enables effective water resources management development, resource allocation, delegation of power and responsibility, and clearly articulated sectoral plans (Folifac, 2007). The South African NWP is internationally regarded as progressive, forward-looking, and ambitious. Moreover, the world has welcomed it because it is based on universal human rights and equality for all people (MacKay *et al.*, 2003). Drastic changes were required to realise the vision and aspirations of the NWP, and to date there have been numerous successes in its implementation.

Among the accomplishments is that clarity was achieved on managing the nation's water resources, which were previously in disarray. The institutional fragmentation that characterised the sector was eliminated. Significant progress has been made in harmonising water services to include rural populations. Among the 28 water principles developed, water services authorities (WSAs) have made significant efforts to promote the values enshrined in the Bill of Rights. Great emphasis has been placed on equity and the sustainable management of water resources.

Despite the documented successes, there are numerous challenges in implementing the NWP. These include failing to achieve equitable access to clean water for all South Africans (Bayliss *et al*., 2016). For example, in the agricultural sector, which has been a bone of contention in addressing the ills of past oppressive regimes, Chikozho (2008) reported that 95% of water resources continue to be in the hands of white commercial farmers. In addition, Viljoen and Van der Walt (2018) report that the agricultural sector remains the largest consumer of freshwater in the country and, in contrast, is also the sector that experiences the greatest water stress. Reports of failures to transfer water use rights to land reform beneficiaries who have been deprived of water rights for centuries reveal some of the implementation shortcomings of the NWP (Chikozho *et al.*, 2020).

Moreover, the lack of clarity on some principles, such as meeting the basic water needs of all South Africans and profitable water pricing models, hinders the realisation of NWP ambitions (Donnenfeld *et al.*, 2018). Like other regions of the world, water demand in South Africa is increasing rapidly due to urbanisation, population growth, and climate change, while supply is shrinking, which requires efficient management of water supply. As a result, the gap in meeting the water needs of the previously disadvantaged black majority is also widening instead of closing, which suggests that the policy is failing to address one of the critical objectives of the NWP (Adom & Simatele, 2021).

2.4.3 The National Water Act (No. 36 of 1998) and the Water Services Act (No. 108 of 1997)

Based on the basic principles and objectives mentioned above, the National Water Act was promulgated in 1998. To enforce the repeal of more than 100 water laws and the abolition of all riparian rights (RSA, 1998a:Schedule 7), the national government must assume full responsibility for managing the country's water resources for the benefit of its people. The rationale was that the government must act as a public trustee of the nation's water resources in the public interest to ensure that water is "protected, used, developed, conserved, managed and controlled for the benefit of all people in a sustainable and equitable manner" (DWAF, 2001).

The enactment of the National Water Act established a legal framework under which water supply and sanitation, water resources management, and water use were to be coordinated. These provisions of the law provide the basis for environmentally sustainable development and natural resource utilisation while promoting reasonable economic and social development (DWAF, 2004; 1996).

The National Water Act works in conjunction with the Water Services Act of 1997, which regulates the provision of water and sanitation services to the residents of South Africa (RSA, 1998a; 1997). This Act provides for establishing water services institutions (WSIs), such as WSAs, and their powers and responsibilities.

The main objectives of the Act are as follows:

- The right to access basic water supply and sanitation;

- The preparation and adoption of water services development plans by WSAs; and
- A regulatory framework, monitoring, and financial support for WSIs.

Section 9 of the National Water Act provides that the minister may prescribe mandatory national standards for water services (RSA, 1998a, 1997; DWAF, 2001; Karodia & Weston, 2001). Section 9 was designed to ensure equity and water security for all South Africans.

2.4.4 The National Water Resource Strategy (NWRS)

The NWRS1 was written as a legal instrument to implement the National Water Act of 1998. It introduced mechanisms and approaches for managing all water sources to achieve the development goals of the national government. Regarding wastewater reuse, the NWRS1 describes direct wastewater reuse, which takes the form of treated wastewater flowing back into natural waters and to consumers downstream. This type of wastewater reuse is used extensively for agricultural activities.

Subsequently, in accordance with the law, the Department of Water Affairs (DWA) prepared and published the second edition of the NWRS (NWRS2) for the period 2013 to 2017 (DWA, 2013). The NWRS2 builds on the progress made by the NWRS1. Following the National Development Strategy, the main objective of NWRS2, which continues to be based on the requirements of the National Water Act, is to respond to the priorities set out in the South African National Development Plan: Vision 2030 (National Planning Commission, 2011). By developing a strategy for protecting, managing, and controlling scarce national water resources, the NWRS2 pursues integrated water resources management (IWRM) with a more holistic view of sustainable water resources management. This strategy includes using alternative water sources such as seawater desalination, rainwater harvesting, and water reclamation (reuse and recycling) to meet South Africa’s current and projected water demands.

The NWRS2 emphasises wastewater reuse as indispensable to meeting the water needs of South Africa’s socioeconomic development. The policy defines reuse as use of treated or untreated wastewater for a process other than the one that generated it, i.e., there is a change of user. For example, the reuse of municipal wastewater for agricultural irrigation. Water reuse can be direct or indirect, intentional or unintentional, planned or unplanned, local, regional, or national, depending on location, scale, and significance. Water reuse may (or may not) involve various types of treatment, and reclaimed water may be used for various purposes (DWA, 2013).

The NWRS2 indicates that wastewater reuse is gaining social acceptance and is proving to be technically feasible (Van Niekerk & Schneider, 2013). However, the controls imposed on wastewater reuse by pieces of legislations such as the National Water Act (No. 36 of 1998), the National Environmental Management Act (No. 107 of 1998), the National Environmental Management: Waste Act (No. 59 of 2008), and the Water Services Amendment Act (No. 30 of 2004) make its implementation cumbersome. This is further compounded by the statutory provision that allows municipalities to enact by-laws for the reuse of municipal wastewater, which can lead to numerous inconsistent wastewater reuse ordinances across the country. This complicates the entire process.

An essential aspect of the NWRS2 related to the reuse of treated municipal wastewater is the establishment of water guidelines that address water quality requirements, treatment technology selection, construction, maintenance, and financing of water reclamation systems. This policy document cites a lack of public participation as a significant reason for the failure of water reclamation initiatives. This is because there are no national classification standards for reclamation (water reuse), and it benefits only a small segment of society, mainly agriculture. It is emphasised that there is a need to explore the governance of water reclamation to set standards for drinking water, recycling, and crop irrigation (DWA, 2013). This study was therefore devoted to reusing municipal wastewater as an alternative water source to augment urban water supplies. The effective management of water resources critically depends on institutional arrangements. The following section describes the current institutional arrangements for water supply in South Africa.

2.5 SOUTH AFRICAN WATER INSTITUTIONS

Section 40(1) of the Constitution states that the spheres of government of the Republic shall be national, provincial, and local, which shall be distinctive, interdependent, and interrelated. Accordingly, the South African water sector is quasi-federal, with the statutory framework for the sector being set by the national department, which governs the water sector institutions. In addition, the Water Services Act of 1997 sets out the powers and responsibilities of the water sector institutions. It highlights the role of the following: WSAs, water services providers (WSPs), water boards (WBs), water services committees (WSCs), and water services intermediaries. Regional water institutions include regional water utilities (RWUs), catchment management agencies (CMAs), catchment management forums (CMFs), and water user associations (WUAs) (Kranz *et al.*, 2005; DWAF, 2009, 2004). A common goal of all these water institutions is to ensure sustainable ecological development and use of natural water resources while promoting defensible economic and social development (DWAF, 1996).

Figure 2.2 depicts the hierarchy of South African water institutions. South African water institutions report to the Minister of Water and Sanitation. The Department of Water and Sanitation (DWS), responsible for managing the nation's water resources, is located in the first branch of government, namely the national branch. The DWS's responsibilities include assisting local authorities by setting norms and standards for the operation, monitoring, and administration of the National Water Act of 1998. The WBs, whose main role is to ensure the efficient supply of bulk treated water to the commercial sector, are located in the second sphere, namely the provincial government. The third sphere is the local authorities, which are responsible for providing clean water and adequate sanitation to residents within their jurisdiction.

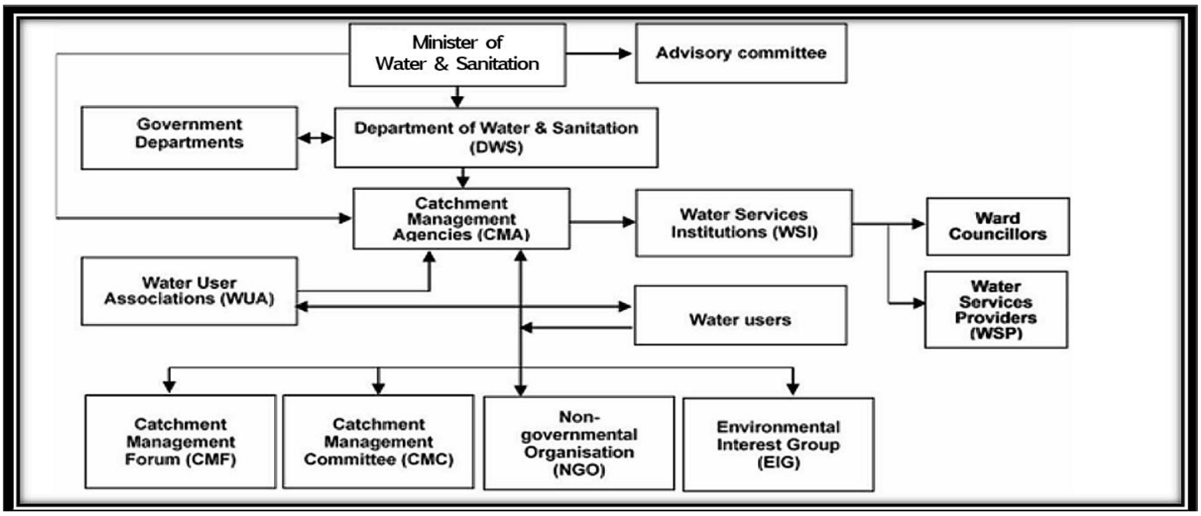


Figure 2.2: Water management institutions in South Africa

Source: Kapfudzaruwa and Sowman (2009)

At the local government level, WSAs are responsible for planning and implementing water services for consumers under the guidance of Integrated Development Plans (IDPs). The local municipality is responsible for maintaining the water distribution network and effectively and efficiently managing the water system to meet consumer needs. Beyond the WSAs, there are several WSIs with different responsibilities. Table 2.1 lists these WSIs and their main characteristics.

Table 2.1: Main features of water services institutions (WSIs) in South Africa

WSIs	Main features
WSAs	• A municipality is responsible for ensuring access to water supply and sanitation services. • Must be a municipality and no other institution – Category A, B, or C municipality (if authorised by the Minister of the Department of Provincial and Local Government). • May itself perform functions of a WSP, or enter into a contract/joint venture with another WSP.
WSPs	• Provide water supply and sanitation services (physically) to consumers under contract to the WSA. • WSP functions can be performed by the municipality, WB, non-governmental organisation (NGO), community-based organisation, private sector company, or any other private or public body. • No person may operate as a WSP without the approval of the WSA.
WBs	• Established by the Minister of Water Affairs and Forestry. • Primary function: to provide water services to other WSIs. • Are public WSPs. • May perform secondary activities if primary functions and financial standing are not compromised. Examples: ○ Provide management services, training, and other support services.

	○ Supply untreated water not for household purposes. ○ Provide catchment management services. ○ Provide water supply and sanitation services in a joint venture with WSAs. ○ Perform water conservation functions. ○ With the approval of the WSA, supply water directly for industrial use, accept industrial effluent, and act as WSPs to consumers.
WSCs	• A statutory committee may be established by the minister should a WSA fail in its duty. • A WSC does not refer to a community-based organisation that performs a WSP function at the community level (rural).
Water services intermediaries	• A person or body providing water to people as a minor part of a contract (e.g., farmer to labourers, landlord of flats to tenants, mining company to employees in housing). • Only applicable where there is an obligation by one party to provide services to another as part of a contract.

Source: DWS (n.d.)

2.6 CURRENT STELLENBOSCH LOCAL GOVERNMENT INSTITUTIONAL ARRANGEMENTS

Stellenbosch Municipality is responsible for the water supply to Stellenbosch and falls under the B category of municipalities. According to section 155(6) of the Constitution, a municipality consists of a political structure, an administration, and the municipality itself (Section 2(b) of the Municipal Systems Act, No. 32 of 2000 [RSA, 2000]). The mayor is the political head of the municipality, is responsible for the annual budget, and exercises certain delegated administrative functions. The Municipal Council has the right to regulate the affairs of the municipality, exercise legislative and executive powers, and finance the affairs of the municipality by collecting taxes and fees for services. In addition, the municipality has the right to levy surcharges for services.

The mayor plays a central role in the management of the municipality's water systems, as well as the preparation, approval, and execution of the municipality's budget. Funds allocated for water services are accounted for in the budget. Stellenbosch Municipality's annual operating budget is currently R2.2 billion, including the capital budget of R519.6 million (Stellenbosch Municipality, 2018a). National government allocations, taxes, service charges, and borrowing cover the annual budget. The budget year is from July to June. The capital budget is funded by revenues, grants, and borrowing. It is at the discretion of the mayor to decide which funds to use to improve or implement novel water management projects such as the municipal wastewater reuse recommended in this study. Funding is the first criterion for the success of any project.

Section 23 of the Constitution requires each municipality to adopt a five-year IDP at the first meeting of a new council. The purpose of the IDP is to guide and inform all planning, development, and management activities of municipalities (RSA, 2000). IDPs must be consistent with annual budgets and capital projects. Capital expenditure should be prioritised accordingly. An IDP is essentially a planning document and by law should be adopted annually along with the budget. The finances of the council are managed by a municipal manager who carries out their functions in line with the Municipal Finance Management Act (No. 56 of 2003 [RSA, 2003]). The council also appoints directors as heads of the respective departments. The term of office of the directors is five years. It is the responsibility of the council to adopt policies to guide the administration in performing its duties. The policies are reviewed annually when the budget is adopted. In the context of water management, such administrative arrangements increase uncertainty in introducing new water management approaches.

Council-appointed directors include the Director of Engineering Services, with several senior managers reporting to him/her, who are responsible for various services, including water supply and sanitation. The role of the Senior Manager of Water and Sanitation is to ensure adequate water supply and sanitation services to consumers. Two components must thus be managed: the supply of sufficient clean water to the utility system and the wastewater generated in the system. Generally, once the raw water is extracted, it is treated and distributed to urban centres, where it is used for various purposes, such as residential, agricultural, and industrial activities. As a result, wastewater is generated, which should be treated before being discharged into natural water bodies. Figure 1.1 illustrated the urban water system, which showed the negative impacts of discharging inadequately treated wastewater from urban environments into natural water bodies. CMAs have been established to monitor and protect water resources from contamination by wastewater discharged from malfunctioning WWTPs. Accordingly, the National Water Act of 1998 provides for the establishment of CMAs, as shown in Figure 2.2. These CMAs are considered statutory bodies that are responsible for protecting development and promoting the sustainable and equitable management of water resources. The main objective of establishing

CMAs is to devolve water resources management to the catchment level and to involve local communities (Kahinda *et al*., 2016). CMAs develop their watershed management strategies at the basin level, according to their needs. Within the CMAs, there are catchment management committees (CMCs), CMFs, NGOs, and environmental interest groups, whose main objective is to promote public participation in the watershed. In addition, there are WUAs, which are composed of water users who want to work together because of common interests (Swatuk, 2010).

Establishing CMAs is considered best practice internationally and is an important tool for implementing the principles of IWRM. Through CMAs, local communities are allowed to participate in water management decision-making processes. This facilitates participatory water resources management and promotes clear accountability. However, the formation and functioning of these water institutions in South Africa have been highly volatile, which negatively impacts water sector performance and leaves stakeholders disillusioned. Ten years after the enactment of the National Water Act, the process of establishing CMAs was suspended for four years, and by 2012, only two CMAs had been established. The establishment of the Inkomati and Breede CMAs in 2005 and 2006 was highly problematic, as was the process of establishing the 1997 CMAs, which was suspended in 2009 and resumed in 2012 (Bourblanc & Blanchon, 2014); thus complicating the watershed water management process.

On most occasions, when the public has been involved in watershed decision making, accusations and counter-accusations have occurred between the DWS, municipalities, and the public (Kapfudzaruwa & Sowman, 2009). For example, the DWS accuses municipalities of poor planning and inadequate water supply. In contrast, the municipalities claim they are open to changes where they can take on additional responsibilities, such as reallocating water for various purposes. This is in contrast to the DWS, which is unwilling to delegate its mandate to other institutions and prefers that water allocation remains in its own hands (Weaver *et al.*, 2017). The discrepancy between policy promises and what local municipalities can deliver complicates public participation as community trust in these institutions diminishes (Sithole & Mathonsi, 2015). However, water management is a dynamic process that evolves with time and needs.

2.7 EVOLUTION OF GLOBAL WATER MANAGEMENT APPROACHES

For more than a century, water has been managed worldwide by state-centred water management systems. This was characterised by a top-down, command-and-control approach (Hassanzadeh *et al.*, 2016; Scoullos, 2012; Pahl-Wostl *et al.*, 2007), which was also technocratic (Pahl-Wostl, 2002) and fragmented (Frantzeskaki & Loorbach, 2010). This water management approach required governments to take full responsibility for all water issues without input from stakeholders and civil society (Walker, 2014). However, as global water challenges increased and assumed greater complexity, water managers and practitioners came to the consensus that the traditional role of governments as the sole decision makers in water resources management and the provision of water and sanitation services was no longer feasible (Georgakakos *et al.*, 2012; Mukhtarov, 2008). A shift in the water governance paradigm from a government-centred one to a more efficient, effective, efficacious, and robust water management approach that is stakeholder-centred was advocated (Walker, 2014; Castro, 2007). The stakeholders comprise government, civil society, NGOs, and the private sector, who, within a democratic framework, participate in shared risks and benefits of the management of water resources and services.

In the wake of this stakeholder-centred water management discourse, various global water initiatives attempting to respond to global water challenges have emerged, as well as global water management principles. For example, the principle of IWRM was coined at a UN conference in Mar del Plata in 1977 (Worthington, 1977). From then on, the IWRM framework became a global principle and practice of water management, underpinning a coordinated approach to managing of water and related natural resources to achieve equity, the efficiency of use, and environmental sustainability. Other notable international water conferences included the International Conference on Water and Environment, held in Dublin, Ireland, in 1992, which produced the Dublin Principles, and the UN Conference on Environment and Development, held in Rio de Janeiro, which produced Agenda 21. Since then, these emerging water governance principles have had a profound impact on the development of IWRM worldwide (Molinos-Senante *et al.*, 2014).

However, the IWRM concept has proven limited because it is only a process with tools to assess and evaluate IWRM programmes. It does not provide specific guidance on how to deal with particular water management problems. It describes a wide range of principles, tools, and guidelines that need to be adapted to the specific context of the country or region at the river basin level (Xie, 2006). In this regard, the Global North, particularly Europe, has made progress in incorporating IWRM principles into its water management approaches. For example, the adoption of the Water Framework Directive (WFD) by European Union (EU) member states has improved the application of IWRM in the region. As a result, scientists have reported improved management of water resources, which has contributed to better economic and social development for European citizens. The result has been equitable water distribution while maintaining sustainable ecosystems (Benson *et al.*, 2015; Safavi *et al.*, 2015; Warner *et al.*, 2008).

In contrast, efforts to implement IWRM principles in the Global South face numerous challenges related to how to integrate different aspects of water management principles (Saravanan *et al.*, 2009). There is confusion about how and by whom certain aspects of water integration should be implemented in the region (Biswas, 2008; 2004). This poses a significant challenge in the implementation phase and complicates efforts to translate the concept into an operational tool that can guide the development of water resources management in the region (Grigg, 2008). In addition, elements such as untrained staff in water resources institutions, weak financial structures, lack of political will (Swatuk, 2005), and problems within institutional arrangements (Grigg, 2008) exacerbate these challenges.

Despite these challenges, the Global South is compelled to apply the IWRM concept, regardless of its ineffectiveness in the context of the Global South. A major obstacle is that countries in the Global South rely on donor funding for development programmes that aim to alleviate poverty and hunger, increase the wellbeing of citizens, and improve sustainable resource management (Ako *et al.*, 2010). Nevertheless, several researchers have pointed out the importance of adapting internationally conceived water governance principles before applying them in the Global South (Gallego-Ayala & Juízo, 2014; Kahinda & Boroto, 2009). Mapedza *et al.* (2016) pointed to the application of internationally conceived water management principles in the context of geographic location. However, the following questions remain: Is there significant progress in implementing water management principles conceived at the international level and transferred for implementation in the Global

South? Which components of the concepts need to be adapted to conditions in the Global South?

Since their inception, IWRM principles have been applicable at the level of regions and river basins (Coelho *et al*., 2012). Accordingly, several parallel concepts have emerged that attempt to address water management challenges from an urban perspective. Among these concepts is IUWM (Furlong *et al.*, 2016).

2.7.1 The IUWM principle

The principle of IUWM is mainly aimed at improving water resources management in an urban environment. The main objective is to diversify resources, achieve efficient water use, conserve water resources through sustainable coordination of all competing water sources and users, and promote stakeholder engagement and public participation in urban water systems (Closas *et al.*, 2012). Bahri (2012) viewed IUWM as a mindset rather than a method and emphasised that there is no one-size-fits-all solution, but a mix of sound water management principles adapted to local sociocultural and economic conditions. The World Bank (2016) defined IUWM as a flexible, participatory, and iterative process that integrates the elements of the urban water system (water supply, sanitation, stormwater management, and solid waste management). The World Bank's (2016) definition of IUWM also includes urban development and river basin management to maximise economic, social, and environmental benefits equitably. Koop and Van Leeuwen (2017) cited the benefits of IUWM as (i) improving environmental protection, (ii) improving the quality of life of the urban poor through the health benefits of a clean environment from improved sanitation and efficient drainage systems, and (iii) improved inclusive urban planning that delivers social, environmental, and economic benefits to the poorest. IUWM is implemented sequentially based on fundamental principles.

2.7.2 Key principles of IUWM

Key IUWM principles include recognising the value of alternative water sources, which promotes purposeful water use. Along with an approach to urban water systems, water storage, distribution, treatment, recycling, and disposal are managed as a complete cycle. IUWM aims to protect, conserve, and efficiently use surface and groundwater sources. It recognises the dependence of rural water users on the same water source as urban residents. Consequently, it aligns formal institutions with informal practices that regulate water in and for urban areas. The guiding principle of IUWM is to efficiently coordinate the relationships among water resources, land use, and energy while striving for economic efficiency, social equity, and environmental sustainability through the meaningful participation of all stakeholders (Bahri, 2012).

2.7.3 Application of IUWM

A wide application of IUWM projects took place from 2006 to 2011, when the EU funded a research programme called SWITCH to initiate a transformation in the management of urban water systems from a fragmented management approach into an integrated paradigm. The philosophy of the IUWM principles is to find sustainable solutions to urban water system challenges. System design and management should therefore be based on analysing the whole system rather than its elements. Another notable application of IUWM is the Cooperative Research Centre for Water Sensitive Cities, based at Monash University in Australia (Commonwealth Scientific and Industrial Research Organisation, 2012). The International Water Association's Cities of the Future Programme, coordinated by the University of South Florida, is an essential avenue for knowledge exchange and dissemination of information on IUWM. The OMEGA (Outil Mèthodologique de Gestion Intègrèe des Eaux Urbaines) project is a current collaboration between three French research institutes, a water supply and sanitation utility (Lyonnaise des Eaux/Suez Environment), and three French municipalities. Another place where IUWM has been implemented is Brazil, through water pollution control projects. The World Bank, through the Water Partnership Programme, has succeeded in introducing IUWM projects in Latin America, Europe, Central Asia, and Africa. Analytical studies on the potential of introducing IUWM in Africa have been conducted in sub-Saharan Africa. In Jacobsen *et al.*'s (2012) reports, in which they studied 31 African countries, Nairobi emerged as the only city where IUWM could be implemented with financial support from the World Bank.

The researcher considers four scenarios in which IUWM has been successfully implemented, drawing on the work of Jacobsen *et al.* (2012).

2.7.3.1 First scenario

The first scenario is Windhoek, Namibia, where water scarcity due to limited water resources was the driving force for implementing IUWM. This scenario introduced the principles of IUWM by recognising the value of alternative water sources and distinguishing the qualities and uses of water sources. The protection, conservation, and use of water at the source and stakeholder participation were emphasised.

2.7.3.2 Second scenario

The second scenario is Melbourne in Australia, where the main reason for introducing IUWM was climate extremes. The IUWM principle that guided this project was recognising the value of water by distinguishing the qualities and uses of water resources. In addition, water storage, distribution, treatment, reuse, disposal, protection, and conservation are considered as a single process. Water use at the source, consideration of non-urban users, and participation of all stakeholders are encouraged.

2.7.3.3 Third scenario

The main reason for implementing IUWM in Rotterdam in the Netherlands was its coastal location, which makes the city prone to flooding. This required strict management of water pollution and environmental health. Fundamental principles of the IUWM included recognising the value of alternative water sources and promoting expedient water sources. In this regard, the storage, distribution, treatment, recycling, and disposal of water are considered a single process. At the same time, it encourages participation by all stakeholders and seeks economic efficiency, social equity, and environmental sustainability.

2.7.3.4 Fourth scenario

The implementation of IUWM in Vitória in Brazil was driven by urbanisation. The core of this project was the recognition of non-urban users that depended on the same water source in the wider catchment area. The driving forces of this project were the pursuit of economic efficiency, social equity, and environmental sustainability to protect, conserve, and use surface water and groundwater. It encouraged participation by all stakeholders and recognised the relationships among water resources, water quality, and other sectors.

This study focused on the IUWM principle that requires that the components of an urban water cycle be managed in an integrated rather than fragmented manner (Fletcher *et al.*, 2007). In this way, natural water systems are mimicked in recycling resources to prevent the depletion of natural water resources. The key goal is to create a total system solution by minimising pollution generated and discharged in an urban environment by using water as close as possible to its point of origin and accurately maintaining the required water quality for its intended use (Heaney *et al.*, 1999). Stellenbosch Municipality, which serves as the case study of this research, still applies conventional technical approaches to water management that are linear and fragmented. Municipal waters are managed by utilities through centralised control and routed through separate infrastructure systems, including drinking water, wastewater, and stormwater. Stellenbosch Municipality remains solely responsible for all water issues within its jurisdiction. This practice follows a top-down management approach with command and control characterised by technocratic solutions. This approach is based on the philosophy that building larger treatment plants, dams, and reservoirs is the main solution to meeting the water needs of citizens. Accordingly, this research was born out of a desire to provide Stellenbosch Municipality with tools and mechanisms to improve the management of its urban water system by shifting from a government-centred to a stakeholder-centred approach.

2.8 SUMMARY

This chapter presented a detailed description of the town of Stellenbosch, from which the name of the case study municipality is derived, to provide context. Following an overview of South African water legislation, the chapter outlined how this has evolved over the century, which led to the passage of the National Water Act of 1998 and the Water Services Act of 1997, which now form the basis of water legislation in South Africa. As a result of the new political order, in which parliament adopted an interim constitution in 1993 and the final constitution in 1996, the Constitution (RSA, 1996) fundamentally changed South African society in relation to water and aimed to eliminate the effects of apartheid on the socioeconomic structure of the country. Although the world admires South Africa’s post-apartheid water policies and laws, this chapter highlighted the inadequacies of the legal framework and the limitations of implementing the democratically elected government's aspirations in the water sector.

The chapter provided an overview of how approaches to water governance have evolved worldwide. The government-centred approach to water governance has given way to a more inclusive, stakeholder-centred approach that is widely recognised as a more appropriate strategy to address the complex and uncertain problems associated with water resources. While the study found that the principles of stakeholder-based water governance are widely adopted in the Global North, the Global South, including South Africa, is lagging behind.

Although the current National Water Act contains provisions to implement internationally designed water governance principles, their implementation remains a challenge. The IUWM approach is the focus of this study, and the chapter showed that Stellenbosch Municipality continues to apply the government-centred approach to water management. This study therefore sought to identify mechanisms and tools to assist Stellenbosch Municipality's water authority in transitioning to stakeholder-centred water management. However, as the main objective of this study is to improve the management of the municipal water supply system, this cannot be achieved if supply does not meet demand at all times. Alternative water sources will thus be explored to improve freshwater supply in an urban setting.

The following chapter presents international perspectives on the reuse of municipal wastewater in irrigated agriculture to build confidence in the Stellenbosch Municipality water authorities. By transferring knowledge from experiences in other countries that have successfully implemented the reuse of treated municipal wastewater to improve their municipal water supply, the potential for the Stellenbosch Municipality to do the same is explored.

CHAPTER 3:
INTERNATIONAL PERSPECTIVE ON MUNICIPAL WASTEWATER REUSE FOR AGRICULTURAL PURPOSES

3.1 OVERVIEW

One of the most essential requirements for the efficient management of urban water systems is adequate water supply to the system to meet the water needs of consumers. Accordingly, various strategies for managing urban water demand and supply without allowing demand to exceed supply are continuously being studied and presented. There are two components, namely demand and supply, that need to be managed in an urban water system. This chapter focuses on the water supply management component.

Initially, conventional water managers employed supply management strategies that focused primarily on developing measures to expand water supply capacity. These measures included upgrading water infrastructure, such as water treatment plants (WTPs) and dams, or building entirely new water infrastructure. This also improved water transfers between states (Khalid, 2018). However, these strategies, which are seen as technocratic, are increasingly inadequate to meet the water needs of cities. This is due to the rapid and continuous population growth that accompanies urbanisation and increasing human economic activities. Of concern are reports that demand for freshwater will exceed supply in several parts of the world by 2030 (Ahmadalipour *et al.*, 2019). A 2015 UN report also predicted a global water deficit of 40% by 2030 (UN, 2015). In addition, the negative impacts of climate change, which irregularly alters and reduces precipitation, exacerbate the water supply deficit problem. These challenges require a shift from conventional, technocratic approaches to water management methods that are capable of addressing these challenges.

To mitigate the emerging water supply shortages, alternative water sources have become a buzzword in water supply management (Domènech *et al.*, 2013; Farooqui *et al.*, 2016; Bichai *et al.*, 2015). Alternative urban water sources that are being used on a large scale include rainwater harvesting, centralised reuse of municipal wastewater, and seawater desalination (Opher *et al.*, 2019; Ghernaout & Ibn-Elkhattab, 2020). Following the philosophy of alternative water sources are measures to reduce freshwater withdrawals, while the agricultural sector, whose freshwater withdrawals account for an average of 65% of total withdrawals worldwide, has become a subject of concern (Chen *et al.*, 2018). Accordingly, research on strategies to reduce freshwater withdrawals for agricultural purposes is highly topical. The most popular strategy is the reuse of treated municipal wastewater in irrigated agriculture, as it is considered one of the most economical strategies, although it comes with some challenges (Arena *et al.*, 2020).

Several developed countries where water scarcity threatens economic activities have made significant progress in reusing treated municipal wastewater in irrigated agriculture. This has primarily been achieved through the drafting and adoption of policies, laws, regulations, and guidelines that explicitly outline procedures and processes for the reuse of treated municipal wastewater in irrigated agriculture, which enable stakeholders to implement this practice effectively (Shoushtarian & Negahban-Azar, 2020). In contrast, in developing countries, particularly in Africa, the reuse of treated municipal wastewater in irrigated agriculture remains largely unplanned, and in some cases untreated wastewater is used. Among the documented reasons is the lack of country-specific policies, regulations, and guidelines that explicitly articulate and promote the reuse of treated municipal wastewater in irrigated agriculture (Kellis *et al.*, 2013).

This chapter reviews literature and government documents on developing policies, laws, regulations, and guidelines that address the reuse of treated municipal wastewater in irrigated agriculture. In the Global North, the State of California in the United States of America (USA) was selected as a case study, given its experience with water scarcity, the negative impacts of climate change, the uneven spatial distribution of water resources, and its pioneering role in promulgating regulations and standards for the reuse of treated municipal wastewater in irrigated agriculture in 1918 (Null & Prudencio, 2016). Spain in the EU was selected due to its asymmetric distribution of water resources and because it is ranked first among EU member states in the reuse of treated municipal wastewater in irrigated agriculture (TYPSA Consulting Engineers & Architects, 2013). In the Global South, Mexico has made significant progress in reusing municipal wastewater in irrigated agriculture in Latin America, which is why the country was selected as a case study. In Asia, China was selected because of the complex water management problems it faces as a result of the pollution of natural waters from extensive economic activities. In addition, China ranks first in the world in using untreated municipal wastewater in irrigated agriculture (Jiménez & Asano, 2008; Ungureanu *et al.*, 2020; Slobodiuk *et al.*, 2021). Egypt in North Africa is among the countries that are making progress in reusing treated municipal wastewater in irrigated agriculture (El-Zanfaly, 2015), while sub-Saharan Africa lags behind with limited data on the reuse of municipal wastewater in irrigated agriculture (Jiménez & Asano, 2008; Niquice *et al.*, 2020).

3.2 EVOLUTION OF MUNICIPAL WASTEWATER REUSE IN IRRIGATED AGRICULTURE GLOBALLY

Because the reuse of municipal wastewater in irrigated agriculture has been practised for centuries, this chapter examines the major events that contributed to its development. Table 3.1 shows the global evolution of municipal wastewater reuse in irrigated agriculture; starting from its beginnings around 3200 BCE, through to the development of guidelines for the reuse of treated municipal wastewater in irrigated agriculture by the State of California in 1918, to the first publications of guidelines for the reuse of municipal wastewater in irrigated agriculture by international organisations such as the World Health Organization (WHO), the Food and Agriculture Organization (FAO), and the International Organization for Standardization (ISO). Currently, planned and unplanned and treated and untreated municipal wastewater is used for reuse in irrigated agriculture in certain regions (Eslamian, 2016).

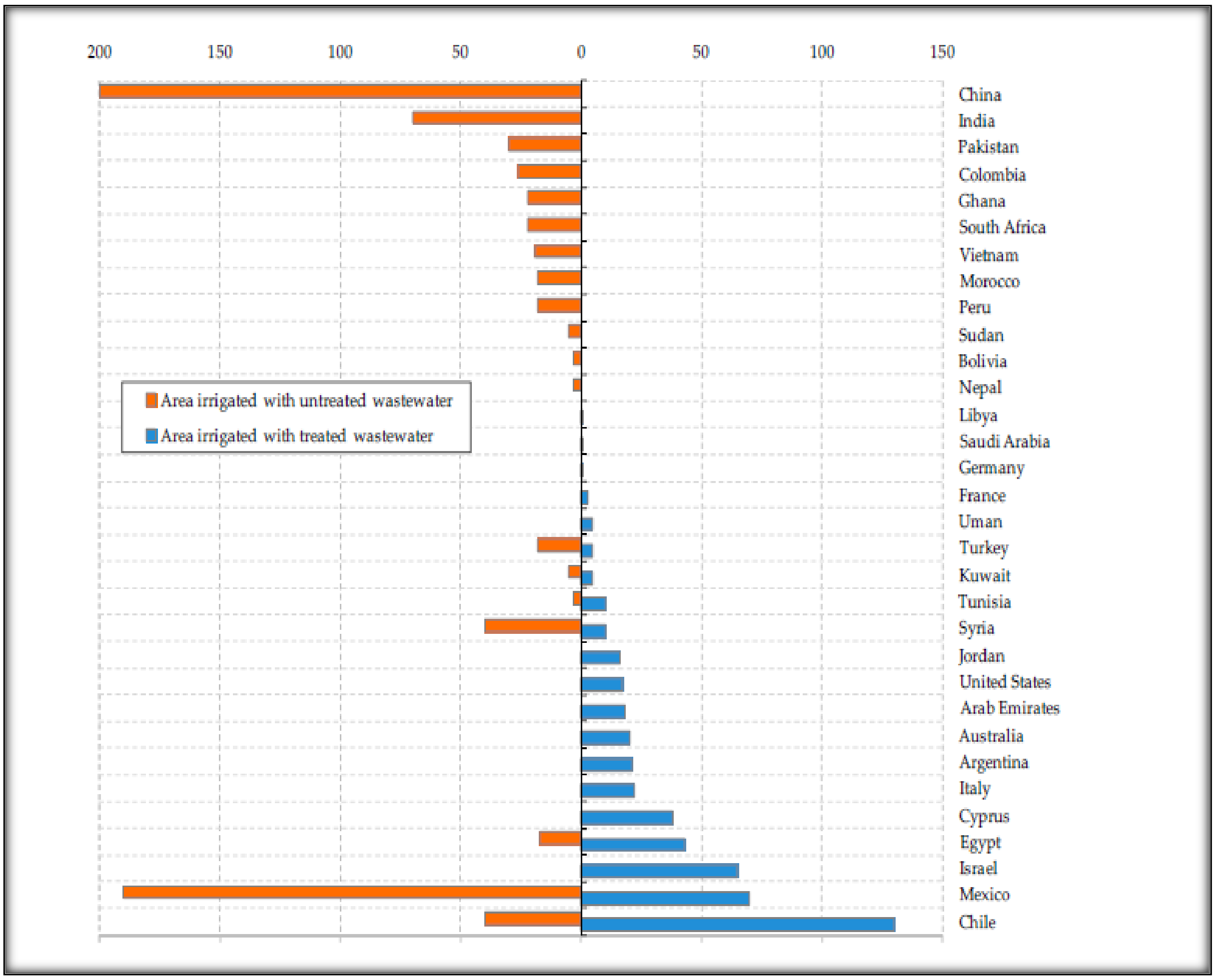
The importance of the reuse of treated municipal wastewater as a measure for sustainable management of water resources is also justified by the fact that 11% of the total global freshwater withdrawal of 3.928×1 012 m^3 in 2010 was reportedly used for municipal water needs. Of this, 3% was used for direct consumption and 8% was discharged as municipal wastewater (UN, 2017). The report also shows that a total of 2.75×106 km^2 of the global land area is irrigated, and 15% of this area could be irrigated with treated municipal wastewater. The estimated area of 5×105 km^2 irrigated with raw or diluted municipal wastewater is of great concern. Figure 3.1 shows the countries and size of areas irrigated with treated or untreated urban wastewater. Except for Egypt, many African countries currently use untreated municipal wastewater in irrigated agriculture. This justifies the need to explore ways to improve and promote the planned and treated reuse of municipal wastewater in irrigated agriculture in Africa, particularly in sub-Saharan Africa. Among the regions of the Global South, Latin America is recognised to have made significant progress, while in Africa, North Africa is making considerable progress. Figure 3.1 shows that Egypt has made the most progress in the planned reuse of treated municipal wastewater in irrigated agriculture. In contrast, several countries in Asia continue to use untreated municipal wastewater in irrigated agriculture, while limited data are available on the practice in sub-Saharan Africa. Instead, available reports indicate widespread use of unplanned, untreated municipal wastewater reuse in irrigated agriculture in the region.

Despite some limitations, history has shown that the reuse of municipal wastewater has several benefits and is an indispensable, viable alternative water source in water-scarce regions. To illustrate the importance and dynamics of this practice, the study examined freshwater sources relative to use and progress in using treated municipal wastewater in irrigated agriculture in the selected case study countries.

Table 3.1: Evolution of municipal wastewater reuse in irrigated agriculture

Period	Municipal wastewater reuse for irrigated agriculture activities	Source
3200 to 1100 BCE	- During prehistoric civilizations, domestic wastewater was deployed for agricultural purposes. - The Greek and Roman civilizations collected and conveyed domestic wastewater to the peripheries of major cities to be used for irrigation and as a source of fertiliser.	Angelakis and Gikas (2014); Tzanakakis *et al.* (2007); Cooper (2001)
1550 to 1700	- Germany, Scotland, and England employed direct municipal wastewater reuse in irrigated agriculture.	Drechsel *et al.* (2010); Tzanakakis *et al*. (2014)
1800s	- Direct municipal wastewater reuse in irrigated agriculture was widely adopted in Europe and the USA. - Direct municipal wastewater reuse in irrigated agriculture was legalised as a method for wastewater disposal in London, Paris, and Boston.	Felizatto (2001); Tzanakakis *et al.* (2007)
1872	- Paris maximised municipal wastewater reuse in irrigated agriculture. - Systematic disposal of municipal wastewater was established in Australia.	Tzanakakis *et al*. (2014)
1897	- Establishment of the first planned municipal wastewater reuse irrigated field in Melbourne.	Tzanakakis *et al*. (2014)
Early 1900s	- Catastrophic cholera and typhoid outbreak due to the disposal of untreated municipal wastewater on open fields resulted in Great Britain's Public Health Act in response to the cholera and typhoid outbreaks. - The emergence of the International Sanitary Movement; sanitary conferences organised; International Office of Public Hygiene established.	Felizatto (2001); Seguí (2004); Barona *et al*. (2008)

Period	Municipal wastewater reuse for irrigated agriculture activities	Source
Mid-1900s	- Underground sewerage systems developed in response to unhygienic conditions created by industrialisation and urbanisation. - Increased interest in indirect municipal wastewater reuse for agricultural purposes as water demand in agriculture increased, while concerns over the risk to public health and the environment over municipal wastewater reuse increased.	Angelakis and Gikas (2014); Jiménez and Asano (2008)
1973	- The WHO drafted the document “Reuse of effluents: Methods of municipal wastewater treatment and health safeguards”.	Carr (2005)
1986	- Extensive epidemiology studies were conducted, which led to the review of the 1973 WHO guidelines on municipal wastewater reuse for agricultural purposes.	Kamizoulis (2008); Mara *et al*. (2007)
1987	- The FAO developed guidelines for reusing municipal wastewater in agriculture.	Mara *et al.* (2007)
1989	- The 1986 WHO guidelines were updated to include parameters on microbiological levels of municipal wastewater reuse for irrigated agriculture.	Mara *et al.* (2007)
1992	- The Environmental Protection Agency (EPA) discovered toxicity in crops irrigated with municipal wastewater stemming from trace elements in the wastewater.	EPA (2004)
1999	- The FAO published the guidelines “Agricultural reuse of treated waters and treatment required”.	EPA (2004)
2004	- The EPA conducted extensive research on wastewater reuse for agricultural purposes.	EPA (2004)
2006	- The WHO drafted guidelines on the handling of wastewater, excreta, and greywater.	Ayers and Wescott (1985)
2000 to 2006	- Over 3 300 wastewater facilities registered worldwide within the framework of the AQUAREC international project.	Wintgens *et al.* (2006)



* Data refer to thousand hectares (ha)

Figure 3.1: Areas of wastewater reuse in agriculture by country

Source: Jaramillo and Restrepo (2017)

3.3 FRESHWATER SOURCES AND PLANNED, TREATED MUNICIPAL WASTEWATER REUSE IN IRRIGATED AGRICULTURE

There are two forms of agricultural practice: rainfed agriculture, which relies on direct rainfall, and irrigated agriculture, which draws freshwater from surface and groundwater sources. In 2016, the FAO-AQUASTAT reported that 275 million ha or 2.75×106 km^2 are irrigated worldwide. Because of their high-water intensity, irrigated agriculture is likely to be curtailed if global freshwater availability declines.

3.3.1 United States of America (USA)

Water sources in the USA are primarily surface and groundwater. Figure 3.2 shows the percentage of water sources in the USA.

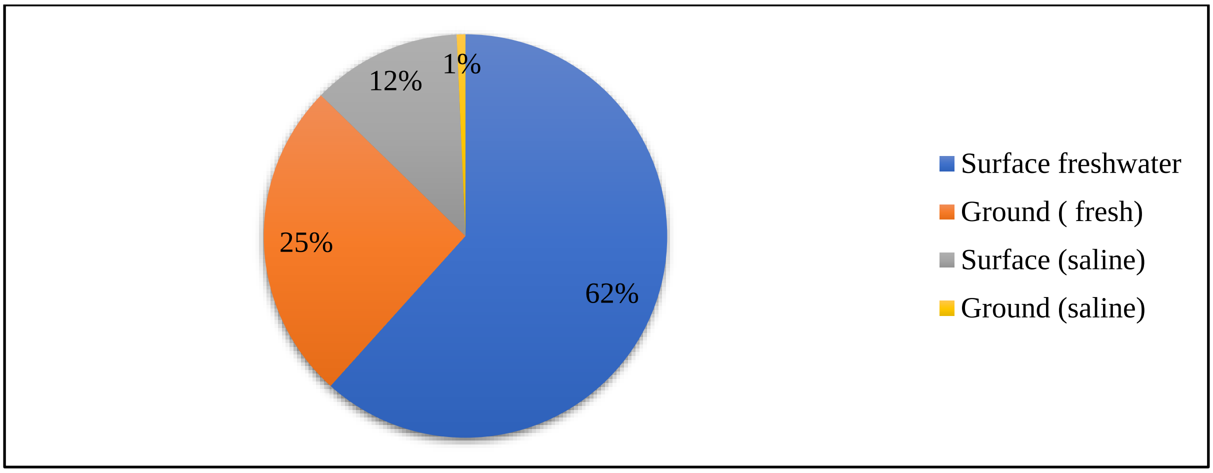


Figure 3.2: Water sources in the USA

Source: Dieter *et al.* (2018)

Dieter *et al.* (2018) reported an estimated water consumption of 322 billion gallons per day (3.79×106 m^3/day) in the USA. However, these withdrawals are the lowest since 1970 and show a continuing downward trend in water withdrawals. The year 2015 had the lowest freshwater withdrawal since and before 1970. It is worth noting that surface and groundwater sources account for 87% of total freshwater sources. From 2010 to 2015, the USA experienced a 14% decrease in freshwater withdrawals from surface water and an 8% increase in freshwater withdrawals from groundwater. These changes can be attributed to efficient water management, as evidenced by an 8% reduction in total freshwater withdrawals from thermoelectric power generation and 7% and 8% reductions in public supply and self-supply respectively. Self-supply for industry and aquaculture also experienced decreases of 9% and 16% respectively, while total water withdrawals for irrigation and mining increased slightly, by 2% and 1% respectively. To illustrate the relevance of the above freshwater withdrawal statistics to water supply practices, Figure 3.3 presents American water demand by sector.

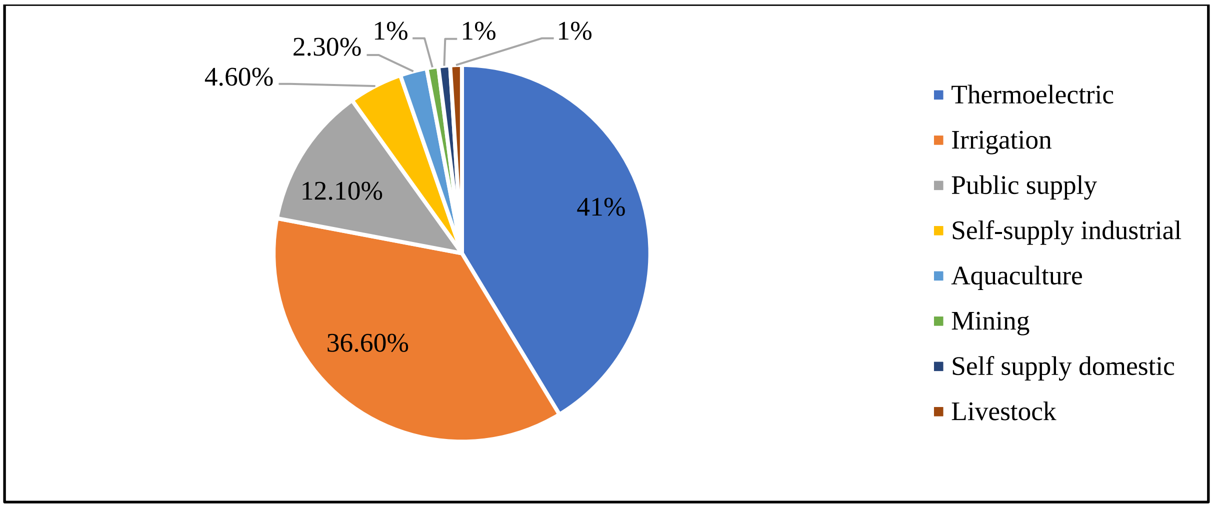


Figure 3.3: Percentage requirements per sector in the USA

Source: Dieter *et al*. (2018)

Despite the declining trend in freshwater withdrawals in some sectors in the USA, the combined factors of increasing population growth and industrialisation, coupled with the adverse effects of climate change, are of great concern, as they are in every other region. With irrigated agriculture being the second largest consumer of available freshwater resources, it is imperative to curb the growth of freshwater withdrawals for this sector, which requires innovative water security measures. Accordingly, water managers continue to work on sustainable water supply management approaches that address existing challenges. The uneven distribution of water resources across the USA requires that each state adopts and practises water management approaches that address the unique challenges of its water resources. One example is the State of California, which is why it was selected as a case study. It is worth noting that agricultural production and manufacturing in California contribute 2% of the state's gross domestic product (GDP) and employ 7.3% of the workforce (Mount & Hanak, 2019). Due to the importance of the agricultural sector in times of water scarcity, California was the first state to explore the reuse of treated municipal wastewater in irrigated agriculture to meet its water needs as early as 1918 (Ritter, 2021).

3.3.1.1 The State of California

Southern California has a high population density with limited water resources. The reuse of municipal wastewater was therefore explored in the early 20th century, which culminated in creating and publishing municipal wastewater reuse guidelines in 1918

(California State Board of Health, 1918). The primary reasons for publishing the guidelines were environmental and public health concerns, and they became a worldwide reference for guidelines on the reuse of treated municipal wastewater in irrigated agriculture. Research to improve these regulations has continued to this day, and the State of California's Department of Public Health expressed confidence in reusing treated municipal wastewater in 2014.

In California, reusing treated municipal wastewater is now an integral part of the strategic plan to meet current and future water needs. The 2015 California State Water Resources Control Board (SWRCB) and Department of Water Resources report on municipal wastewater reuse showed that a total of 714 000 acre-feet per year (rounded to the nearest 1 000 acre-feet per year) of Title 22-compliant recycled water was used for various purposes in California. Of the total recycled municipal wastewater, 220 000 acre-feet per year was used for agricultural irrigation, as shown in Figure 3.4 (Olivieri *et al*., 2020).

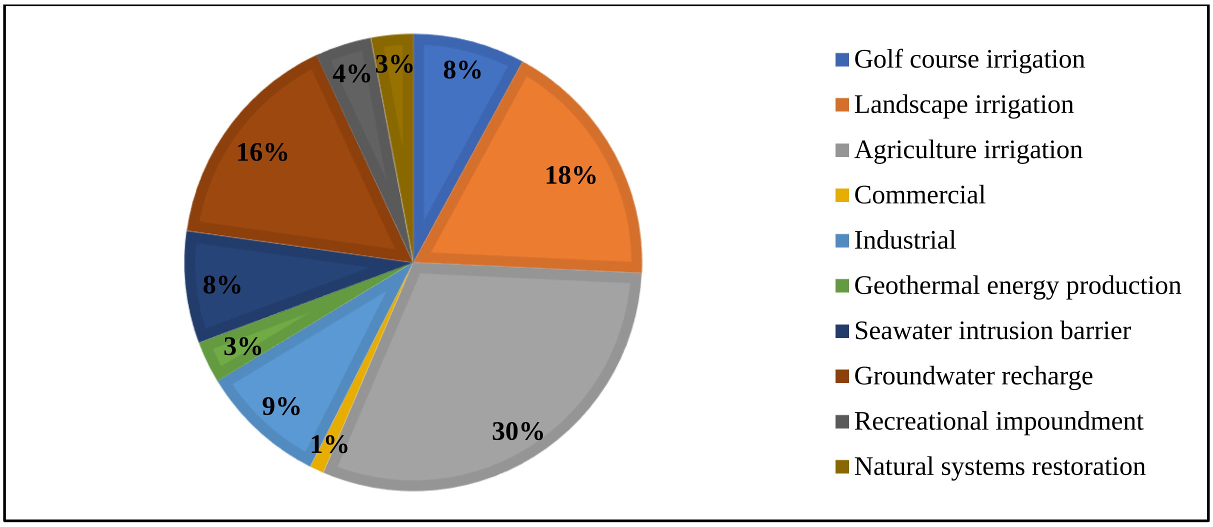


Figure 3.4: Municipal wastewater reuse by sector in the State of California in 2015

Source: Olivieri *et al*. (2020)

Between 2009 and 2015, there was generally an increase in the reuse of treated municipal wastewater for golf course irrigation and groundwater recharge but a decrease in agricultural irrigation. This was attributed to a reduction in state and federal water supplies due to the unprecedented drought in the region (Olivieri *et al*., 2020). The State of California has adopted a water recycling funding programme to promote the efficient reuse of treated municipal wastewater. This programme provides financial assistance to agencies and stakeholders involved in municipal wastewater reuse projects. It also developed well-articulated guidelines that are accompanied by continuous review and updating of regulations and standards to ensure that desired outcomes are achieved. In essence, the recycling funding programme created an enabling environment for optimal treated municipal wastewater reuse.

3.3.2 Europe

Abundant water resources characterise Europe and, for centuries, water was considered an infinite commodity until recent decades (Bixio *et al.*, 2006). However, there are predictions of impending water scarcity in Europe due to the deterioration of surface water quality, which threatens freshwater availability. Currently, several European countries, where 70% of the European population lives, are facing water problems. To provide context for water resources in Europe, Figure 3.5 shows annual freshwater withdrawals by source.

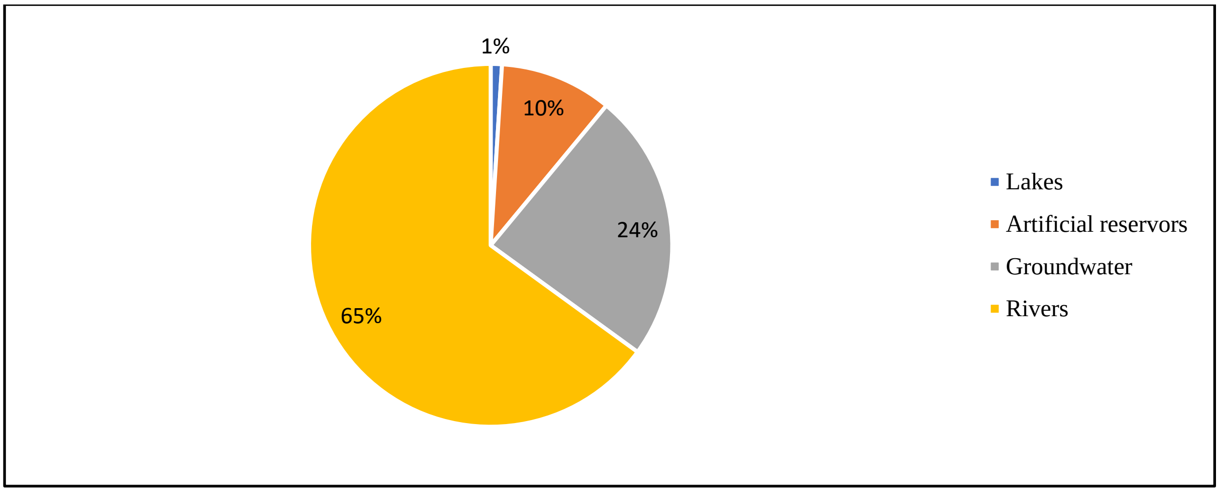


Figure 3.5: European annual freshwater abstraction by source

Source: European Environment Agency (2015)

Figure 3.5 shows that rivers account for 65% of freshwater withdrawals, which has necessitated the introduction of stringent measures to protect them. The EU has managed to protect its waters by adopting the WFD (2000/60/EC) in 2000. This is a single piece of legislation for EU member states that aims to achieve uniform management of waters in the EU to remove the complexity of fragmented water policies of member states. Previously, member states developed their water policies, standards, and regulations independently and without consultation with the EU supranational body. This resulted in incoherent policies in a region where transboundary rivers are the main source of freshwater. The objectives of the WFD (2000/60/EC) aim to protect all European waters (surface water and groundwater) and to ensure efficient water management in river basins (European Parliament and Council, 2000). Strict limits are imposed as directives on pollutant emissions from wastewater that may be discharged into natural waters. Figure 3.6 shows the annual water consumption by sector in the EU.

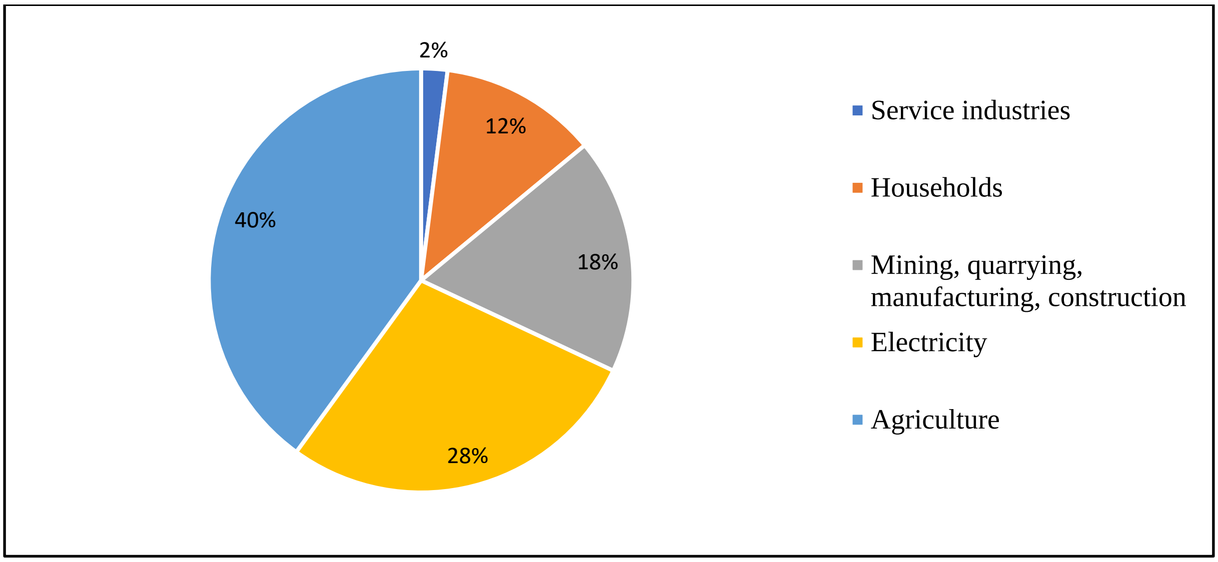


Figure 3.6: Annual water usage by sector in the EU

Source: European Environment Agency (2015)

Since the introduction of the WFD (2000/60/EC), freshwater abstraction in the EU has declined between 2000 and 2017. However, some regions, including southern Europe, which has intensive agriculture, continue to suffer from freshwater scarcity due in part to extreme drought and flood events. Despite these challenges, agriculture remains the largest consumer of freshwater resources and account for 59% of freshwater withdrawals (European Environment Agency, 2017). Reports predict serious water scarcity, which will force water agencies to consider alternative water sources (European Commission, 2007).

Regarding the above challenges, treated municipal wastewater is a viable alternative water source and has been used for centuries. It is currently embedded in European and national water management strategies. The communication “A Blueprint to safeguard Europe’s water resources” (European Commission, 2012) emphasises the maximum reuse of water. The rationale lies in the environmental, social, and economic benefits associated with the reuse of treated municipal wastewater. Quantitative environmental benefits include reductions in freshwater withdrawals and fewer wastewater discharges from municipal WWTPs. In addition, the reuse of treated municipal wastewater compares favourably with other alternative water sources in terms of cost, energy consumption, and greenhouse gas emissions (Paneque, 2015).

Another advantage of reusing treated municipal wastewater in irrigated agriculture is its reliability, as it is not affected by seasonal drought or weather fluctuations (European Commission, 2014). This is highly profitable for farmers, as interruptions in the water supply are avoided. Furthermore, the nutrients in treated wastewater are an added benefit, as less fertiliser needs to be used, which results in savings. An estimated 2.4% of treated municipal wastewater is used annually in the EU, which represents 0.5% of the EU's annual freshwater withdrawal. Several member states use treated municipal wastewater in irrigated agriculture, and Spain has made significant progress in this regard (European Commission, 2015). The following subsection discusses freshwater sources and requirements and developments in reusing treated municipal wastewater in irrigated agriculture in Spain.

3.3.2.1 Spain

Spain has a very large spatial variability of water resources distributed throughout the country. The northern basin receives 18 times more water than the south-eastern basin, which is an intensive irrigation region that serves as a tourist destination in the summer (Hernández-Mora & Del Moral, 2015). Both these activities are water intensive and require Spanish authorities to adopt innovative approaches to managing the water resources of the southern region.

Spanish water resources mirror those of the EU and require efficient protection of surface water and groundwater to ensure freshwater security. Figure 3.7 shows the percentage of water demand by sector.

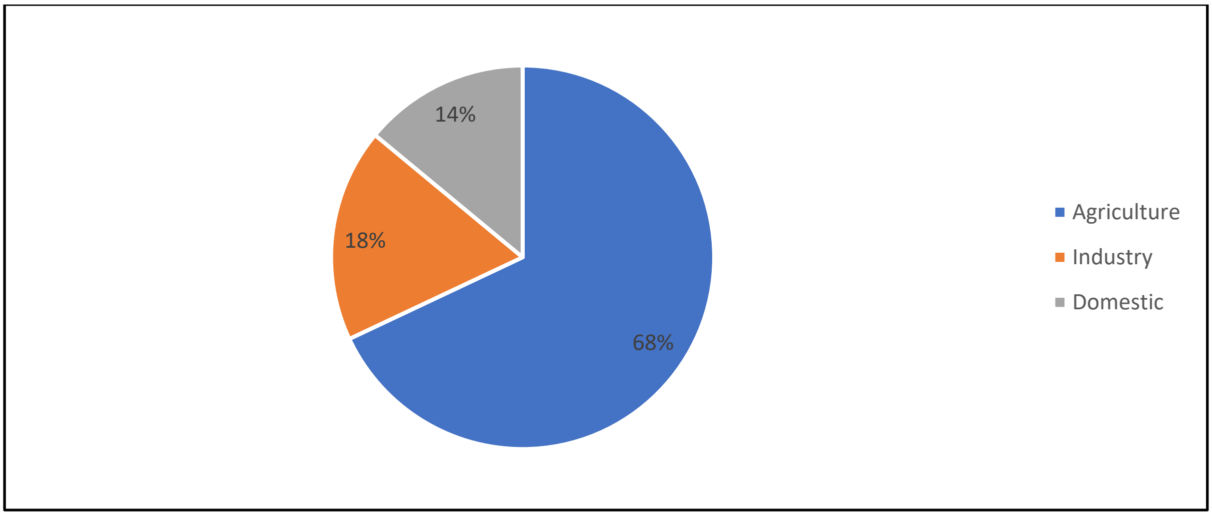


Figure 3.7: Spanish water requirements by sector

Source: FAO-AQUASTAT (2016)

As in other regions, agriculture is the largest consumer of freshwater, which accounts for 68%. Future water shortages therefore pose a high risk to the agricultural sector. To reduce freshwater withdrawals by agriculture while protecting the agricultural industry, which contributes 2.6% of the GDP and employs 4.091% of the labour force (Trading Economics, 2020; The Global Economy.com, 2019), new water management strategies are being developed. These include exploring alternative water sources, including seawater desalination and reuse of treated municipal wastewater. Since the introduction of the WFD (2000/60/EC), Spain, like other EU regions, has placed a higher value on water and achieved a decrease in consumption. However, climate change poses a major threat to the availability of water resources in southern Spain, which requires a continuous revision of national policies in line with the WFD, which promotes optimal water use in agriculture and the development of alternative water sources, such as the reuse of treated municipal wastewater. Currently, large amounts of treated municipal wastewater are used in the south-eastern region of Spain, with 75.8% used for irrigated agriculture in the Júcar and Segura river basins. In general, treated municipal wastewater covers 5.4% of water demand in Spain, with some regions such as the Canary Islands reaching 25% (Ricart & Rico, 2019).

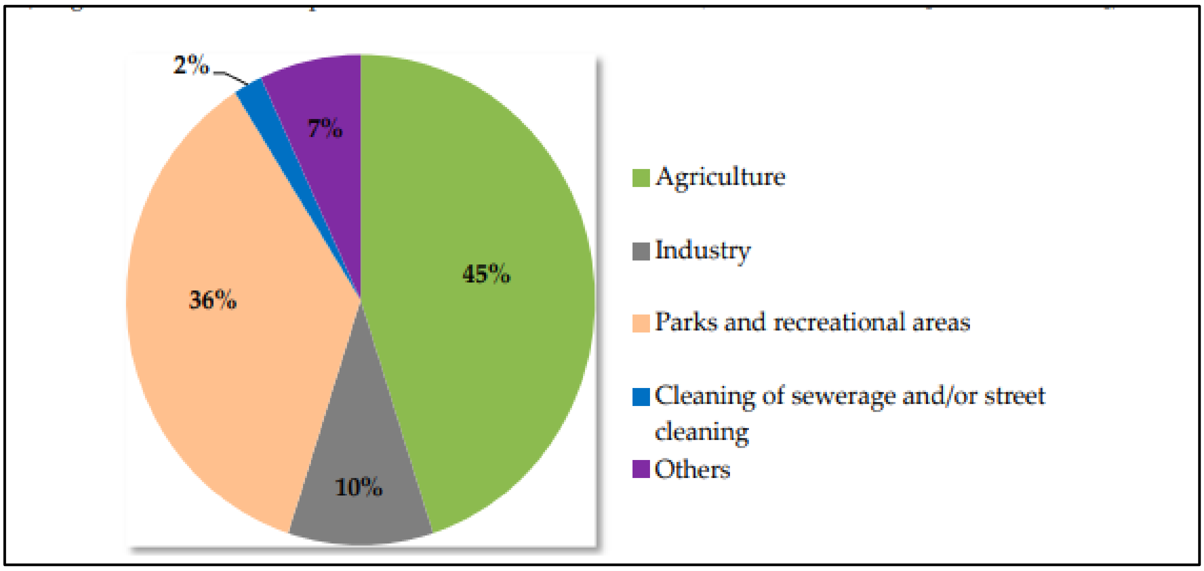


Figure 3.8: Uses of reclaimed water in Spain (%) considering a total volume of 268 hm³ per year

Source: Melgarejo-Moreno (2019)

Figure 3.8 shows that agriculture is the largest consumer of reclaimed water, followed by the irrigation of parks and recreation areas. Spain is a testament to the importance of wastewater reuse in meeting water needs by reducing freshwater withdrawals, which should be emulated by other regions of the world.

3.3.3 Mexico

As in most of Latin America, water resources in Mexico are asymmetrically distributed. An estimated two-thirds of Mexico's population lives in the northern and central regions, which are arid or semi-arid (Freedman *et al.*, 2015). Consequently, meeting the water needs of these areas with limited water resources in the face of increasing population and economic growth is a significant challenge. To ensure sustainable development, which includes meeting the water needs of the agricultural sector, prudent water management is central to the central government's development strategies. The importance of the agricultural sector is underscored by the 2019 statistics, which indicate its contribution to Mexico's GDP as 3.47% (Plecher, 2020b) and it employs 12.6% of the labour force (Plecher, 2020a). Figures 3.9 and 3.10 illustrate Mexico's water sources and demand by sector.

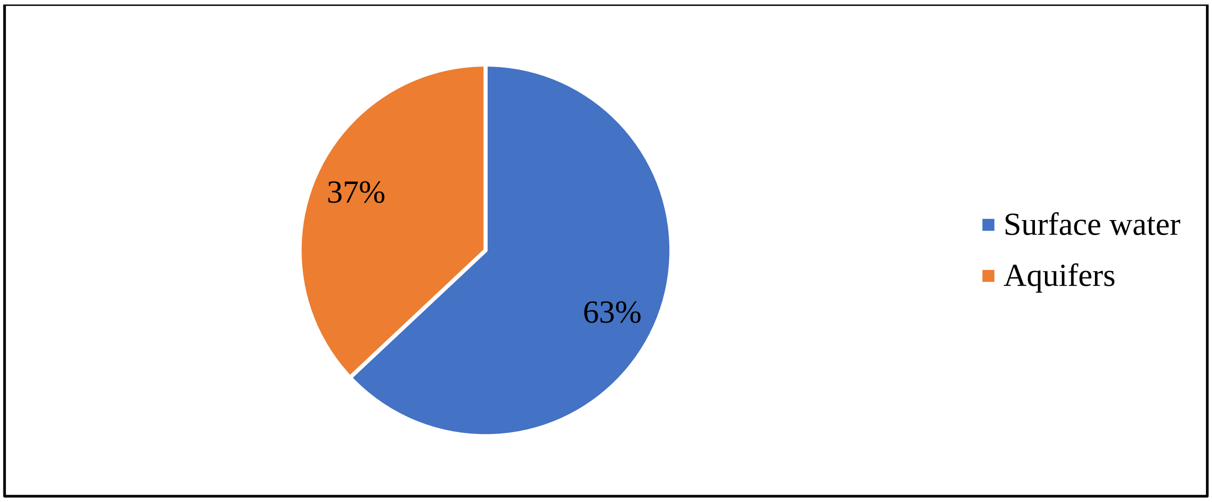


Figure 3.9: Mexican water sources

Source: Almendarez-Hernández *et al*. (2016)

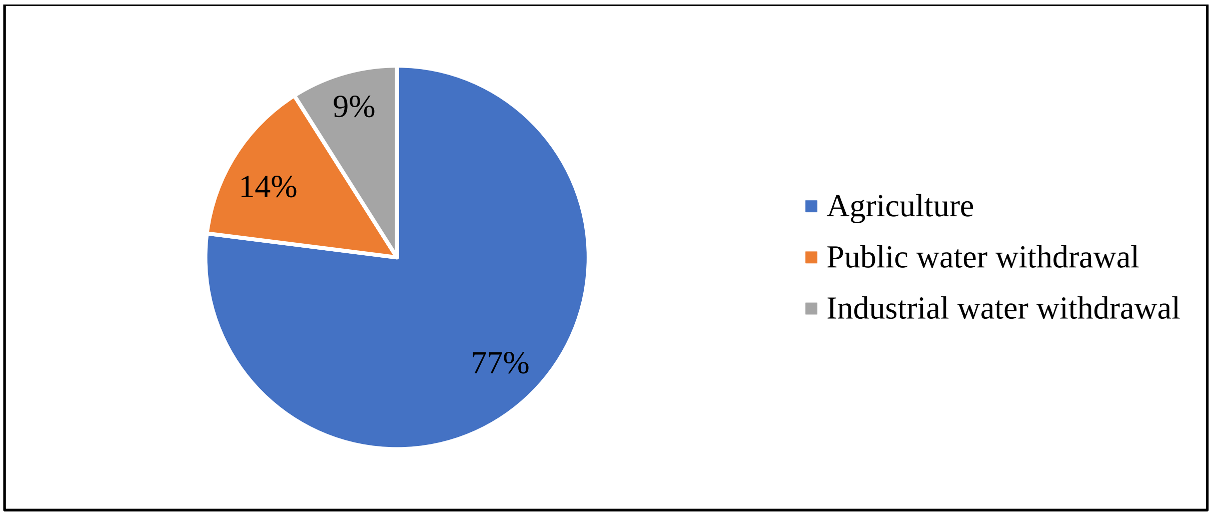


Figure 3.10: Mexican water requirements by sector

Source: Almendarez-Hernández *et al*. (2016)

Figure 3.10 shows that agriculture is the main consumer of freshwater and accounts for a high 77% of total water demand. During high rainfall seasons, 45% of cultivated land in Mexico was rainfed. However, over the past decade, rainfall has declined by 30 mm per decade in southern and south-eastern Mexico (Munroe *et al.*, 2014). Currently, predictions indicate impending droughts around the Gulf of Mexico, triggered by climate change, which will place significant stress on the management of irrigation systems (National Institute of Statistics and Geography, 2013). Considering that the primary water source is surface water (see Figure 3.9), measures to mitigate the negative impacts of water deficits are being explored. These include farming methods that explore postponing the start of harvesting, selecting new seed varieties, and optimising water use (Santillán-Fernández *et al.*, 2021).

One of the strategies that the Mexican government prioritises to optimise water use and to avoid affecting crop production is the reuse of treated municipal wastewater in irrigated agriculture. To implement this strategy, the Mexican government approved the National Water Programme 2007-2012 (Comisión Nacional del Agua [CONAGUA], 2007). Then, in April 2014, another National Water Programme was launched for 2014-2018 (CONAGUA, 2018). The main objective of these programmes is to strengthen integrated and sustainable water management by focusing on the reuse of treated municipal wastewater and the treatment of municipal wastewater to fit-for-purpose standards (CONAGUA, 2018). During the implementation and evaluation of these programmes, it was reported that the reuse of treated municipal wastewater was widespread throughout the agricultural sector (Mexico Now, 2013). It is worth noting that the government has allocated significant resources to municipal wastewater treatment infrastructure in Mexico to realise these positive developments. Between 2007 and 2011, a 132% increase in investment in wastewater treatment was reported (CONAGUA, 2010). In 2012, over 90% of the population was connected to the sewage network (Global Water Intelligence, 2012). According to Jiménez (2006), Mexico ranks first in the world in terms of the reuse of treated wastewater in irrigated agriculture, with a consumption of 4 492 800 m^3/d. Despite some persistent challenges, Mexico remains a pacesetter in the reuse of treated municipal wastewater in irrigated agriculture in Latin America (Gilabert-Alarcón *et al.*, 2018).

3.3.4 China

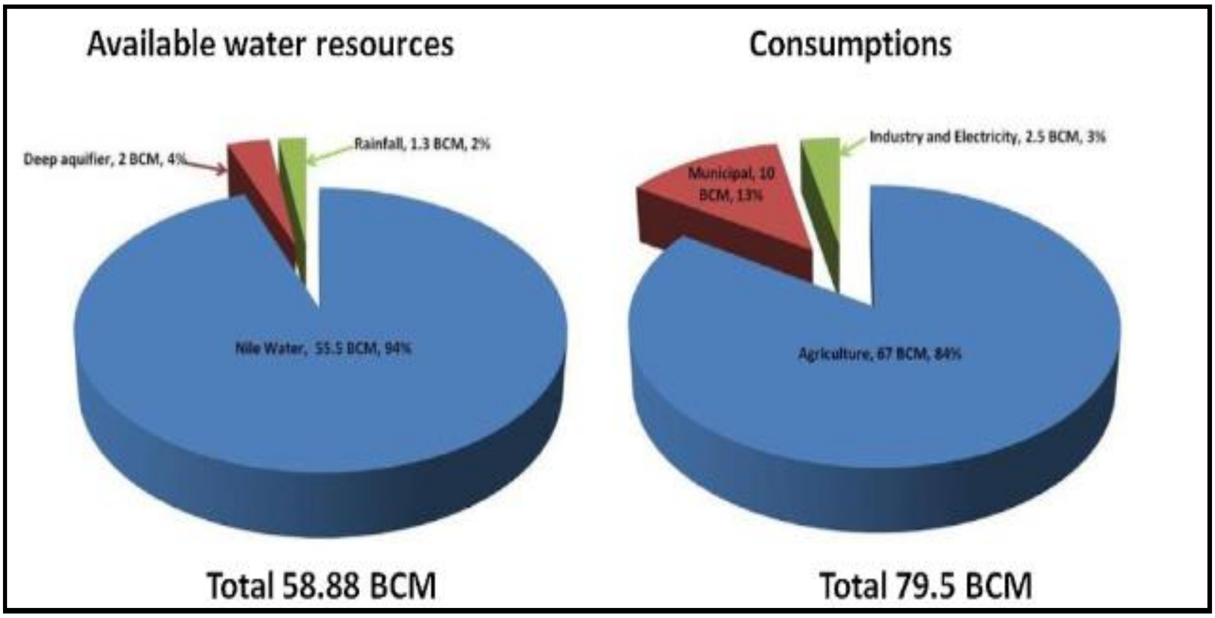
Over the past three decades, China has experienced rapid economic growth that has dramatically changed the country's socioeconomic landscape (Liu & Speed, 2009). However, the flipside of these development gains is the detrimental impact on water resources, as natural water bodies are highly polluted. It is estimated that one-third of lakes and rivers are so polluted that the water can no longer be used for human consumption (Gleick, 2009). Turner (2006) reported on the effects of pollution of natural waters on the ecosystem, citing the drying up of water bodies and the extinction of aquatic life. China is considered the country with the most polluted waters in the world, which exacerbates the existing disparity between population size and available water resources. China is home to 20% of the world's population, yet only 7% of the world's freshwater resources are in China (Shao *et al.*, 2009; Udimal *et al.*, 2017).

Bandyopadhyay (2017) lamented the pollution of natural water bodies and its impact on the agricultural sector and reported how local farmers in some regions resort to using contaminated water for their agricultural activities. He *et al.* (2014) reported that per capita water resources in China are decreasing yearly, which threatens economic activities and the ecological environment. This poses complex challenges for Chinese water managers to sustainably manage their water resources to adequately meet the country's current and future water needs.

Since surface water and groundwater are the main sources of water in China, the pollution of natural waters has significant negative impacts on freshwater availability. As a result, groundwater supplies are declining due to over-exploitation as rivers are heavily polluted. According to Zhang *et al.* (2022), lakes and reservoirs are the most important sources of freshwater supply in China. They supply 40.6% of China's water needs, followed by rivers and groundwater at 30.8% and 28.6% respectively. Similar to most regions of the world, agricultural water use in China is estimated to account for over 70% of the total water supply (Cai *et al.*, 2020). Agricultural activities include irrigation of agricultural fields, forestry, orchards, and grasslands; replenishment of fisheries; and water for livestock. While agriculture has the disadvantage of high-water intensity, it has the advantage of contributing 7.1% of China's GDP and employing an estimated 25.1% of China's labour force (Textor, 2020a; 2020b). Accordingly, water supply security is critical for agriculture to thrive, which necessitates the protection of surface water, the optimisation of water use, and the search for alternative water sources (Udimal *et al.*, 2017). The Ministry of Housing and Rural Development and the National Development and Reform Commission have highly recommended the reuse of treated municipal wastewater as an alternative water source. These two ministries predicted that wastewater reuse would reach 30% in some parts of China by 2020 if it is well managed. Currently, industry and landscaping are the main consumers of treated wastewater, with percentage consumption varying from province to province (Zhu & Dou, 2018). It is reported that an estimated 10% to 29% of treated wastewater is used in agriculture, and this percentage also varies from province to province (Hashem & Qi, 2021).

3.3.5 Egypt

Egypt is an arid country with an estimated area of 1 000 000 km^2 and has experienced continuous rapid population growth over the past 50 years from 19 million inhabitants in 1949 to 102.3 million in 2022 (O'Neill, 2022). Abdel-Lateef *et al.* (2011) forecasted that Egypt's population will reach the 100 million mark by 2025. This exponential population growth poses significant challenges for Egyptian authorities in managing water resources. Since the Nile River is the main source of water and Egypt receives a fixed share of 55.5×109 m^3 of water annually, 95% of Egypt's population lives on the banks of the Nile Valley and Delta, which account for only 4% of Egypt's land area (Rassoul, 2006). Figure 3.11 shows Egypt's sectoral water demand. Rainfall is very sparse as Egypt receives at most 200 mm of rain annually. Most of this rain is concentrated in the northern parts of the country (Elmenoufy *et al.*, 2017).



*BCM = Billion cubic metres

Figure 3.11: Egyptian water sources and requirements by sector

Source: Abukila (2015)

To meet current and future water needs, Egyptian water authorities are developing several strategies. Meeting agricultural water needs, which account for over 80% of the total water supply, is at the core of their strategy. Since the agricultural sector contributes 11.05% of Egypt's GDP (Plecher, 2020c) and employs 23.79% of Egypt's labour force (Plecher, 2020d), it is essential to ensure adequate water supply to the sector. To safeguard the water-intensive agricultural sector, Egyptian water authorities are exploring alternative water sources (Jussah *et al.*, 2020). In 2014, the Ministry of Water Resources and Irrigation stated that reusing municipal wastewater as an alternative water source was a viable option to augment Egypt's water supply. However, as in most developing countries, wastewater infrastructure in Egypt is underdeveloped. In order to develop municipal wastewater as an alternative water resource, Egyptian authorities have made significant investments in developing wastewater infrastructure. Abdel-Kader and Abdel-Rassoul (2010) reported that the wastewater network was expanded by 6 000 km between 2005 and 2010. Although these achievements are notable, the goal of connecting all citizens to the sewerage system is still far off. The government has recognised that it is practically impossible to connect all citizens to the municipal sewerage network for various reasons, including the availability of financial resources. Water and sanitation authorities are opting for decentralised municipal WWTPs (Nasr *et al.*, 2022). Elbana *et al.* (2017) estimated the volume of municipal wastewater generated annually at 5.5 to 6.5 billion m³, with 3.7 billion m³ being treated. A volume of 0.7 billion m³ of treated municipal wastewater is reused in irrigated agriculture (Abdel-Shafy & Mansour, 2013).

Table 3.2: Summary of water sources, needs, and wastewater reuse in California, Spain, Mexico, and Egypt

Region	Major water sources	%	Major water requirements	%	Treated municipal wastewater reuse	%
California	Surface freshwater	62	Thermoelectric	41	Agriculture irrigation	30
	Ground freshwater	25	Irrigation	36.6	Landscape Irrigation	18
	Surface saline	12	Public supply	12.2	Groundwater recharge	16
Spain	Surface freshwater	65	Agriculture	40	Agriculture irrigation	45
	Ground water	24	Electricity	28	Parks and recreational areas	36
	Artificial reservoirs	10	Households	12	Industry	10
Mexico	Surface water	63	Agriculture	77	Total	86
	Aquifers	37	Public supply	14		
			Industrial supplies	9		
China	Lakes and reservoirs	40.6	Agriculture	60	Landscape irrigation	60
	Rivers	30.8	Industry	24	Industrial usage	30

Region	Major water sources	%	Major water requirements	%	Treated municipal wastewater reuse	%
	Groundwater	28.6	Public supply	12	Agriculture irrigation	10-29
Egypt	Nile River	94	Agriculture	84	Agriculture irrigation	10
	Deep aquifer	4	Municipal supply	13		
	Rainfall	2	Industry and electricity	3		

Table 3.2 shows that the reuse of treated municipal wastewater is a viable alternative water source. Regions such as California and Spain are systematically and progressively implementing municipal wastewater reuse and have demonstrated its positive impact on reducing freshwater withdrawals. South Africa, which currently makes limited use of treated municipal wastewater reuse, can adopt some of the principles followed by California and Spain to achieve effective and efficient reuse of treated municipal wastewater. Best practices include having a well-developed strategy with goals and timelines to achieve them and providing sufficient funding for treated municipal wastewater reuse projects. Following the examples of the case study regions, different uses for treated municipal wastewater should be considered. There are some uses that require minimal treatment of municipal wastewater that South Africa can focus on initially. These include irrigation of parks and recreation areas and landscape irrigation before considering agricultural activities that are more sensitive to pollutants in the water and may require highly engineered treatment of wastewater.

3.4 INTERNATIONAL GUIDELINES ON MUNICIPAL WASTEWATER REUSE IN IRRIGATED AGRICULTURE

Research points to gaps in the uniformity of policy development and the formulation of regulations and guidelines that create an enabling environment for the widespread reuse of municipal wastewater in several regions of the Global South (Hanjra *et al.*, 2012). The lack of universal guidelines and standards has been pointed out to significantly weaken stakeholder confidence in the adoption of wastewater reuse around the world (Qadir *et al.*, 2010). However, non-binding guidelines have been published by international organisations such as the WHO, the FAO, and the ISO. These guidelines can be useful for the Global South, where the reuse of treated municipal wastewater in irrigated agriculture is still in its infancy and appropriate guidelines have not yet been developed.

The first international organisation to publish guidelines on the reuse of municipal wastewater in irrigated agriculture was the WHO in 1973. The document was titled "Reuse of effluents: Methods of wastewater treatment and health safeguards". Its main objectives were to protect public health and to ensure the safe application of wastewater reuse and excreta handling in agriculture. However, the document failed to achieve these goals because it did not include explicit preventive measures for the public health risks associated with agricultural municipal wastewater reuse and was not supported by epidemiological studies. After extensive epidemiological studies, the 1973 WHO guidelines were updated in 1989 in a document titled "Health guidelines for the use of wastewater in agriculture and aquaculture". This document focused on microbiological limits permissible in irrigated agriculture and emphasised protecting public health and the environment (WHO, 1989). The latest guidance document published in 2006 by the WHO, titled "Guidelines for the safe use of wastewater excreta and greywater", is well informed by extensive research. Public health issues are explicitly addressed through health risk assessment, health-related objectives, and health protection measures. Surveillance and system assessment measures are formulated, and social, cultural, financial, and environmental issues are considered (WHO, 2006). The WHO guidelines highlight human parasites as a major risk factor and identify their elimination as a priority. The FAO followed the WHO and published guidelines in 1987, which were updated in 1999, which focus on quality standards for effluents for various uses (Shoushtarian & Negahban-Azar, 2020). The limits of trace elements allowed in the irrigation of certain crops are described. Regarding microbial requirements, the guidelines are less restrictive for reusing municipal wastewater for unrestricted irrigation, while more stringent water quality values apply to the irrigation of fruit trees. It is important to note that the physicochemical parameters of the FAO guidelines have been incorporated into the standards, criteria, guidelines, and regulations of several organisations and government agencies (FAO, 1999).

There are several other internationally recognised guidelines, including the ISO standard for Treated Wastewater Use for Irrigation Projects, ISO/TC 282, WHO guidelines (2006), the Australian Guidelines for Water Recycling (2006), the Israeli Agricultural Irrigation Regulations (1978, 1999, and 2005), and the California Code of Regulations (Title 22, Division 4, Chapter 3, Water Recycling Criteria, 2000) (Shoushtarian & Negahban-Azar, 2020). The aforementioned guidelines formed the basis for the creation of the ISO standard 16075-2:2015. In 2015, the ISO 16075 series on guidelines for the use of treated municipal wastewater in irrigated agriculture was published.

3.5 DEVELOPMENT OF POLICIES, REGULATIONS, AND GUIDELINES FOR MUNICIPAL WASTEWATER REUSE IN IRRIGATED AGRICULTURE

Although the reuse of treated municipal wastewater is gaining momentum worldwide, its widespread application is hampered by a lack of mandatory universal policies, regulations, and guidelines. As a result, several countries have developed their own country-specific policies, regulations, and guidelines that prioritise the protection of public health and the environment. Geographic, economic, and social circumstances are critical to the development of these policies, regulations, and guidelines. Accordingly, there are also differences in the permissible limits for microbial and physicochemical parameters (Brissaud, 2008). In this regard, developed countries have many years of experience in developing their regulations and guidelines.

Although the development of water reuse regulations and standards in the USA is the responsibility of the states, the EPA has also developed comprehensive water reuse guidelines that work in concert with the guidelines formulated by the states and all agencies involved in water reuse projects to mitigate any inconsistency between federal and state guidelines (EPA, 2012).

3.5.1 State of California

The State of California is known worldwide for being the first to promulgate regulations and guidelines for the reuse of treated municipal wastewater, in 1918. These regulations are explicit and comprehensive and set strict limits on wastewater reuse that include parametric limits on the irrigation of certain crops and the use of specific irrigation techniques. While many states in the USA were considering what to do with effluent from their WWTPs after congress passed the Clean Water Act in 1972, the EPA was required to set minimum standards for effluent from these plants. California was well ahead of the curve with its water recycling projects. To institutionalise and strengthen the reuse of treated municipal wastewater, the California State Legislature enacted the Waste Water Reuse Law in 1974 (Brown & Weinstock, 1980). From the 1918 guidelines to the water quality standards and treatment reliability criteria contained in the California Department of Public Health's (CDPH) Regulations Related to Recycled Water (Title 22, Division 4, Chapter 3 of the California Code of Regulations), California boasts more than a century of safe use of treated municipal water for food crop irrigation. These standards and guidelines have evolved dynamically with improved wastewater treatment technologies, increasing knowledge of pathogen behaviour and its impact on human health, as well as changes in agricultural and irrigation practices. A recent review of these CDPH criteria for water recycling by the National Water Research Institute (2013) provided data on annual recycled water from 1989, as shown in Figure 3.12. The three largest users of recycled water are agriculture (37%), landscape irrigation (17%), and groundwater recharge and barrier to seawater intrusion (19%). With the federal Clean Water Act and the state Waste Water Reuse Law, coupled with Title 22 of the California Code of Regulations, extensive wastewater reclamation projects have been implemented. These projects were funded by enormous state and federal grants and included farms with large areas irrigated with treated wastewater.

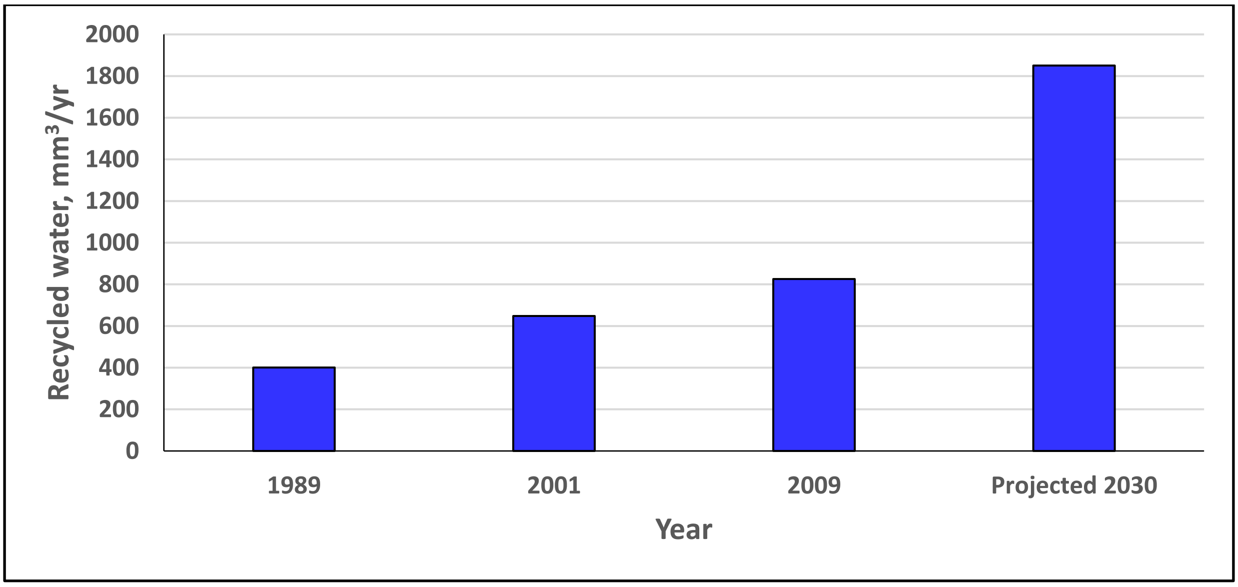


Figure 3.12: Treated wastewater in the State of California

Source: National Water Research Institute (2013)

The role of institutions in successfully implementing wastewater reuse in irrigation in the case of California cannot be understated. The CDPH, the SWRCB, and nine Regional Water Quality Control Boards are involved in reusing treated municipal wastewater. While the state and regional WBs oversee the environmental health of the state's waters, the SWRCB manages water rights. The CDPH is charged with establishing public health criteria for wastewater reclamation, including groundwater recharge, and reviewing all proposals for such projects in the state. A Memorandum of Understanding between these agencies ensures cooperation in implementing successful projects. Although champions are required for successful agricultural projects that use treated municipal wastewater for irrigation, these mandated institutions' overarching policies, regulations, and guidelines ensure success from one project to another. This approach sets a precedent for countries in the Global South when it comes to successfully implementing such agricultural projects.

3.5.2 European Union (EU)

The potential for reusing treated municipal wastewater in the EU is increasingly recognised; to the extent that it had to be enshrined in the EU WFD. The EU's recognition of the importance of reusing treated municipal wastewater was reflected in the 2012 European Innovation Partnership for Water, which supports innovative solutions to water challenges, and the European Commission's (2012) report, which sets out a plan to protect Europe's water resources. The WFD promoted the creation of legal frameworks between member states to protect public health, the environment, and natural waters within their jurisdiction. Spain, one of the EU member states, is considered a pioneer in reusing treated municipal wastewater (TYPSA Consulting Engineers & Architects, 2013). For this reason, the development of the Spanish legal framework for the reuse of treated municipal wastewater in irrigated agriculture was studied.

3.5.2.1 Spain

Using treated municipal wastewater in agriculture began in Spain in 1970 at a WWTP in Las Palmas (Jódar-Abellán *et al.*, 2019). This practice was extended to other cities and regions before the Water Law came into force in 1985 and Spain joined the European Communities in 1986. These years, 1985 and 1986, were particularly significant for the reuse of wastewater in agriculture. The 1985 Water Law stated: "The government shall establish the basic conditions for the reuse of water based on the purification process, its quality, and its intended uses" (Article 101) and served as the basis for regulations and guidelines for the reuse of wastewater. With Spain's accession to the European Communities in 1986, its regulations and guidelines had to be amended later on to align with EU environmental directives included in the WFD and other directives for habitats, birds, the sea, and floods. The strategy to bring Spain into compliance with EU directives required a permit for effluents from WWTPs that ensured that mitigation measures were in place, coupled with penalties for non-compliance with the directives. Between 1986 and 2007, various laws were enacted and repealed. The culmination was Royal Decree 1620/2007, which stated that "the government shall establish the basic conditions for the reuse of water and determine the required quality of treated wastewater based on the expected uses". This is the current legislation that regulates the reuse of wastewater for agricultural production. It contains permissible microbiological and physicochemical parameters of treated wastewater used for crop irrigation. The parameters in question are classified according to the end users, e.g., consumed raw or not consumed raw. Other end users include those that may undergo industrial processes and pasture for dairy- or meat-producing animals. For tree crops, the following issues are considered: Does the treated effluent come into contact with fruit that humans may consume? For other crops such as ornamental flowers, nurseries and greenhouses, silage, grains, and oilseeds, special criteria apply. In essence, the regulations and guidelines are comprehensive and consistent with those of the State of California. The total amount of treated wastewater reused in Spain varies significantly depending on the data source; ranging from 370 million m^3 per year to 500 million m^3 per year (Jódar-Abellán *et al.,* 2019). The distribution of the use is shown in Figure 3.13. The fact that agriculture is the largest user of treated wastewater highlights its importance to the Spanish economy.

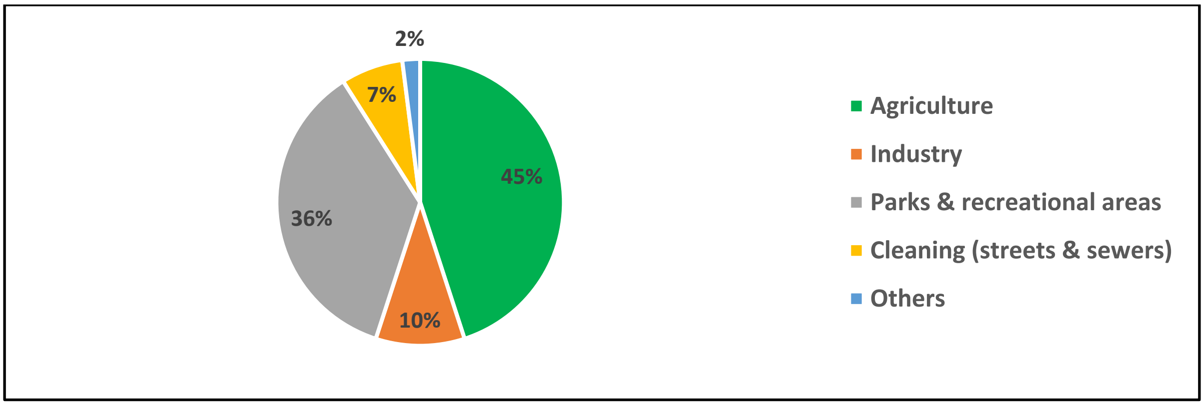


Figure 3.13: Users of treated wastewater in Spain

Source: Jódar-Abellán *et al*. (2019)

As noted earlier, legislation does not exist in a vacuum. Laws require institutions and adequate human capacity to translate them into successful municipal wastewater reuse projects. In the case of Spain, the Spanish Ministry of Agriculture, Food and Fisheries, together with the Ministry of Health, issued Royal Decree 1620/2007 as the legal framework. Project proposals for the use of treated municipal wastewater in agriculture must be approved by health authorities to ensure that they comply with the provisions of the decree in terms of technical and water quality aspects, and that self-monitoring and risk management programmes are in place (Navarro-Caballero, 2018).

3.5.3 Mexico

In Latin America, Mexico has made significant progress in reusing treated municipal wastewater in irrigated agriculture. Mexico’s success is partly due to its policy process and detailed regulations and standards, as shown in Table 3.2. The policy process began with the passage of the Water Law of 1871, which focused on the prevention and control of water pollution. This was followed by the publication and revision of various standards and regulations from 1991 until NOM-001-ECOL-1996 was issued. Table 3.3 lists the updated Mexican standards.

Table 3.3: Mexican recommended revised microbiological guidelines for treated wastewater reuse in agriculture (NOM-001-ECOL-1996)

Category	Reuse conditions	Exposed group	Irrigation technique	Intestinal nematodes (arithmetic mean no. of eggs per litre)	Faecal coliforms (geometric mean no. per 100 ml)	Wastewater treatment is expected to achieve the required microbiological quality
A	Unrestricted irrigation. A1 vegetable and salad crops eaten uncooked, sports fields, and public parks	Workers, consumers, and the public	Any	≤0.1	$\leq 10^3$	Well-designed series of waste stabilisation ponds, sequential batch-fed wastewater storage and treatment reservoirs, or equivalent treatment (e.g., conventional secondary treatment supplemented by either polishing ponds or filtration and disinfection)
B	Restricted irrigation. Cereal crops, industrial crops, fodder crops, pasture, and trees	B1 workers (but no children < 15 years) and nearby communities	(a) Spray/ sprinkler	≤1	$\leq 10^5$	Retention in waste stabilisation pond series including one maturation pond or in sequential wastewater storage and treatment reservoirs or equivalent treatment (e.g., conventional secondary treatment supplemented by either polishing ponds or filtration)
		B2 As B1	(b) Flood/furrow	≤1	$\leq 10^3$	As for Category A
		B3 workers, including children <15 years and nearby communities	Any	≤0.1	$\leq 10^3$	As for Category A
C	Localised irrigation of crops in Category B if exposure of workers and public does not occur	None	Trickle, drip, or bubbler	Not applicable	Not applicable	Pre-treatment as required by irrigation technology, but less than primary sedimentation

Source: Peasey *et al.* (2000)

The Mexican microbiological guidelines emphasise faecal coliform levels for determining pathogenic contamination, and the proposed changes to the guidelines, shown in Table 3.4, reflect the WHO guidelines. The NOM-001-ECOL-1996 guidelines establish microbiological levels for municipal wastewater reuse in unrestricted, restricted, and localised irrigation, considering the health hazards that may result from public exposure to pathogens during irrigation.

Table 3.4: Proposed changes to Mexican Standard NOM-001-ECOL-1996

Irrigation	**Mexican standards**		**Proposed standards for Mexico**		**WHO guidelines**	
	FC/100 ml	Ova/litre	FC/100 ml	Ova/litre	FC/100 ml	Ova/litre
Restricted	≤ 10^3	≤ 5	≤ 10^3 – 10^4	≤ 0.1 -1.0	Not required	≤ 1
Unrestricted	≤ 10^3	≤ 1	≤ 10^3	≤ 0.1 – 1.0	≤ 10^3	≤ 1

Source: Peasey *et al.* (2000)

<u>Note:</u> Where there is a range of standards, the level of acceptable health risk will determine the standard adopted.

3.5.4 China

Although China ranks first in the world in the reuse of untreated municipal wastewater in irrigated agriculture, the Chinese government has promoted the reuse of treated municipal wastewater since 1958. This was promoted by including it in the national key scientific and technological projects of the Seventh, Eighth, and Ninth Five-Year Plans. At the beginning of these projects, the major drawback was that there was no infrastructure for collecting and treating municipal wastewater, which resulted in untreated wastewater being widely used in irrigated agriculture. In the tenth year, the reuse of municipal wastewater in irrigated agriculture was comprehensively and systematically studied. The National Hi-Tech Research and Development Program (863) of the Eleventh Five-Year Plan (2006-2010) produced research results to inform policy makers and to improve technical conditions for using treated municipal wastewater in irrigated agriculture (Yi *et al*., 2011). To improve infrastructure, ¥30.4 billion was invested in upgrading existing infrastructure and building municipal WWTPs capable of producing fit-for-purpose wastewater. Through this improvement of WWTPs, cities with water shortages should achieve the goal of using recycled water for at least 20% of treated wastewater (Andrew Leung International Consultants and Investments Limited, 2015).

To achieve the set targets, several strategies have been developed at the national level to promote the reuse of municipal wastewater, as shown in Table 3.5. To institutionalise and strengthen the reuse of treated municipal wastewater, the Chinese government has issued and published several decrees, as shown in Table 3.6. In addition, standards for the reuse of treated municipal wastewater for specific purposes have been set by the Ministry of Construction and the Standardization Administration of China (Lyu *et al.*, 2016). It is worth mentioning the inclusion of plans, maintenance, risk management, and the accepted quality of municipal WWTPs that must be built to achieve the desired goals. This is to achieve the maximum efficient and effective reuse of treated municipal wastewater. Table 3.7 lists effluent quality standards for various uses. Despite all these efforts, the reuse of treated municipal wastewater is still in its infancy and faces several challenges that limit its use (Zhu & Dou, 2018).

Table 3.5: Chinese wastewater reclamation and reuse policies at the national level

Government sector	Wastewater reclamation and reuse policies	Wastewater reclamation and reuse policies prescriptions
The State Council	The 12th Five-Year Comprehensive and Emission Reduction Plan (2011); The 12th Five-Year National Urban Sewage Treatment and Recycling Facilities Construction Plan (2012)	1. Adopting reasonably the price of reclaimed water, which should be lower than that of conventional water, providing the privileged policies of tax and fee reduction for reclaimed water producers. 2. Encouraging reclaimed water to be used in industries, carwashes, urban facilities, and landscaping, forcing certain water users to use reclaimed water.
Ministry of Housing and Urban-Rural Development and the Ministry of Science and Technology	The Interim Procedures of Reclaimed Water Facilities Management in Urban (1995); Regulation of Saving Water Management in Urban (1998); Policy of Wastewater Reclamation and Reuse Technology in Urban (2006); and the 12th Five-year Development Plan of National Science and Technology (2011)	1. Using actively reclaimed water, issuing the technology policy of wastewater reclamation and reuse. 2. Considering preferentially the landscaping use of reclaimed water, using the secondary effluent from municipal WWTPs in agriculture irrigation. 3. Making policies to encourage wastewater reclamation and reuse by related central and local governments, offering financial support for wastewater recycling by local government. 4. Establishing a gradually reasonable water price system and water utilisation structure.

Government sector	Wastewater reclamation and reuse policies	Wastewater reclamation and reuse policies prescriptions
Ministry of Environmental Protection and General Administration of Quality Supervision, Inspection and Quarantine	The 12th Five-Year National Environmental Protection Regulation and Environmental Economic Policy Construction Plan (2011), and a series of water quality standards for different reclaimed water reuse	1. Setting water quality standards for different reclaimed water uses.
Ministry of Finance and the National Development and Reform Commission	The Notice of Implementing the Policy Without Value-Added Tax for Reclaimed Water and Others (2008), and The Notice of Suggestion about Supporting the Investment and Financing Policy of the Circular Economy Development (2011)	1. Reaching to wastewater reuse rates of 20-25% for cities with water scarcity in North China and 10% to 15% for coastal areas of South China in 2015. 2. Encouraging wastewater reclamation and reuse to increase water resources development efficiency.

Source: Lyu *et al.* (2016)

Table 3.6: Chinese government decrees on treated municipal wastewater reuse

Decree	Contents of the decree
GB/T 189198-2002	Divided municipal wastewater reuse into five categories and seven corresponding national standards.
GB/T 18920-2002	Specified standards for urban miscellaneous water quality, sampling, and analysis methods.
GB/T 18921-2002	Provided water quality and used patterns of reclaimed water for the landscape environment.
GB/T 19923-2005	Stipulated water quality and use patterns of reclaimed water for industrial water.
GB/T 19772-2005	Formulated control projects, limits, sampling, and monitoring of reclaimed water for groundwater recharge.
GB 20922-2007	Stipulated water quality control programmes, requirements, and analysis methods of reclaimed water for farmland irrigation.

Source: Yi *et al.* (2011)

Table 3.7: China's water quality standards for municipal wastewater reuse in irrigated agriculture

Values	Agricultural irrigation			
Indexes	Fibre crop	Dry land grain and oil crop	Paddy grain	Field vegetables
Chroma	-	-	-	-
Turbidity (NTU)	-	-	-	-
pH	5.5-8.5	5.5-8.5	5.5-8.5	5.5-8.5
TDS (mg/L)	≤ 1000, ≤ 2000	≤ 1000, ≤2000	≤ 1000, ≤2000	≤ 1000
SS (mg/L)	≤ 100	≤ 90	≤ 80	≤ 60
DO (mg/L)	-	-	≥ 0.5	≥ 0.5
BOD_5 (mg/L)	≤ 100	≤ 80	≤ 60	≤ 40
COD_{Cr} (mg/L)	≤ 200	≤ 180	≤ 150	≤ 100

Values	Agricultural irrigation			
Indexes	**Fibre crop**	**Dry land grain and oil crop**	**Paddy grain**	**Field vegetables**
NH_3-N (mg/L)	-	-	-	-
LAS (mg/L)	≤ 8.0	≤ 8.0	≤ 5.0	≤ 5.0
Hg (mg/L)	≤ 0.001	≤ 0.001	≤ 0.001	≤ 0.001
Cd (mg/L)	≤ 0.01	≤ 0.01	≤ 0.01	≤ 0.01
As (mg/L)	≤ 0.1	≤ 0.1	≤ 0.05	≤ 0.05
Cr (mg/L)	≤ 0.1	≤ 0.1	≤ 0.1	≤ 0.1
Pb (mg/L)	≤ 0.2	≤ 0.2	≤ 0.2	≤ 0.2
Fe (mg/L)	≤ 1.5	≤ 1.5	≤ 1.5	≤ 1.5
Mn (mg/L)	≤ 0.3	≤ 0.3	≤ 0.3	≤ 0.3
Faecal Coliform (MPN/L)	≤ 40 000	≤ 40 000	≤ 40 000	≤ 20 000

Source: Lyu *et al.* (2016)

3.5.5 Egypt

On the African continent, this study examined Egypt. Because Egypt has problems with freshwater availability, using treated municipal wastewater in irrigated agriculture is an innovative way to replenish water supplies. Egypt ranks first in Africa in the reuse of treated municipal wastewater in irrigated agriculture. Egypt’s National Water Resources Plan – 2017 included the possibility of using 1.4 billion m³ of treated wastewater annually in irrigated agriculture (Ministry of Water Resources and Irrigation, 2005). However, Egyptian standards and regulations do not currently address the reuse of treated municipal wastewater for irrigated edible crops. To this end, Gabr (2018) presented draft proposed standards for wastewater reuse in irrigated agriculture in Egypt and the recommended maximum allowable concentrations of heavy metals compared to other globally recognised organisations (see Tables 3.8 and 3.9).

Table 3.8: Draft proposed wastewater reuse standards for irrigated agriculture in Egypt

Parameter	Wastewater reuse for agriculture irrigation	
Coliform (/100 ml)	Cooked vegetables	ND FC (median)
	Cooked crops for human consumption	FC (cfu) ≤200 (median)
Turbidity (NTU)	Cooked vegetables	≤10
	Cooked crops for human consumption	-
Suspended solids (mg/L)	Cooked vegetables	TSS ≤15
	Cooked crops for human consumption	TSS ≤ 35
BOD (mg/L)	Cooked vegetables	≤15
	Cooked crops for human consumption	≤30
COD (mg/L)	≤ 30	
Odour	-	
T-N (mg/L)	≤ 15	
T-P (mg/L)	≤ 2	
pH	6.5 to 8.5	
EC (uS/cm)	Food crops	≤ 700
	Processed food crops	≤ 2 000
Aluminium (Al)	5.0	
Arsenic (As)	0.1	
Beryllium (Be)	0.1	
Boron (B)	0.75	
Cadmium (Cd)	0.01	
Chromium (Cr)	0.1	
Cobalt (Co)	0.05	
Copper (Cu)	0.2	
Fluoride (F)	1.0	
Iron (Fe)	5.0	
Lead (Pb)	5.0	
Lithium (Li)	2.5	
Manganese (Mn)	0.2	
Molybdenum (Mb)	0.01	
Nickel (Ni)	0.2	
Selenium (Se)	0.02	
Vanadium (V)	0.1	
Zinc (Zn)	2.0	
Mercury (Hg)	0.01	
Cyanide (CN)	0.001	
Tin (Sn)	0.005	
Thallium (Ti)	-	
Phenolates	0.02	
Detergents	0.02	

Source: Gabr (2018)

Table 3.9: Comparison of recommended maximum concentration of trace elements in irrigation water

Parameter	WHO (2006)	USA EPA (2012)	Spanish Royal Decree (2007)	Egypt Decree 92 (2013)
Al	5.0	5.0	-	-
As	0.1	0.1	0.1	0.01
Be	0.1	0.1	0.1	-
B	0.7	0.75	-	0.5
Cd	0.01	0.01	0.01	0.001
Cr	0.1	0.1	0.1	0.05
Co	0.05	0.05	0.05	-
Cu	0.2	0.2	0.2	0.01
F	1.0	1.0	-	0.5
Fe	5.0	5.0	-	0.5
Pb	5.0	5.0	-	0.01
Li	2.5	2.5	-	-
Mn	0.2	0.2	0.2	0.2
Mb	0.01	0.01	0.01	0.07
Ni	0.2	0.2	02	0.02
Se	0.02	0.02	0.02	0.01
V	0.1	0.1	0.1	-
Zn	2.0	2.0	-	0.01
Hg	-	-	-	0.001
CN	-	-	-	0.005
Sn	-	-	-	-
Ti	-	-	-	-
Phenolates	-	-	-	0.02
Detergents	-	-	-	0.5

Source: Gabr (2018)

3.6 CHALLENGES WITH TREATED MUNICIPAL WASTEWATER REUSE

Globally, the main barriers to the reuse of municipal wastewater, particularly in irrigated agriculture, can be categorised as technical, institutional, economic, and implementation challenges. Extensive research is being conducted in the Global North to overcome these barriers, which are at an advanced stage and can be leapfrogged by countries in the Global South. The State of California in the USA and Spain in Europe are precedents, and their progress is highlighted in this study.

3.6.1 Institutional arrangements

The lack of a universally applicable international legal framework for reusing municipal wastewater in irrigated agriculture is perceived as a major shortcoming worldwide. The consequences are inconsistent water reuse guidelines and dysfunctional institutions. To this end, the EPA in the USA has successfully developed regulatory frameworks that work in concert with those of the states. The EU has also followed suit and issued guidelines for its member states. However, the implementation of these guidelines by member states is often fraught with problems. To mitigate these inequities, platforms such as the NORMAN Network have been established for interdisciplinary research and the development of effective tools to improve the reuse of municipal wastewater in agriculture.

In developing countries, there is no overarching municipal wastewater management system, and an overarching cause of problems is the involvement of multiple ministries without clearly defined roles in projects to reuse treated municipal wastewater. Gilabert-Alarcón *et al.* (2018) reported challenges in Mexico resulting from the lack of shared responsibility and effective communication among ministries responsible for the reuse of treated municipal wastewater in irrigated agriculture.

Similar trends have been reported in China, where there is no consensus among ministries on the definition and statistical scope of treated municipal wastewater reuse. This has led to inconsistencies in the statistical data published by these ministries. In addition, Liu and Persson (2013) cited incomplete regulations, lack of supporting policies and laws, coupled with inconsistent standards, as major drawbacks to the adoption of the reuse of treated municipal wastewater for irrigation purposes in China.

In Egypt there are also problems regarding institutional organisations that result from the involvement of multiple ministries without clearly defined responsibilities that work in silos. In addition, there is a lack of political will and policy that explicitly provides for the reuse of treated municipal wastewater in irrigated agriculture.

3.6.2 Technical issues

Technical issues in reusing treated municipal wastewater in irrigated agriculture depend on the effectiveness of wastewater collection and treatment, followed by the ability of the treatment process to bring the wastewater to the required standard. The technical issues vary depending on the level of development in a region, state, or political jurisdiction. In jurisdictions such as the State of California, issues related to the effective collection and treatment of municipal wastewater have been extensively addressed. Current efforts are directed towards reclamation processes (Asano & Pettygrove, 1987). Various treatment technologies have been and are being developed.

The EU manages the WWTPs of its member states, although the infrastructure is developed differently by each member state. An example is Spain, where issues of effective and efficient collection and treatment of urban wastewater in accordance with the EU's Urban Wastewater Treatment Directive 91/271/EEC (UN Environment Programme, 1991) are prevalent. The Spanish water authorities are striving to fully comply with this directive through the construction of additional WWTPs and the upgrading of existing plants. Wastewater destined for reuse undergoes more advanced treatment in a water reclamation plant (WRP). Various treatment technologies have been and are being developed to treat wastewater to meet the quality standards required for reuse. To assist water authorities in selecting appropriate treatment technologies, the concept of best available technology was published in the Industrial Emissions Directive 2010/75/EU (European Parliament and Council, 2010). However, due to the changing composition of municipal wastewater over time, the selection of appropriate technologies remains a challenge and significantly limits the reuse of treated municipal wastewater in irrigated agriculture.

In the Global North, there is growing concern about the presence of "contaminants of emerging concern" (CECs) in municipal wastewater, the primary sources of which include pharmaceuticals and personal care products (Schwarzenbach *et al.*, 2006). The challenge with CECs is that they are not regulated and their long-term effects on the environment are unknown. In addition, scientists agree that recycled wastewater releases antibiotic-resistant bacteria. These findings make municipal wastewater treatment and recovery processes extremely complex and expensive (Berendonk *et al*., 2015).

In the Global South, the development of basic sanitation infrastructure is a widespread problem. Mexican authorities are working to develop their wastewater treatment infrastructure and to introduce treatment technologies that treat wastewater from these facilities according to the standards outlined in the WHO guidelines. However, severe restrictions, particularly on the irrigation of vegetables, fruits, and root crops that are eaten raw, make the reuse of treated municipal wastewater in irrigated agriculture economically unfeasible. Farmers are unwilling to invest in high-quality wastewater treatment technologies to meet the required water quality standards.

Following the Chinese government's decision to promote the reuse of treated urban wastewater, the challenges associated with municipal wastewater collection and treatment have been examined in detail in conjunction with high-tech research and development of water reclamation technologies (European Commission, 2020). Several technologies are currently available to produce wastewater with quality standards that meet the intended use. However, the challenge is that public agencies do not have the financial resources to bear the high capital and maintenance costs of these treatment technologies.

Egypt continues to struggle with sewerage network and treatment plant problems that negatively impact the effective and efficient collection of municipal wastewater. There are reports of large volumes of untreated wastewater flowing into natural water bodies (Abdel-Shafy & Mansour, 2013). In addition, current large centralised municipal WWTPs are not suitable for the reuse of wastewater in irrigated agriculture (Abdel-Shafy & Mansour, 2013). This is due to differences in the operation of several WWTPs, which result in differences in the quality of wastewater produced by these plants that complicate any plans for standardised wastewater reuse. In addition, most residents are not yet connected to the wastewater network.

3.6.3 Economic feasibility

The deployment of treated municipal wastewater reuse projects is highly dependent on economic feasibility and is usually a trade-off between costs and benefits. A large component of operation and maintenance costs is energy consumption, which typically accounts for 30% to 55% of total costs (Melgarejo *et al.*, 2016). Due to the global energy shortage, this cost component has become an important factor in evaluating economic feasibility. There is thus a need to find cheaper and cleaner energy sources for treated municipal wastewater reuse projects. Another important cost component is addressing problems encountered during the water reuse process.

A comprehensive cost calculation for reclaimed water remains problematic because multiple and changing wastewater components must be considered (Melgarejo *et al.*, 2016). Apart from the capital costs of developing infrastructure for treatment, storage, and distribution, there are additional costs such as operation and maintenance, and economic and environmental externalities that are usually ignored because in many cases it is difficult to quantify them; water agencies are thus unwilling to internalise them (Ayers & Wescott, 1985). It is therefore imperative to formulate a cost structure that accounts for treatment costs and reclaimed water management to provide incentives that encourage the maximum use of treated municipal wastewater. Farmers can be persuaded if agencies provide financial incentives for using reclaimed water while assuring that it meets water quality standards that ensure the safety of their agricultural products. However, in cases where reclaimed water is used only to supplement water supplies, incentives may not be necessary.

The EU funding model, where 50% of the upfront costs for municipal wastewater reuse projects can be secured through grants and the balance comes from the water reuse project, as provided for in the WFD, raises the issue of sustainability, as wastewater reuse prices are not guaranteed and depend on demand and supply scenarios. Another way to encourage farmers to use reclaimed water is to introduce subsidies. However, subsidies also pose another challenge because they only cover the costs of design, technical assistance, research, and construction, and do not consider externalities such as the financial, social, and environmental impacts of sanitation.

The Mexican financing model for municipal wastewater reuse projects is complicated by the Organisation for Economic Co-operation and Development's (2013) variable and non-transparent water budget, which makes it difficult for local authorities to plan and implement initiatives to reuse treated municipal wastewater. In addition, the arrangement of regional and local levels of government coordinating and mobilising investments in water infrastructure and then negotiating with Mexico's national water authority, CONAGUA, at the national level to approve funds for sanitation is complicated and limits their economic viability. Also, according to the Organisation for Economic Co-operation and Development (2013), the arrangement whereby CONAGUA collects revenues and channels them to the federal treasury, after which only 38% of revenues are transferred to local authorities for the construction, operation, and maintenance costs of WWTPs, limits the use of treated municipal wastewater for reuse in irrigated agriculture. The lack of well-structured water pricing policies that encourage farmers' use of reclaimed water exacerbates the situation.

In China, variability in financing negatively impacts the development of water reclamation facilities, which in turn impacts the success of treated municipal wastewater reuse projects. Chinese pricing of reclaimed water is not comprehensive, as current pricing only considers the economic and operational costs of treatment plants (Chu *et al*., 2004). The financial challenges in Egypt are due to the fact that public agencies are not adequately funded to cover the high capital and operating costs of treatment and network infrastructure for municipal wastewater facilities (Abdel-Shafy & Mansour, 2020). The pricing of treated municipal wastewater is still controversial in Egypt.

3.6.4 Implementation procedures

To encourage the reuse of treated municipal wastewater, the State of California has established a Water Recycling Funding Program (WRFP). This programme has contributed significantly to the success of municipal wastewater reuse projects. Because water management challenges are dynamic, continuous monitoring and evaluation of these projects are essential to ensure improvement in their implementation, which is embodied in the WRFP. There are several well-designed projects to reuse treated municipal wastewater in agriculture in Spain, of which one example is the Rincón de León WWTP-WRP.

The Mexican model of reusing treated municipal wastewater is based on the principles of IWRM, which emphasise stakeholder involvement and public participation. To this end, WUAs have been created that bring together various stakeholders. However, water authorities refer to them as civil society, which limits the participation of these WUAs in water decision-making processes at the local level because their contributions are not recognised by law (Mendoza-Espinosa *et al.*, 2004). In some cases, agreements on the reuse of treated municipal wastewater in irrigated agriculture are concluded between farmers and the authorities without consulting local communities and WUAs. As a result, public knowledge is not considered in the planning and implementation of these projects (Leach & Pelkey, 2001).

In China, the perception of municipal wastewater reuse is currently very low (Hanjra *et al.*, 2012) due to a lack of awareness of water resource issues and systemic risk management that could promote stakeholder confidence. Consequently, the uptake of municipal wastewater reuse is relatively low.

In Egypt, the implementation process is twofold. Firstly, there is an administrative component that is the responsibility of the water authorities. Secondly, a social component involves stakeholder engagement and public participation. There is currently no well-designed implementation plan. Oertlé *et al.* (2020) suggested that water managers should identify case studies as potential wastewater reuse sites that can be studied to document the advantages and disadvantages of reusing treated municipal wastewater. The lack of qualified personnel and appropriate equipment results in a lack of monitoring and evaluation of treated municipal wastewater reuse programmes. This is highly problematic as both the positive and negative risks of such programmes cannot be promoted or averted (El-Zanfaly, 2015).

3.7 SUMMARY

Although the disadvantages of using treated municipal wastewater in irrigated agriculture have been reported, the advantages cannot be denied. From a water management perspective, several benefits were identified in this study, including increased freshwater availability, the sustainable use of water resources, reduced freshwater withdrawals, and an economically viable alternative water source. Benefits to agriculture include reduced crop production costs due to reduced amounts of fertiliser applied, while agricultural production increases due to nutrients from treated municipal wastewater combined with a more reliable alternative water source, which increases employment in the agricultural sector and contributes to GDP. Environmental protection is also improved by reducing nutrient loading to natural waters. The use of treated urban wastewater as an alternative water resource should thus be given serious consideration and attention in Africa and especially in sub-Saharan Africa, starting with institutional arrangements that encourage it.

For the Global North, this study found that effective and systematic involvement of a supranational or regional institution in the use of treated municipal wastewater in irrigated agriculture promotes its use. An example is the EU, which has adopted the WFD, through which directives are issued to regulate certain water matters, including the use of treated municipal wastewater in irrigated agriculture in member states. These directives work with member states' national policies, regulations, and guidelines.

In the USA, while each state is responsible for formulating and promulgating its water policies, laws, regulations, and guidelines for various water uses, the federal government's role supports the states' institutional frameworks through the EPA. The institutional arrangements of the EU and American federal governments have been shown to promote consistency in water management across their regions. This promotes trust and institutional support among stakeholders in the management of treated municipal wastewater in irrigated agriculture. In addition, platforms such as NORMAN in Europe provide interdisciplinary knowledge and research and development of effective tools for reusing treated municipal wastewater in irrigated agriculture.

At the national (Spain) and state (State of California) level, the study found that policies for the reuse of treated municipal wastewater in irrigated agriculture explicitly articulate the “what”, “where”, “when”, “who”, and “how”. Current water laws flesh out regulations and guidelines for the reuse of treated municipal wastewater in irrigated agriculture, as well as the political will of local jurisdictions to support and implement specific projects. Spain has institutionalised supramunicipal administrative units that directly or indirectly manage the operation and maintenance of WWTPs through competent bodies to ensure uniformity and compliance with EU directives.

In developing countries, the reuse of treated municipal wastewater in agriculture is hampered by uncoordinated multiple ministry involvement without clear roles, policy gaps, inconsistent guidelines, and incomplete regulations. However, important institutional reforms have been undertaken in Mexico, which resulted in adopting policies, laws, regulations, and standards. In addition, political will is expressed by the national government, which has led to the recognition of treated municipal wastewater as an alternative water source. In China, on the other hand, institutional challenges do not encourage the widespread use of treated municipal wastewater in agriculture.

In Africa, Egypt in North Africa was considered. In the absence of a regional authority, Egypt is entirely responsible for its water management. The study found deficiencies in institutional arrangements that result from the uncoordinated responsibilities of several ministries to reuse treated municipal wastewater in irrigated agriculture. Policy gaps and the lack of strict regulations and guidelines hinder the reuse of treated municipal wastewater in irrigated agriculture in Egypt.

As mentioned earlier, the technical basis for the reuse of treated urban wastewater is the effective collection and treatment of wastewater. In the Global North, the State of California has effectively mastered these fundamentals and developed tertiary treatment technologies that produce effluent quality that meets mandated standards for reuse. In the EU, compliance with the WFD guidelines is mandatory for municipal wastewater reuse. Several member states meet these requirements. While still on track for 100% compliance, Spain has made progress in reusing treated municipal wastewater in irrigated agriculture. Spanish authorities continue to conduct intensive research on water reclamation technologies to improve economic efficiency, reduce energy costs, and decrease the amount of waste disposed into the environment.

The Global North is aware of the growing concern about CECs, and appropriate treatment technologies for their removal continue to be intensively researched. While countries can import these technologies to the Global South, this step may not be feasible due to affordability and appropriateness for the local landscape.

CHAPTER 4:
THE APPLICATION OF MACHINE LEARNING TECHNIQUES IN URBAN WATER SYSTEM MANAGEMENT, A REVIEW.

4.1 OVERVIEW

The sustainable and efficient management of urban water supply systems requires an efficient and continuous supply of water in sufficient quantity and good quality at an acceptable pressure and price. At the same time, a reliable water distribution network must be maintained. Meeting these consumer expectations requires comprehensive and accurate planning and good decision-making processes. Over the past two decades, decision-making challenges in managing urban water systems have increased. Almandoz *et al*. (2003) stated that the accurate prediction of water demand is the most important aspect in the sustainable management of any urban water system because water demand depends on factors that greatly impact on creating an efficient urban water system. To this end, researchers have demonstrated the benefits of accurate short-, medium-, and long-term water demand forecasts and projections for managing an urban water system. According to Zhou *et al.* (2002), accurate short-term water demand forecasts and predictions promote precise process design and implementation. Bougadis *et al.* (2005) claimed that accurate medium-term water demand forecasts and projections enable accurate water demand projections that account for population change. Ghiassi *et al.* (2008) pointed out that long-term water demand plays a vital role in planning and shaping future water infrastructure development and formulating related water policies.

Accordingly, there has been a strong demand for powerful models that can accurately forecast and predict the water demand of an urban system over the years. Various conventional models have been studied and successfully used in the past. Models developed using the Autoregressive Integrated Moving Average (ARIMA) and its extension, the Seasonal Autoregression Integrated Moving Average (SARIMA), along with Vector Autoregression (VAR) techniques, are among the powerful conventional models that have been formally used (Box *et al.*, 2015; Smolak *et al.*, 2020). However, due to unabated population growth, rapid urbanisation, and the adverse effects of climate change on precipitation, models developed using conventional techniques to predict water demand are becoming increasingly inaccurate due to unquantifiable uncertainties and increasing variables that enter the urban water system and have negative impacts on the system. Shabani *et al.* (2017) reported some conventional models that overestimated the actual water demand of a city by up to 100%.

To address the problems with model performance, Herrfahrdt-Pähle (2013) suggested that urban water system management should shift to a more adaptive management paradigm. As a result, machine learning techniques are gaining popularity over conventional modelling techniques. Tiwari and Adamowski (2017) advocated for machine learning modelling techniques by citing their ability to create robust models that progressively reduce risks to urban water systems as uncertainties are quantified and all variables that affect the system are considered. This allows water agencies and policy makers to understand and interpret variability in the urban water system, which improves their decision-making processes. This chapter describes and reviews machine learning algorithms used in urban water demand forecasting and prediction.

4.2 MACHINE LEARNING ALGORITHMS

The information age has spawned the field of AI, and machine learning is a subfield of AI coined by Arthur Samuel in 1959. He defined machine learning as a "field of study that gives computers the ability to learn without being explicitly programmed". To date, machine learning continues to evolve as a refined scientific tool for describing future scenarios that maximises possible outcomes while minimising the risk (uncertainty) that the outcome will be falsified. A notable milestone in machine learning was reached when Robert Nealey, a self-proclaimed chess master, lost a game to the IBM 7094 computer in 1962. This was seen as a precursor to the superiority of machines over humans in performing intelligent tasks.

Tom Mitchell formalised Samuel's definition in 1997 through the following scenario: "A computer programme learns from an experience E with respect to a task T and a performance measure P if its performance on T, as measured by P, improves with experience E." However, the researcher adopted Michalski *et al.*'s (2013) definition, which states that machine learning is a subfield of AI in which statistical methods are adopted and used in computer algorithms to teach (learn) machines to make classifications or predictions and to extract from large datasets important insights about how to perform certain tasks or make certain decisions. Essentially, there are two main machine learning categories: inductive and deductive learning. Deductive learning uses available facts and knowledge to reach a valid conclusion. Inductive learning, on the other hand, involves creating computer programs to find relationships or patterns in large datasets (Pelaccia *et al.*, 2019). These two categories have evolved into supervised and unsupervised machine learning.

Unsupervised learning uses neither classified nor labelled data for training without human supervision. The goal is to make inferences from the input data and to then model the hidden or underlying structure and distribution in the data to learn more about the data (Khanum *et al.*, 2015).

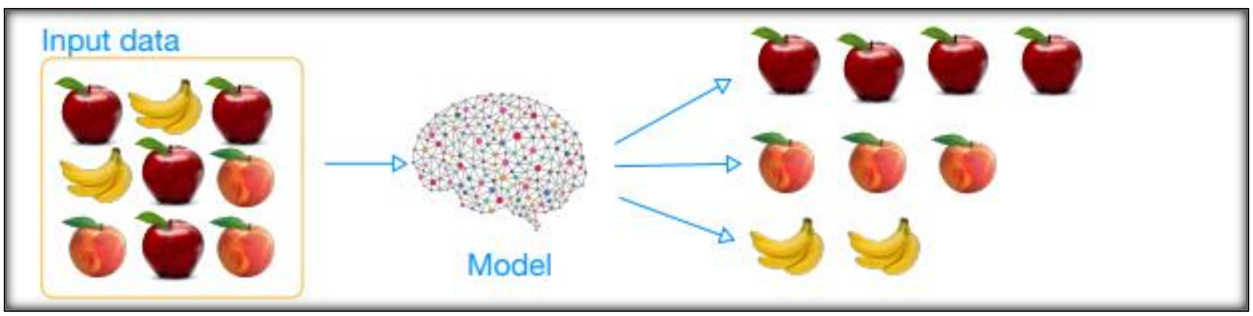


Figure 4.1: Unsupervised machine learning schematic diagram

Source: Ma *et al.* (2018)

Figure 4.1 shows a simplified version of the unsupervised machine learning process. Here, only unlabelled input data and no corresponding output data are provided to the model. The model can independently discover patterns and information hidden in the data. Unsupervised learning is used in clustering and association problems using algorithms such as K-Means, Gaussian mixture models, and principal component analysis. Unsupervised learning is widely used in building recommender systems, anomaly detection, and customer segmentation (Li *et al.*, 2021; Yassine *et al.*, 2021; Ray, 2019; Pu *et al.*, 2020).

Accordingly, supervised learning is a machine learning process that occurs under human supervision. The learning process mainly applies what has been learned in the past to new data and uses labelled examples to predict future events. In this process, "ground truth" data are available for training. From the analysis of a known training dataset, a learning algorithm generates a derived function to make predictions about output values. The system can provide targets for each new input after sufficient training and can improve its estimates using ground truth and repetition until the algorithm reaches a desired level of accuracy (Shanthamallu *et al.*, 2017). Supervised machine learning has been applied in various fields, including water management (Bata *et al.*, 2020).

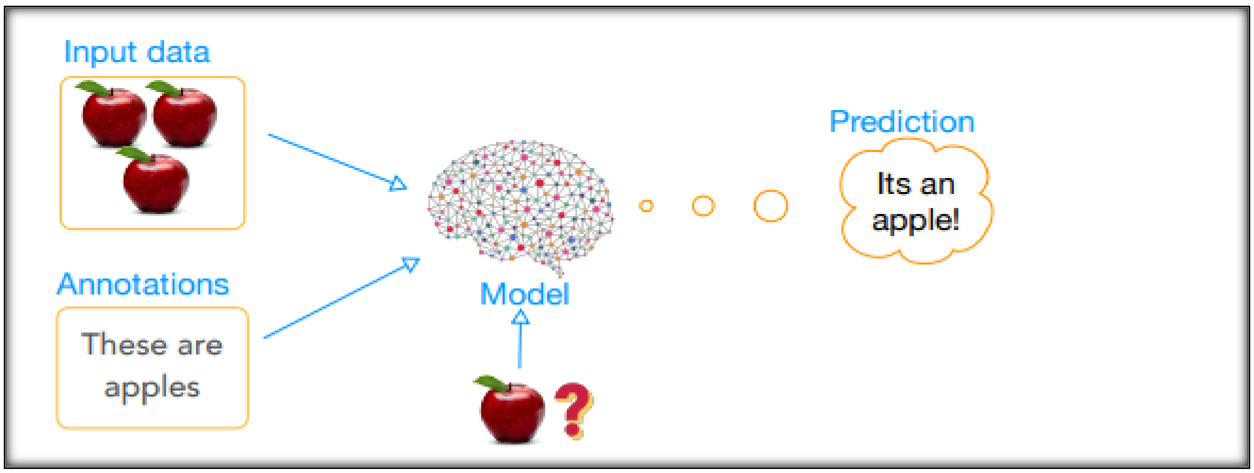


Figure 4.2: Supervised machine learning schematic diagram

Source: Ma *et al.* (2018)

Figure 4.2 shows a simplified process of supervised learning. In this process, the model is provided with input data and annotations (known responses to the data) that form the training dataset with which the model is trained to produce the desired predictions for the response to new data; in this case, "It's an apple!" An algorithm is then used to measure the accuracy of the prediction using the loss function, which is adjusted until the error is minimised (Nasteski, 2017). Supervised learning is applied to classification and regression problems and provides more accurate results compared to unsupervised learning. Algorithms used in the classification category include neural networks and Support Vector Machines (SVMs). In regression, linear regression, logistic regression, Support Vector Regression (SVR), and ensemble methods, algorithms form a subset of the various algorithms available (Jiang *et al.*, 2020).

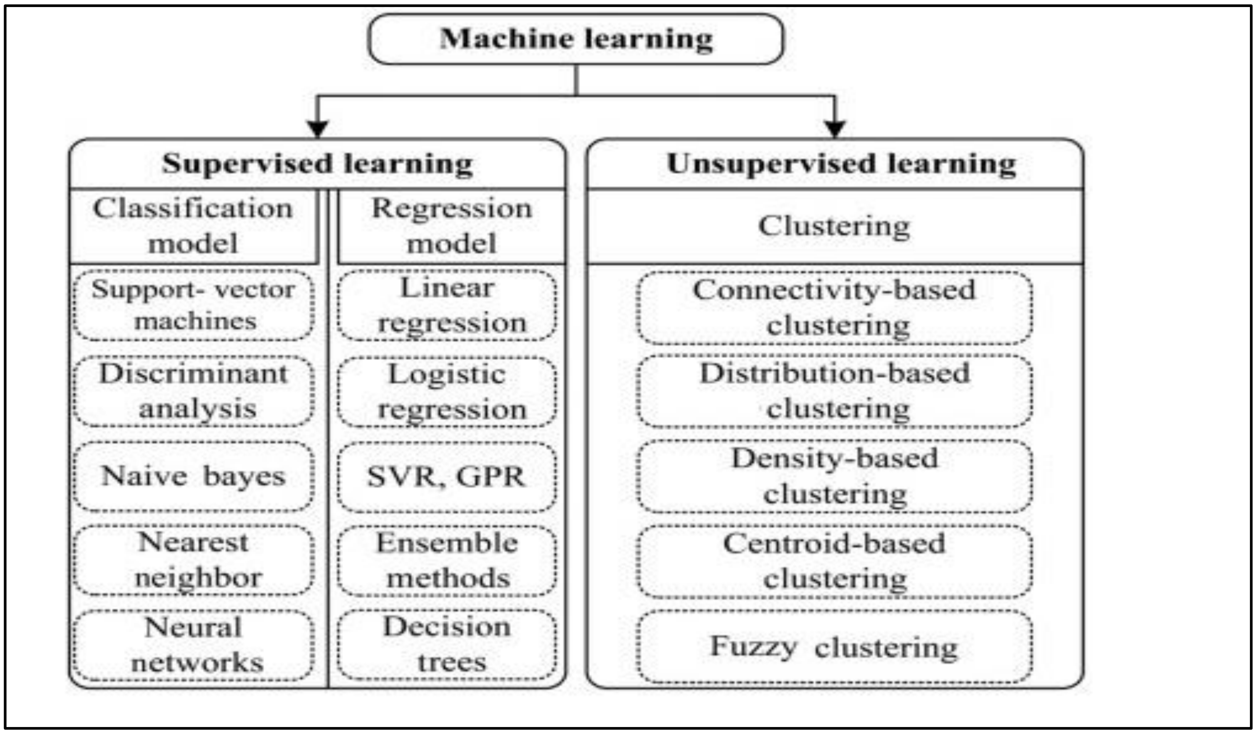


Figure 4.3: Classification of the machine learning algorithm

Source: Kumar and Singh (2022)

Figure 4.3 shows the two main categories of machine learning as described above: supervised and unsupervised learning. Unsupervised learning has only one category, clustering, and several algorithms with which to apply it. Supervised learning consists of two categories, namely classification and regression, and the corresponding algorithms are presented. Since supervised learning was used in this study, an overview of the following regression models was made. Thus, linear regression, logistic regression, SVR and ensemble methods, along with the ANN a classification algorithm that is gaining popularity and producing very powerful models when applied in various domains including water demand prediction (Kumar & Singh, 2022).

4.3 REGRESSION SUPERVISED MACHINE LEARNING ALGORITHMS

Machine learning algorithms derive their behaviour from statistics. The difference lies in the language used in each domain and in the ability of machine learning algorithms to draw insights from large datasets. Accordingly, in the next subsections, the researcher presents the basics of linear, logistic, ridge, Least Absolute Shrinkage and Selection Operator (LASSO), and polynomial regression techniques as they form the basis of most machine learning algorithms used in urban water demand modelling and forecasting.

4.3.1 Linear regression

The supervised machine learning algorithm performs regression tasks when developing models. A simple linear regression model represents the relationship between a single independent predictor variable X and a dependent variable Y, the response variable (Bangdiwala, 2018). When a single independent variable is considered, the model created is a simple univariate linear regression model that is mathematically captured by the following equation:

$$Y_i = \beta_0 + \beta_1 X_i \tag{4.1}$$

Where the dependent variable $Y_i = \{y_1, y_2, \ldots, y_n\}$, the independent $X_i = \{x_1, x_2, \ldots, x_n\}$, n is assumed large to be considered to represent the population of the data, and β_0 and β_1 are parameters of the model. In addition, a comprehensive linear regression analysis considers the residual standard deviation. It is the measure used to assess how well a linear regression model fits the data. The equation below incorporates the residual standard deviation component into the univariate linear regression equation.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \sigma_{\text{res}} \tag{4.2}$$

Where the hat denotes the fitted values from n data points, and σ_{res} is the residual standard deviation. The standard deviation of residuals σ_{res} characterises the variability around the regression line; that is the smaller the σ_{res}, the better the fit and is estimated by the following equation:

$$\sigma_{\text{res}} = \sqrt{\frac{\sum (\text{residuals})^2}{n-2}} \quad \text{or} \quad \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}} \tag{4.3}$$

In addition to the residual standard deviation, the confidence interval is also considered when evaluating a regression line. This indicates the percentage probability that an estimated range of possible values includes the actual estimated value. Most often, a confidence level of 95% is used. The calculation is based on the standard error of β_1 using the following mathematical equation:

$$\text{se}(\beta_1) = \frac{S_{\text{res}}}{\sqrt{S_{xx}}} = \frac{S_{\text{res}}}{\sqrt{\sum x_i^{\,2} - \frac{(\sum x_i)^2}{n}}} \tag{4.4}$$

Source: Stijnen and Mulder (1999)

When two or more independent variables and one dependent variable are considered, multivariate linear regression analysis is used. In a multiple linear regression model, Y is the response variable ("dependent") which is considered to depend on p predictor variables ("independent"), X_1, X_2, X_3, …, X_p. The estimate for Y is obtained from the following equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \; .... \; + \beta_p X_p \tag{4.5}$$

Source: Alexopoulos (2010)

Where the model parameters the β_0, β_1 …, β_p are obtained from that fit to n sample data points that estimate the dependent variable by the equation.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \; .... \; + \hat{\beta}_p X_p \tag{4.6}$$

One of the approaches commonly used to calculate the parameters β_0, β_1 …, β_p is the least square method which minimizes the square of the deviation of the residuals, $(Y - \hat{Y})$.

4.3.2 Logistic regression

The logistic regression algorithm is used to develop models that predict the categorical dependent variable based on a given set of independent variables. It provides probability values that range from 0 to 1. It is used to solve classification problems. In a logistic regression model, instead of a regression line, an S-shaped logistic function is fitted that predicts two maximum values (0 or 1), as shown in Figure 4.4 and a probability of occurrence that is given by Equation 4.7.

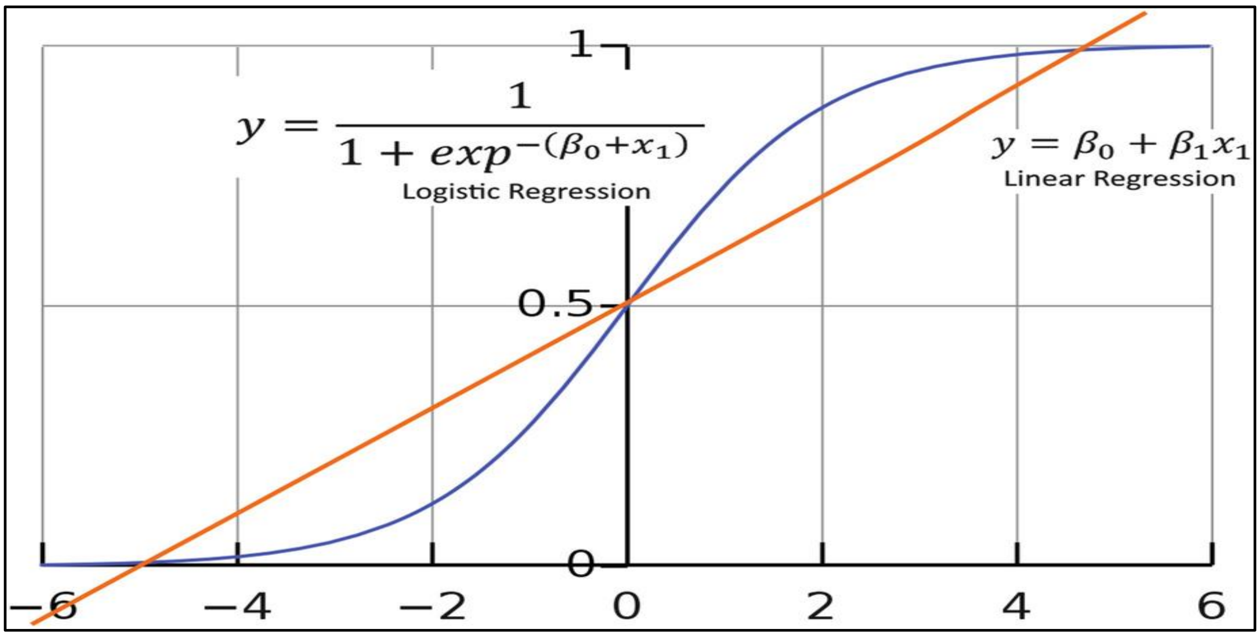


Figure 4.4: Standard logistic regression model

Source: Adopted from Yadav *et al*. (2019)

$$p(x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}} \qquad (4.7)$$

Where β_0 is referred to as the intercept and β_1 the rate parameter.

4.3.3 Least Absolute Shrinkage and Selection Operator (LASSO) and ridge regression

LASSO and ridge regression algorithms are used to reduce model complexity and prevent overfitting, which can occur with simple linear regression. This method uses a cost function that calculates the error between the predicted and actual values and is represented as a single real number. The LASSO regression algorithm is used over regression methods to improve the accuracy of the prediction through L1 regularisation (it adds an L1 penalty equal to the absolute value of the magnitude of the coefficient). This technique facilitates the process of shrinkage, where the data values are shrunk to a central point as the mean, and the coefficient of the less important features is shrunk to zero, removing some features completely; hence the use in feature selection. The LASSO procedure promotes simple, sparse models and is suitable for models that have a high degree of multi-collinearity or where certain parts of the model selection need to be automated. The cost function for LASSO regression is calculated using the following equation:

$$\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{m}\left[y_i - \sum_{j=0}^{p}(w_j \times x_{ij})\right]^2 + \lambda\sum_{j=0}^{p}\left|w_j\right| \tag{4.8}$$

Where m and p represent the realizations and features, respectively, and λ is a penalty function on the weights, w_j. LASSO regression is highly effective in reducing overfitting.

The ridge regression algorithm is used in model tuning by performing L2 regularisation (i.e., adding an L2 penalty equal to the square of the magnitude of the coefficient) on data that suffer from multi-collinearity. This modifies the cost function by adding a penalty equal to the square of the coefficient, as shown in the following equation:

$$\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{m}\left[y_i - \sum_{j=0}^{p}(w_j \times x_{ij})\right]^2 + \lambda\sum_{j=0}^{p}w_j^{\,2} \tag{4.9}$$

Ridge regression reduces the coefficients and never sets the value of the coefficient to absolute zero. Essentially, it helps to reduce model complexity and multi-collinearity. Figure 4.5 delineates how ridge regression operates geometrically.

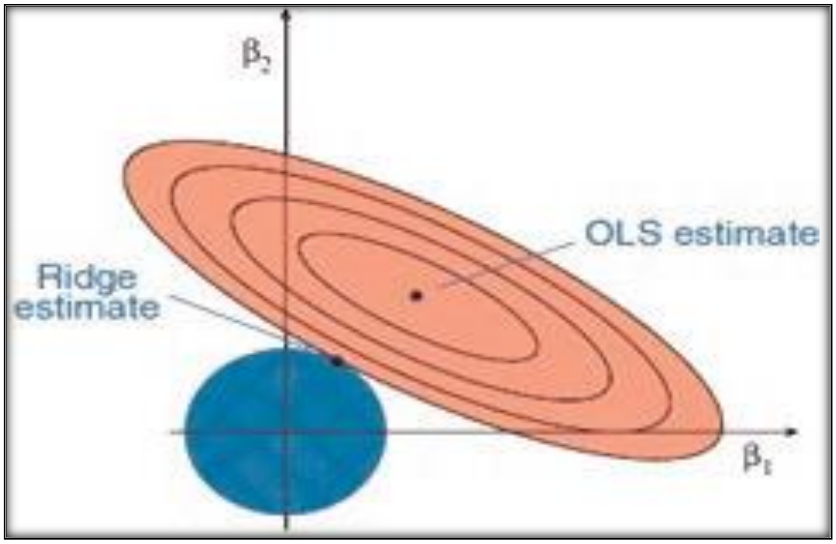


Figure 4.5: Geometrical representation of ridge regression

Source: Adopted from Yadav *et al.* (2019)

In a geometric ridge operation, the main objective is to minimise the ellipse size and the circle at the same time. The ridge estimate is given by the point where the ellipse and the circle touch. There is a trade-off between the penalty term and Residual Sum of Squares (RSS)

4.3.4 Polynomial regression

Polynomial regression is a special case of multiple regression that is used when there is a nonlinear relationship between dependent and independent variables. Accordingly, some polynomial terms are added to linear regression to convert it into polynomial regression. The polynomial regression of order $k > 1$ is given by.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \ \dots + \beta_k x_i^k + \varepsilon_i, \text{ for } i = 1, 2, \dots, n \tag{4.10}$$

Machine learning algorithms have emerged as superior, efficient, multifunctional, data-driven tools capable of handling the increasing complexity of urban water demand forecasting. This is in contrast to conventional statistical methods, which have limitations in processing huge datasets and are unable to extract detailed insights from large datasets. The efficiency of using machine learning algorithms in predicting urban water demand has increased the importance of machine learning as a reliable method for understanding, planning, and ultimately developing a better water management strategy. The following section describes the specific machine learning algorithms used in this study to model the water demand of the Stellenbosch Municipality; considering that numerous machine learning algorithms are continuously being developed for specific sectors, including water management in general.

4.4 REGRESSION ALGORITHMS TO BE DEPLOYED

Several regression algorithms have been developed using the principles of regression, and they have been successfully applied in various fields. However, in this study, the focus is on urban water demand forecasting and related predictions. Of the extensively researched regression-based algorithms, the researcher gives an overview of the ensemble algorithms SVR and Extreme Gradient Boosting (XGBoost) in the following subsections.

4.4.1 Support Vector Machine (SVM) and Support Vector Regression (SVR) algorithms

SVR is an SVM algorithm used in regression tasks. Since the SVR algorithm is based on the principles of the SVM, it is essential to provide an overview of the SVM algorithm before further discussing SVR. The SVM algorithm is considered one of the best machine learning algorithms proposed in the 1990s for pattern classification, including image and speech recognition, because it can be used for both classification and regression tasks (Cortes & Vapnik, 1995). Procedurally, each data item is plotted as a point in an n-dimensional space, where n represents multiple features considered and the value of each feature is the value of a particular coordinate. Classification is then performed by finding the hyperplane that uniquely classifies the two classes; hence the name discriminative classifier. Figure 4.6 presents a schematic diagram of SVM graphs for two-group classification problems.

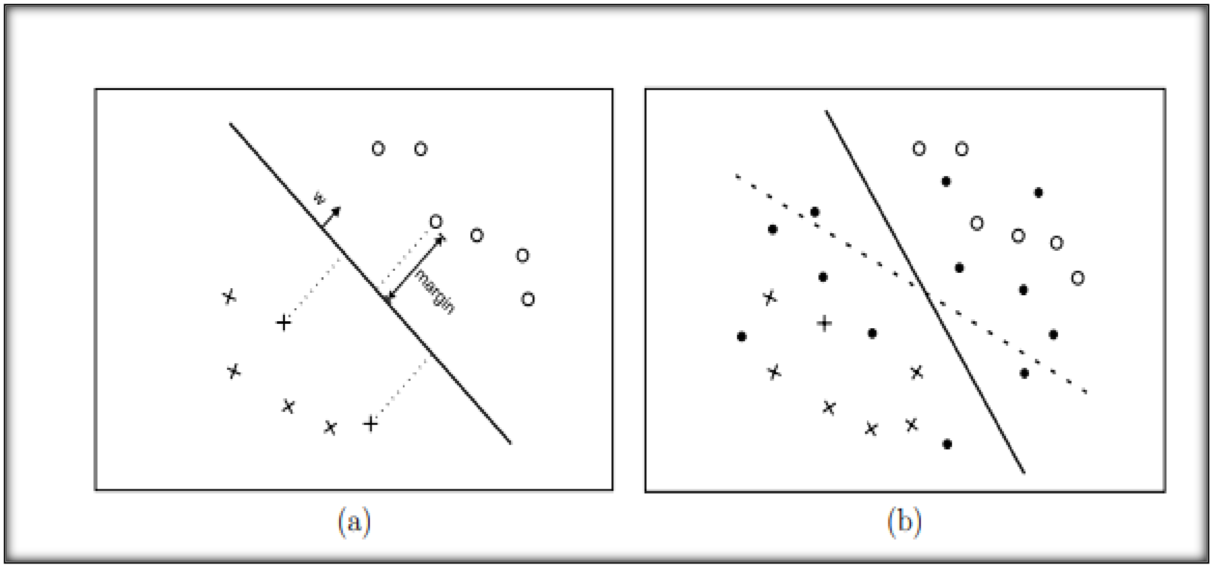


Figure 4.6: (a) A simple linear SVM; (b) An SVM (dotted line) and a transductive SVM (solid line)
Source: Burges (1998)

In Figure 4.6(b), solid circles represent unlabelled instances.

For a given training dataset denoted by $\{X_1,...,X_n\}$ that are vectors in some space $X \subseteq R^d$ with labels $\{Y_1,\ldots,Y_n\}$ where $Y_i \in \{-1,1\}$. The SVM is the hyperplane (the dividing line between two classes of data) that separates the training data by a maximum edge, as shown in Figure 4.6 (a). Vectors on one side of the hyperplane are labelled -1 and all vectors on the other side are labelled 1. The training instances closest to the hyperplane are the support vectors. Essentially, SVMs allow the original training data in space X to be projected onto a higher dimensional feature space F via a Mercer kernel operator K.

This can be expressed in an equation where a set of classifiers of the form

$$f(x) = \left(\sum_{i=1}^{n} \alpha_i K(x_i, x) \right) \quad (4.11)$$

are considered. When K satisfies Mercer's condition, the following expression can be captured (Burges, 1998):

$$K(u,v) = \Phi(u).\Phi(v) \tag{4.12}$$

Where Φ: $X \rightarrow F$, and "·" denotes an inner product. The expression for $f(x)$ is denoted by the following equation:

$$f(x) = w.\Phi(x), \quad \text{where} \quad w = \sum_{i=1}^{n} \alpha_i \Phi(x_i) \tag{4.13}$$

Thus, by using K, the training data are implicitly projected into another (often higher dimensional) feature space F. The SVM then computes the α corresponding to the maximum margin hyperplot in F (Tong & Koller, 2001; Pradhan, 2012). An applied kernel helps to reduce the computational cost when the dimension of the data increases. A higher dimension is required when a separating hyperplane cannot be created in a given dimension. SVR uses the same principle as SVM for regression problems.

The regression employs the tasks of SVR, using kernels, sparse solution, and controlling the margin and number of support vectors through Vapnik-Chervonenkis theory. The robustness of SVR comes from its effectiveness in estimating real value functions. It trains with a symmetric loss function that penalises high and low misestimates equally. A minimum radius flexible tube is formed symmetrically around the estimated function using Vapnik's ε-insensitive approach. This allows the absolute values of errors smaller than a certain threshold ε to be ignored both above and below the estimate. In this way, points outside the tube are penalised, but points inside the tube, either above or below the function, receive no penalty. The SVR is considered to have excellent generalisation ability with high predictive accuracy (Awad & Khanna, 2015). Figure 4.7 presents a schematic diagram of a one-dimensional linear SVR.

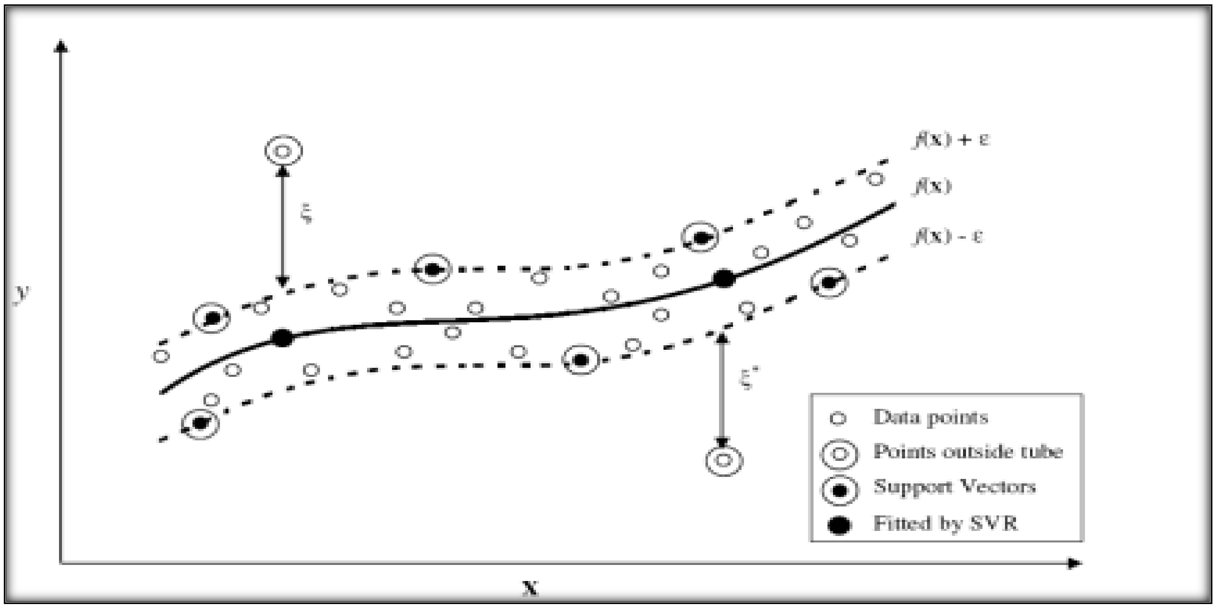


Figure 4.7: A schematic diagram of the SVR using ε sensitive loss function

Source: Lahiri and Ghanta (2008)

Equation 4.14 captures the formula for determining the estimated continuous value function, while Equation 4.15 presents the formula for obtaining the multivariate regression.

$$Y = f(x) = \langle w, x \rangle + b = \sum_{j=1}^{m} w_j x_j + b, \quad y, b \in \mathbb{R}, \quad x, w \in \mathbb{R}^m \tag{4.14}$$

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = w^T x + bx, w \in \mathbb{R}^{m+1} \in \mathbb{R}^m \tag{4.15}$$

Essentially, SVR focuses on optimisation by finding the narrowest tube centred around the surface while minimising the prediction error. This expression is reflected by the function in Equation 4.16, where $||w||$ is the magnitude of the normal vector to the surface being approximated:

$$\min_w \frac{1}{2} \|w\|^2 \tag{4.16}$$

Source: Awad and Khanna (2015)

In the last decade, SVR has gained popularity in water demand forecasting and prediction tasks. Advantages include resilience to overfitting and lower error on previously unseen data, which are important attributes for noisy water use data (Ghalehkhondabi *et al.*, 2017). However, the disadvantage of the SVR algorithm is that its performance depends on the choice of parameters. This study investigated the use of a hybrid model, which is a combination of the Prophet model and SVR (Bai *et al.*, 2015). The details of the Prophet-SVR hybrid model are presented in Section 4.7.4.

4.4.2 Extreme Gradient Boosting (XGBoost) ensemble model

Because multiple machine learning models underperform when used individually, ensemble learning has emerged, in which the predictive power of multiple weak learners is systematically combined to create a single powerful model that provides the combined results of multiple models. Currently, there are three ensemble learning methods, namely bagging, stacking and boosting. The researcher investigated the boosting learning method, which is used to build more powerful models to predict water demand. The XGBoost algorithm developed by Chen and Guestrin (2016) serves as a good example. The advantages of the XGBoost algorithm include its robustness as a tree-based ensemble learning algorithm, its high effectiveness in minimising overfitting, and its ability to increase computational and memory capacity while handling missing values well (Fan *et al.*, 2018; Chen *et al.*, 2015).

The main objective function of the XGBoost algorithm is regularisation using the expression Ω to control model complexity. The following equations illustrate the process.

$$obj = \sum_{i} (y_i F(x_i) + \sum \Omega(f_t) \tag{4.17}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j}^{T} w_j^2 \tag{4.18}$$

Where:

T = leaf count of the tree;

F = the computed score of the j^{th} leaf of tree f;

$f(x)$ is a function; $f(x) = wq(x)$;

$q(x)$ = a tree that plots sample x to the corresponding leaf;

λ = optimisation parameter for rigid regularisation; and

y = the threshold for the score function for splitting the tree (Kim *et al*., 2022).

Although the XGBoost algorithm offers several advantages, there are also disadvantages, such as its weak performance on sparse and unstructured data and its high sensitivity to outliers.

4.5 ARTIFICIAL NEURAL NETWORKS (ANNs) ALGORITHM

In addition to regression supervised machine learning models, ANNs have also been investigated. Typically, ANNs attempt to mimic the human network of neurons in order to train computers to learn patterns by which they can make decisions in a human-like manner. An ANN consists of processing nodes, or neurons, that are connected in a specific order to perform simple numerical manipulations. Structurally, they are divided into three layers: the input layer, the hidden layer, and the output layer.

The input layer receives input from the outside world, which the network processes. The nodes on the input layer are passive and only receive a single value on their input, duplicate the value to their many outputs, and send it to all hidden nodes. The hidden layers perform nonlinear transformations on the inputs that have entered the network. The hidden layers vary depending on the function of the neural network. A neural network can consist of one or more hidden layers. The simplest network consists of a single hidden layer, such as a perceptron. These hidden layers perform various types of mathematical computations on the input data and recognise the patterns of the data. The hidden layer is then connected to an output layer that receives connections from the hidden layer or the input layer. Within an ANN, each node, i.e., each artificial neuron, is connected to another and has a corresponding weight and threshold. If the output of a single node is above the specified threshold, that node is activated and sends data to the next layer of the network. Otherwise, no data are forwarded to the next layer of the network. This results in the output of one node becoming the input of the next node. Passing data from one layer to the next defines the neural network as a feedforward network. Finally, the output layer provides the results of the rigorous computations performed by the middle layer (Adejo & Connolly, 2018). Figure 4.8 presents a schematic diagram of the basic structure of an ANN.

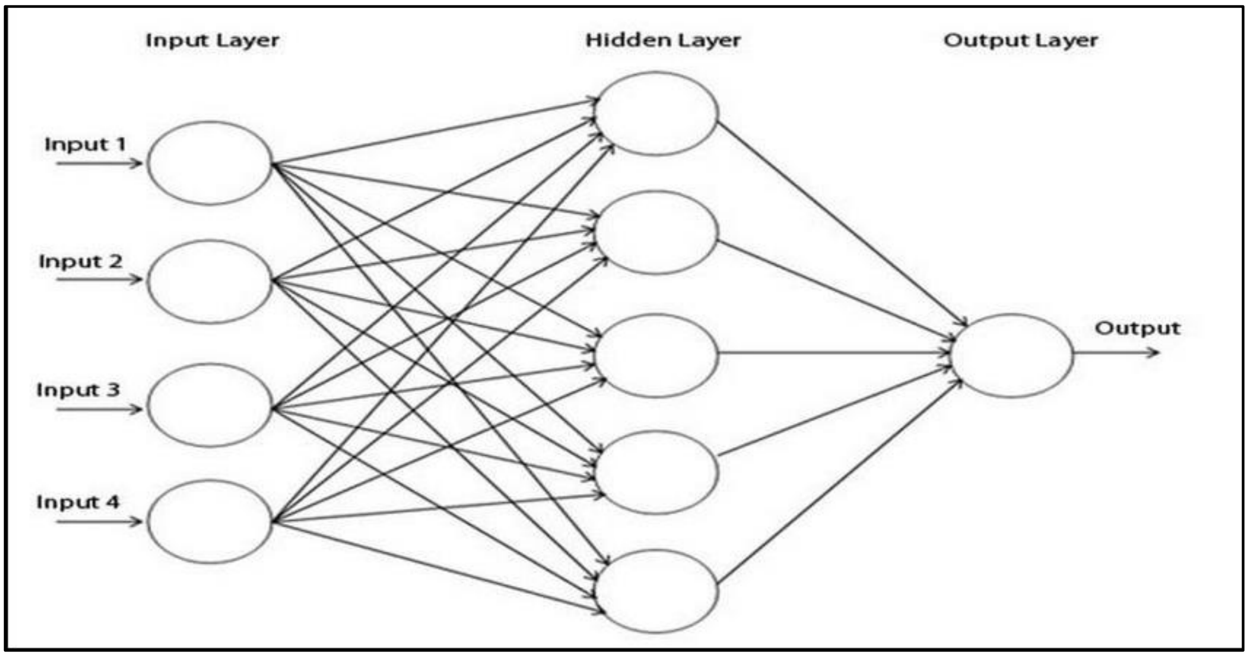


Figure 4.8: Schematic diagram of the structure of an ANN

Source: Adejo and Connolly (2018)

However, the ability of an ANN to perform useful data manipulations depends on the proper selection of weights. A single neuron-like node in the hidden layer is expressed in terms of a linear combination of weights and input data that incorporates a bias, as given by Equation (4.19).

$$z_i = \sum_{j=1}^{m} w_{ij} x_j + b_i \tag{4.19}$$

Where w_{ij} are the weights, x_j the input variable, b_i the bias and z_i is the output from the hidden layer. The output from the hidden layer is obtained from applying an activation function to z_i that is given in Equation (4.20):

$$y_i = g(z_i) \tag{4.20}$$

Where $g(z_i)$ is the activation or transfer function (for ANN models that have more than one hidden layer). The ANN algorithm is becoming increasingly popular and is being used in various fields. Although it also has disadvantages, the advantages have continued to increase while the disadvantages have been reduced by extensive scientific research to improve ANN applications. Advantages such as the ability to deal with complex and nonlinear relationships between inputs and outputs, as well as high fault tolerance, make ANNs increasingly important as a machine learning algorithm with the ability to build powerful water demand models (Zubaidi, Ortega-Martorell, Al-Bugharbee *et al*., 2020; Liu *et al.*, 2018).

The ANN algorithm is becoming increasingly popular and is being used in various fields. Although it also has disadvantages, the advantages have continued to increase while the disadvantages have been reduced by extensive scientific research to improve ANN applications. Advantages such as the ability to deal with complex and nonlinear relationships between inputs and outputs, as well as high fault tolerance, make ANNs increasingly important as a machine learning algorithm with the ability to build powerful water demand models (Zubaidi, Ortega-Martorell, Al-Bugharbee *et al*., 2020; Liu *et al.*, 2018).

4.6 THE PROPHET ALGORITHM

Several researchers have taken up the use of hybrid models to predict water demand as a method to improve model performance (Altunkaynak & Nigussie, 2017; Pandey et al., 2021). Accordingly, the researcher investigated a hybrid model comprising the Prophet algorithm and SVR (Prophet-SVR). The Prophet algorithm, developed by Facebook for time series prediction, is presented as the main technique. It offers many advantages, such as robustness in predicting time series data by using an additive algorithm that compensates for non-linear trends with annual, weekly and daily seasonality, as well as holiday effects. It is also robust in dealing with missing data and trend shifts and copes well with outliers. Since the Prophet algorithm is a univariate forecasting algorithm, its combination with SVR (described in section 4.7.1) optimises its performance. The result is a powerful model that can handle different climate factors and gain deep insights from the dataset, which in turn would improve decision-making processes.

The following formula describes the Prophet model:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \tag{4.21}$$

Where:

$g(t)$ = piecewise-linear trend;

$s(t)$ = various seasonal patterns;

$h(t)$ = captures the holiday effects; and

ε_t = white noise error term.

4.7 DEPLOYMENT OF MACHINE LEARNING ALGORITHMS IN URBAN WATER SYSTEM MANAGEMENT

For over a decade, machine learning algorithms have been gaining popularity in urban water demand forecasting, compared to traditional stochastic algorithms across all time horizons (Antunes *et al.*, 2018; Xu *et al*., 2019). This is due to their ability to produce high-performance models (Smolak *et al.*, 2020; Kang *et al.*, 2015). Since model accuracy is critical to improving the operation of urban water systems (Bata *et al*., 2020), the search for algorithms capable of developing high-performance models continues, which has led to extensive research and the development of several machine learning algorithms (Pacchin *et al*., 2019; Pesantez *et al*., 2020; Adamowski *et al*., 2012). However, for more than a decade, research on the use of machine learning algorithms for water demand forecasting was mostly pronounced in the Global North (Dogo *et al*., 2019; Sundui *et al.*, 2021). In contrast, it is still in its infancy in the Global South (Carvalho *et al*., 2021; Raj & Kumar, 2022). A major shortcoming is the lack of large datasets needed to use machine learning techniques. However, the data problem is gradually being solved as several government agencies are collecting data thanks to the invention of the Internet. In addition, the looming global water scarcity is forcing the Global South to explore and deploy machine learning algorithms for urban water demand forecasting. An overview of the developments in the application of the machine learning algorithms proposed in this study for urban water demand forecasting is therefore provided, starting with SVR.

4.7.1 SVR

In the Global North, the SVR algorithm has been widely used for short-term water demand forecasting. Researchers who have used the SVR algorithm in the Global North include Herrera *et al.* (2010), who studied the performance of different models in short-term water demand forecasting for a city in south-eastern Spain. These researchers found that models developed with the SVR algorithm outperformed those developed with Multivariate Adaptive Regression Splines, Project Pursuit Regression, and Random Forest. Herrera *et al.* (2014) and Candelieri and Archetti (2014) confirmed the satisfactory performance of the models developed with the SVR algorithm. However, Braun *et al*. (2014) applied the SVR algorithm in predicting the short-term water demand for a district in Berlin and, based on their results, they suggested improving the SVR algorithm's performance.

The inadequate performance of the SVR algorithm in predicting urban water demand has increased over time. Mouatadid and Adamowski (2017) investigated several machine learning algorithms for short-term water demand forecasting in the Canadian city of Montreal, which included SVR. Based on the squared coefficient of determination, mean squared error, and an examination of the residuals, the SVR algorithm did not provide good accuracy compared to other algorithms. Candelieri (2017), on the other hand, applied SVR to detecting short-term water demand and anomalies for the city of Milan in Italy and obtained satisfactory results. However, with the advent of more advanced machine learning algorithms, models developed using SVR show increasingly poor performance. Smolak *et al.* (2020) demonstrated the inadequacy of the SVR algorithm compared to other algorithms in Wroclaw, Poland.

Researchers in the Global South are overhauling machine learning algorithms for predicting water demand by building on work in the Global North. One example is the work of Brentan *et al.* (2017) in Brazil, which highlighted the shortcomings of SVR algorithms and presented a proposal to combine SVR with another algorithm to improve model performance. A high-performance hybrid SVR+AFS model was proposed; that is, a combination of SVR and Adaptive Fourier Series (AFS). Although advanced algorithms for short-term water demand forecasting are being investigated in countries such as China, the SVR algorithm continues to serve as a benchmark. It continues to be used either individually and its performance compared with other algorithms or as a hybrid model in combination with other algorithms to improve its performance (Yan & Yang, 2018; Xu *et al.*, 2019; Mu *et al.*, 2020). Researchers in India also use hybrid models to predict water demand, using SVR in combination with other algorithms (Vijai & Sivakumar, 2018).

In South Africa, Oyebode and Ighravwe (2019) developed SVR in short-term water demand forecasting models for the city of Ekurhuleni in the Gauteng province. The SVM, an SVR used for regression tasks, was used and its performance was compared with single and hybrid models.

The performance of the models was as follows:

ANN-DE > SVM > MLR > ANN-CG

DE = Differential evolution
CG = Conjugate gradient
ES = Exponential smoothing
ANN = Artificial Neural Network
MLR = Multiple linear regression

In this study, the researcher proposed to deployed both a single SVR and a hybrid version of it; thus, combining SVR with the Prophet to give the Prophet-SVR hybrid model; given that in certain circumstances, the single SVR algorithm performance was satisfactory and other hybrid models in which SVR was one of the algorithms exhibited better performance.

4.7.2 XGBoost ensemble model

The use of the XGBoost algorithm in urban water demand forecasting is still in its infancy. However, its robustness in improving model performance is attracting the attention of researchers in various fields, including water management (Xenochristou & Kapelan, 2020). Osman *et al.* (2021) reported the excellent performance of the XGBoost algorithm in groundwater prediction, and Lu and Ma (2020) had similar experiences when the hybrid version Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)-XGBoost was used for short-term water quality prediction. Accordingly, the XGBoost algorithm was proposed in this study.

4.7.3 ANN algorithm

The ANN algorithm, as a data-driven, self-adaptive, and nonlinear forecasting tool, has been widely used for short-term urban water demand forecasting for over two decades (Zhang *et al.*, 2018) because of its ability to capture nonlinear relationships among variables that constitute complex urban water systems and the possibility of its application in constructing a deterministic model of a system about which insufficient process knowledge is available (Gernaey *et al.*, 2004). Since the inception of the basic ANN algorithms in modelling short-term water demand forecasting models, researchers have repeatedly demonstrated their superior performance compared to other algorithms (Jain & Ormsbee, 2002; Jain & Kumar, 2007; Adamowski, 2008; Caiado, 2010; Tiwari & Adamowski, 2013; Huang *et al.*, 2014; Vijai & Sivakumar, 2018). There has been an increased interest in research on ANN algorithms worldwide, which has led to the development of various ANN hybrid algorithms.

Initially, traditional gradient-descent feedforward backpropagation ANNs were widely used in the early 2000s, and they outperformed the popular regression and time series models used at the time to model short-term urban water demand (Jain *et al.*, 2001; Jain & Ormsbee, 2002; Pulido-Calvo *et al.*, 2003; Bougadis *et al.*, 2005; Jain *et al.*, 2001; Adamowski, 2008). Gradually, various configurations of ANN-based algorithms came to the fore, coupled with other modelling techniques. These ANN hybrid algorithms outperformed traditional gradient-descent. Hybrid ANN algorithms included cascade correlation ANNs, Chebyshev ANNs, particle swarm optimisation ANNs, and dynamic ANNs (Heller & Thind, 1994; Chen *et al.*, 2005; Yue *et al.*, 2007; Ghiassi *et al.*, 2008). Thereafter, the search for powerful hybrid ANN-based algorithms for urban water demand became the focus.

Adamowski and Karapataki (2010), in an effort to find solutions to a water crisis in Cyprus, used various forms of ANN-based algorithms and found that the Levenberg-Marquardt ANN model provided the most accurate results. Subsequently, Adamowski *et al.* (2012) applied coupled wavelet – artificial neural networks (WA-ANN) to predict the short-term water demand for Montreal in Canada. These researchers demonstrated the superiority of the coupled WA-ANNs models as they outperformed all models developed by normal ANNs, multiple linear regression, multiple nonlinear regression, and ARIMA. Previously, the hybrid WA-ANN model dominated short-term water demand modelling (Mohammed & Ibrahim, 2012; Campisi-Pinto *et al.*, 2012; Tiwari & Adamowski, 2013; Tian *et al.*, 2016; Ghalehkhondabi *et al.*, 2017). The work of Zubaidi, Dooley *et al.* (2018), Zubaidi, Gharghan *et al.* (2018), Dooley *et al.* (2018), and Zubaidi, Ortega-Martorell, Kot *et al*. (2020) contributed immensely to the development of ANN-based hybrid models for short-term urban water demand forecasting in the Global North.

Recently, the use of ANNs and ANN-based hybrid models for short-term urban water demand forecasting is gaining momentum in different parts of the world. Al-Ghamdi *et al.* (2021) applied ANN algorithms to short-term water demand forecasting in Saudi Arabia and obtained satisfactory results. In Iraq, Rezaali *et al.* (2021) proposed an ANN and compared it with various algorithms such as Least Squares Support Vector Machines, Regularised Extreme Learning Machines, and Random Forest to improve model accuracy. In developing countries, China is leading the way with numerous researchers focusing on improving the performance of ANN and ANN-based hybrid models. Guo and Liu (2018) and Salloom *et al.* (2022) confirmed the high performance of hybrid ANN-based models in predicting urban water demand, which outperform conventional ANNs. Hybrid models based on ANN are also gaining momentum for urban water demand prediction in regions such as India (Vijai & Sivakumar, 2018), Nepal (Shrestha *et al.*, 2020), and Brazil (Carvalho *et al.*, 2021).

In South Africa, the application of ANN-based algorithms for urban water demand forecasting is still in its infancy. There are few reports on the use of ANN algorithms in water demand forecasting. However, it is worth mentioning the work of Msiza *et al.* (2008), which compared the performance of an ANN algorithm with SVM and found better performance of ANNs compared to the SVM algorithm. Since 2019, the use of ANN algorithms or their hybrids has been closely followed in South Africa (Oyebode & Ighravwe 2019; Zubaidi, Ortega-Martorell, Kot *et al*., 2020). The researcher thus deemed it appropriate to include the ANN algorithm in the list of algorithms used in this study.

4.7.4 The Prophet-SVR hybrid algorithm

To the researcher's knowledge, there are currently no reports of the Prophet-SVR hybrid model being used in urban water demand forecasting. However, wherever the Prophet algorithm has been used, the models' performance has been exceptional. For example, Ivanko *et al.* (2020) used the Prophet algorithm to predict the heat consumption of hotels in Norway and obtained exceptionally good results. Also, in China, Guo *et al.* (2021) applied the hybrid Prophet-SVR algorithm to predict time series demand in the manufacturing industry with seasonality. Compared to other algorithms, the Prophet-SVR performed better. Since seasonality greatly impacts water demand prediction, the researcher highly recommends deployment of a Prophet-SVR hybrid in urban water demand forecasting.

4.8 SUMMARY

In urban water management, forecasts and predictions of water demand are critical. The benefits of accurate short-, medium-, and long-term forecasting of urban water demand have been demonstrated. In the past, conventional models were very effective in forecasting and predicting urban water demand. However, as the variables of the urban water system increased in number, becoming highly interconnected and interdependent, an extremely complex system gradually emerged. Coupled with the need to quantify the uncertainties in the system caused by climate change, the inadequacy of conventional modelling techniques in forecasting urban water demand was exacerbated. As a result, algorithms from machine learning have emerged as a preferred option for predicting water demand in the urban water supply system to build high-performance models. Since the introduction of machine learning algorithms, high-performance urban water demand forecasting models have been increasingly produced. The effectiveness of machine learning algorithms stems from the ability to use them as either stand-alone or hybrid models. As a result, developing and studying various machine learning algorithms to accurately predict urban water demand has increased rapidly.

The study found that the development and use of machine learning algorithms for urban water demand prediction are strong in countries of the Global North. Countries in the Global South, such as China, Brazil, and India, are seizing the opportunity presented to researchers in the Global North on this topic. This allows a leap forward as numerous machine learning algorithms have been developed, tried, and tested. There is thus an opportunity for researchers in the Global South to simply build on the work that has already been done in the Global North. In addition, the use of machine learning algorithms is also receiving attention in Middle Eastern countries where there are severe water shortages. Numerous single and hybrid machine learning algorithms are currently being developed and the search for high-performance models is increasing. Researchers have repeatedly demonstrated the superiority of the machine learning models over conventional models.

In South Africa, the use of machine learning algorithms in managing urban water systems is still in its infancy. Researchers working in South Africa’s Gauteng province pointed to the need for further research in other regions with different climatic and socio-economic factors, and recommended that other machine learning algorithms be used either as stand-alone or hybrid models. In a study conducted during the famous "Day Zero" in Cape Town, the researchers, who used machine learning algorithms to predict droughts, pointed to further research focusing on the negative impact of climate change on rainfall, especially in the Southern Hemisphere. The researcher therefore developed several machine learning models and compared their performance with conventional models for predicting urban water demand in the Stellenbosch municipality in chapter 7.

CHAPTER 5: RESEARCH METHODOLOGY

5.1 RESEARCH PHILOSOPHY

This chapter describes the research methodology that formed the basis of this study. Although the terms "research methodology" and "methods" are sometimes used interchangeably, the researcher believes that the two terms should be distinguished in the context of this research project. The chapter begins by reviewing what other researchers have presented. For example, Harding (1987) described research methodology as "the epistemology and interpretive framework that guide a particular research project" and research methods as "techniques for gathering empirical evidence". Checkland (1985) understood research methodology as the principles of methods and procedures formulated and elaborated to solve research problems. In this study, research methodology is understood as a general term that refers to the overall logic and theoretical perspective of the research project. Research methods, on the other hand, are the techniques and procedures used to collect and analyse the research data.

It is essential for any research project that the researcher establishes the philosophy that determines the research methodology. According to Žukauskas *et al.* (2018), a research philosophy is a set of fundamental beliefs that guide the selection of the research strategy, the formulation of the research problem, and the way data should be collected, processed, and analysed. There are four main categories of research philosophies: positivist, interpretivist, pragmatist, and realist (Tamminen & Poucher, 2020). For this study, the researcher chose the pragmatist research philosophy, which allowed the researcher to choose the methods, techniques, and procedures that best meet the requirements and scientific research objectives of the research project being pursued (Alghamdi & Li, 2013). The objectives that facilitated the achievement of the main goal were presented in Chapter 1. To introduce the methodology chapter, Figure 5.1 summarises the research problem and objectives pursued to achieve the main goal of the research.

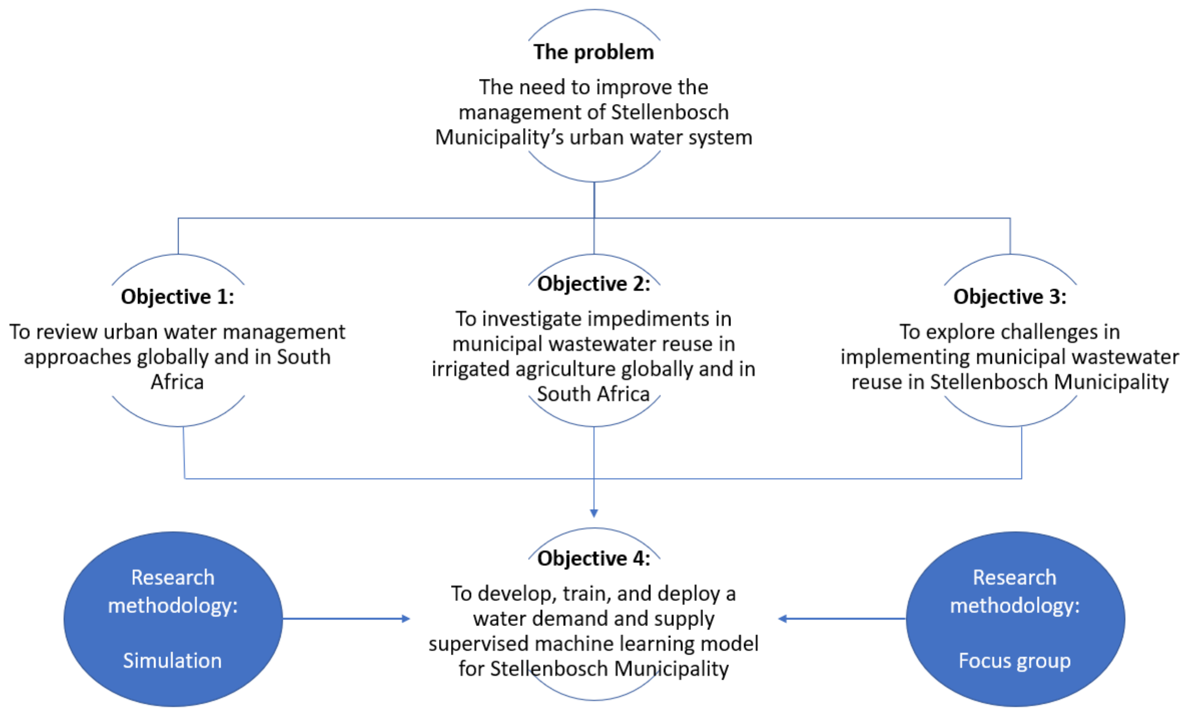


Figure 5.1: Summary of research objectives

The problems associated with urban water system management in general were discussed in Chapter 1, which was followed in Chapter 2 by a literature review of the evolution of the approach to urban water management systems at both the global and local (South African) levels. Reusing treated municipal wastewater in irrigated agriculture at the international level was the focus of Chapter 3; the reason being that to achieve a sustainable urban water system, firstly, both water demand and supply must be managed efficiently. Secondly, the unprecedented impacts of climate change on freshwater availability necessitate a shift to alternative water sources to reduce dependence on precipitation. To this end, the reuse of treated municipal wastewater has emerged as a feasible alternative water source. Chapter 3 therefore examined the use of treated municipal wastewater in irrigated agriculture internationally – considering that irrigation is the largest consumer of freshwater in the world. The increasing complexity of managing urban water systems was highlighted in Chapter 4. This highlighted the need for a management approach that captures and interprets the growing number of interrelated and interdependent variables that make up an urban water system. This includes the ability to draw insights from large datasets and the quantification of uncertainties created by climate change.

In Chapter 4, machine learning models capable of addressing the above requirements were reviewed. In this study, the use of machine learning in urban water system management was demonstrated through a case study. The objective of this chapter is twofold. Firstly, the research design for the study is discussed and, secondly, the research methods and procedures to achieve the main goal of the study are presented, namely to develop a strategy using technology for sustainable management of Stellenbosch Municipality's urban water system. Stellenbosch Municipality is a water authority whose jurisdiction includes several small towns, including the town of Stellenbosch, as a case study. These towns are located in the Western Cape province of South Africa.

The rationale for using model development and case study methods stems from the transdisciplinary nature of this study, which was justified in Chapter 1. This chapter thus first discusses the research design of this study. Next, all methods and procedures used in each phase of the research are discussed, and finally, a summary of the chapter is presented.

5.2 RESEARCH DESIGN

As described in previous chapters, the management of urban water systems is currently characterised by a variety of challenges that occur simultaneously and aggressively. To effectively understand these challenges and find appropriate solutions, the researcher employed a transdisciplinary research methodology. It emerged in the early 1970s and was coined by Piaget (1972) (López-Huertas, 2013; Nicolescu, 2010), at a time when scholars were critiquing the standard configuration of knowledge in the disciplines in their curriculum, including moral and ethical concerns (Bernstein, 2015). The focus was on issues of epistemology and the planning of future universities and educational programmes (Mahan, 1970; Kockelmans, 1979). Thereafter, the transdisciplinary research discourse fell dormant for two decades and did not re-emerge until the 1990s (Kessel & Rosenfield, 2008), which was due to the emergence of highly complex global problems resulting from the adverse effects of climate change and the growing need for sustainable development approaches to mitigate malignant problems. These challenges could not be addressed within disciplinary boundaries or with traditional empirical methods. They required research methods capable of providing solutions that could be sustainably integrated into the triad of science, technology, and society and sequentially inform policy formulation and decision making to achieve sustainable development (Klein, 2001).

Over the years of using multi- and interdisciplinary research methods, a gap has likely emerged that Westley *et al.* (2011) referred to as the "ingenuity gap", which is the gap between the world's ever-growing challenges and the effort to find timely and appropriate solutions. McGregor (2012) referred to the transdisciplinary research methodology as a solution to bridge the "ingenuity gap". Hadorn *et al*. (2008) also highlighted the complexity of real-world problems as a driving force for rethinking the transdisciplinary research methodology. These researchers emphasised the need to use the transdisciplinary research methodology in situations where knowledge about a societal problem is uncertain and contested and if not resolved in a timely manner could have catastrophic consequences. It was believed that the transdisciplinary research methodology would enable researchers to holistically grasp, recognise, and take note of multiple worldviews and scientific perceptions regarding a real-world problem. In this way, researchers would be able to find appropriate and relevant solutions to particular, case-specific, real-world problems.

However, in the search for the research methodology, transdisciplinary and extensive research has been conducted on its definition and application. Accordingly, several researchers made efforts to define transdisciplinary research. In this study, the researcher took Mittelstraß (1992) as a starting point, to consider transdisciplinarity as a research methodology that is not bound to a specific discipline. It can define and provide solutions to real problems, regardless of the disciplines associated with the problems. In 1994, at the First World Congress on Transdisciplinarity in Portugal, transdisciplinary research was adopted as a methodology informed by the new sciences of quantum theory, chaos theory, and living systems theory (Klein, 2004; Nicolescu, 2006).

The researcher acknowledges Nicolescu's (2008; 2007; 2005; 2004; 2002) work on transdisciplinary methodology, which explicitly discussed the characteristics of this research methodology. Several descriptions are given with terms such as "trans", which include zigzag, cross-over, and crossing boundaries. In addition, meanings of terms such as "mono" for one, "multi" for more than one, and "inter" as between were pronounced. This led to the explanation of what mono-, multi-, and interdisciplinary research methods were. Accordingly, these were defined as research methods in which the research project is pursued strictly within the boundaries of the particular discipline, either exclusively within a particular discipline or by forming collaborations between different disciplines within the disciplinary boundaries without interacting with the rest of the world. In this case, all research activities would be confined to the walls of the university (Nicolescu, 1997).

Similarly, McGregor (2004) also sought to distinguish between multi- and interdisciplinarity, and described multidisciplinary research as a research approach characterised by members of a research team working within the boundaries of their specific disciplines on a problem of common interest. On the other hand, interdisciplinary research methodology allows multidisciplinary research teams to collaborate, communicate with one another, and integrate the team's research findings without removing the boundaries between the disciplines involved. Subsequently, Nicolescu (2007) extended his work to transdisciplinary methodology by building on McGregor's (2004) definition of multi- and interdisciplinary research approaches to explicitly define the transdisciplinary methodology. He emphasised that the team's work is highly organised and guided by broad constructs and methods that transcend disciplinary structures and conventions, along with the understanding that transdisciplinary teams evolve into a community of researchers working for a common cause, not just a collective (as in multi- and interdisciplinary research approaches). Nicolescu (2008; 2007) also presented his transdisciplinary description, which included three axioms, multiple levels of reality and the hidden third, the logic of the included middle, and epistemology, which is knowledge as an emergent complexity.

However, Cicovacki (2009; 2004) argued for a fourth axiom, namely value theory. He alluded that value can provide an axis of orientation for life, attitudes, and actions in decision making. To support his argument, Cicovacki (2004) echoed Nicolescu's (1997) assertion that the transdisciplinary methodology is "a path of self-transformation directed toward the knowledge of the self and the creation of a new art of living". Similarly, Glasser (2006) argued that because of the concern for the reality level in transdisciplinarity, it is necessary to pay attention to what people see as valuable to themselves.

Furthermore, Scholz *et al.* (2006), in their definition of transdisciplinary research methodology, summarised its facets and explicitly articulated its advantages over other research methods, which include:

- increased likelihood of finding relevant solutions to real, complex societal problems;
- the ability to complement traditional disciplines and interdisciplinary scientific activities by integrating non-disciplinary stakeholders (society); and
- effective facilitation of mutual learning processes between science and society (mutual learning) – science is thus done with society and not for society.

Pohl and Hadorn (2008), in their description of the transdisciplinary research methodology, highlighted how it transcends and integrates disciplinary paradigms, incorporates participatory research, and seeks unity in knowledge. In an effort to strengthen the transdisciplinary research methodology, Hadorn *et al.* (2008) argued that it does not aim to reject scientific knowledge. Instead, they argued for the unity of knowledge among multiple subject matter experts and non-disciplinary stakeholders to reshape the concept of science and the distinctions of science in solving complex real-world problems. McGregor and Volckmann (2013) categorised transdisciplinary research methodology into two categories: firstly, as an exclusive focus on collaborative problem solving rooted in the science-technology-society triad, and, secondly, recognised as a distinct methodology.

This study was guided by McGregor and Volckmann's (2013) basic classification of the transdisciplinary methodology, namely exclusive focus on joint problem solving anchored in the science-technology-society triad. The study intended to use transdisciplinary methodology not only as a method but rather as a research approach to explicitly address a contemporary and complex societal problem involving multiple disciplines and non-academic stakeholders. Given the transversal nature of urban water system management and its ever-increasing complexity, mono-, inter-, and multidisciplinary research approaches are insufficient. In contrast, the transdisciplinary research approach has proven to be robust in studying complex real-world problems because it allows for a collaborative process between scientists and non-scientists on specific real-world problems, such as urban water system management. It represents a more robust research approach that can fill the gap left by multi- and interdisciplinary research approaches (Blättel-Mink & Kastenholz, 2005). Solving problems related to urban water system management also requires solutions outside of academia (Decker & Fleischer, 2010). The transdisciplinary research approach is critical as it allows for the integration of "inside" (academics), "outside" (non-academics), and the researcher to participate equally and actively in the research process.

In the collaboration between academics and non-academics in a transdisciplinary research approach, Mobjörk (2010) identified two types of transdisciplinary collaborations, namely consultative and participatory transdisciplinarity. In consultative transdisciplinarity, the contributions of non-scientific actors in knowledge production are limited. In participatory transdisciplinarity, on the other hand, the contributions of non-scientific actors are fully included in the process of knowledge generation. Following Mobjörk (2010), participatory transdisciplinarity means involvement in the entire research process, while consultative transdisciplinarity means involvement in a part of the research process. This involvement can take place during problem definition and problem posing.

5.2.1 Ontology

Following the great philosophers of antiquity, such as Aristotle, ontology was considered a branch of metaphysics concerned with the nature of being. Subsequently, several scholars endeavoured to define ontology in their respective fields. However, for this study, the simple definition of ontology, namely "the science of being", is used, while acknowledging the variability of the definition across fields. In the context of the transdisciplinary research approach, ontology seeks to answer the questions posed by the researcher in conceptualising the problem under study (Scholz *et al.*, 2006). As a result, the characteristics of transdisciplinary ontology remain highly contested (McGregor, 2012).

However, by its very nature, transdisciplinarity is used in research to find relevant solutions to real-world problems, but classifying these problems within disciplinary science is inadequate. The ontological axiom that justifies the use of transdisciplinary research approaches is based on the understanding that there are different levels of reality in nature and knowledge about nature that correspond to varying levels of perception (Nicolescu, 2006). These different levels of reality and perception are considered complementary by Nicolescu (2006). The transdisciplinary research approach therefore uses this complementarity to realise comprehensive perspectives on the reality of a problem. Considering the different levels of reality and perception of a given grouping, it is impossible to use conventional expertise and professional knowledge and arguments. This is because they have no value and affect the ability to determine the causes of the problems in the world. This is a scenario that leads to conflicting perceptions or understanding of problems under study and consequently limits the ability to find relevant and timely solutions to problems (Funtowicz & Ravetz, 2008). However, Hartman (1967) argued that giving priority to values in finding solutions to a problem does not mean that people do not have values, but rather provides an opportunity to be aware of confrontational values in solving a social problem. In this way, the problem-solving process is strengthened as the unique patterns of individuals are observed and compared to the patterns of others.

In essence, the transdisciplinary approach is well suited to deal with complex, “wicked” real-world problems (Pohl & Hardon, 2008). In the context of this study, the challenges of managing urban water systems qualify as “wicked” problems (Batie, 2008). Because multiple variables that make up the system are recurrent and increasingly interconnected and interdependent, solving one of the variables, if addressed in isolation, can affect other system components. For example, if demand management is given great importance, insufficient water sources could eventually occur due to climate change. Both the demand and supply sides must be managed in a sustainable manner. To achieve this desired outcome, the question is what tools are robust enough to enable water agencies to sustainably manage ever-increasing urban water demand in the face of shrinking supply, with increasing uncertainties.

In addition, ontological considerations allow the nature of the phenomenon that the researcher is studying to unfold (Scholz *et al.*, 2006). This study used a case of urban water management in Stellenbosch Municipality in the Western Cape province of South Africa. Supervised machine learning modelling techniques were used to develop models for predicting and forecasting water demand and supply over short- and medium-term time horizons. Statistical data on Stellenbosch Municipality were obtained from various government agencies on the following main areas: water demand, water supply and distribution, demographics, and weather patterns.

5.2.2 Epistemology

Cohen *et al.* (2007) defined epistemology as assumptions about "the foundations of knowledge – its nature and form, how it can be acquired, and how it can be communicated to others". These authors pointed out that the epistemological assumptions we make or hold about knowledge fundamentally affect the methodology used to decode social behaviour; that is, researchers decide which methods to use depending on their epistemological assumptions; for example, if knowledge is viewed as hard, objective, and tangible, the researcher adopts an observer role while using scientific methods such as tests and measurements to answer the research questions. However, when knowledge is seen as personal, subjective, and unique, the researcher is forced to reject natural science methods and use methods that allow for the greater involvement of their research subjects (Al-Saadi, 2014).

In this study, the researcher closely followed Scholz *et al.*'s (2006) definition of epistemology, which refers to the science of generating, integrating, and using knowledge with particular attention to structure, scope, and validity. Given the above definition, three forms of knowledge characterise the transdisciplinary research approach: systems knowledge, target knowledge, and transformation knowledge (Pohl & Hadorn, 2008). Figure 5.2 summarises these forms of knowledge and the respective research questions they seek to answer.

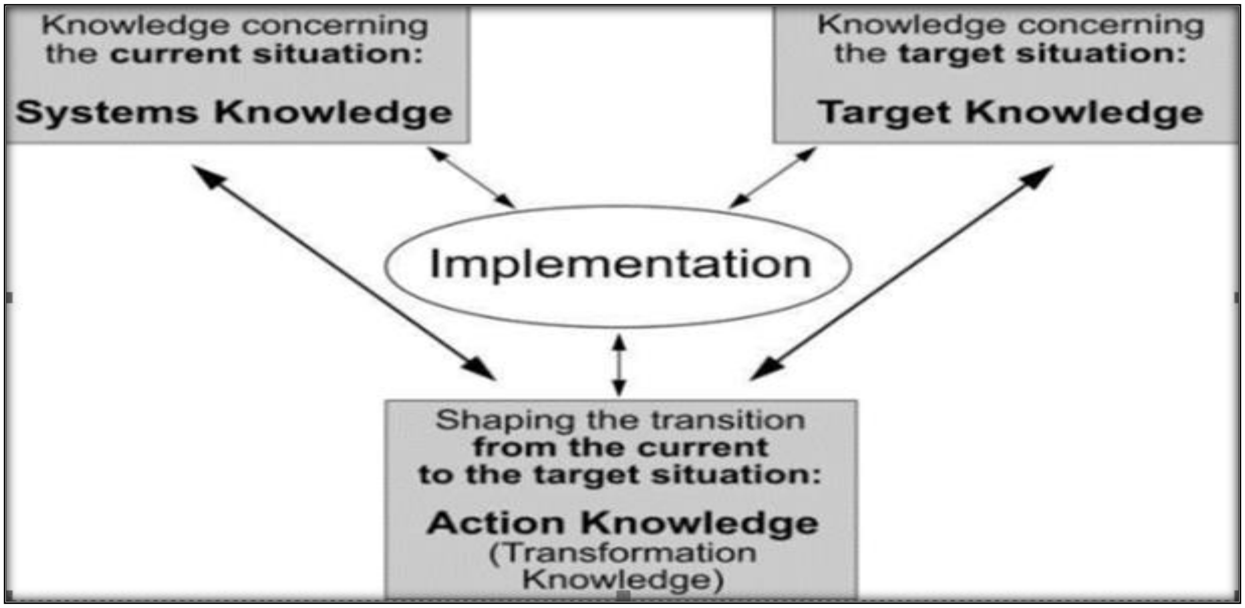


Figure 5.2: Types of knowledge in a transdisciplinary context

Source: Messerli and Messerli (2008)

In the context of this study, the three types of knowledge depicted by Messerli and Messerli (2008) in Figure 5.2 were examined in the transdisciplinary domain. Thus, in order to generate systems knowledge, questions were formulated about the origins of current problems, the possible evolution of these problems, and their interpretation in the context of Stellenbosch Municipality's urban water system. The result was the identification of variables that are assumed to influence the sustainable management of the urban water system. In determining the effects of the independent (predictor) variables on the dependent (response) variable, the researcher was provided with the systems knowledge of Stellenbosch Municipality's urban water system. Once the systems knowledge was established, the researcher was in a good position to develop questions that would explicitly identify the severity of problems in the system. This allowed the researcher to communicate to the water authorities the need for change and what better practices should be put in place for the sustainable management of the urban water system.

Based on the target knowledge needed to solve the identified system management challenges, the use of supervised machine learning modelling techniques could be advocated. The result is the development of supervised machine learning models for forecasting and predicting the water demand and supply of Stellenbosch Municipality. By using the developed model, Stellenbosch Municipality will be provided with the transformative knowledge required to improve its decision-making processes in managing the urban water system (transformative knowledge).

To implement the transdisciplinary research approach, the researcher involved multiple disciplines, including water law and policy, socioeconomics, agriculture, public health, engineering, natural sciences, and machine learning (as a subfield of AI). This was accomplished by incorporating concepts and methods from the aforementioned disciplines, along with non-academic experts who participated in the research. Figure 5.3 provides an overview of the disciplines and non-disciplines involved in this study.

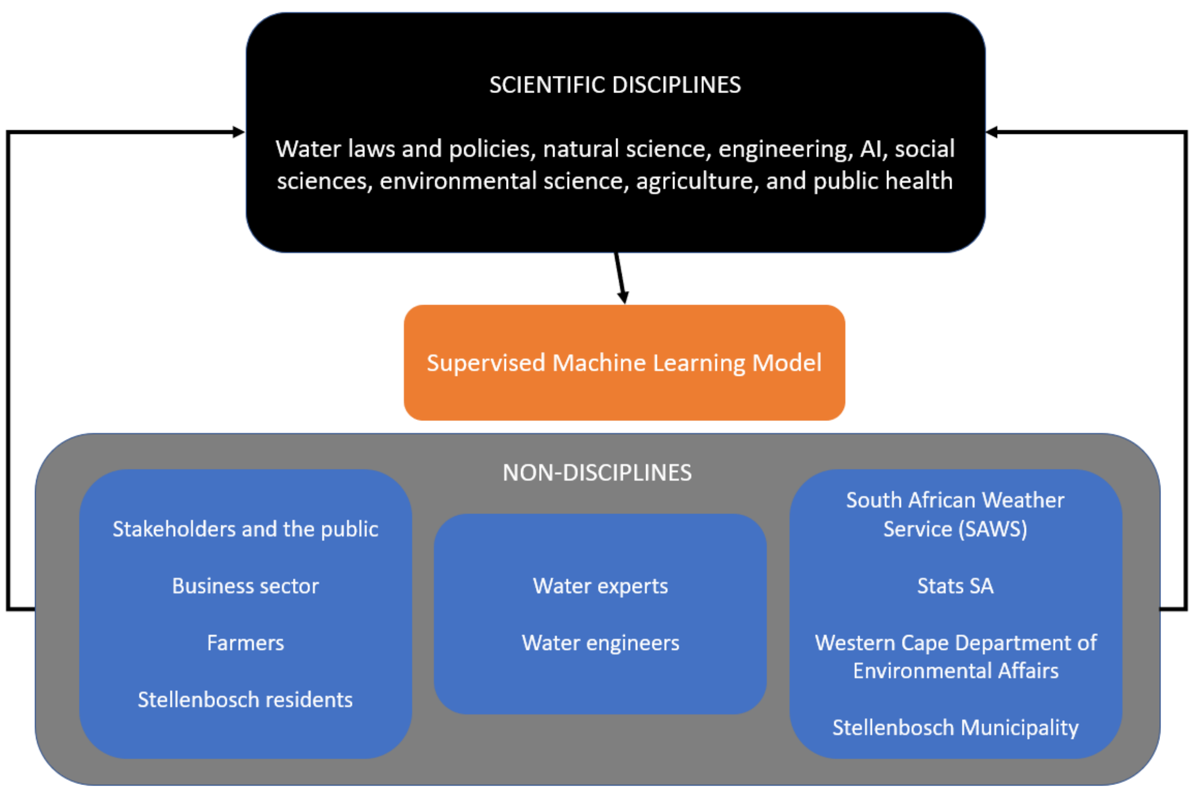


Figure 5.3: Summary of expertise and disciplines involved in this study

Since the research took a transdisciplinary approach, the researcher involved the following disciplines: law, politics, natural sciences, technology, machine learning, social sciences, environmental sciences, and agriculture. The reason for this is that Stellenbosch Municipality is a water authority that is supposed to enforce South African water laws and policies at the local level, manage the urban water system and deliver water services to its communities efficiently according to the South African Constitution. (Saleth & Dinar, 2004). However, deficiencies in water policies and laws usually have a negative impact on the management of urban water systems. Water laws and policies were examined and participants knowledgeable in these areas were part of the focus group and consultations. In addition to water institutions, the management of an urban water system is also affected by the infrastructure and management of water demand and supply. To this end, the researcher involved water engineers, used machine learning techniques, and considered environmental aspects. The water management approach examined in this study involved extensive stakeholder and public consultation in the case study area. These are reflected in the non-disciplinary groupings that participated in the research, including several government agencies that influence the management of Stellenbosch Municipality's urban water system.

This research approach typically requires multiple modes of explanation, as knowledge about the research question(s) is gained from various stakeholders; ranging from multiple scientific disciplines to non-disciplines. It is believed that these explanatory approaches complement one another and allowed the researcher to gain a more comprehensive perspective on the reality of the research problem. A significant drawback of transdisciplinary research approaches is the challenge of validating and integrating non-scientific knowledge into a scientific enterprise. To this end, the researcher employed technology (Concept Star decision-making tools for professionals) to facilitate the capture and integration of knowledge from both scientific and non-scientific sources. The details of the method are presented in the next section.

5.2.3 Methodology

The transdisciplinary research approach was the overarching research methodology in this study because it allowed the researcher to choose methods from different research traditions. Other methods used in the transdisciplinary research process included a critical systematic literature review, interactive management, simulation, standard cross-industry process for data-mining research, and a case study. These methods were considered robust enough by the researcher to achieve the main objective of the study, which is to develop a supervised machine learning model capable of predicting and forecasting the urban water demand and supply of Stellenbosch Municipality under different scenarios. This will provide water supply authorities with a toolkit to manage their urban water system sustainably. In addition, the goal of using the developed model ensured that the researcher turned the research project from a purely academic exercise into a transformation of the management practices of Stellenbosch Municipality's urban water systems. The following subsections describe how the above methods were used in this research.

5.2.4 Organisation

Scholz *et al.* (2006) described the organisation of a research project as the general procedures and organisational framework of the research project. Although the transdisciplinary organisational framework is not fully developed, a general interactive and constructive procedure has been developed to guide the process of inclusivity during transdisciplinary research projects (i.e., disciplinary, and non-disciplinary actors). Accordingly, Flinterman *et al.* (2001) presented a procedure that should be used in all transdisciplinary research projects, which is as follows:

- Definition of a research field;
- Identifying and contacting all relevant stakeholders;
- Literature review;
- In-depth interviews with participants;
- Discussion rounds or focus groups;
- Interactive workshops;

- Repeated feedback on all types of outcomes by all participants; and
- Development of shared constructs and a holistic vision.

In this research, the above elements of the transdisciplinary process were carried out through participatory and consultative collaboration. The research questions were formulated during the transdisciplinary summer school, which was led by the Transdisciplinary, Sustainability, Analysis, Modelling and Assessment (TSAMA) Hub 15 at Stellenbosch University. In addition, the researcher took six modules during the first year of research to understand the transdisciplinary research approach and to improve the formulation of the research problem. During the same period, the researcher identified and contacted all relevant stakeholders, including non-discipline stakeholders, consisting of community associations, water experts, businesses, and farmers, in Stellenbosch, as well as the Western Cape Government Department of Environmental Affairs, Statistics South Africa (Stats SA), and the South African Weather Service (SAWS). Specific multidiscipline participants came from the following fields (engineering, law and policy, natural scientists, AI, social sciences, environmental science, agriculture, and public health. This was accomplished through public meetings, formal workshops, and face-to-face discussions. The discussions at these meetings formed the contributions of disciplinary and non-disciplinary stakeholders to the problem formulation and research processes. The following transdisciplinary activities, i.e., literature review, focus group, interactive workshop, repeated feedback on all types of findings by all participants, and the development of shared constructs and a holistic vision, were also conducted. The details of these activities are described in Chapters 2, 3, 6, and 7.

Stellenbosch University, like any other traditional university, remains strictly organised in disciplinary structures and this study was conducted within the TSAMA hub. According to Stellenbosch University's rules and policies, a PhD student must be enrolled in a specific faculty or department where specific supervisors are willing to participate in transdisciplinary research projects. This study was therefore enrolled in the Faculty of Military Sciences. TSAMA supports transdisciplinary researchers by providing a platform that facilitates the process of crossing disciplinary boundaries and transcending disciplines through the organisation of core pedagogical modules, of which the researcher took six, as mentioned above. These modules were Development Theory and Practice, Facilitation for Sustainability Transitions, Sustainable Cities, Complexity Theory and Systems Thinking, Applied Economics, and Policy and Law. TSAMA also hosts forums where transdisciplinary PhD students can discuss and share the successes and challenges of their research projects. The researcher participated in these forums and improved her understanding of the transdisciplinary methodology. Other research methods used in this study as part of the transdisciplinary approach are considered below.

5.2.4.1 Case study research methodology

Since its beginnings in 1900, the case study research method has evolved greatly from its exclusive use in anthropology, to its use in various disciplines, to its explicit evolution toward eclecticism and pragmatism (Patton, 1990). After World War II, however, the case study methodology was heavily criticised; only to re-emerge in the 1990s as an explicit and comprehensive research methodology. Accordingly, several scholars have sought to define a “case study”, and despite differences in definition, Yin (1994), Merriam (1998), and Gillham (2000) have reached consensus on the following: a “case” is an object of study characterised by a complex functional unit that is to be studied in its natural context through multiple methods in its present form. Accordingly, the researcher adopted Yin’s (1993) definition of a “case study,” which is “an empirical investigation that examines a contemporary phenomenon in its real-world context and deals with a situation in which the boundaries between phenomenon and context are not obvious”.

Several drawbacks are associated with the case study methodology. These include being limited to one or a handful of examples and, in the statistical domain, the limited number of data points that can be used (Yin, 2003; Flyvbjerg, 2006). Gummesson (1988) also mentioned another disadvantage, which is that the researcher must spend a great deal of time collecting basic information due to the lack of prior understanding of the case. Despite numerous disadvantages, the case study research methodology is still considered to be highly robust in answering the “how” and “why” questions related to a range of contemporary events under study (Leonard-Barton, 1990; Meyer, 2001).

Despite the drawbacks, there are also several advantages that led to the use of the case study method in this study. These include the advantages outlined by Gummesson (1988), which include that it allows the researcher to examine multiple aspects of the research problem and to consider the factors that influence the problem in the context of the problem's environment. Other advantages include the ability to examine the context and other complex conditions related to the specific case, which is essential for a comprehensive understanding of the unit of analysis (Yin, 2003). In addition, the case study is likely to obtain data from multiple sources, which is desirable because it strengthens the validity of the research findings (Yin, 2009). In short, applying the case study methodology in this study provided an in-depth and holistic understanding of a single complex unit of analysis in its real-world context, i.e., the Stellenbosch urban water system (Feagin *et al.*, 1991; Bromley, 1986).

5.2.4.2 The use of a case study approach and challenges

In this section, the researcher discusses the case study research approach and addresses some of its challenges. The case study research approach is generally considered robust because it requires a holistic and in-depth understanding of the research problem (Feagin *et al*., 1991). This research approach is widely used in all disciplines and several procedures for its use have been discussed (Yin, 1994; Stake, 1995). Yin (1993) presented explanatory, exploratory, and descriptive case study methodologies, while Stake (1995) also presented the following case study methodologies: when the researcher has an interest in the case study (intrinsic), when more than one case study is used in the study (collective), when a group of case studies is studied, and when the case helps the researcher gain a better understanding than is apparent to the observer (instrumental).

This study used Yin's (2003) case study methodology, which can be either a single case or multiple cases. A single case was used in this research. It focused on Stellenbosch Municipality, a water authority that serves Stellenbosch as the main town and other surrounding smaller towns in its jurisdiction. The unit of analysis was the management of the urban water system. Eisner (1998) pointed out the use of a case study approach to understand complex real-life problems involving social phenomena even though the case study approach does not involve sampling (Feagin *et al.*, 1991; Yin, 1994). The reason for using a case study approach in this research was to evaluate the interconnectedness of variables in the system and to be able to determine the effects of an intervention (Yin, 1994). The result is solutions that can effectively improve the sustainable management of Stellenbosch Municipality's urban water system.

Dyer and Wilkins (1991) expressed concerns about the case study approach because it offers generalised explanations. In this regard, Yin (1994) outlined the difference between statistical and analytical generalisation and stated that it is unusual to generalise scientific data. Flyvbjerg (2006) argued that the knowledge gained from a case study should be included in the collective process of knowledge accumulation in a particular field. Eisner (1998) argued that knowledge transfer occurs during a process of critical engagement when ideas appear to the reader. Its use in this study is therefore well warranted.

5.2.4.3 Participatory and consultation approach

In Section 5.2.3, the researcher reports on the participatory and consultative processes used in the transdisciplinary research approach of the study. The participatory component of the research was achieved in part through the organisation and delivery of workshops, including a water indaba held at Spier on 13 November 2015. The workshop's main objective was to understand the water management concerns within the case study delineation. It was also intended to assist the researcher to formulate research questions and to gain a deeper understanding of the phenomenon in the case study context. Details of the indaba are attached as Appendix A.

During the workshop and further consultations with the proposed stakeholders from the workshop resolutions, key challenges in the management of Stellenbosch Municipality's urban water system were identified and these findings served as the basis for the formulation of the research problem and objectives of this study. The challenges identified included the detrimental effects of inefficient municipal wastewater management in the city. A significant portion of the study was thus dedicated to municipal wastewater management and reuse in Stellenbosch. In addition, a global perspective on municipal wastewater reuse in irrigated agriculture was considered in Chapter 3, which culminated in the publication of a book chapter. In summary, the workshop participants discussed the following issues that should be considered in the strategy of improving the management of Stellenbosch Municipality's urban water system:

- The negative impacts of climate change that lead to a change in the rainfall cycle were discussed.
- The need for alternative water sources was highlighted and two options were presented: groundwater extraction and the reuse of treated municipal wastewater.
- Reducing water losses in the network.
- Issues of water policy, laws, and administration were addressed.
- It was emphasised that robust forecasting and predictive tools for water demand and supply are needed to prevent catastrophic water shortages as demand increases and supply shrinks due to climate change.

Overall, the workshop and consultations highlighted the phenomenon of the reuse of municipal wastewater. This was also advocated in the discourse on alternative water sources during the research project, which gained momentum due to the devastating drought that prevailed at the time of the research project. In addition, predictions that indicated an emerging significant water shortage by the year 2040, triggered by climate change in the region, led to extensive consideration of municipal wastewater reuse as a possible alternative water source to augment the water needs of Stellenbosch Municipality, given its geographic location. Accordingly, farmers showed great interest in reusing municipal wastewater but also expressed great concern about the quality of the treated municipal wastewater produced by local authorities at the time. As for the water supply, authorities were still struggling with the general challenges of municipal wastewater treatment and management. However, the farmers were offered two options by the water authority:

- The water authority takes care of the whole process and provides the infrastructure for further treatment of the treated municipal wastewater to meet the farmers' quality requirements.

- The water authority delivers standard treated municipal wastewater to the farmers and the farmers provide their infrastructure and further treat the municipal wastewater to meet the water quality standards of their respective operations.

Subsequently, an interactive management workshop was held in March 2018 at the main campus of Stellenbosch University with selected stakeholders and the public. The workshop addressed the research objective of investigating the barriers to the implementation of municipal wastewater reuse in Stellenbosch Municipality. Details of the methodology of how the workshop was conducted are provided in the following subsection.

5.2.4.4 Interactive management research methodology

A soft systems thinking methodology was used to achieve the above objective of the interactive management study. This research methodology is well suited for addressing complex challenges such as those encountered in managing urban water systems. The relevance of interactive management to this study stems from its ability to facilitate the grouping of disciplinary and non-disciplinary participants and allowing for brainstorming, engagement, and knowledge sharing among participants. The process allows for a comprehensive diagnosis of fundamental issues to be addressed by the research question(s), including their interrelationships and interdependencies, as participants from different fields interact.

The interactive management methodology includes two components that must be effectively managed to achieve the desired outcome, namely the topics to be studied and the focus group. To effectively manage a diverse focus group, the interactive management methodology provides a structured method by creating an environment that allows for collaboration and modelling of ideas from each individual within the group (Warfield & Cárdenas, 1994). Figure 5.4 shows the interactive management triad, which is composed of Issues (the research problem), Team (the focus group participants and the management team), and Tools (Concept Star decision-making tools for professionals).

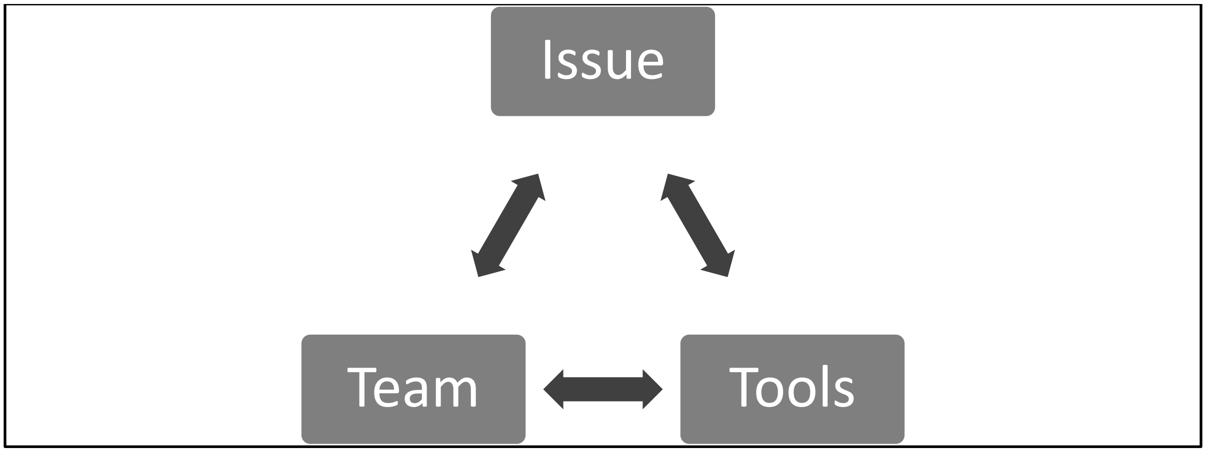


Figure. 5.4: The interactive management triad

Source: Researcher

Warfield and Cárdenas (1994) presented the interactive management process, which consists of the following three phases:

(1) Planning phase

In this phase, the researcher organises the following elements:

- Finding a suitable location for the workshop.
- Finding suitable participants.
- Organising interactive management staff, budget, equipment to be used, and a schedule for the workshop.
- Formulating the research question(s) and context statement to guide the process and to create the model.
- Defining the objectives of the interactive management process and the process flow.
- Formulating triggering and generic questions.

(2) Workshop phase

In the workshop phase, the identified participants meet; in this case, the researcher was the interactive management practitioner and organiser. The interactive management practitioner and interactive management staff generate, capture, structure, and interpret the participants' key ideas (Jackson, 2003) using Concept Star decision-making tools for professionals. Modelling techniques such as interpretive structural modelling, idea writing, and the nominal group technique may be used. An odd number of participants was preferred in this study to allow for voting decisions during the modelling phase.

(3) Follow-up phase

The follow-up phase may involve repeating the interactive management process, implementing recommended designs and alternatives, or combining both. However, in this study, a follow-up phase was unnecessary because the results obtained in the first phase answered the research questions and achieved the objectives.

The interactive management process has several strengths and weaknesses. The interactive management method used in this study provided a platform for integrating knowledge from different disciplines and non-disciplines. It described a guide for integrating and implementing the ideas that emerged from the focus group. In addition, interactive management is a well-structured process that allows the researcher and a group of several discipline and non-discipline participants to work together synergistically. During the workshop phase of the interactive management process, there is an opportunity to discuss the participants' unique perceptions or opinions on the topic. Consensus-building tools and methods can be used to create a shared understanding of the issues, their implications, and their relevance to the study. Generally, interactive management is recommended as a well-articulated methodology that promotes learning and generates genuine participant engagement (Jackson, 2007).

Interactive management has also come under criticism. Researchers argue that the methodology's recommendation to work with small groups is limiting, and that the entire process is cumbersome (Jackson, 2007). The quality of the interactive management outcome depends largely on skilled interactive management process leadership in the form of the facilitator (Warfield & Cárdenas, 1994), which could be difficult to achieve.

5.2.4.5 Sampling size and technique

The sample size recommended for the interactive management process is between seven and 13 participants (Janes, 1988). For this study, the sample size was 11.

5.2.4.6 Decreasing non-sampling error

It is assumed that non-sampling errors stem from the instruments used for data collection, not from the sample itself. For the intent and purposes of this study, the researcher was required to complete training on modelling tools and facilitating sustainable transitions offered by the Institute of Sustainability at Stellenbosch University. This training enhanced the researcher's previous experience with facilitating academic events and workshops, which allowed her to conduct a successful interactive management workshop.

In addition to the workshops conducted, the researcher held one-on-one meetings with relevant stakeholders, communities, and water agencies from 2015 until the end of the study. The combination of participatory workshops and consultations with relevant stakeholders helped the researcher to understand the research problem and the attitudes and perceptions of different stakeholders, the public, and water authorities regarding water issues in Stellenbosch Municipality.

Researchers such as Moran-Ellis *et al.* (2006) and others recommend triangulation of data, i.e., using multiple approaches when examining a study, to validate the research findings. Accordingly, in addition to the above activities, statistical data were collected from various government agencies for the study's simulation. The research methodology used for the simulation is considered below.

5.2.4.7 Simulation

Computer simulation is becoming increasingly important as a methodological approach to organisational research. Simulation research assumes the intrinsic complexity of a system under study. While other research methods attempt to answer the “what,” “how”, and “why” of a research question, simulation attempts to answer the “what if” question. According to Bradley *et al.* (1987), simulation involves developing a model of a system with appropriate inputs to produce the desired outputs. The simulated data come from a source where they are extensively processed according to specific rules (data wrangling), rather than being measured directly in the real world. Among the advantages of simulation is its robustness in supporting intuition and in studying complex systems. This is because it is capable of making predictions and forecasts under certain conditions (Dooley & Lenihan, 2005).

The simulation research methodology was used to achieve the fourth objective of this study, which was to develop, train and deploy a highly accurate water demand and supply prediction and forecasting model for Stellenbosch Municipality under certain conditions. The prediction simulation uses the created, trained, and deployed model using old data to produce output results (observed results under the specified conditions). By comparing multiple outputs for given inputs, researchers can infer possible real-life outcomes if certain actions were taken. However, the accuracy of the output results depends solely on the performance of the model (accuracy).

The selection of the algorithms to be used is crucial as it greatly affects the performance of the model. Accordingly, the conventional algorithms proposed were VAR and SARIMA. For supervised machine learning, the SVR, XGBoost, and Regular Neural Networks algorithms were proposed including a hybrid model combining Prophet with SVR subject to the data that would finally be provided. The proposal of these algorithms was influenced by the nature of the study, which included the prediction of continuous values and the desired high accuracy in producing the output values. The literature review on the use of supervised machine learning algorithms in urban water management in Chapter 4 also contributed to the proposal of these particular algorithms for this study. To achieve the fourth objective, the Cross-Industry Standard Process for Data Mining (CRISP-DM) research methodology was applied, which is briefly summarised below.

5.2.4.8 The Cross-Industry Standard Process for Data Mining (CRISP-DM) research methodology

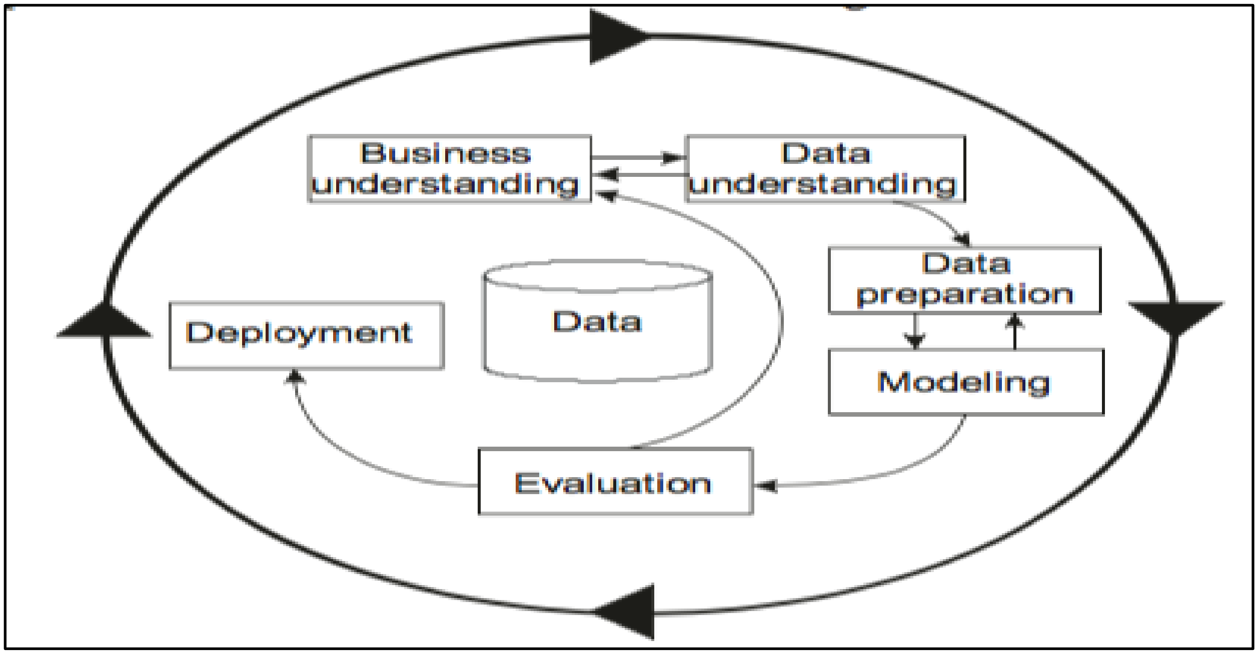


Figure 5.5: The CRISP-DM research methodology

Source: Wirth and Hipp (2000)

This methodology involved collecting statistical data from appropriate entities and processing that data according to the CRISP-DM methodology shown in Figure 5.5. The CRISP-DM entails six phases that describe the life cycle of a machine learning project. It describes the process of planning, organising, and implementing a machine learning project as follows:

1) Understanding the research question: What is the research problem to be addressed? This phase includes determining the research objectives, success criteria, research plan, and deliverables.
2) Understanding the data: What data are available, what data are needed, and what are the data-cleaning requirements? Initial data collection, data descriptions, and explorations are conducted during this phase.
3) Data preparation: The researcher examines how the data are organised for modelling. This includes data cleaning, sampling, normalisation, and feature selection.

4) Modelling: This phase addresses the question of which modelling algorithms to use. The modelling process includes modelling algorithm selection, model creation and training, and prediction.
5) Evaluation: Which model best meets the research objectives? The evaluation process includes model validation, review of results, and evaluation of success criteria.
6) Deployment: How do stakeholders access the results? This includes visualisation of results and report generation (Burhanuddin *et al.*, 2018).

5.3 RESEARCH METHODS

To achieve the research objectives, an exploratory, sequential mixed-methods research design was used, as shown in Figure 5.6. Qualitative and quantitative data were collected. Mixed-methods research has gained prominence over the past four decades. Cook and Reichardt (1979) demonstrated the advantages of combining quantitative and qualitative research. In the 1980s to 1990s, researchers such as Fetters *et al.* (2013) systematised mixed methods. The exploratory sequential mixed-methods research design used in this study was characterised by two phases, with qualitative data collected in the first phase. The qualitative data collected in the first phase formed the basis for the quantitative data collection in the second phase. The results from both phases informed the construction of the supervised machine learning model for forecasting the water demand and supply of Stellenbosch Municipality. The methodology formed the basis for developing models, and the consistently applied approach was synergistic. It allowed the benefits of quantitative and qualitative research methods to be complementary to achieve a more comprehensive, in-depth assessment of the research problem. In addition, the researcher could leverage the strength of multiple data sources, which resulted in the verification and validation of the collected data, while complementing similar data.

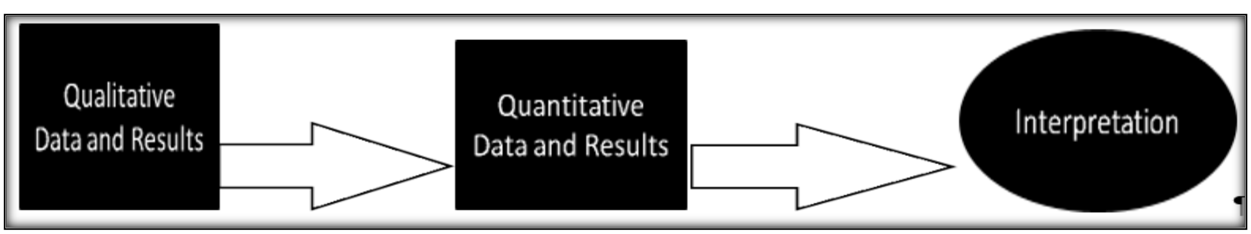


Figure 5.6: Exploratory sequential mixed-methods design

Source: Researcher

The first two objectives of the study were:

- To review urban water management approaches globally and in South Africa.
- To investigate impediments in municipal wastewater in irrigated agriculture worldwide and in South Africa.

To achieve these objectives, the following research methods and tools were used. Since the data collected to achieve the above objectives were secondary, the literature review satisfactorily achieved these objectives. The literature review was conducted using online databases such as Google Scholar, ScienceDirect, SCOPUS, government and international community websites, and government directories and documents. Key terms such as “water management”, “municipal wastewater reuse”, “municipal wastewater institutions”, “IUWM”, “municipal wastewater reuse policy”, “legislation”, “guidelines”, and a combination of two or more of these terms were used to obtain relevant studies.

Qualitative data collection was conducted to achieve the following objective: To explore challenges in implementing municipal wastewater reuse in Stellenbosch Municipality. This objective was achieved through the collection of qualitative data by conducting interactive management with a focus group composed of water legislators, policy makers, water service administrators, and private sector water professionals and wastewater infrastructure development specialists.

The following subsections describe the research methods and instruments used to achieve the specific objectives of the study.

5.3.1 Identification of the non-academic target population

In the interactive management process, it was not difficult to find non-academic participants, as the public and certain stakeholders were eager to participate in water-related discussions given the drought and the negative impacts of inefficient management of municipal wastewater in Stellenbosch Municipality. Participants were drawn from various community groups in the delineated case study, the private and public sectors, agriculture, and politics. Public sector stakeholders included Stellenbosch Municipality technical and water managers, and, at the provincial level, Department of Environmental Affairs representatives. Policy makers had council members representing specific districts. From the private sector, representatives of water consulting firms were present. Representatives from international water consulting firms were also present to provide international perspectives on the water issues in the case study. Among these groups, it was difficult to deal with the politicians because they usually seek a platform to push their own political agendas. In addition, poor communities complained that their contributions were never recognised. Both the public and private sectors showed great enthusiasm and worked very well together. They showed great interest in the research findings and were willing to take them into account in their activities.

In order to strengthen the research findings, attention was paid to the position of certain participants in society when identifying them. Specific stakeholders who have a special interest in the water sector and influence the formulation of water policy were considered. For confidentiality reasons, the names of the representatives are not published, but consent was given to disclose their positions as this would increase confidence in the data obtained from the focus group.

Once the specific representatives were identified, they needed to be contacted, which was accomplished through the following process:

- Beginning of the workshop: A request for an appointment with a potential participant was made via email or phone (see Appendix A).
- Once the appointment was made, a face-to-face interview was conducted with the potential participant to informally assess their understanding of the research problem and availability, as well as to determine whether they were willing to participate in the study.
- An invitation with a date, time, and venue was then emailed to confirmed participants. The email included details of the day’s programme and the attachment of the informed consent form, which was explained to and signed by the participants at the workshop.

5.4 QUALITATIVE DATA ANALYSIS

The qualitative data collected during the focus group at the interactive management workshop were modelled using Concept Star decision-making tools for professionals, version 3.64, to capture the ideas from the focus group. This is because the program reliably captures all ideas from the focus group and incorporates them into the model. The method is able to provide a strategic "roadmap" for resolving complex situations where there are numerous issues to consider. Its robustness stems from its ability to provide a basic understanding of complex situations while designing a course of action for solving the challenges under study.

5.4.1 Quantitative data collection and analysis

To achieve the fourth research objective, namely "To build, train, and deploy a water demand and supply prediction and forecasting supervised machine learning model for Stellenbosch Municipality", data-driven supervised machine learning modelling techniques were used to model the interactions of variables in the urban water system under study. The selection of these techniques stems from their ability to capture the relationship between variables in a system without requiring a description of the physical processes within the system. They are also easier and faster to apply and have been shown to be robust in quantifying uncertainty (Tiwari & Adamowski, 2017). The main drawback, however, is that since they are data-driven, extensive useful data are needed to achieve good prediction results. However, extensive data on water use, population growth, weather, and precipitation forecasting are increasingly available as various agencies actively collect data through surveys, reports, and multiple techniques (Chini & Stillwell, 2018). Accordingly, data-driven machine learning techniques are becoming more popular and superior to traditional techniques.

5.4.1.1 Supervised machine learning modelling method

Chapter 4 provided a literature review of supervised machine learning algorithms used in urban water management. Figure 5.7 shows the methodological framework followed in the creation, training, and deployment of the developed model, which is detailed in the Jupyter notebook (see Appendix E).

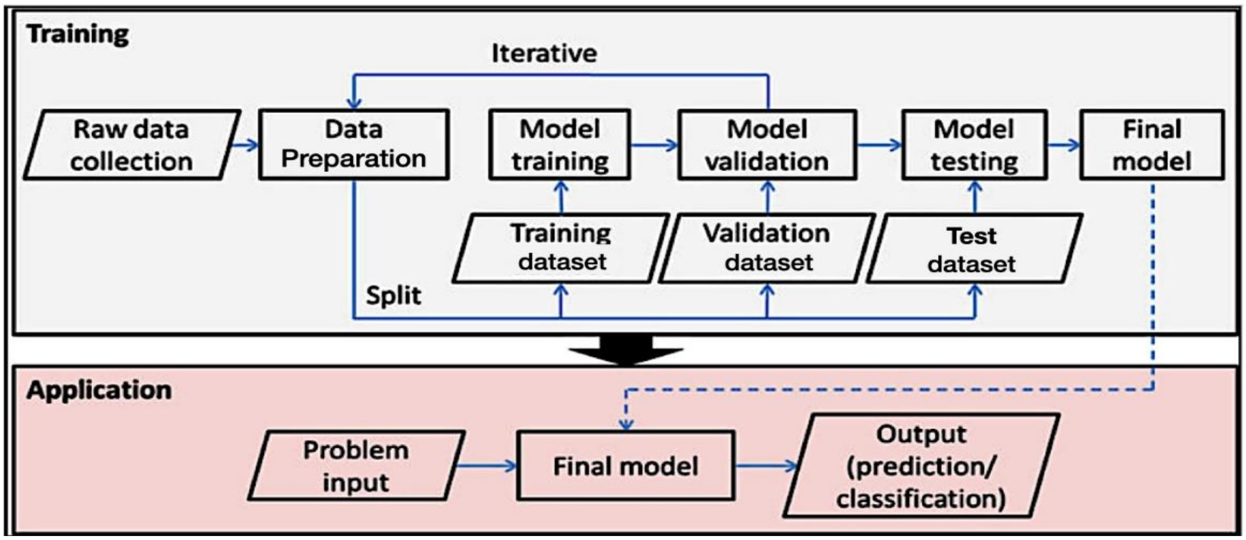


Figure 5.7: Methodological framework

Source: Researcher

This study collected raw data from the Western Cape Government's Department of Environmental Affairs, Stats SA, Stellenbosch Municipality, and the SAWS. The data cover 15 years (June 2006 to December 2021) of monthly water use, weather statistics, and annual demographic data (January 1995 to December 2021). The model development process began with the pre-processing step, and the resulting comma-separated values (CSV) file was uploaded to a Jupyter notebook. Scripts were written on the Jupyter notebook in the Python programming language (version 3.8) installed in the Anaconda environment. Modelling was performed on a local computer (Windows 10). A Jupyter notebook that details the modelling steps can be found in Appendix E. Figure 5.8 illustrates the steps followed during modelling.

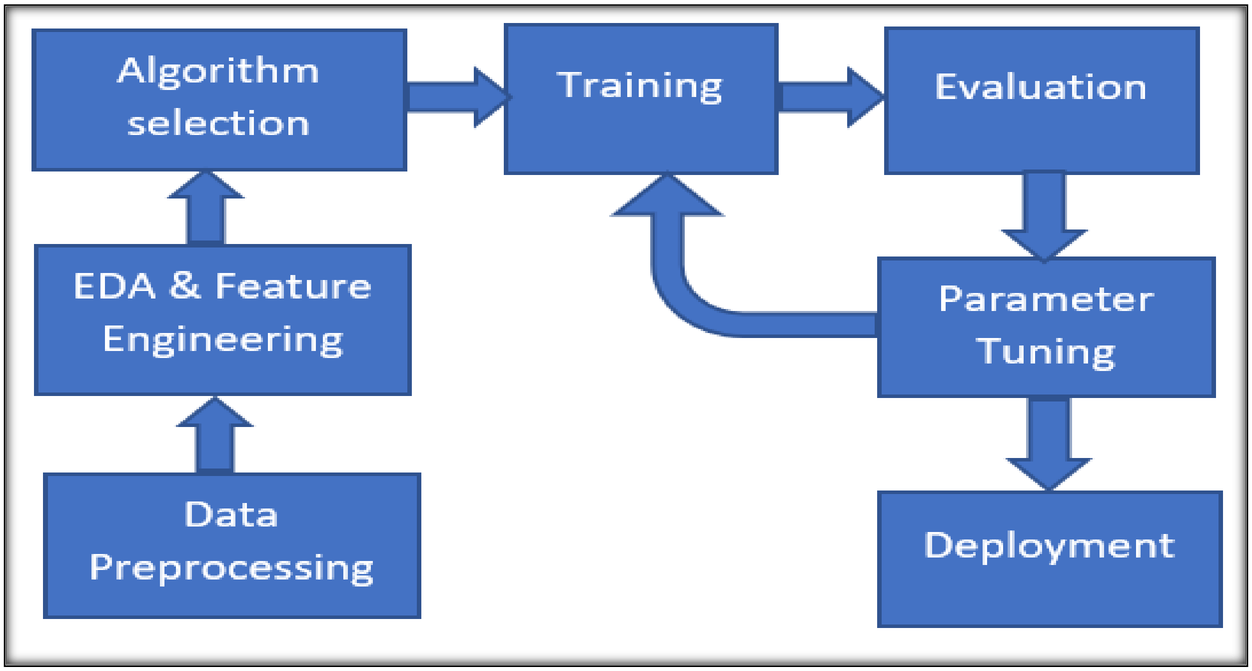


Figure 5.8: Supervised machine learning process work flow

Source: Researcher

(a) Data pre-processing

The pre-processing step, which includes data cleaning, integration, reduction, and transformation, was used to eliminate problems related to inconsistent formatting, human error, and missing values. Since the data in this study were collected from various government agencies, they are inevitably susceptible to the above problems. Although the pre-processing step is estimated to take 70% of the project time, its rigorous implementation is extremely important for the successful delivery of powerful models. The resulting CSV file was initially created on Microsoft Excel but was later uploaded to a Jupyter notebook for the EDA step.

EDA is a crucial step in model development. This is because it facilitates the acquisition of insights and statistical measures about the dataset that are essential for model development. In this case, the EDA process provided the researcher with deeper insights into the dataset that helped to interpret the results from the developed models. This was accomplished through data profiling, which produced descriptive statistics of the dataset and allowed the researcher to query and visualise the data in a variety of ways. Consequently, interesting features and relationships between features became apparent and the researcher could decide what in the dataset to correct, discard, or treat differently. In addition, multiple variables could be examined using different techniques to search for and find systematic patterns. The details of the EDA process are described in the Jupyter notebook and Microsoft Excel spreadsheet (see Appendices D and E).

(b) Modelling procedure

The model-building process consisted of four sequential steps: (1) training, (2) validation, (3) testing, and (4) final model for comparing the predicted output with the desired output. The whole process is shown in Figure 5.8, and a detailed description of the algorithms proposed was provided in Chapter 4. These are the conventional algorithms SARIMA and VAR. The performance of these conventional models was to be compared to models developed using extensively researched and powerful supervised machine learning algorithms, namely SVR, XGBoost, and Regular Neural Networks. Considering that hybrid models are highly recommended for water demand prediction, the Prophet-SVR hybrid model was also proposed.

5.4.1.2 Overview of algorithms to be deployed

(a) Conventional algorithms

- Seasonal Autoregressive Integrated Moving Average (SARIMA (p, d, q) (P, D, Q))

A widely used conventional model for predicting water demand and supply is the ARIMA (Kofinas *et al*., 2014; Oliveira *et al*., 2017). However, in this study, the extension of ARIMA, the SARIMA, was proposed instead (Mombeni *et al*., 2013). That is because in developing the models, time series data are presented that contain a seasonal periodic component; the SARIMA model should therefore be preferred since it is inherently multiplicative and can accept additional parameters $(P, D, Q)\ m$ that specifically describe the seasonal components of the model (Braun *et al.*, 2014). Here, P, D, and Q represent the seasonal regression, differencing, and moving average coefficients, respectively, and m represents the number of data points (rows) in each seasonal cycle. The notation for the SARIMA model parameters was captured by Guo *et al.* (2018) with the following expression:

SARIMA (p, d, q) $(P, D, Q)\ m$

The modelling processes and procedures were captured in Python and described in the associated Jupyter notebook (see Appendix E1).

(b) Supervised machine learning algorithms

Considering that supervised machine learning regression algorithms are still widely used in developing water demand and supply forecasting models, specific regression algorithms were proposed. Chapter 4 provided a general overview of the regression algorithms used, describing the basics of the regression techniques and a description of the specific regression algorithms widely used in urban water demand forecasting. This section provides an overview of the rationale for their use in urban water demand forecasting. The regression algorithms in question are SVR and XGBoost.

- SVR

SVR is an SVM technique in which kernel functions have been added to add another dimension to perform linear regression analysis (Shabani *et al.*, 2017). According to Awad and Khanna (2015), the SVR algorithm has great generalisation capabilities, can account for nonlinearity in a system, and has high predictive accuracy, which make it ideal for urban water demand modelling. Since urban water demand prediction is affected by many variables, accuracy is critical. The most important feature of the SVR algorithm is its ability to specify a margin ε within which errors in the sample data can be accepted without affecting the predictive accuracy of the model. It is also resistant to overfitting and has a minimal error on the previously unseen data, which is ideal for urban water demand forecasting (Ghalehkhondabi *et al*., 2017; Smolak *et al.*, 2020), hence its wide application in the sector.

- XGBoost ensemble model

Ensemble methods are machine learning methods that combine multiple weak learners to produce a strong learner. Currently, bagging, stacking, and boosting ensemble techniques are attracting the attention of researchers in various fields. In this study, boosting was investigated using gradient boosting machines. This is in light of research that shows that individual models underperform in terms of accuracy, bias, and predictions for peak periods (Ghalehkhondabi *et al.*, 2017; Xenochristou & Kapelan, 2020). Chen and Guestrin (2016) developed the XGBoost algorithm to implement gradient boosting machines. Its effectiveness as a tree-based ensemble learning algorithm, coupled with its ability to minimise overfitting and its high execution speed, is the reason for its wide application. Considering the properties of ensemble regression models and the data finally provided, the researcher decided to develop the following gradient boosting models: Adaptive Boosting (AdaBoost), Gradient Boosting, Stochastic Boosting and Random Forest. Since the main objective of the simulation process was to develop a powerful model. The description of its functioning and the modelling process can be found in Chapter 7 and in the related Jupyter notebook (see Appendix E2).

(c) ANNs

The application of machine learning algorithms (ANNs) has been extensively researched and implemented in predicting urban water demand due to their high predictive accuracy, reliability, ability to handle large datasets, and efficiency in handling nonlinearities and discontinuities if any are found in a dataset (Vozhehova *et al.*, 2019; Vijai & Sivakumar, 2018). In addition, researchers have found that models developed using the ANN algorithm have, on average, high accuracy in forecasting and predicting short- and long-term urban water demand (Tiwari & Adamowski, 2015; Brentan *et al.*, 2017). This study used the ANN algorithm in the form of a neural network with a single hidden layer (feedforward), which was described in detail in Chapter 4.

(d) Hybrid models

- Prophet-SVR

Hybrid models consist of two or more algorithms; one acting as the main algorithm and the other(s) serving to integrate and optimise the main algorithm. Given their robustness in dealing with variability in climate factors and ability to gain deep insights into the dataset, hybrid models are becoming increasingly popular in municipal water demand modelling (Altunkaynak & Nigussie, 2017). To improve the performance of the model in predicting and forecasting the municipal water demand and supply of Stellenbosch

Municipality, the researcher first proposed a hybrid model combining the Prophet model with SVR, i.e. Prophet-SVR, as proposed by Guo et al. (2021). However, the requirements of the study and the data provided indicated that the use of the ensemble models was sufficient and could be compared with the conventional SARIMA model.

5.5 SUMMARY

This study employed an overarching transdisciplinary research methodology that spans disciplines. Both consultative and participatory approaches were used, which included disciplinary and non-disciplinary participants in workshops and one-on-one discussions. The ontology, epistemology, methodology, and organisational characteristics of transdisciplinary research methodology were presented. To address the four objectives of this study, an exploratory sequential mixed-methods research design was used. In doing so, the first and second objectives were achieved through a literature review, which formed Chapters 2 and 3 of this study. This chapter discussed and described the research philosophy, overarching research methodology, sub-research methodology, and methods used to achieve the third and fourth objectives.

CHAPTER 6:
QUALITATIVE RESEARCH FINDINGS

6.1 OVERVIEW

To satisfy water users in an urban environment, all components of an urban water system require a well-coordinated management strategy. To this end, the researcher examined South African water policy, laws, and administration, as these are considered critical to the formulation of water management policy. Chapter 2 provided an overview of the evolution of water laws in South Africa from 1910 to the present. An analysis of the 1996 constitutional provisions on water management was conducted, with the aim of fully understanding the objectives of the NWP and the national water Acts enacted by the democratically elected government of 1994. A report was then prepared on the policy responsiveness of the NWP and the effectiveness of the new water institutions as required by the Constitution. To understand the approaches to water management used by Stellenbosch Municipality, two major global approaches to water management were presented, namely government and governance water management. Principles of water management, such as IWRM at the river basin level and IUWM in an urban setting, were presented. Several concepts have evolved from the IUWM principle. The researcher advocated alternative water sources with emphasis on the reuse of treated municipal wastewater. The reason for this is that it is suitable for Stellenbosch Municipality. Chapter 5 provided an overview of the interactive management methodology used in the study.

This chapter aims to provide an overview of the urban water management framework and the challenges faced in Stellenbosch Municipality’s water system. It also presents the data-collection process and analysis and a discussion of the modelling results that emerged from the workshop that was conducted to investigate barriers to the reuse of treated municipal wastewater in Stellenbosch Municipality. The results are derived from the interpretive structural model that was developed during the workshop.

6.2 STELLENBOSCH MUNICIPALITY WATER CYCLE

Stellenbosch Municipality's water supply system consists of a freshwater system. In this system, raw water is obtained from river abstraction within the jurisdiction of Stellenbosch Municipality and the DWS, which is taken from Theewaterskloof and sent to the WTP for purification before being fed into the freshwater system. To make up for any shortfall in the water supply, treated water is purchased in large quantities from the City of Cape Town. All treated water is then fed into the network for distribution within Stellenbosch Municipality's jurisdiction. Wastewater from the urban centres is piped through the sewer system to a central treatment plant, where it is treated and discharged to a natural receiving water body. Figure 6.1 shows the current general water cycle of Stellenbosch Municipality.

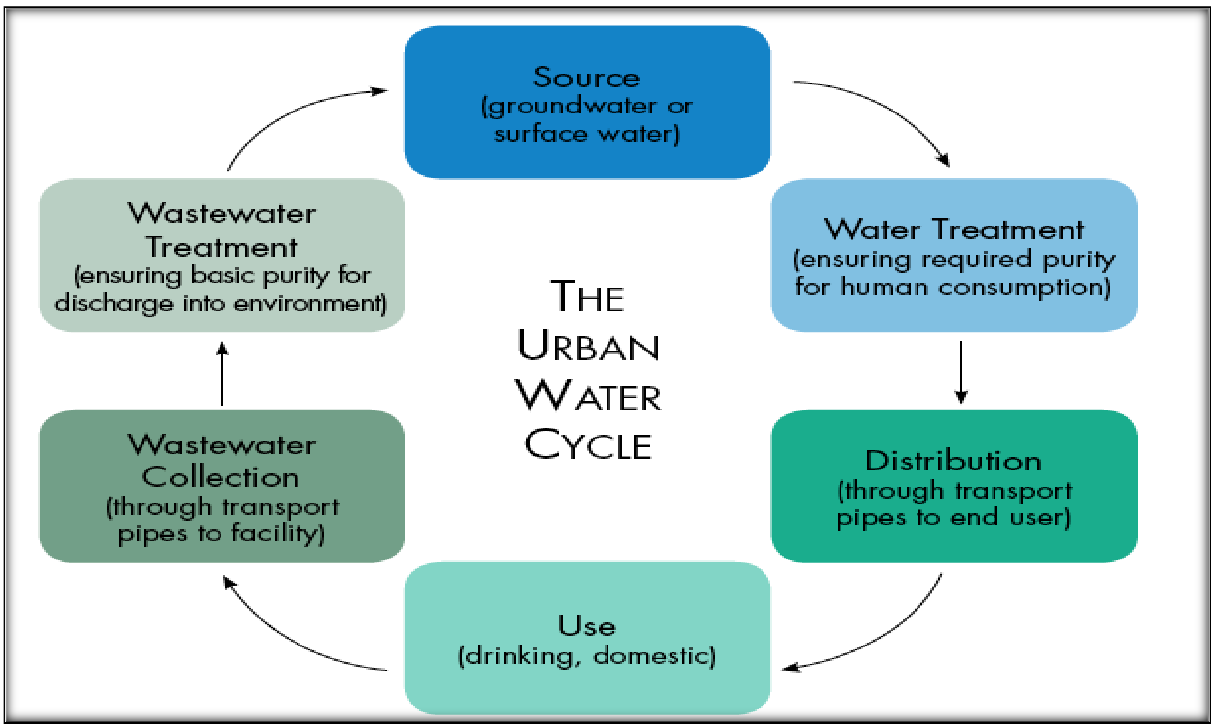


Figure 6.1: Stellenbosch Municipality's urban water cycle

Source: Baker *et al.* (2017)

Figure 6.2 shows water consumption in a typical South African household, which is also applicable to Stellenbosch Municipality. Only 3% is used for the most important aspects of life, cooking, and drinking. Sixty-two percent of the water used for toileting, washing, and bathing can be reused, which increases the amount of water that flows into the water system and decreases the amount of water drawn or purchased from natural water sources. Reusing treated municipal wastewater requires extensive changes to current infrastructure. The state of current freshwater and wastewater infrastructure must thus be considered.

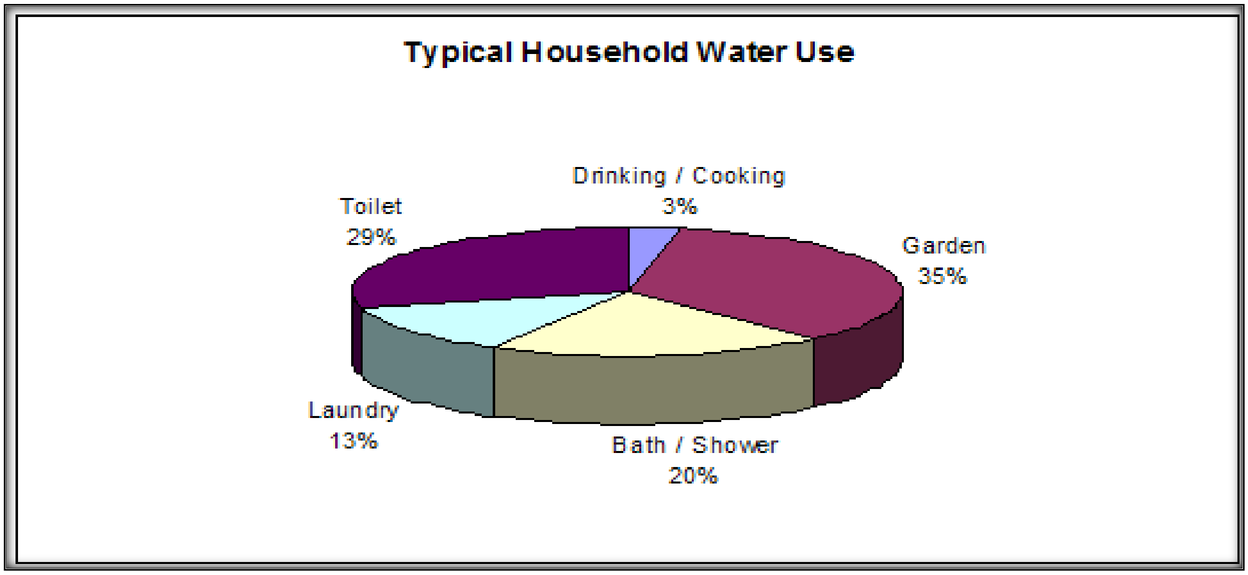


Figure 6.2: Typical household water consumption per utility in South Africa

Source: Ramulongo *et al*. (2016)

6.3 STELLENBOSCH WATER INFRASTRUCTURE

Regarding water infrastructure development in general, the national government is responsible for major water infrastructure projects, while the provincial government plays a monitoring and supportive role in ensuring that municipalities fulfil their water supply mandate. The national government provides funding for major projects through municipal infrastructure grants. Operation and maintenance are the responsibility of the municipalities and are funded through taxes and levies, as well as national operating and capital grants.

In 2012, the Sustainability Institute reported that most of Stellenbosch's major water supply facilities are desolate, which is exacerbated by a backlog of maintenance and repair of existing infrastructure (Stellenbosch Municipality, 2012). Due to population growth, demand has outpaced water infrastructure development, which negatively impacts municipal water services. Water infrastructure is estimated to have lost 52.3% of its value, which has resulted in reactive rather than proactive measures being taken to ensure the efficient delivery of water services. In order to improve the steadily deteriorating infrastructure, Stellenbosch Municipality has created a Water Supply Master Plan.

6.3.1 Drinking water infrastructure

The Stellenbosch Municipality Water Supply Master Plan of 2011 envisaged the improvement of the water supply infrastructure by upgrading the freshwater supply systems in terms of the WTP, water pumping stations, and reservoir infrastructure (Stellenbosch Municipality, 2018b). In its 2015/2016 annual report, Stellenbosch Municipality (2017) outlined several water conservation and demand management initiatives. These included the replacement of water mains due to main water leaks. Particular attention was paid to water supply systems in need, as 25% of the water fed into the supply system is unaccounted for (Stellenbosch Municipality, 2018b). According to the report, efforts to improve revenue collection include conducting meter audits to improve billing accuracy. In addition, the master plan describes the strategy to secure fresh water supplies for the urban water system, which includes the construction of high-volume reservoirs in Kayamandi, Groendal, Franschhoek, Klapmuts, Cloetesville, and Idas Valley (Stellenbosch Municipality, 2018b).

It can be inferred that water conservation and demand management continue to follow a linear engineering approach. Such a water management approach is concerning because Stellenbosch Municipality is located in a region that is predicted to experience severe drought by 2040. Due to climate change, there is no certainty that the region will continue to receive the rainfall it has received in recent years. New methods of managing available freshwater supplies will be needed to address the multiple emerging challenges.

Due to the 2015/2016 drought, Stellenbosch Municipality had to draw additional raw water from reservoirs such as the Kleinplaas Dam to supplement the supply from Jonkershoek. There were plans to intensify drilling to obtain more water from underground sources to reduce the water deficit (Stellenbosch Municipality, 2017; 2012). However, from an environmental management perspective, such a strategy would threaten the ecological reserve as people increasingly extract freshwater from groundwater resources while natural water resources are only replenished to a limited extent by rainfall. The principle of water management proposed in this study aims to manage water sustainably by reducing the amount of water withdrawn from natural sources, encouraging the reuse of water in the municipal wastewater system as often as possible, and optimising the operation of the water supply system.

6.3.2 Wastewater infrastructure (wastewater treatment works)

The Stellenbosch sewage treatment plant was commissioned in 1924, built with a hydraulic capacity of 20 ML/d, and uses conventional activated sludge treatment technology. However, rapid population growth combined with ageing infrastructure contributed to the overutilisation of the plant and regular discharges of inefficiently treated effluent from the treatment plant into the Eerste River affecting the freshwater system. The consequences are contamination of groundwater, eutrophication of the river, degradation of the ecosystem, and the spread of waterborne diseases. In 2012, the Stellenbosch WWTP was documented as having the following attributes: a design capacity of 20.2 ML/d, with a plant utilisation rate of 102.9%. At the time of the assessment, effluent quality compliance was 65.8%, and the effluent risk score was 74.1% (DWA, 2012; 2011), which placed the Stellenbosch WWTP in a high-risk category. As a result, inefficiently treated wastewater continued to be discharged via the Veldwachters River into the nearby Eerste River, which negatively impacts irrigation and tourism downstream.

As a result, Stellenbosch Municipality authorities agreed to undertake a project to expand the treatment plant capacity (hydraulic and process) to 35 ML/d and improve the plant's treatment process to accommodate future growth in the catchment. A technical feasibility study on data collection and design parameters was conducted for the project. An analysis of technology options for treatment was also conducted. Factors such as capital and life cycle costs, space requirements, ease of operation, quality of treated effluent, potential for effluent reuse, and sludge treatment were considered. Membrane bioreactor technology was chosen because it was expected to produce high-quality, reusable wastewater. These developments also contributed to the motivation for this study, as the quality of treated municipal wastewater currently produced represents an opportunity for reuse. To this end, the researcher conducted an interactive management workshop to develop a strategy for the reuse of treated municipal wastewater to augment Stellenbosch's urban water supply. The framework of the interactive management workshop and the process of identifying and inviting focus group members were presented in Chapter 5. The following sections present the data-collection process, research findings, and discussions.

6.4 DATA COLLECTION AND RESULTS

After reaching consensus on the date for convening the workshop, 11 participants took part in the deliberations. Descriptions of these participants were presented in Chapter 5, and Table 6.1 provides a descriptive summary of the participants. As recommended by Warfield and Cárdenas (1994), the researcher sought to assemble a diverse group of participants to achieve high-quality research results by capturing the perceptions and needs of key stakeholders. At the same time, this reduces the risk that differing perspectives on particular issues will be influenced by different cultures, roles, and interests in the areas being discussed (Schmidt *et al.*, 2001). A total of 11 participants were convened because an odd number would facilitate decision making in the voting phase of the modelling. The number of 11 is within the recommended group size for interpretive structural modelling, although Janes (1988) recommended a maximum number of eight. According to Janes (1988), group size plays a crucial role in the interactive management process, as an increase in group size leads to a decrease in the quality of the debate. Considering that any member can talk to any other member, the possible communication between individuals is described as follows. For n individuals in the focus group, there are n(n-1) possible communications and if n is increased from six to ten participants, the number of communications can increase from 30 to 90. However, in this study, 11 participants were considered adequate to achieve high-quality research results. Since Stellenbosch Municipality contains a very heterogeneous community, water management issues affect these groupings differently, as mentioned in the description of Stellenbosch Municipality in Chapter 2.

Table 6.1: Participants in the focus group

Stakeholder group	Description	Designation targeted for this research	Number of participants
Western Cape Department of Environmental Affairs official	Environmental practitioner	The official contributed to environmental laws and policies in relation to urban wastewater management.	1
CEO of a South African state-owned water enterprise	Public water practitioner	The CEO provided guidance on water and sanitation laws and policies and administration in the country.	1
Professor at the Council for Scientific and Industrial Research	Chief scientist specialising in wastewater (urban, mine, and industrial effluents)	The professor provided insights into what has transpired in the water research field on urban wastewater management relating to the case study.	1
Private sector urban wastewater practitioners	1. Technical director of a wastewater engineering firm 2. Managing director of wastewater at an international consulting firm	The practitioners contributed towards wastewater treatment works infrastructural development relating to the case study.	2
Municipal officials	1. Director of Engineering Services of the case study 2. Municipal manager from the Netherlands	The practitioners contributed to urban wastewater administration on the case study. The municipal manager from the Netherlands provided an international perspective on urban wastewater management.	2
Former city councillor (Stellenbosch)	Politician	The former city councillor contributed to urban wastewater management in the delimited area from a political perspective.	1
Academia	Professor of Public Administration	The professor contributed to governance pertaining to urban wastewater relating to the case study.	1
Community members	Members from previously disadvantaged communities	The community members contributed to the perceptions of and the actual urban wastewater management landscape in their areas.	2

6.4.1 Generation and clarification of ideas

The workshop began with a brief introduction to the research objective, which was to investigate barriers to the reuse of treated municipal wastewater in the Stellenbosch Municipality jurisdiction. This was followed by a discussion of the key issues related to the objective: water policy, legislation, and governance. In this case, the facilitator of the workshop, the researcher, explained in detail the principles of IUWM and alternative water sources to address the importance of reusing municipal wastewater as an alternative water source for Stellenbosch Municipality. The methodology of the workshop was then explained. To initiate the interactive management process, the participants were asked to discuss the following questions to gather ideas on the factors that hinder the reuse of treated municipal wastewater in Stellenbosch Municipality:

- Do water policies support and enable the reuse of treated municipal wastewater?
- To what extent do current water laws enable water utilities to implement treated municipal wastewater reuse projects?
- What are the challenges of reusing treated municipal wastewater?

During the brainstorming session, the participants were asked to identify and describe factors that hinder the implementation of treated municipal wastewater reuse in Stellenbosch Municipality. These factors were water policy, legislation, and administration. These three components were discussed independently, and the factors and their descriptions were compiled. A total of 41 factors were identified. Table 6.2 summarises these factors and their descriptions.

Table 6.2: Factors that impede treated municipal wastewater reuse in Stellenbosch Municipality

No.	Factor	Factor description	Reference
1	Legislation synergy	Lack of legislative synergy due to inconsistent responsibilities of the three levels of government in managing water resources as provided for in the Constitution.	Adewumi *et al*. (2010)
2	Policy	The NWP does not articulate the reuse of municipal wastewater in a way that encourages this practice.	Malisa *et al.* (2019)
3	Governance	The three levels of government should play an important role in ensuring that the reuse of treated municipal wastewater is successfully implemented, but they are not currently doing so.	Malisa-Van der Walt and Taigbenu (2022)
4	Administration	The reuse of treated municipal wastewater is perceived as complex.	Bixio *et al.* (2005)
5	Capacity in administration	There is a lack of qualified personnel to implement the practice at the municipal level.	Edokpayi *et al.* (2020)
6	Moving water services	For projects such as the reuse of treated municipal wastewater, municipalities should contract organisations with qualified personnel to manage the process, which they are not prepared to do.	
7	Water cost tariffs	Lack of a water pricing model that would encourage the use of treated municipal wastewater.	
8	WTPs	The WTP was overutilised at the time.	

No.	Factor	Factor description	Reference
9	Integrated water management	Lack of implementing IUWM principles.	
10	Disruptive events	Disruptive events such as drought can hinder water reuse as the ecological balance of water should be maintained.	Adewumi *et al*. (2010); Bixio *et al.* (2005)
11	In-migration	Rapid in-migration strains water infrastructure, including WWTPs. This makes it difficult to produce high-quality wastewater for reuse.	Stellenbosch Municipality (2010)
12	Climate change	Has both positive and negative impacts on treated wastewater recycling projects.	
13	Politics	Politicians tend to misinform citizens depending on what their constituency wants to hear to gain political advantage.	Radingoana *et al*. (2020)
14	Informal settlements	The lack of proper water infrastructure defeats the idea of reusing treated municipal wastewater.	Ntombela *et al*. (2016); Seeliger and Turok (2014)
15	Unaffordability	The introduction of reusing treated municipal wastewater requires restructuring the urban water system and high-tech wastewater treatment technologies to produce wastewater of reusable standards.	Kumarasamy and Dube (2016)
16	Capacity	There is a lack of highly qualified personnel for such special projects. Wastewater treatment is a very sensitive undertaking in terms of health. Lack of capacity erodes the trust in the municipality to undertake such an initiative.	Valdes Ramos *et al*. (2019)
17	Population growth	Population growth emanating from informal settlements and lack of adequate sanitation in these communities make it difficult for the municipality to provide clean water in sufficient quantities. Implementing water reuse projects would be a mammoth task.	Seeliger and Turok (2014)
18	Skills drainage	Highly qualified young people do not stay long in the public sector. There are several reasons for this, including compensation and incentives.	Edokpayi *et al.* (2020)
19	Procurement challenges	Local government supply chain management is very bureaucratic.	
20	Populism (fake news) and ineffective public participation	It is easy for politicians to hijack such projects and work against the ruling party.	
21	Natural risk assessment	Where there is appropriate risk assessment, such as the availability of fresh water, it may be necessary to apply water management principles that include the reuse of wastewater.	
22	Reality and what is perceived	Varying perceptions of using treated wastewater.	
23	Theft and vandalism	South Africa suffers from theft and vandalism of water infrastructure, which can affect the implementation of expensive new projects, as new infrastructure would be required.	
24	Emotions trump evidence	The emotional effect of introducing a phenomenon such as the reuse of treated wastewater without using the mind to understand the whole process and objectives.	
25	Policy change in by-laws	It is a complex mandate for a single municipality to enact its by-laws for high-sensitivity wastewater reuse. Also, multiple by-laws from several municipalities erode confidence among users.	Malisa-Van der Walt and Taigbenu (2022)

No.	Factor	Factor description	Reference
26	Overambitious public perceptions of service quality	Public expectations of water services are very high; any project that could harm the public would thus be disastrous for the municipality.	
27	Undue influence in the community	Some parts of the community might not approve of the concept and, in turn, influence others.	
28	Redundancy of service delivery	Some constituencies may favour outsourcing water services because they are not satisfied with the services they currently receive.	
29	Disruptive political conflict	Different political attitudes during elections may negatively impact public attitudes toward the reuse of treated municipal wastewater. Politicians could misuse the idea for their political goals.	
30	Perceptions of wastewater reuse	Different views on wastewater reuse by different communities. Poor communities may not welcome the idea because they are suspicious of the water services provided by the municipality. Wealthy communities may doubt the ability of the municipality to handle such sensitive projects; considering that the municipality has failed for years to provide wastewater services in their jurisdiction.	Maryam and Büyükgüngör (2019)
31	Wastewater versus clean water preference	If there are no incentives, the public will naturally prefer fresh, clean water to recycled water.	Al-Saidi (2021)
32	Lack of water awareness	In some communities, there is a general lack of water awareness.	Radingoana *et al.* (2020)
33	Perceptions of the value of water	Different groups in the community have different views on water, which is also influenced by different levels of education and economic participation.	
34	Technology	The use of technologies in such projects is a critical factor, as the community, stakeholders, and water agencies should have a high level of acceptance of the technologies to achieve the desired results in treated municipal wastewater reuse projects. Appropriate technologies and their acceptance are critical.	
35	Budget	The five-year term for each administration was raised as a problem since water projects of this nature need more time to mature.	
36	Environmental ethics	There are several environmental issues that Stellenbosch Municipality is still struggling with in terms of wastewater management.	
37	Lack of prioritisation of wastewater reuse	There is no one to champion initiatives such as wastewater reuse in all areas of government.	
38	Impact on vested interests for agricultural activities	Stellenbosch is an agricultural centre of the region and most agricultural products are exported; this sector therefore has a major impact on activities such as wastewater reuse as they are prone to water shortages.	
39	Borehole drilling	Due to water rights, drilling wells is hardly replaceable by the possibility of reusing wastewater for farmers.	
40	Big business influence	Large water-intensive businesses can hurt wastewater reuse projects.	
41	Trade-off of businesses	Some of the business community may be willing to embrace the idea if incentives are offered.	

The identified factors reflect the different perceptions of the participants, depending on their expertise, role, status, and ethnicity within the Stellenbosch community. Twenty of the factors have already been mentioned in the literature. Consensus was easily reached during the discussions of the factors presented, and the entire interactive management process proceeded without interruptions.

6.4.2 Interpretive structural modelling

Concept Star software, developed by Sorach International, was used to create a hierarchical model of the identified factors that impede the reuse of treated municipal wastewater in Stellenbosch Municipality. The 41 identified factors were loaded into the software to develop a model through pairwise comparison prompts. The participants discussed each pairwise comparison to reach an agreement on whether the answer to the pairwise comparison statement was a yes or no. Figure 6.3 describes the process and Table 6.3 describes the pairwise relationship between items.

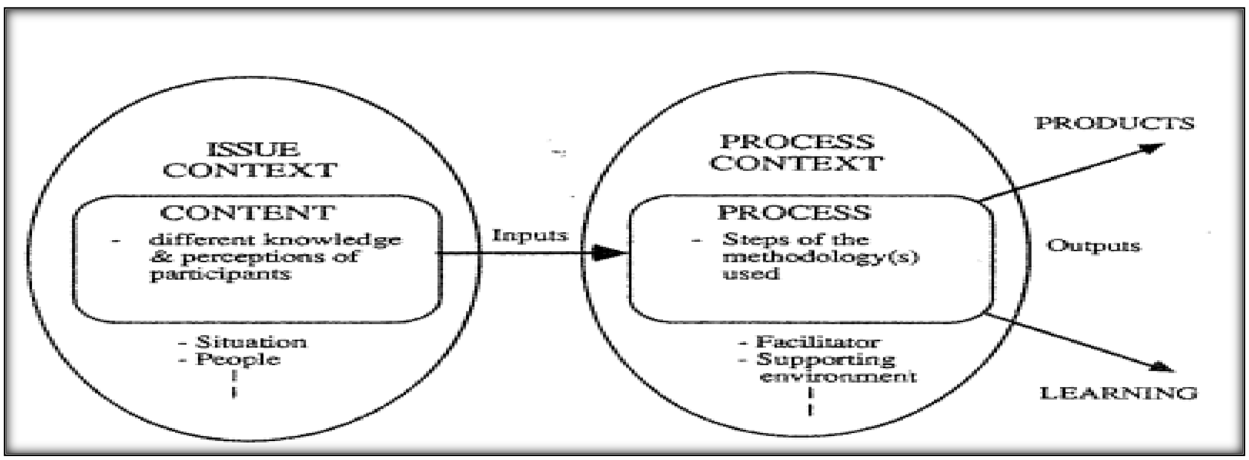


Figure 6.3: Content, context, process, and product the IM process

Source: Janes (1988)

Table 6.3: Example of pairwise statement relation between elements

Elements	Relation
1. Does law/legislation synergy in water	“Significantly influence” policy in treated municipal wastewater reuse initiatives, yes or no?
2. Does policy	“Significantly influence” governance of treated municipal wastewater reuse initiatives, yes or no?
3. Does governance	“Significantly influence” administration of treated municipal wastewater initiatives, yes or no?

The participants discussed each pairwise comparison to agree on whether the answer to the statement of the pairwise comparison was yes or no. The model created was reviewed for inconsistencies and accepted as the final interpretive structural model for factors that impede the reuse of treated municipal wastewater in Stellenbosch Municipality.

6.5 RESULTS AND DISCUSSION

6.5.1 Model presentation

The relationship model in Figure 6.4 shows the relationship between the factors that hinder the reuse of treated municipal wastewater in Stellenbosch Municipality. The factors are shown in the boxes and the direction of the arrows indicates the direction of the relationship between the factors and the order in which they affect each other. In this model, the relationship shown is “significantly affect”. Most of the factors are shown in the individual boxes. There is no “circular relationship” in the model; e.g., A significantly influences B and B significantly influences C. The model is hierarchical, with the elements on the far left forming the base of the hierarchy. The elements on the far right of the diagram represent the factors that do not influence other factors.

The model shown in Figure 6.4 consists of two hierarchical levels. The first level on the left side of the model shows two factors as primary key factors that in turn influence the other four factors that hinder the reuse of treated municipal wastewater in Stellenbosch Municipality, and these are:

- overambitious public perceptions of service quality;
- lack of water awareness;
- disruptive events;
- population growth;
- populism (fake news) and public participation; and
- perceptions of the value of water.

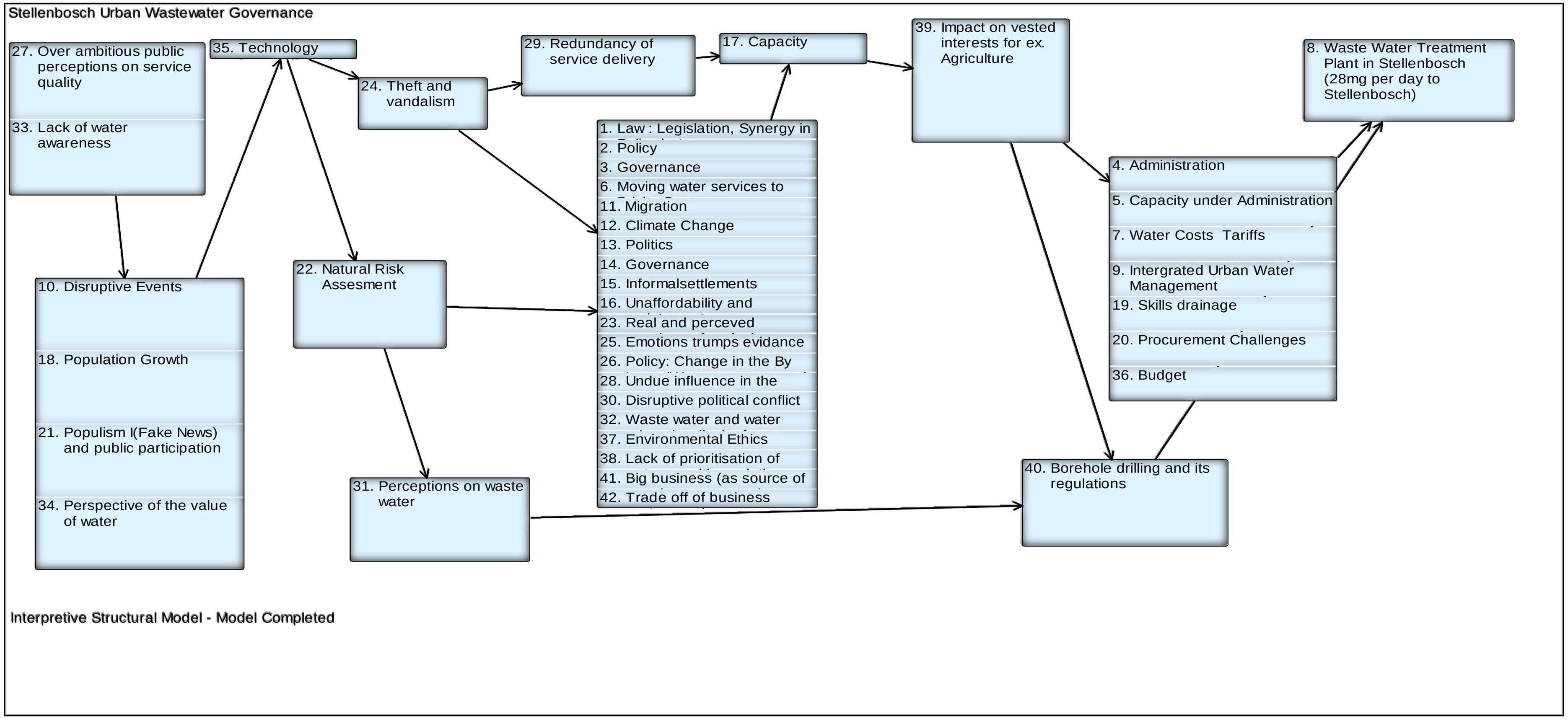


Figure 6.4: Interpretive structural model

Source: Researcher

Although the first two identified factors influence the other four, the researcher applied Senge's (1997) systems thinking philosophy, which states that small actions focused on the right things or areas will lead to large and lasting improvements in any endeavour. Applying Senge's (1997) statement to the factors that hinder the reuse of treated municipal wastewater in Stellenbosch, the hierarchical model suggests that by focusing on the six key factors identified in the interactive management process, the potential for successful implementation of treated municipal wastewater reuse in Stellenbosch Municipality increases significantly and better results can be achieved if the strategy described by the interpretive structural model is followed.

The results of this study indicate that Stellenbosch water managers need to focus on certain key factors (as discussed in the following subsections) and must develop an implementation strategy according to the interpretive structural model in order to successfully implement treated municipal wastewater reuse initiatives. These results also reflect the philosophy of design thinking, which puts people at the centre. When solving a problem or implementing a new initiative that involves consumers, high priority is given to issues that affect them.

6.5.1.1 First-level elements

(a) No. 26: Overambitious public perceptions of service quality

An overly ambitious public perception of service quality is the biggest obstacle to implementing the treated municipal wastewater reuse initiative in Stellenbosch Municipality. The reasons for this lie in the different perceptions of water services by different groups in the community. This is influenced by economic inequality, as Stellenbosch has the highest Gini coefficient in the country. Historical imbalances in water infrastructure development and services provided are still firmly entrenched in Stellenbosch Municipality. One example is that more affluent people demand high-quality water services from the municipality because they pay their fees and taxes.

In contrast, the poor's perception of water services is strongly influenced by the South African Constitution, which states that access to water is a human right. As a result, they believe that the water authority must provide them with water and sanitation services, regardless of any challenges they face. If these expectations are not met, they might significantly convince the population to consider reusing treated wastewater when a disruptive event such as drought occurs.

In Stellenbosch Municipality, the African population was heavily controlled during apartheid (which was relaxed at the beginning of democracy), which resulted in an influx of Africans from rural areas. If Stellenbosch Municipality were to embark on a project to reuse treated wastewater, there would likely be opposition from both the rich and poor communities. The rich, i.e., the mostly white minority, would not welcome the idea because they have always been entitled to first-class services and should not mind Africans coming to a predominantly white area. The poor Africans, who are in the majority, will feel that the municipality wants to give them recycled water because they are poor and deserve second-class services, just as it was during the apartheid era. Unless this attitude is eliminated through education and information about droughts or other potential disruptions, the perception of services in these communities is the biggest obstacle.

Perceptions of municipal services also have a significantly negative impact on how different communities value water. Regardless of population growth, which is highest in poor areas, an initiative such as recycled water reuse would not be adopted. The lack of trust between the community and Stellenbosch Municipality and the underdeveloped water infrastructure in these poor communities exacerbate the situation. Moreover, a disgruntled community is fertile ground for populism, as it can be easily persuaded by anyone who addresses its grievances. Politicians tend to do this especially in poor communities because water is a highly political issue.

(b) No. 32: Lack of water awareness

Lack of water awareness was identified as a major barrier to the implementation of treated municipal wastewater reuse projects in Stellenbosch Municipality. Society needs to support the initiative and fully understand the “why”, “what”, “when”, and “where” of the treated municipal wastewater reuse project. If communities, both rich and poor, do not have the same understanding of the issue, such an initiative can never get off the ground, regardless of a drastic event like a drought or climate change that negatively impacts rainfall. Even if population growth increases without a comprehensive understanding of water issues within municipal jurisdictions, populism would easily take root and views on water will vary widely. Stellenbosch water managers therefore need to fully engage the various communities in the municipality on the reuse of treated wastewater and clarify the “what”, “why”, “where”, and “when” of the treated municipal wastewater reuse project. They must go through the process with all communities involved and let them take the initiative.

(c) No. 10: Disruptive events

Another important element that emerged was the different responses of different communities when presented with scientific data on climate change issues and their negative impact on freshwater availability. People tend to react only when a life-threatening event is unavoidable or when they are injured. When these two elements seem far-fetched, it is difficult to convince authorities and society to take preventive measures. Overly ambitious public perceptions of service quality and a lack of water awareness will negatively impact the management of disruptive events such as a drought, which may require the reuse of treated municipal wastewater to augment the city’s water supplies. For this reason, water education is very important to address such issues.

(d) No. 17: Population growth

Population growth triggered by natural increase, the rural exodus, and the growing number of poor people living on the outskirts of the city are major problems for the implementation of the reuse of treated municipal wastewater. These problems are exacerbated by the community’s overambitious public utility requirements and lack of water awareness. To address this problem, Stellenbosch Municipality should implement clear guidelines for water supply and infrastructure development in poor communities before considering a project such as reusing treated municipal wastewater. It should also be noted that an overly ambitious public perception of service quality and a lack of water awareness will affect the resolution of water issues related to the reuse of treated municipal wastewater in Stellenbosch Municipality. To this end, educational campaigns should be conducted that describe the impact of population growth on water resources, are unbiased, and target all residents of Stellenbosch Municipality.

(e) No. 20: Populism and ineffective public participation

Because of its diversity and disproportionate socioeconomic distribution, the population of Stellenbosch Municipality is prone to populism and would not be able to contribute significantly to public participation in water matters that are normally managed by municipal officials. The irony is that while the poor are receptive to rhetoric that addresses their concerns, the affluent resist any populist message that addresses only the plight of the poor. Stellenbosch society is ranked as the most unequal society in South Africa. Populism and unbalanced public participation in society contribute greatly to the spread of false messages on water issues in the various communities. This would have a very negative impact on the implementation of a treated municipal wastewater reuse project. It is imperative that municipal officials work on their relationship with poor communities by addressing and implementing the concerns of these communities. After all, trust is earned through action. At the same time, rich communities should be assured of continued satisfactory water supply.

(f) No. 33: Perspectives of the value of water

A correct estimate of the value of water is extremely important. It is exacerbated by the public's overambition in terms of receiving municipal services and lack of awareness of water. To address this problem, municipal officials should constantly speak truthfully to their communities about water issues and conduct education campaigns aimed at improving the perception of the value of water in Stellenbosch communities. A society with a correct perception of water will meaningfully cooperate in water initiatives and understand why expensive technology must be purchased and implemented in the community, including technology usage in communication and educational procedures devised for the communities.

Table 6.4: Summary of first-level elements and a number of elements influenced by the first level

First-level element	Number of elements influenced
Overambitious of public perceptions of service quality	Four elements
Lack of water awareness	Four elements
Disruptive events	One element
Population growth	One element
Populism and ineffective public participation	One element
Perspectives of the value of water	One element

6.5.1.2 Second-level elements

(a) No. 34: Technology

In the interpretive structural model, the arrow pointing to technology from the first four key elements indicates that failure to address disruptive events, population growth, populism, and ineffective public participation will greatly influence the choice of technology needed to implement the reuse of treated municipal wastewater in Stellenbosch Municipality. The use of technology can range from communications to education campaigns to wastewater treatment technologies. Failure to provide appropriate technology for the project will result in theft and vandalism, which are common in poor communities. A wide range of technologies is available for natural risk assessment. However, if the right technology is not chosen for the task, it can mislead the community, which will undermine trust between community officials and the community. In addition, if the wrong wastewater treatment technology is used and the public is aware of the adverse effects of treated municipal wastewater, this would affect the public's perception of treated municipal wastewater. Several factors were summarised, all related to water policy and laws, and they were listed in the order in which they should be addressed. The most interesting part is how the interpretive structural model points to capacity. This means that if the treated municipal wastewater reuse project is implemented in Stellenbosch Municipality, there will be capacity issues if the block on legislation and governance is not treated well.

(b) No. 16: Capacity

Capacity issues show how they would affect the ability of the treated municipal wastewater to be received by a very important group in Stellenbosch Municipality, namely the farmers. This is very important because they are the main water consumers, for their agricultural activities, and the main actors in water management. Their participation is very important in a project to reuse treated municipal wastewater. The agricultural group also has a greater impact on the last block, which includes issues related to the management or implementation of the treated urban wastewater reuse project. These include administrative capacity, water rates, procurement challenges, budget, and specific skills needed for such an initiative. Farmers, for example, have the potential to make the most use of treated municipal wastewater and can influence the price of treated water. In addition, this community has specific water standards that meet their agricultural needs and would require highly trained staff to meet their needs, without whom the initiative could fail. Unnecessary bureaucracy in the procurement processes is subject to the demands of farmers and other businesses that must use the initiative to reuse treated municipal wastewater.

Figure 6.5 shows the three main components of problems to consider in the meaningful implementation of a treated municipal wastewater reuse project, namely societal, institutional, and implementation problems. From the interpretive structural model, it can be seen that strong drivers in the implementation of treated municipal wastewater reuse form the left side of the model. These elements include primarily social elements. Once the social issues in implementing treated municipal wastewater reuse are successfully addressed, the institutional arrangements around such a project form the middle block and were listed according to their degree of impact and the order in which they should be addressed. Once the social and administrative issues are resolved, a treated municipal wastewater reuse project can be implemented by addressing the elements in Step Three of Figure 6.5 in the order in which they are listed. The performance of the WWTP is the final part of the process. Also fascinating about the perceptions of the reuse of treated wastewater was the influence that groups with vested interests such as agriculture and business have. This suggests that the agricultural and business communities of Stellenbosch are the key stakeholders in implementing a treated municipal wastewater reuse project. The dangers of unregulated drilling were pointed out, including that aquifers will not be recharged and that the underground water system will not be replenished as rainfall decreases due to climate change. The need for alternative water sources was expressed and the farmers and business people in the focus group highly supported the project to reuse treated municipal wastewater.

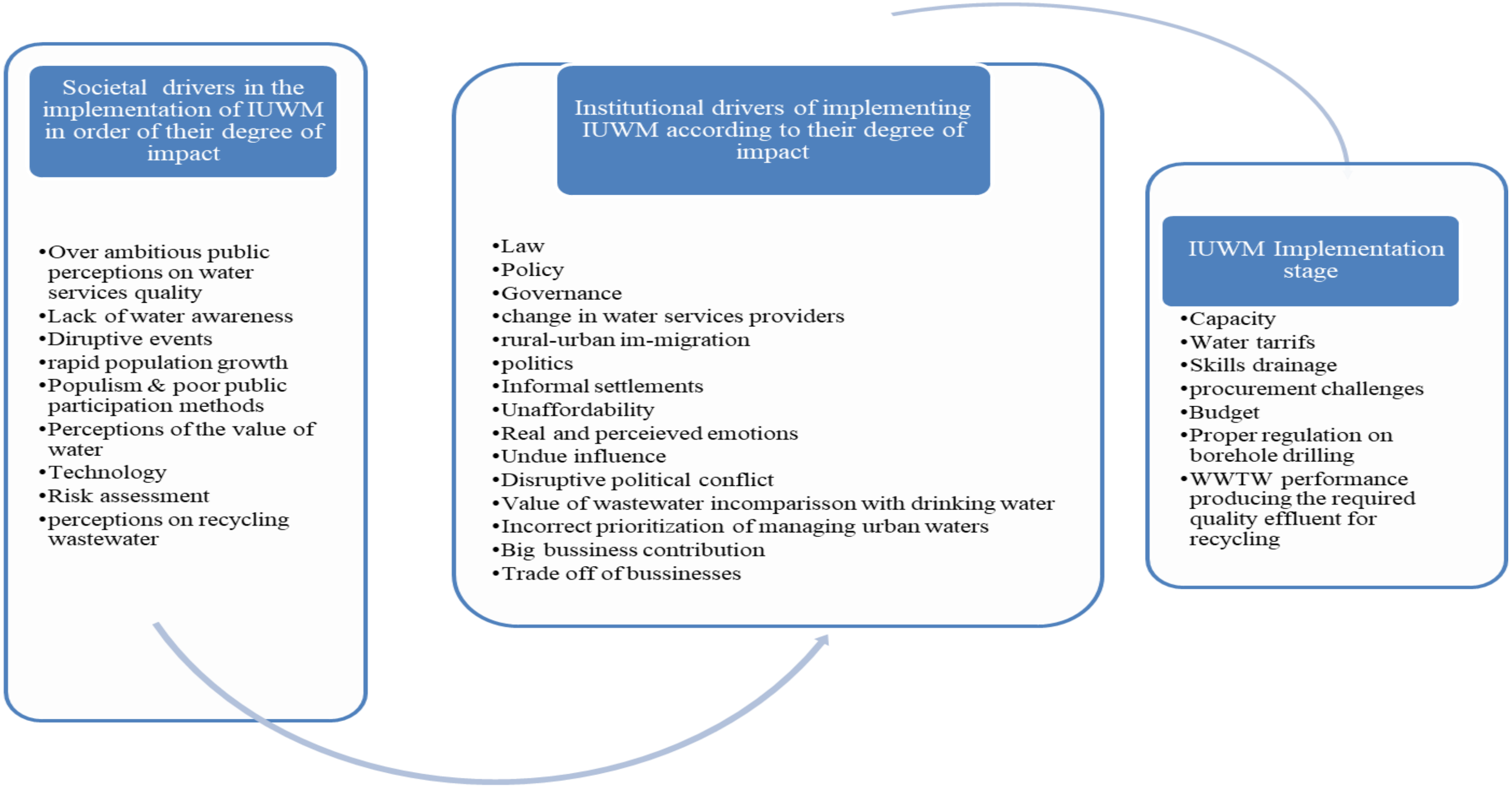


Figure 6.5: Summary of the major components of the issues to be addressed in the implementation of treated municipal wastewater reuse in Stellenbosch Municipality

Source: Researcher

6.6 SUMMARY

The study found that Stellenbosch Municipality follows a water management approach that is top-down, command-and-control, technocratic, and linear. As a result, 62% of the abstracted freshwater is used for toileting, washing, and bathing. The resulting wastewater is piped through the sewer system to a central treatment plant where it is treated and discharged into natural waters. This scenario sparked the researcher's interest in investigating the possibility of reusing the municipal wastewater collected and treated in Stellenbosch Municipality.

Several elements that are important in implementing a project such as the reuse of treated municipal wastewater were investigated. The condition of the water infrastructure in Stellenbosch Municipality was determined. The condition of the water infrastructure was described as catastrophic, having lost 52.3% of its value and reactive rather than proactive water management. In order to improve the steadily deteriorating infrastructure, Stellenbosch Municipality water managers have developed a Water Supply Master Plan to upgrade the fresh water supply system. This programme is intended to reduce problems associated with water leaks, as currently an estimated 25% of the water that enters the system is unaccounted for. Plans also call for large storage tanks to increase raw water storage capacity.

A project to upgrade the wastewater treatment facilities from 20 ML/d to 30 ML/d has been approved. The most interesting aspect of the upgrade is the treatment technology selected, namely the membrane bioreactor, which produces high-quality effluent that can be reused for various purposes without further treatment. These findings led the researcher to investigate the barriers to implementing the treated municipal wastewater reuse initiative in Stellenbosch Municipality. Qualitative data collection was conducted through a focus group workshop using the interactive management research methodology to achieve this goal.

The data-collection methodology, results, and discussions at the interactive management workshop were presented. From the research findings, the key drivers for implementing treated municipal wastewater reuse were identified as lack of water awareness, disruptive events, population growth, populism, and ineffective public participation, which are human-centred. The interpretive structural model outlined the strategic roadmap that must be followed sequentially to successfully implement the treated municipal wastewater reuse initiative in Stellenbosch Municipality. In essence, a people-centred approach to implementing the reuse of treated municipal wastewater in Stellenbosch is strongly recommended. Figure 6.6 describes this approach.

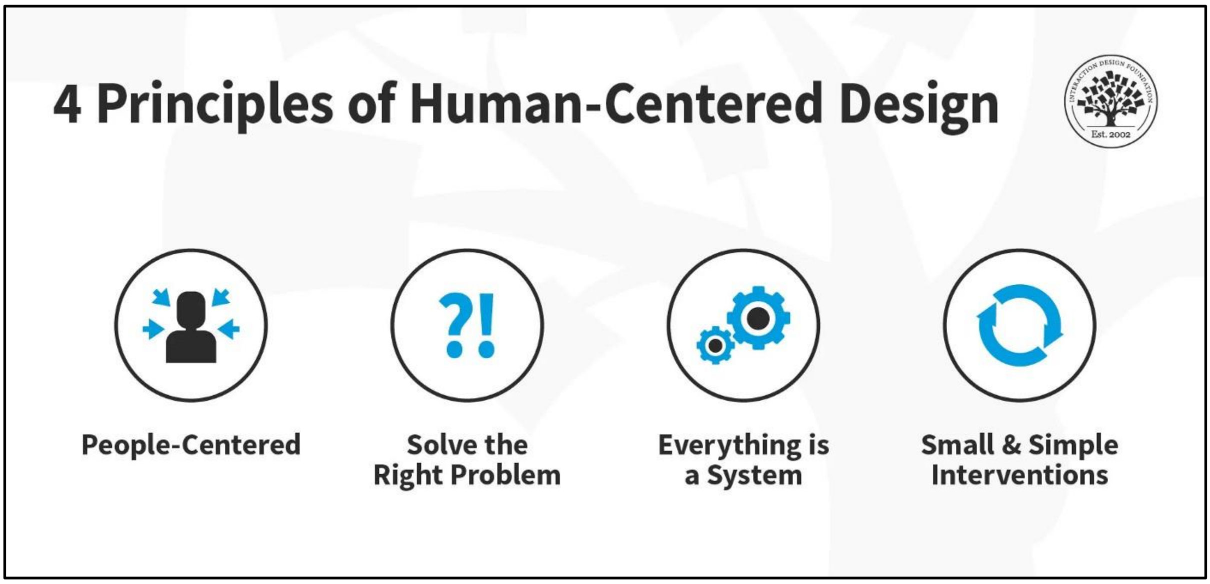


Figure 6.6: The four principles of human-centred design

Source: Norman and Draper (1986)

It is worth noting the role that can be played by the agricultural sector, which is very water-intensive, yet is the largest economic contributor in the region. To address some of the challenges in Stellenbosch's urban water system, such as water storage capacity, deteriorating water infrastructure, and water leakage, the following chapter presents water demand and supply forecasting models to help water managers solve their water system challenges.

CHAPTER 7:
MODEL DEVELOPMENT

7.1 OVERVIEW

This chapter presents the detailed processes and procedures of the models that were developed to support Stellenbosch Municipality in effectively addressing the water system challenges outlined in Chapter 6. The equations of the specific conventional time series and supervised machine learning algorithms are presented along with the assessment metrics used. The approach, results, and discussion of the modelling process are the focus of this chapter.

As mentioned in Chapter 1, the fourth objective of this study was to develop, train, and deploy a highly accurate supervised machine learning model that can support water policy and decision makers in Stellenbosch Municipality. This will lead to the sustainable management of current and future water demand and supply in their jurisdiction. This chapter builds on the work presented in Chapters 4 and 5. Chapter 4 explored the application of supervised machine learning algorithms in managing urban water systems, which led to the description of the conventional and supervised machine learning algorithms proposed to be used in this study. Chapter 5 discussed the rationale for the specific urban water demand and forecasting algorithms proposed and presented the methodological framework for modelling and process flow. This chapter draws on these descriptions and discussions.

7.2 PROBLEM FORMULATION

According to Martin *et al.* (2020), problem formulation is critical in modelling. A well-formulated problem can improve the performance of the model and allow end users to draw more insights from the model to answer various research questions (Passi & Barocas, 2019). The critical component of problem formulation is the identification of the target variables and the solutions that the model should provide. This allows the modeller to make an appropriate selection of independent variables. This study focused on the following identified problems in Stellenbosch Municipality's urban water system: (1) water supply capacity, (2) dilapidated infrastructure, and (3) leakage in the supply system.

The thesis of the modelling process is therefore based on the recognition that accurate short- and medium-term forecasting of urban water demand and supply will enable water agencies to:

- accurately plan and manage current and future water demand;
- efficiently manage the operation of the urban water system;
- plan appropriately for water infrastructure upgrades;
- formulate strategies that respond to the issues at hand; and
- accurately budget operating costs and water rates.

However, the following factors reduce the accuracy of forecasting and predictive models developed using conventional algorithms:

- unchecked population growth and rapid urbanisation;
- negative impacts of climate change, which abruptly alter precipitation cycles;
- increasingly complex factors that affect water demand; and
- unquantifiable uncertainties in the water system.

These factors contribute to models that either perform too well, which can lead to the construction of oversized water systems, or perform too poorly, which can underestimate future water needs and lead to freshwater shortages. As a result, inadequate models have negative implications for the operation and management of water infrastructure and services to provide sufficient clean water to consumers.

To achieve the main goal of this chapter, the following objectives were pursued:

- Collect and clean the data collected from various government agencies.
- Perform EDA using the compiled CSV file.
- Create, train, test, deploy, compare, and contrast the performance of conventional and machine learning developed models to predict water demand and supply in Stellenbosch Municipality.
- Investigate the impact of treated wastewater reuse on water supply in Stellenbosch Municipality.
- Provide the best-performing model to Stellenbosch Municipality's water authority to enable them to more accurately predict and forecast short- and medium-term water demand and supply for their jurisdiction.

In formulating the research problem, a target variable was identified and, in this case, the EDA step described in the supervised machine learning workflow (see Figure 5.8) facilitated this process.

7.2.1 Hypothesis

The hypothesis of this study was that:

Null hypothesis (H_0): Supervised machine learning models can accurately predict and forecast urban water demand compared to conventional models.

Alternative hypothesis (H_A): Supervised machine learning do not accurately predict and forecast urban water demand compared to conventional models.

7.3 EDA RESULTS AND DISCUSSION

This section presents the EDA results obtained from sifting through the data collected in the Microsoft Excel spreadsheet, which were examined using pivot tables. This process provided the researcher with deep insights into the data and enabled the researcher to identify the target variable, which is also described in this section. The rationale for the choice of the target variable is also explained.

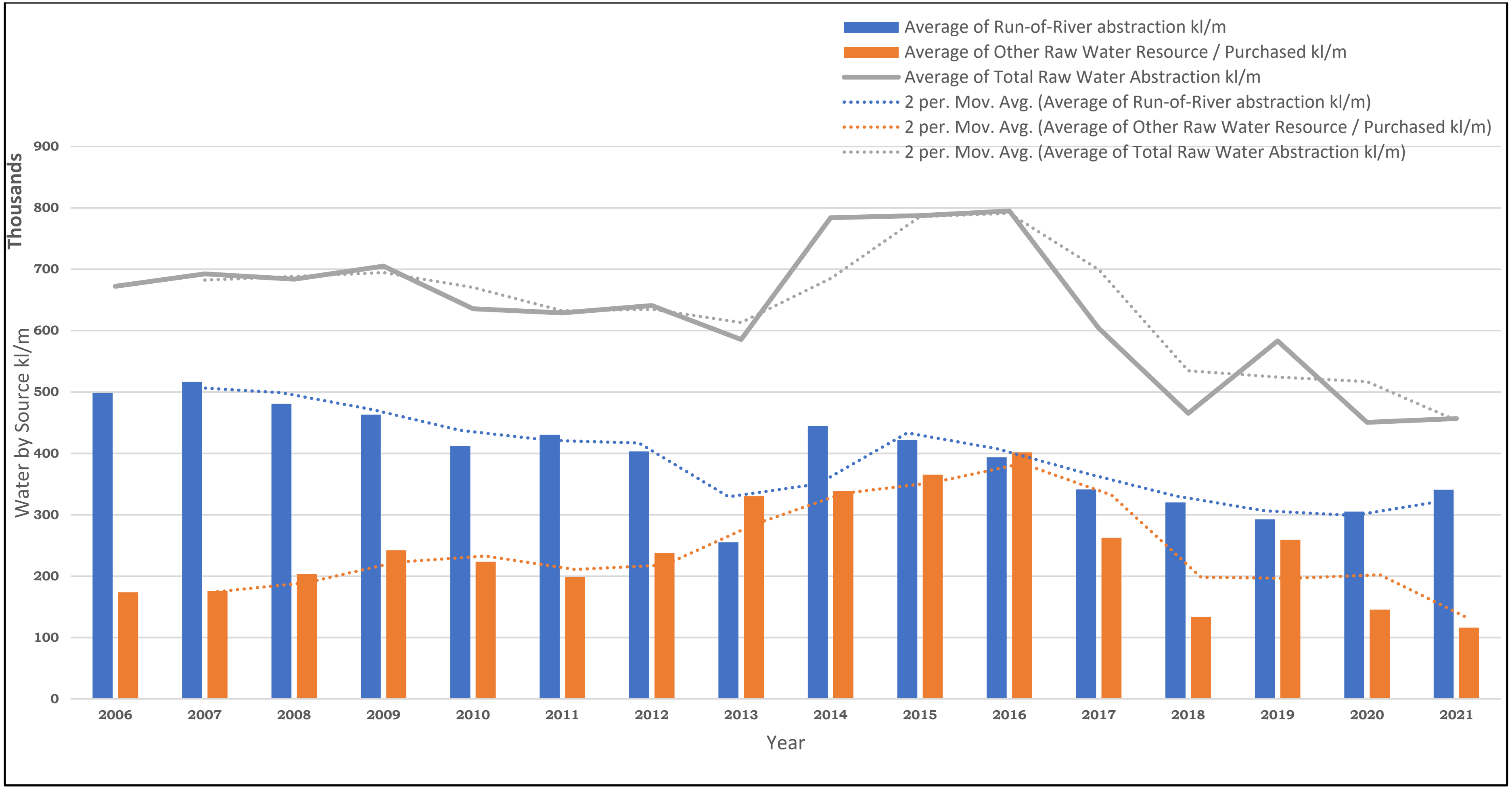


Figure 7.1: Average of run-of-river abstraction (RoRabs) and average raw water purchased over the years

Figure 7.1 shows a decline in average withdrawals from river flows from 2007 to 2013, while average withdrawals from other raw water resources increased steadily. From 2014 to 2016, both withdrawals from river flows and withdrawals from other raw water resources increased. In 2016, withdrawals from the river and purchases from other raw water resources balanced each other. Thereafter, the general trend is for withdrawals from the river and purchases from other raw water resources to decline, which reached the lowest level in 2021. The overall increase in purchases from other raw water resources triggered by the drought, leading to parity with river water withdrawals, is of concern because reliance on external sources poses a high risk to Stellenbosch Municipality as a WSA. Firstly, reliance on external sources to meet its urgent water needs is unsustainable and does not guarantee an efficient water supply in its jurisdiction. The scenario described above indicates that Stellenbosch Municipality is not in a position to deal with issues that could affect its mandate to ensure a sustainable water supply in its jurisdiction. At this point, the idea of alternative water sources needs to be explored in depth. To this end, the study explored the reuse of treated municipal wastewater as an alternative water resource to improve Stellenbosch Municipality's water supply, rather than purchasing large quantities of raw water from external sources, which cannot be guaranteed, nor is it best practice in water resources management. To the researcher's knowledge, Stellenbosch Municipality has not yet considered the practical reuse of treated municipal wastewater as an alternative water source. In addition, water abstraction from rivers is generally declining; there is thus a great need for Stellenbosch Municipality to explore and utilise alternative water sources.

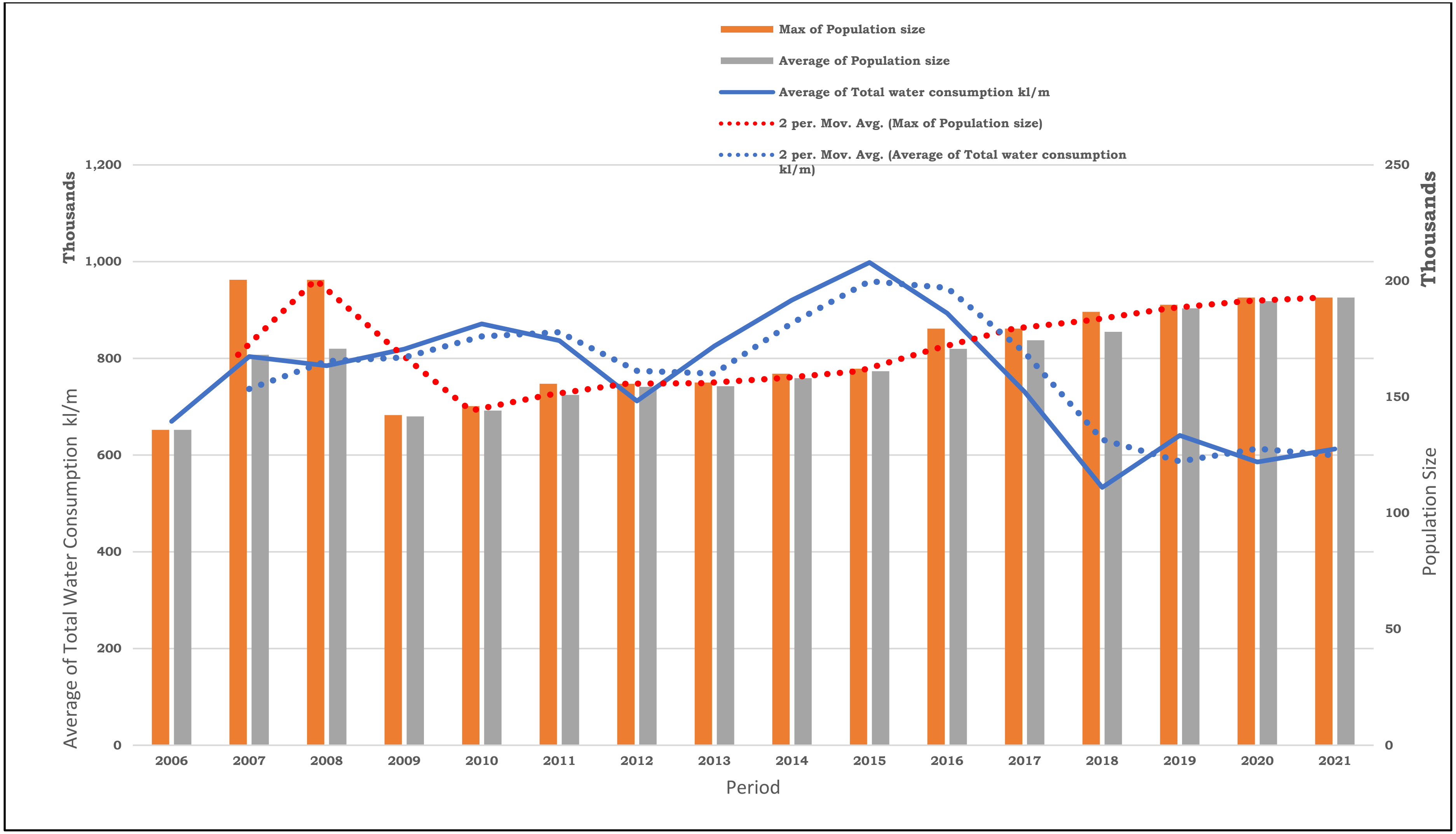


Figure 7.2: Maximum of population size, average of population size, and average of total water consumption versus period

Figure 7.2 shows the trend in average total water use compared to average population growth. The graph shows a general increase in the population during the study period, with the exception of 2007 and 2008, for which the researcher was unable to determine exactly what happened based on the document analysis. To verify the quality of the data, an average population size was calculated for each year. In contrast, total water use, which is assumed to increase as population size increases, jumped from 2006 to 2015. A sharp decrease in total water consumption was recorded for the period 2015 to 2018. Thereafter, the trend levels off to the lowest values during the study period while the population is at its highest. From this analysis, the disparity between population growth and total water consumption, which are not synchronised, is due to reactionary water policies that were very effective. For example, the sharp decline from 2015 to 2018 was the result of a strict water restriction policy formulated to avert the threat of Day Zero in the region. This policy was formulated in the context of a regional crisis, which is an example of reactive rather than proactive management of water resources. The lack of correlation between population growth and total water consumption is also indicative of the lack of adequate planning and strategy to meet water needs in Stellenbosch Municipality's jurisdiction. Instead, water resources management is determined by the dictates of external policies – in this case national and provincial policies. This institutional arrangement poses a high risk to the management of water resources by Stellenbosch Municipality, as it is by law the water authority, yet appears to have minimal control over the management of water resources within its jurisdiction. This analysis also suggests that total water use is not a good candidate as a target variable.

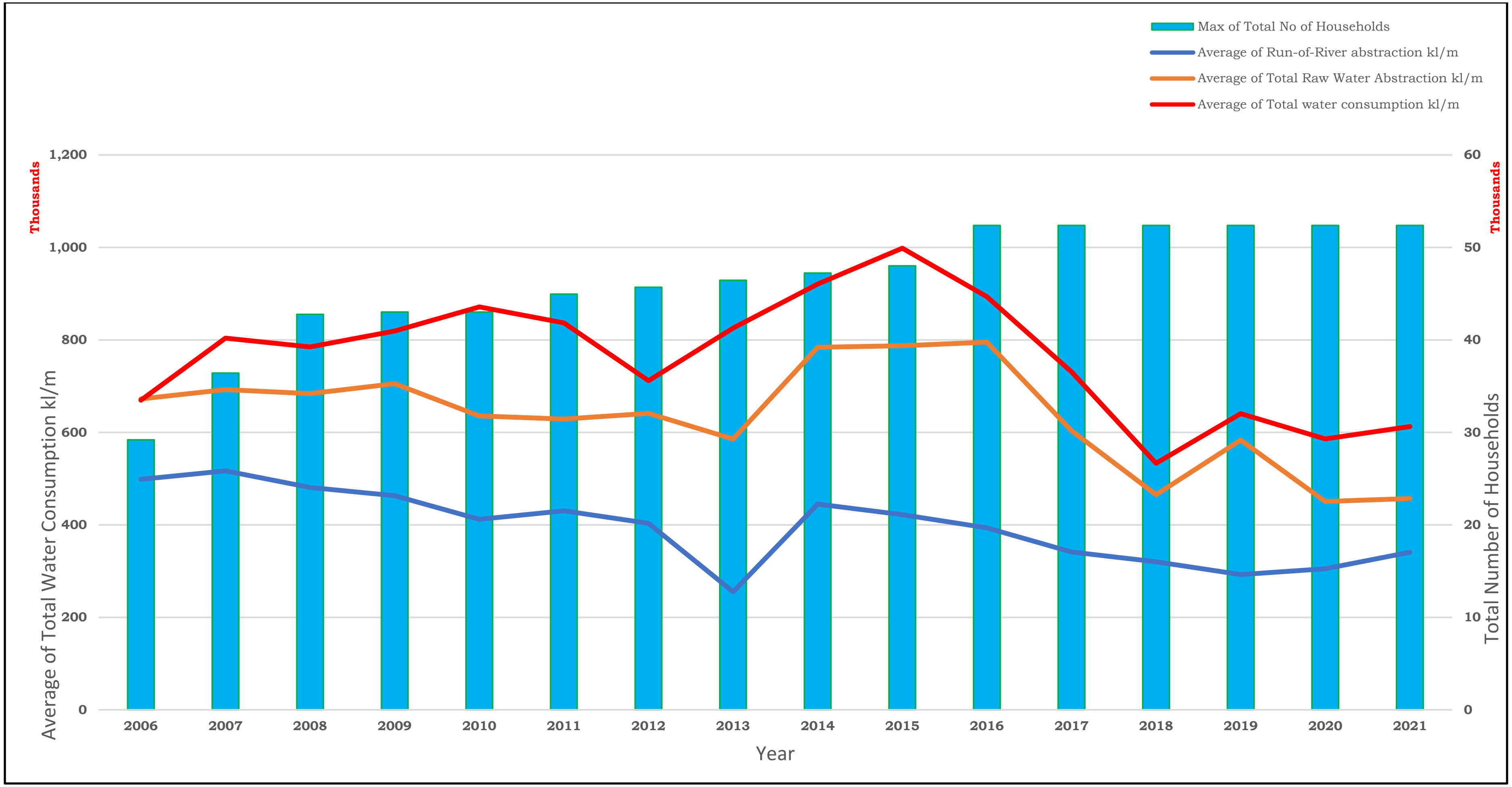


Figure 7.3: Maximum of total households, average of RoRabs, average of total raw water abstraction, and average of total water consumption versus period

Figure 7.3 also confirms the finding of a mismatch between total water use, river withdrawals, total raw water withdrawals, and number of households. The best practice for water resources management would be that as the number of households increases, total water withdrawals should also increase to meet the water demands of the growing number of households. It is worth noting that water withdrawals from rivers have continued to decline since 2014, even though the total number of households to be served has increased. It is worth noting that water abstraction from rivers in 2021 was lower than abstraction in 2006, which is a scenario that requires urgent consideration of alternative water sources within Stellenbosch Municipality's jurisdiction.

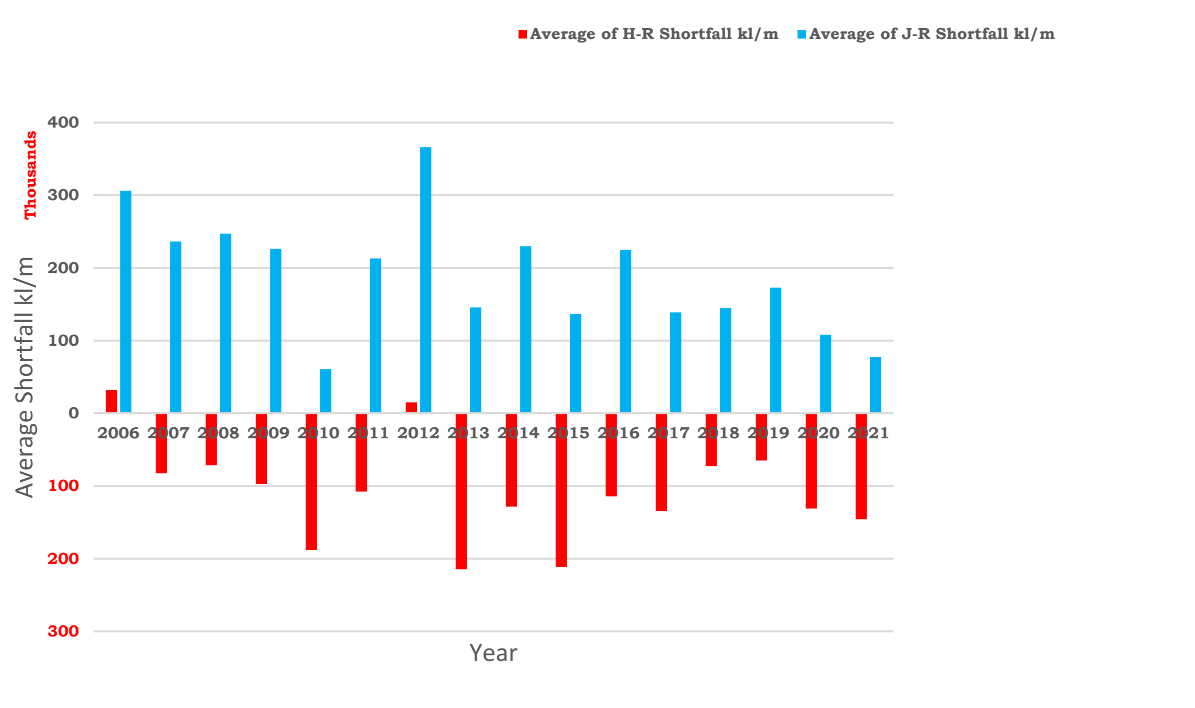


Figure 7.4: Average of treated water from all water treatment works (WTWs) (H) minus total water consumption R (shortfall), average of systems input volume (J) minus total water consumption (R) (shortfall) versus period

The researcher investigated whether Stellenbosch Municipality provided an adequate supply of clean water in its jurisdiction during the years of the study. Figure 7.4 shows that water shortages occurred in Stellenbosch Municipality that were mitigated by the procurement of additional water from external sources. Only in 2006 and 2012 were there no water supply shortages prior to bulk purchases of untreated and treated water. In the other years studied, additional raw and treated water had to be purchased to meet the community's water needs. This is problematic as water prices will inevitably increase and there is an accompanying increased risk that external water resources may not be available for Stellenbosch Municipality.

7.3.1 Target variable: RoRabs

As part of the EDA process, the researcher identified RoRabs as an appropriate target variable. This is because all decisions about the following quantities depend on the RoRabs quantity: total raw water quantity to be purchased, total raw water withdrawal quantity, total raw water quantity to be injected into all WTWs, total raw water quantity before treatment, and total raw and treated water quantity to be purchased. This is the only volume over which the Stellenbosch municipal water authority has some control. It should be noted that weather is also a critical factor in the withdrawal of water from the river. Consequently, decisions on water treatment budgets and reservoir capacities are also highly dependent on RoRabs quantity. Accurate forecasting of RoRabs volume is therefore essential for decision making by Stellenbosch Municipality regarding the management of the entire water supply system. This includes considering what steps should be taken to reduce water withdrawals from the river, such as the reuse of treated municipal wastewater. In addition, water losses should be reduced and deteriorating water infrastructure should be addressed. In essence, forecasting future water withdrawals from the river will improve the efficient operation and management of the Stellenbosch urban water system, facilitate effective planning for the growing water demands of the population, meet the water needs of agriculture and businesses in Stellenbosch, and improve planning for water infrastructure development. Accurate forecasting of RoRabs volume will also inform policy and decision makers of Stellenbosch Municipality's water needs.

7.4 TARGET VARIABLE MODELLING

7.4.1 Time series modelling the target variable (RoRabs)

A conventional (traditional) SARIMA time series model was developed to predict the target variable (RoRabs).

7.4.1.1 Methodology

The researcher began by defining the model to be developed as follows. The basis of the SARIMA model is the ARIMA model, to which three new hyperparameters were added, namely autoregression (AR), differentiation or integration (I), and moving average (MA) for the seasonal component of the series, as well as an additional parameter to indicate the period of seasonality. The notation for the SARIMA model parameters was captured by Guo *et al.* (2018) with the expression SARIMA (p, d, q) (P, D, Q) m and formulated by Bata *et al.* (2020) as follows:

- p and seasonal P indicate the number of autoregressive terms (lags of the stationary series).
- d and seasonal D indicate the differencing that must be performed to stationarise the series.
- q and seasonal Q indicate the number of terms of the moving average (lags of the forecast errors).
- m indicates the seasonal length in the data.

The lower- and upper-case letters refer to the non-seasonal and seasonal components of the model respectively. The following steps were followed to develop the model:

(a) Step 1: Visualising the data

The original dataset was split into a training set and a test set in a 60:40 ratio. The data were then analysed for stationarity, seasonality, and trend. Once stationarity was confirmed, which is a prerequisite for time series modelling, modelling could be performed with both the training and test datasets.

(b) Step 2: Model selection

The best ARIMA and SARIMA models were found using the "auto.arima()" function developed by Hyndman and Khandakar (2008). This function finds the best model by using the unit root test to evaluate the non-seasonal and seasonal degrees of difference required to make the time series stationary and by minimising the Akaike Information Criterion (AIC). To achieve the optimal SARIMA models, separate non-seasonal and seasonal models are first computed and then combined. Using the "auto.arima" function, the best (optimal) SARIMA models were obtained and assessed with the test data.

(c) Step 3: Model fitting

A 66-step forecast was compared with actual trends (real-time data) to evaluate the overall performance of the models. The researcher forecast the period from 2016-01-01 to 2021-06-01 and used two measures of model accuracy. These are the mean absolute percentage error (MAPE) and the root mean square error (RMSE). Predictions were also made for a 36-level step of the SARIMA (3, 1, 0) models and the measure of their confidence interval was presented.

(d) Step 4: Model performance evaluation

- Evaluation metrics

In model development, the performance evaluation of the developed models is crucial, considering that old data are used to develop the predictive model, which in turn is supposed to make predictions for new data where the answer is unknown. Moreover, the results of these models are probabilistic; it is therefore imperative to evaluate the accuracy of the model performance. Several evaluation metrics have been developed for this purpose. However, in this study, the RMSE and MAPE were primarily considered.

i. Root mean square error (RMSE)

RMSE is equal to the square root of the average of the squared difference between the target value and the value predicted by the regression model. Mathematically, it is expressed by the following equation:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\breve{y}_i - y_i)^2}$$

Where:

N is the number of data points;

$\breve{y}_i$ represents the predicted values; and

y_i is the observed values.

The RMSE addresses some weaknesses of the mean squared error (MSE) by retaining the differentiable property of the MSE, as follows:

- It handles the punishment of minor errors by the MSE by rooting them quadratically.

- Error interpretation can thus be smooth since the scale is now identical to the random variable.

ii. Mean absolute percentage error (MAPE)

The RMSE is a measure of the magnitude of error in the regression and does not indicate the explained component of the regression fit (De Myttenaere *et al.*, 2016). This study therefore also used MAPE as an evaluation measure, which is considered the most useful measure for comparing prediction accuracy across different items or products (Ostertagová, 2012). It is commonly used in quantitative forecasting methods. The following equation describes the measures:

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\breve{y}_i - y_i}{y_i}\right|$$

When the calculated MAPE value is less than 10%, it is interpreted as excellent prediction accuracy, between 10% and 20% as good prediction, between 20% and 50% as acceptable prediction, and above 50% as inaccurate prediction (De Myttenaere *et al.*, 2016).

7.4.1.2 Results and discussion

(a) Step 1: Visualising the data

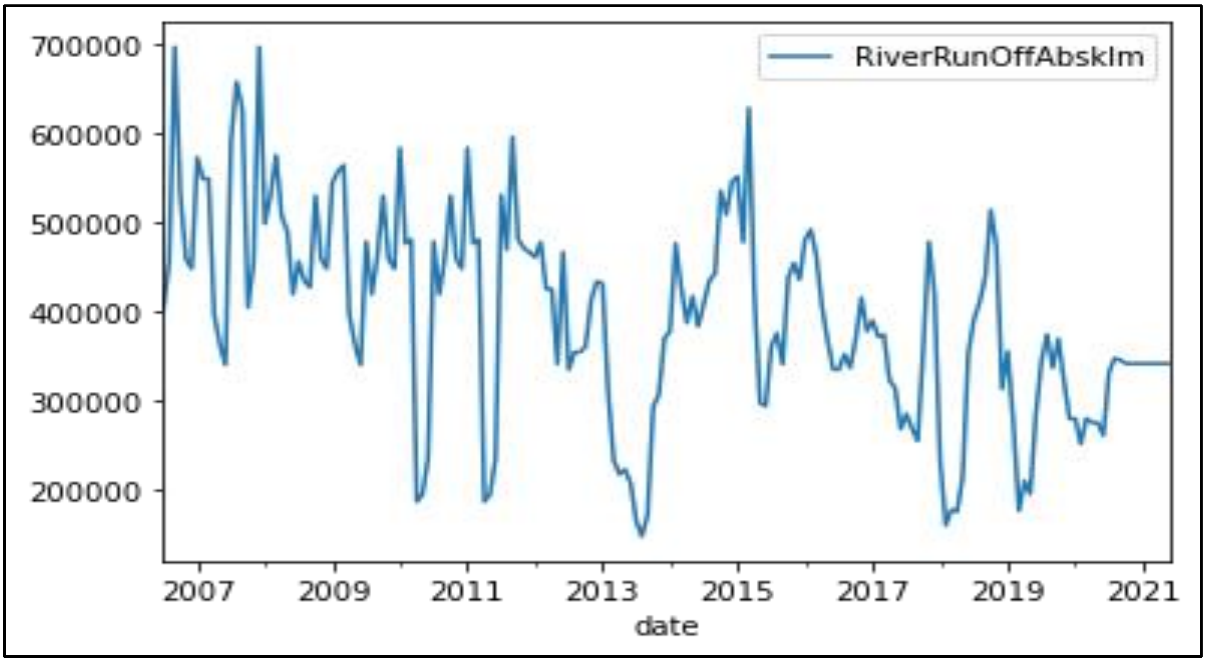


Figure 7.5: Trend and seasonality graph of RoRabs

Figure 7.5 shows the RoRabs graph, which exhibits some seasonality, trend, and non-stationarity of the original dataset.

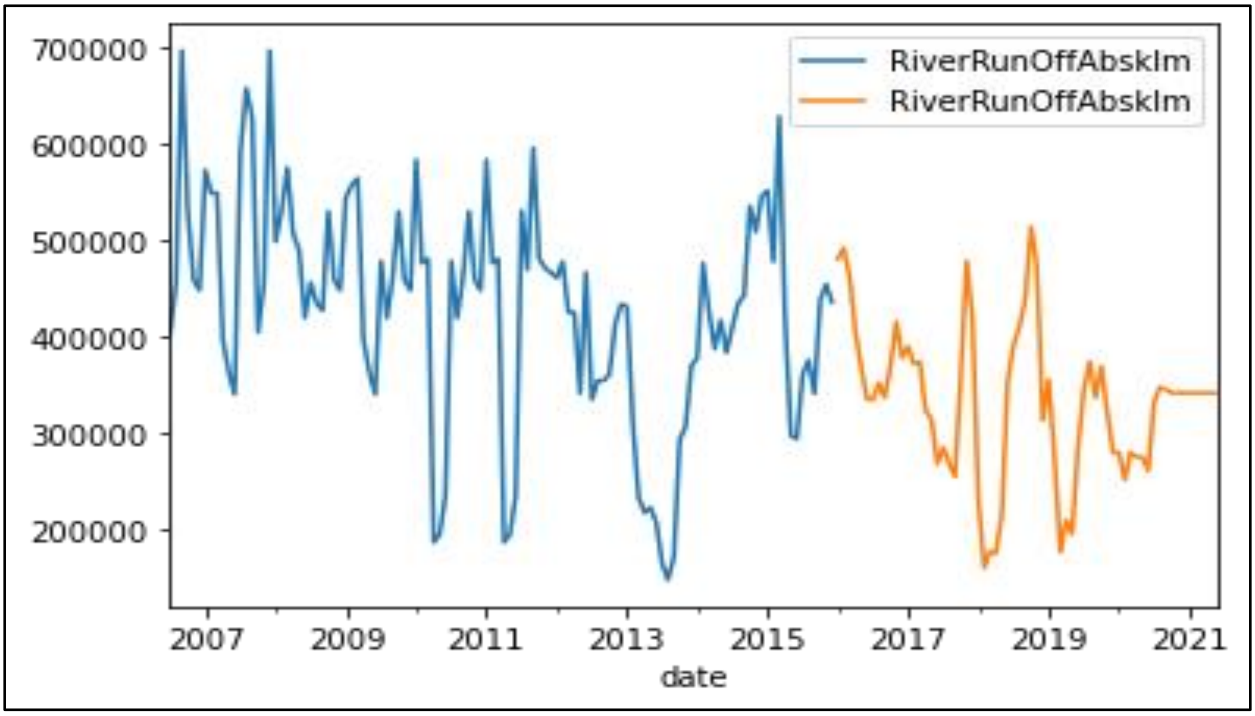


Figure 7.6: The RoRabs model

Figure 7.6 shows the RoRabs model, with the blue series describing the training set and the amber series describing the test set to which the predictions are compared.

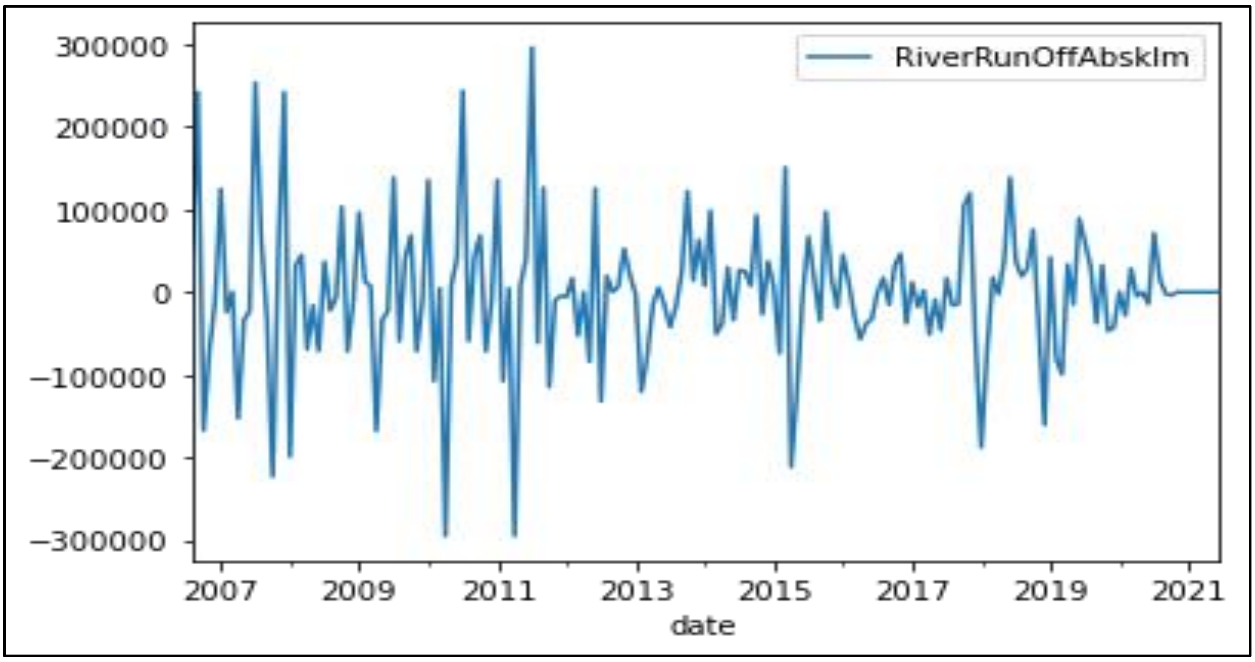


Figure 7.7: RoRabs time series of first-order differenced dataset

After performing first-order differencing, the dataset was stationary, as shown in Figure 7.7, and a p-value of 4.51 x10^{-6} was obtained. Three levels of differencing were performed for model development; i.e., single first-order differencing and two seasonal differencing. The reason was that models with two seasonal differentiations have a lower AIC value than models with single seasonal differencing.

(b) Step 2: Model selection

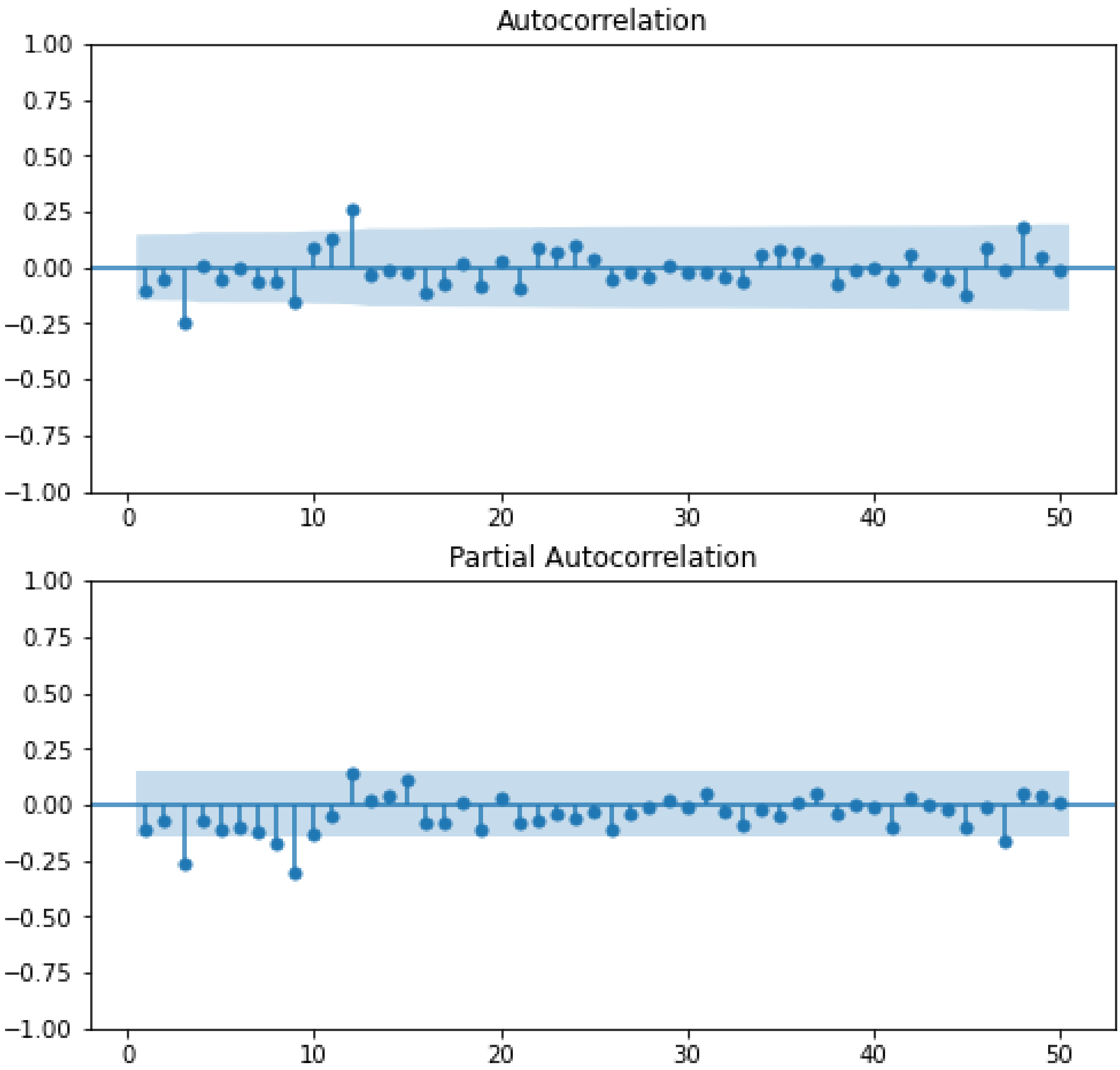


Figure 7.8: The auto correlation function (ACF) correlogram of the original data and partial auto correlation function (PACF) correlogram of the original data

One of the many ways to determine the correct model order in time series is to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The plots are shown in Figure 7.8. See Appendix E1 Conventional Models for details on the derivation of the model order. By comparing the ACF and the PACF, the researcher can infer the model order. However, in this example, both ACF and PACF run out and the values of p and q cannot be derived from the plots of ACF and PACF, as shown in Figure 7.8.

The next option considered by the researcher was using the grid search, several ARIMA models were identified as high-performing models based on their AIC value, of which the researcher selected ARIMA (1, 2, 4), ARIMA (0, 2, 4), and ARIMA (3, 2, 4). Finally, the best SARIMA models selected were SARIMA (1, 2, 4), SARIMA (3, 1, 0), and SARIMA (3, 2, 4).

(c) Step 3: Model fitting

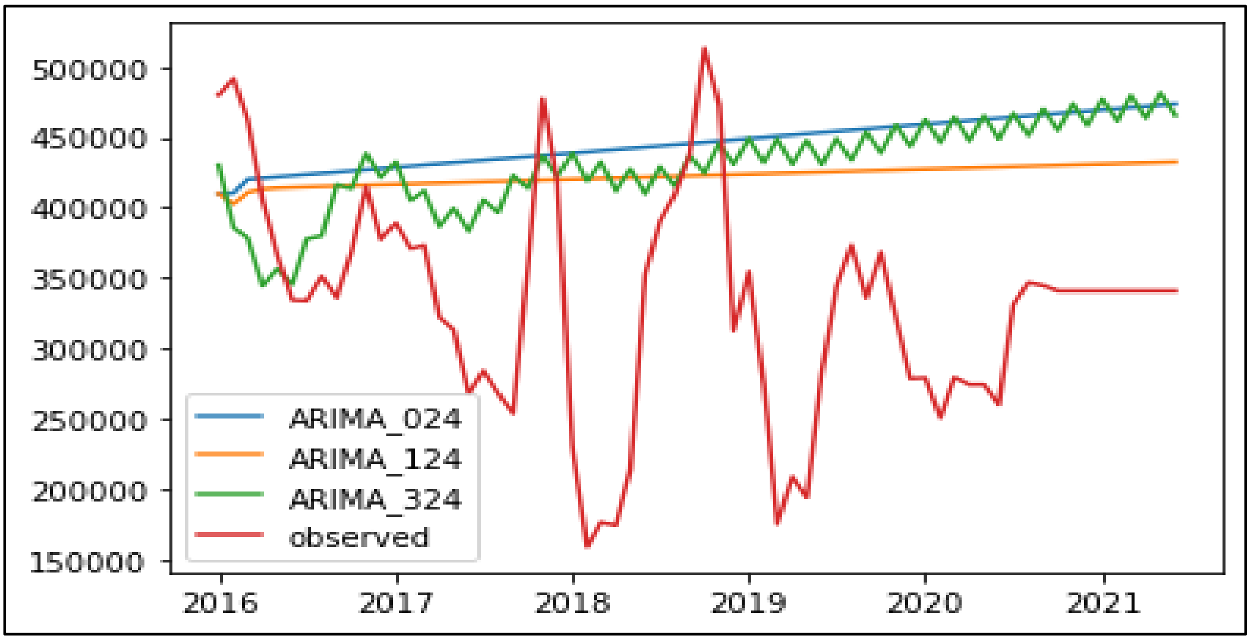


Figure 7.9: Time series plots of 66 months' time step of the ARIMA (3, 2, 4), ARIMA (0, 2, 4), and the ARIMA (1, 2, 4) models in comparison to the observed model

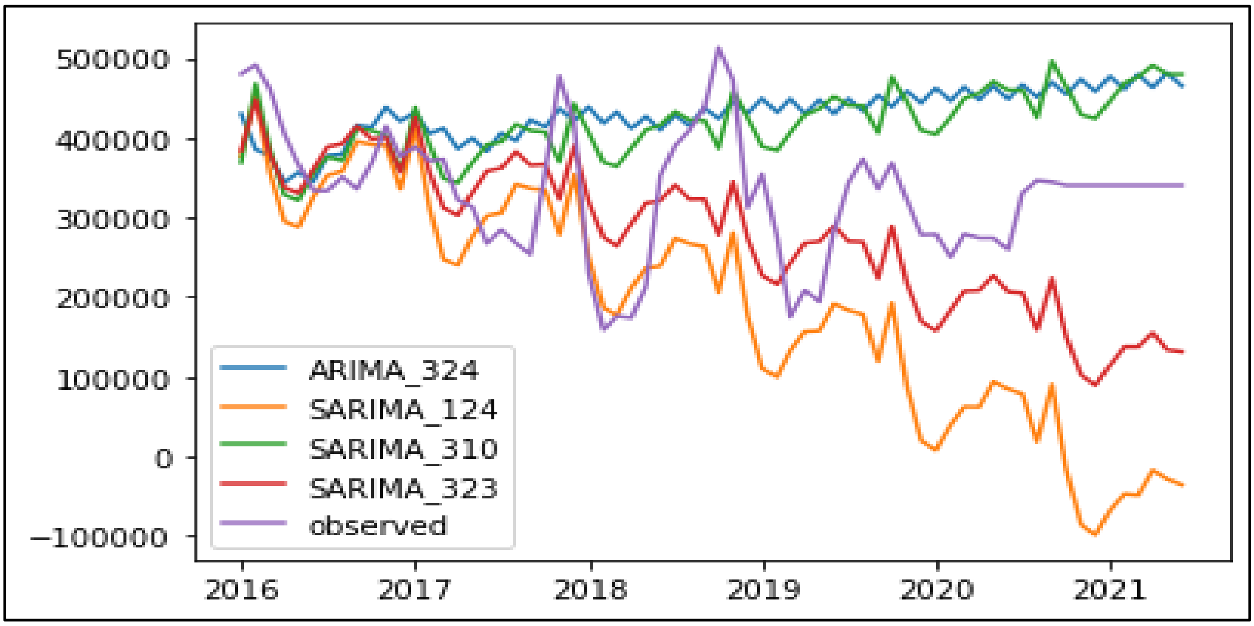


Figure 7.10: Time series plot of 66 months' time step of ARIMA (3, 2, 4), SARIMA (1, 2, 4), SARIMA (3, 1, 0), and SARIMA (3, 2, 3) models in comparison to the observed model

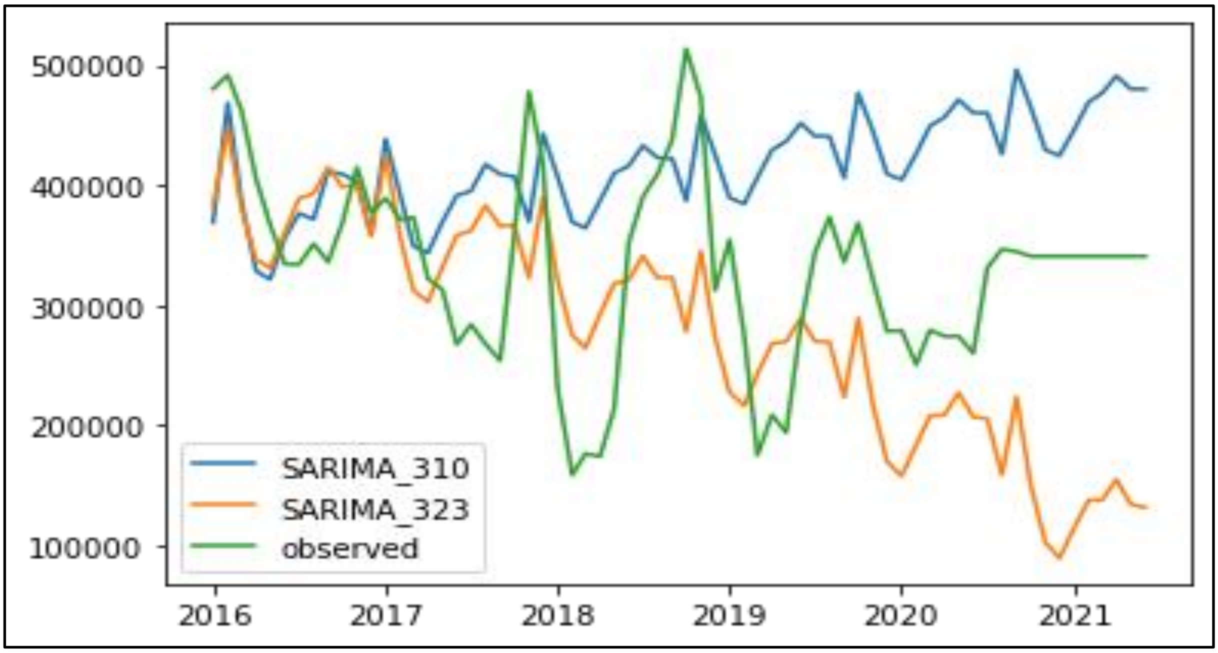


Figure 7.11: Time series plot of 66 months' time step of SARIMA (3, 1, 0) and SARIMA (3, 2, 3) models in comparison to the observed model

Figures 7.9, 7.10, and 7.11 show a comparison of ARIMA and SARIMA models with the observed time series. The SARIMA models exhibit a better trend of the observed time series than the ARIMA models. However, the SARIMA (1, 2, 4) model produced negative values of the RoRabs that are meaningless. SARIMA (3, 1, 0) and SARIMA

(3, 2, 3) were found to be the best-performing models, as confirmed by the evaluation metrics.

(d) Step 4: Model performance evaluation

- Error measures

The accuracy of the prediction was evaluated using error measurements. For this purpose, the predicted values from the training set and the test set were compared. As mentioned earlier, the MAPE and RMSE were considered more appropriate for this study compared to other metrics. The results in Table 7.1 show that the ARIMA (1, 2, 4) model performed reasonably well for the 66-month step. However, the SARIMA (3, 1, 0) and SARIMA (3, 2, 3) models performed the best. The prediction window was from 2016-01-01 to 2021-06-01, but from 2021-02-01 the SARIMA predictions were negative, which is meaningless. This is one of the major drawbacks of SARIMA time series models when they make predictions that extend far into the future. Therefore, the predictions should be as close as possible to the data points used to train the model. For this purpose, a 33-step prediction was performed and the confidence interval obtained showed better results. See Appendix E1. Conventional models for details on 33-step predictions.

Table 7.1: Evaluation metrics of models developed

Performance Index	ARIMA (3, 2, 4)	ARIMA (0, 2, 4)	ARIMA (1, 2, 4)	SARIMA (3, 1, 0)	SARIMA (3, 2, 3)	SARIMA (1, 2, 4)
RMSE (%)	40.0	42.9	36.7	37.1	34.4	60.1
MAPE (%)	41.9	46.0	38.9	38.4	30.3	46.7

7.4.2 Machine learning modelling procedure, results, and discussion

The results of the EDA in Section 7.3 and the time series modelling of the target variables in Section 7.4.1.2 served as the basis for the machine learning modelling.

7.4.2.1 Methodology

In general, the machine learning process consists of four main phases, namely EDA, model building, model fitting, and evaluation. During EDA, the following steps are performed: organisational and structural data analysis, feature engineering, feature selection, feature scaling, and feature relationship analysis.

(a) Step 1: EDA

- Organisational and structural data analysis

The main goal of EDA is to provide the researcher with deep insight into the organisational structure of the dataset. In this phase, the researcher analysed and examined the following:

- Missing values
- All numerical values
- Categorical values
- Cardinality of categorical variables
- Outliers
- Relationships between independent variables and dependent variables.

- Feature engineering

The Date column was found to have a high cardinality, and feature engineering was performed to reduce the cardinality, as described in the Jupyter notebook.

The dataset also contained missing values. To deal with missing values, the package “fancyimpute” was imported. This is a package that contains several advanced imputation methods that use machine learning algorithms to impute missing values. Normally, the simplest approach for dealing with missing values is to use imputation techniques such as mean, median, and mode imputations or interpolation. However, these techniques only use the respective columns to calculate and impute missing values. In contrast, the advanced imputation technique “fancyimpute” is superior in that it also uses other columns to predict the missing values and impute them. There are two very important “fancyimpute” techniques, namely KNN (K Nearest Neighbour) imputation and MICE (Multiple Imputation by Chained Equations) imputation. This study used KNN imputation.

- Feature selection

The data features used in training a machine learning model have a significant impact on the performance of the model created. Irrelevant or partially relevant features reduce the model’s performance. To solve this problem, feature selection is used. This is a process that automatically selects the features from the dataset that contribute the most to the target variable. However, in this study, feature selection was unnecessary due to the number and characteristics of the features under consideration.

- Feature scaling

When building machine learning models, feature scaling, also known as standardisation, is a procedure applied to independent variables to normalise the data within a certain range. This is because this procedure helps to speed up the calculations in an algorithm. However, when specific algorithms are used, this exercise is not necessary, which was the case with this study.

- Feature relationships analysis

At this stage, the researcher conducted an analysis of the trends, patterns, and relationships between the independent and dependent features or variables to gain deeper insight into what is happening in the system under study. The relationships between the variables were plotted graphically and the graphs are presented and discussed in Section 7.4.2.2.

(b) Step 2: Model building

In supervised learning, a dependent variable or target variable and independent variables or predictor variables or features must be identified at the beginning of the modelling procedure. As described in Section 7.3.1, the target variable was identified as the RoRabs. After further rigorous data analysis, the independent variables identified to predict the target variable were monthly minimum temperature (mtmin), monthly maximum temperature (mtmax), sum precipitation (spre), and month extracted from date.

Given the final dataset, which exhibited nonlinearity between independent and dependent variables, and its limitations, the researcher considered ensemble machine learning algorithms as suitable modelling tools; specifically, ensemble training with decision trees since the target variable is continuous. Although the researcher originally intended to use the algorithms recommended in Chapter 5, the demands of the research led to a change in the algorithms that would be used to obtain the desired high-performance models. To this end, the researcher decided to use three ensemble boosting methods, namely Adaptive Boosting (AdaBoost), the Gradient Boosting

Model (GBM), and Stochastic Gradient Boosting (SGB), which are currently used in various fields and have proven to be sufficiently powerful models, especially when limited data are available. The researcher also employed the Random Forest Ensemble Bagging method and ANNs. The modelling process is described in detail in the Jupyter notebook (see Appendix E). Since the algorithms that were eventually used were not described in Chapter 4, an overview of these algorithms is provided in the following subsections.

- Adaptive Boosting (AdaBoost) Model

In AdaBoost, each predictor pays more attention to the instances incorrectly predicted by its predecessor by constantly changing the weights of the training instances. In addition, each predictor is assigned a coefficient alpha that weights its contribution to the final ensemble prediction. Alpha depends on the training error of the predictor. An important parameter used in training is the learning rate, eta, which lies between 0 and 1; it is used to reduce the coefficient alpha of a trained predictor. It is important to note that there is a trade-off between eta and the number of estimators. A smaller value of eta should be compensated by a larger number of estimators. Schapire (2013) captured the pseudocode of the AdaBoost algorithm.

- Gradient Boosting Model (GBM)

Gradient boosting is a popular boosting algorithm that has been a winner in many machine learning competitions. In gradient boosting, each predictor in the ensemble corrects the error of its predecessor. Unlike AdaBoost, the weights of the training instances are not changed. Instead, each predictor is trained using the residual errors of its predecessor as labels.

An important parameter used in training gradient boosted trees is shrinkage. In this context, shrinkage refers to the fact that the prediction of each tree in the ensemble is shrunk after being multiplied by a learning rate eta, which is a number between 0 and 1. Similar to AdaBoost, there is a trade-off between eta and the number of estimators. A decrease in the learning rate must be compensated by an increase in the number of estimators in order for the ensemble to achieve a given performance. Essentially, the gradient boosting regression tree builds the model incrementally and updates the model by minimising the expected value of a given loss function. As many trees are added to the model, the fitted model is likely to have a very small training error. The pseudocode for the generic gradient boosting considered in this study was presented by Zhang and Haghani (2015).

- Stochastic Gradient Boosting (SGB)

Gradient boosting involves an exhaustive search procedure. Each tree in the ensemble is trained to find the best split points and the best features. However, this procedure can result in classification and regression trees (CARTs) (decision trees) that use the same split points and possibly the same features. To mitigate these effects, the SGB algorithm is used. In SGB, each CART is trained on a random subset of the training data. This subset is dropped without replacement. In addition, at the level of each node, features are sampled without replacement when selecting the best split points. This adds further diversity to the ensemble and the net effect is greater variance in the ensemble of trees.

In SGB training, not all training instances are provided to a tree, but only a subset of these instances is sampled without replacement. The sampled data are then used to train a tree. However, not all features are considered in the partitioning process. Instead, only a specific, randomly selected subset of these features is used for this purpose. The stochastic gradient tree boosting algorithm was well described by Gutmann and Kersting (2007).

- Model tuning: Hyperparameter tuning

In this study, an Exhaustive Grid Search of Scikit Learn was also performed to create a grid with all possible hyperparameter combinations and to train the model with each of these combinations. These hyperparameters are parameters that are not learned from the data but are specified prior to fitting the model to the training set. During a grid search, the hyperparameters are adjusted to achieve better model performance. The details of the procedure are described in the Jupyter notebook presented in Appendix E2.

- Random Forest

The Random Forest algorithm, an ensemble bagging method, was also used in this study. In this method, decision trees are created for different samples and their majority vote is used as the average. The researcher considered its use in this study for the following reasons:

- It can handle datasets that contain continuous variables.
- It is able to solve overfitting problems just by the fact that the output is based on averaging.
- It has the property of parallelisation, since each decision tree created is independent of the others and is very stable, since the average responses of a large number of trees are used.

However, Random Forest is very complex compared to decision trees where decisions can be made by following the path of the tree, and the training time is higher compared to other models due to its complexity.

- ANNs

A detailed description of ANNs was presented in Chapter 4. In short, an ANN is a machine learning algorithm that consists of a network of neurons arranged in layers. During the training of an ANN, the input layer is fed with training data from the outside world. The network then begins processing the data from the input layer and passes the information it receives to hidden layers that convert the input data into something the output layer can use to predict a value. Each connection from one neuron to another in the hidden layer has an associated weight, w. Each neuron, with the exception of the input layer, which contains only the input value, also has an additional weight, called the bias weight, b. In feedforward, the input is transformed by multiplying and adding the weights in each layer, and the output of each neuron can also be modified by applying an activation function. Basically, learning in neural networks consists of adjusting the weights or parameters to achieve the desired output. One way to accomplish this is to use the famous gradient descent algorithm and gradually update the weights through a process known as backpropagation.

(c) Step 3: Model fitting

Once the models were created, predictions were made for each model, and the results are presented in Section 7.4.2.2.

(d) Step 4: Model evaluation

The RSME and the MAPE metrics were used for model evaluation.

7.4.2.2 Results and discussion

(a) Step 1: EDA

- Organisational and structural data analysis

Missing values were detected only for numeric features and were handled in the feature engineering stage. The date feature was identified as the only categorical feature with high cardinality, which was also handled in the feature engineering stage.

- Feature engineering

Missing values and the date feature were treated accordingly.

- Feature selection

In this study, the dataset contained only a few relevant features that did not exhibit multicollinearity; feature selection was therefore not necessary.

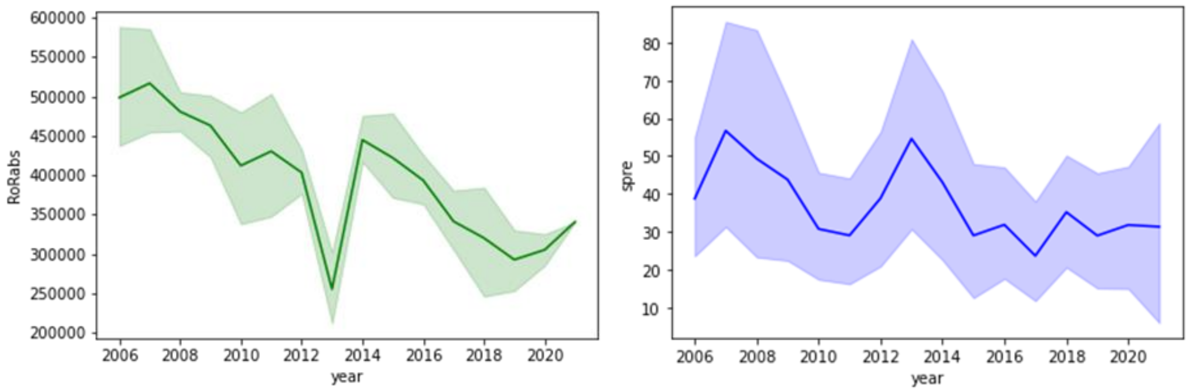
- Feature scaling

Not all algorithms require this procedure, which was the case in this study. Standardisation of the data was thus not necessary.

- Feature relationships analysis

Bi-variant analysis results for yearly RoRabs, annual monthly temperatures, and RoRabs versus temperature are presented in following sections.

i. Yearly RoRabs

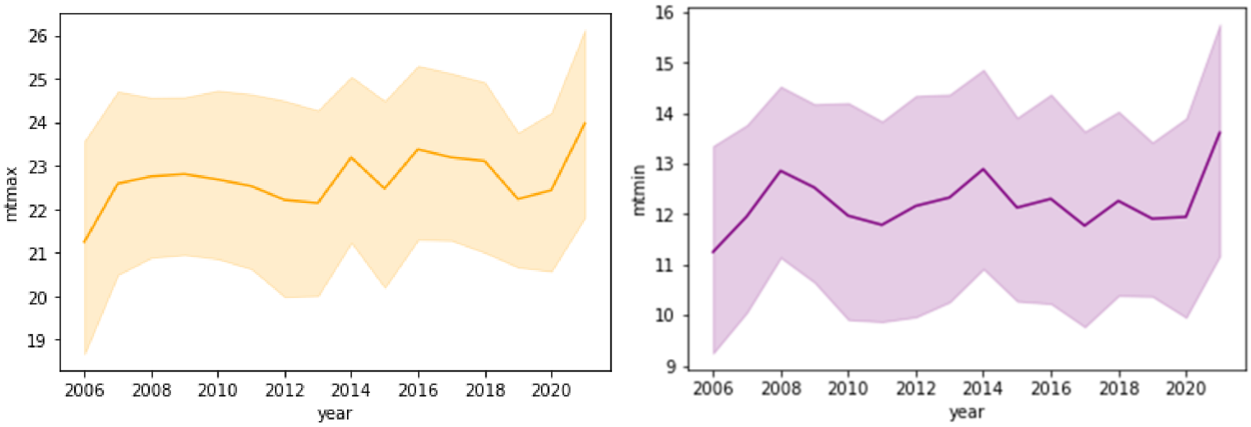


(a) **(b)**

Figure 7.12: (a) Line plot of RoRabs versus year; (b) Line plot sum precipitation (spre) versus year

Figure 7.12(a) shows a decline in RoRabs over the years with a low point in 2013, while Figure 7.12(b) shows that precipitation totals have declined over the years, as have water withdrawals from rivers. Of concern when comparing the figures is the fact that water withdrawals from rivers were the lowest in 2013, but precipitation totals appear to have peaked during the same period. This is contrary to expectations, as one would expect higher withdrawals from rivers in years with high precipitation totals. The question remains as to what happened to the water from the rivers in 2013.

ii. Annual monthly temperatures



(a) (b)

Figure 7.13 (a) Line plot of monthly maximum temperature (mtmax) versus year; (b) Line plot of monthly minimum temperate (mtmin) versus years

Figure 7.13(a) shows a general trend of increasing mtmaxs over the years, perhaps indicating the concern about climate change that currently dominates the discourse. Figure 7.13(b) also shows a general trend of increasing mtmins over the years.

iii. RoRabs

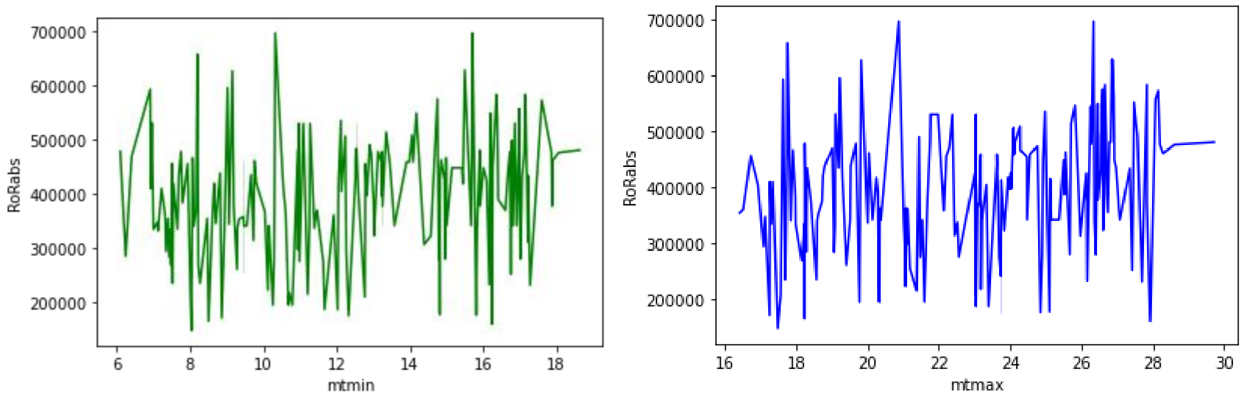


Figure 7.14: (a) Line plot of RoRabs versus mtmin; (b) Line plot of RoRabs versus mtmax

Figure 7.14(a) shows that as minimum temperatures increase, river withdrawals remain at an average rate of 400 000 kl/m. Figure 7.14(b) shows that despite the increase in maximum temperatures, run-of-river withdrawals remain at an average rate of 400 000 kl/m.

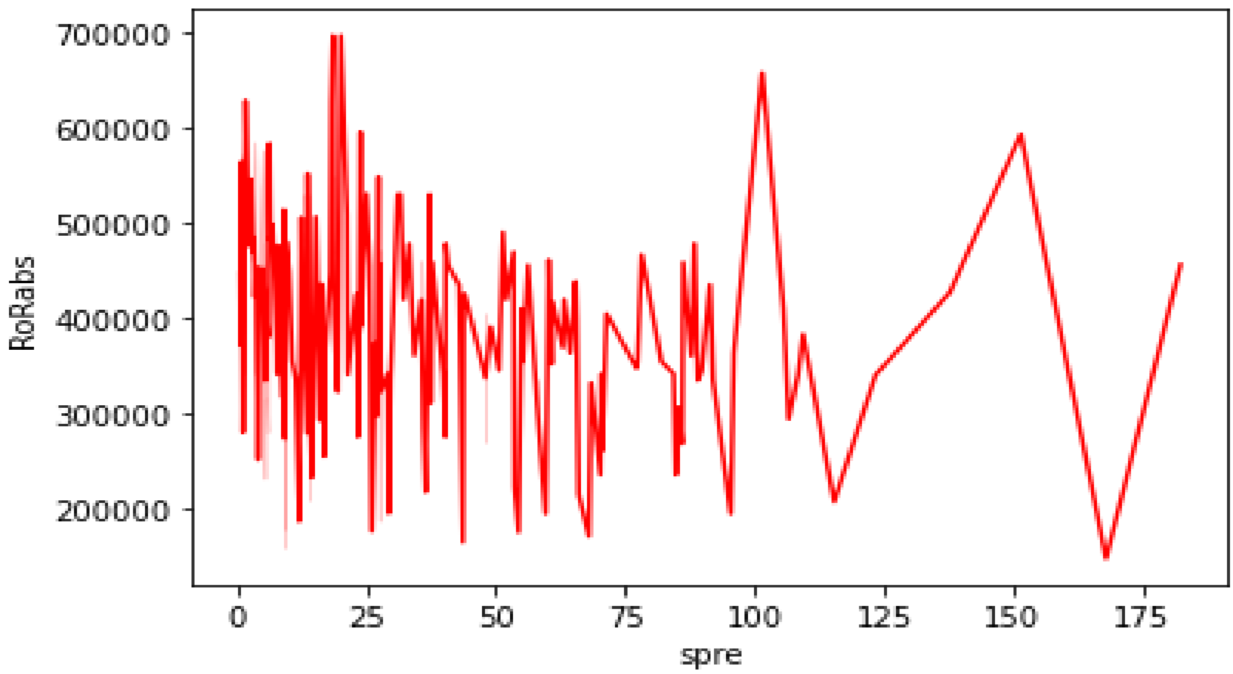


Figure 7.15: Line plot of RoRabs versus spre

Figure 7.15 shows that when spre is high, water withdrawal from rivers is highly irregular. On the other hand, when the rainfall totals are low, the withdrawal from the rivers remains at an average of 400 000 kl/m.

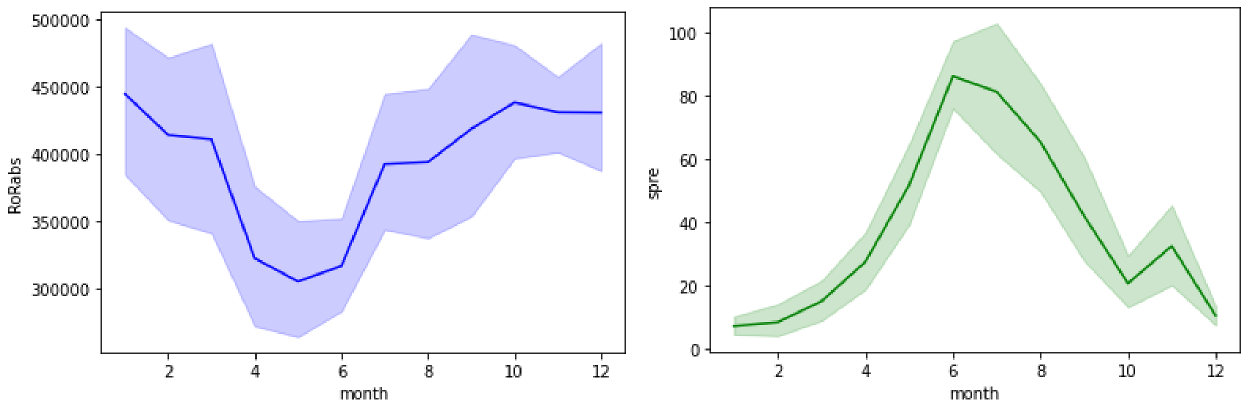
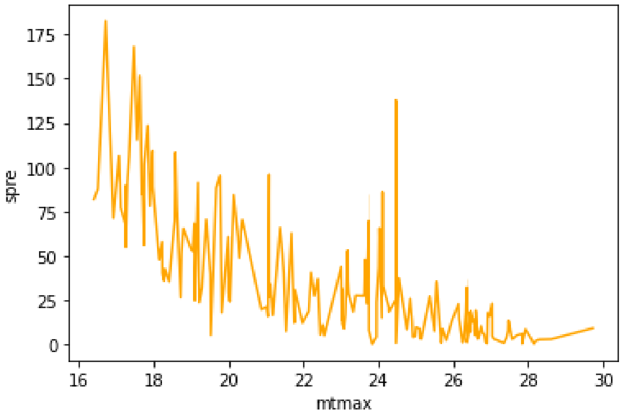
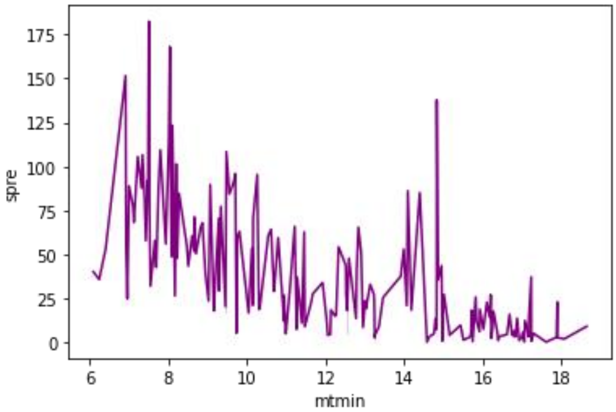


Figure 7.16: (a) Line plot of RoRabs versus month; (b) Line plot of spre versus month

Figure 7.16(a) shows a plot of RoRabs that is quite unexpected, and even more disturbing when interpreted in conjunction with Figure 7.16(b), the plot of precipitation total versus month. It can be seen that run-of-river withdrawals are the lowest in months when precipitation is the highest. This is contrary to expectations. One would expect run-of-river withdrawals to be the highest in the wet season and the lowest in the dry season. Even if the withdrawal methods were inefficient, more water would certainly be withdrawn in the wet season than in the dry season, unless the water is diverted and thus not accounted for in the high rainfall periods.





(a) **(b)**

Figure 7.17: (a) Line plot spre versus mtmax; (b) Line plot spre versus mtmin

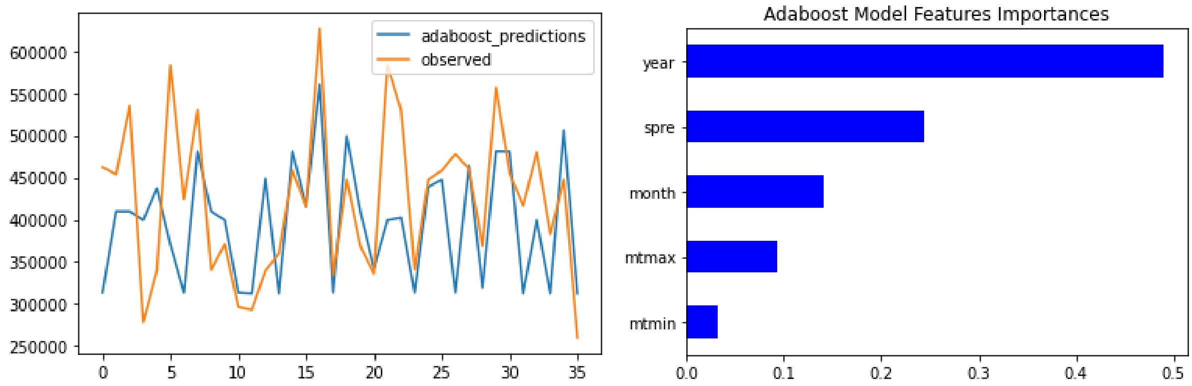
The two plots in Figure 7.17(a) and (b) show a continuous decrease in precipitation totals as temperatures become warmer. This is consistent with the climate change phenomenon of precipitation totals decreasing and becoming erratic as temperatures increase. This observation should be of great concern to the Stellenbosch Municipality water authorities, and they should consider alternative water sources to improve their water supply in the near future.

(b) Step 2: Model building

The following models were successfully developed:

- AdaBoost;
- GBM;
- SGB;
- Random Forest; and
- ANNs.

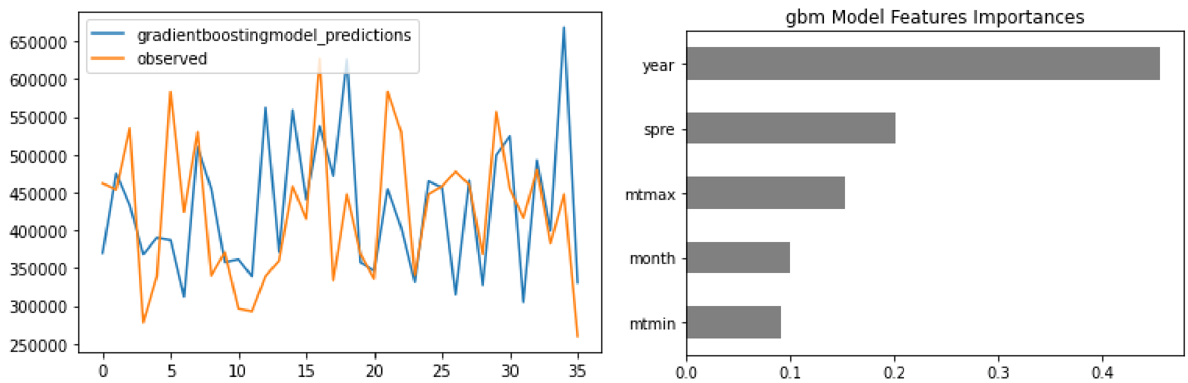
(c) Step 3: Model fitting



(a) **(b)**

Figure 7.18: (a) AdaBoost model; (b) AdaBoost model features importance

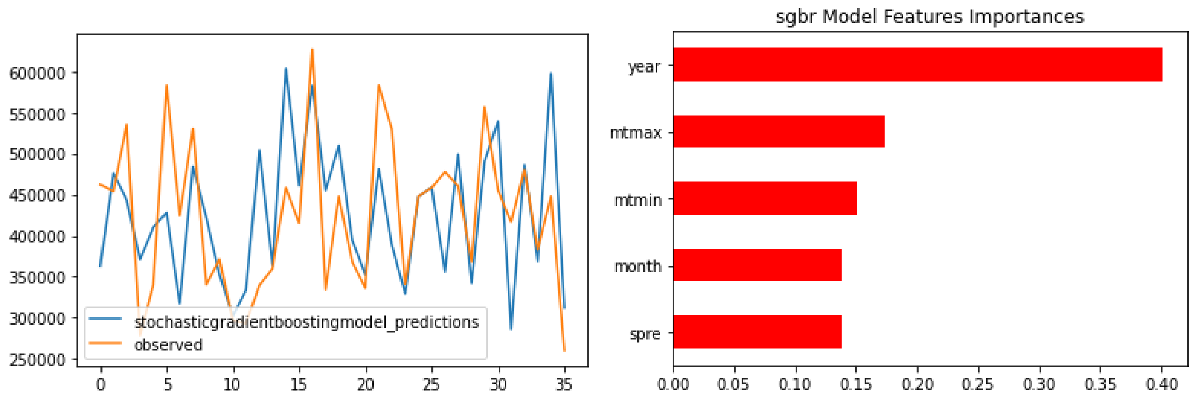
The AdaBoost model in Figure 7.18(a) shows that the model performed quite well. Figure 7.18(b) shows that the AdaBoost model ranks year as the most important feature, followed by rainfall total in modelling RoRabs. The monthly feature ranks third and maximum temperature still has some importance, while the influence of minimum temperature is minimal.



(a) **(b)**

Figure 7.19: (a) GBM; (b) GBM features importance

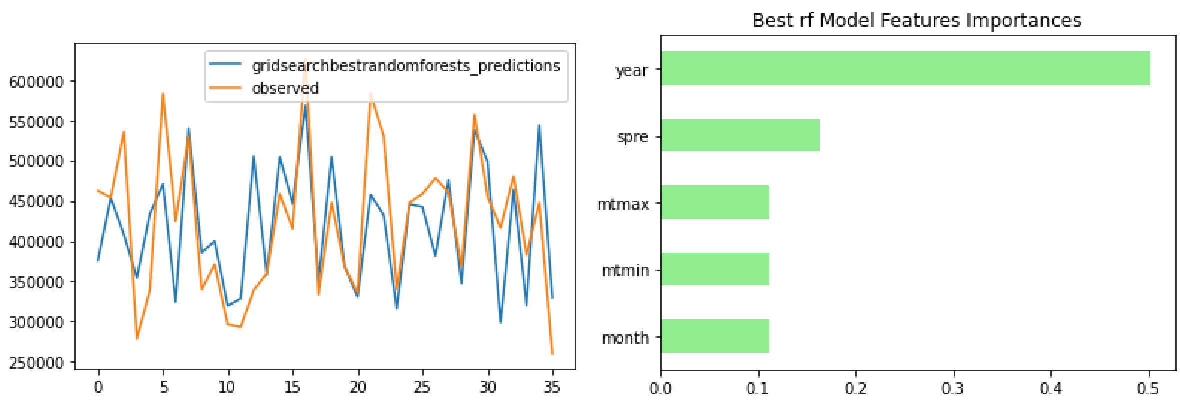
Figure 7.19(a) shows that the GBM also worked quite well. Figure 7.19(b) shows that, similar to the AdaBoost model, the GBM ranked the annual feature as the most important in modelling RoRabs, followed by spre. This is followed by maximum temperature, while monthly and minimum temperatures are almost equally important.



(a) **(b)**

Figure 7.20: (a) SGB model; (b) Stochastic model features importance

Figure 7.20(a) shows that the SGB model performed better than the AdaBoost model and GBM. Similar to the other models, year was the most important feature for SGB. Maximum temperature was second. Minimum temperature, month, and precipitation had more or less the same weighting. Minimum temperature was given considerable importance compared to the weighting by the AdaBoost model and GBM. Figure 7.20(b) shows the ranking of importance of the stochastic model features.



(a) **(b)**

Figure 7.21: (a) Random Forest model; (b) Random Forest feature importance

Figure 7.21(a) shows that the Random Forest model is the best model so far. Year was ranked as the most important feature, followed by the spre in second place. Month, minimum temperature, and maximum temperature were ranked in that order as shown in Figure 7.21(b).

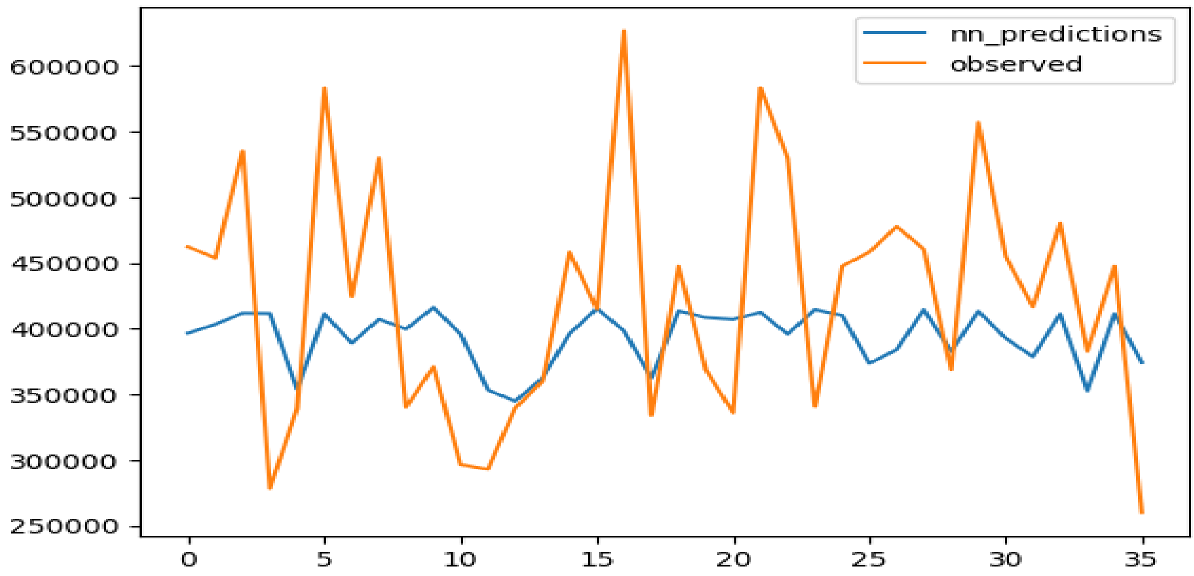


Figure 7.22: ANN model

Figure 7.22 shows that the ANN model performed poorly. This is due to the dataset at hand. In cases where a large dataset is available, ANNs generally perform better than most algorithms currently used to predict urban water demand.

(d) Step 4: Model performance evaluation

Based on the metrics RMSE and MAPE, the best developed model was Random Forest, which was achieved by tuning the hyperparameters of the Random Forest model with cross-validation. This was closely followed by the SGB model, and in third place was AdaBoost. The researcher saved the top two models and recommends them for production. As new data become available, these models can be reloaded to evaluate new data and obtain RoRabs predictions.

Table 7.2: Model evaluation metrics

Model name	Metric	Training set	Test set	Comment
AdaBoost model	RMSE	21.2%	20.4%	Good
	MAPE	17.8%	16.1%	
GBM	RMSE	8.3%	23.5%	Possible signs of overfitting
	MAPE	6.5%	18.1%	
SGB model	RMSE	10.7%	19.7%	Good
	MAPE	8.6%	15.8%	
Decision Tree Regressor Gridsearch	RMSE	23.6%	22.7%	Good
	MAPE	18.7%	18.9%	
Random Forest Regressor Gridsearch	RMSE	10.6%	16.3%	Perfect
	MAPE	8.0%	12.7%	
Neural Networks	RMSE	30.6%	24.2%	
	MAPE	24.7%	19.5%	Limited data

7.5 SUMMARY

The main objective of the study was achieved in this chapter, namely to develop powerful urban water demand models through supervised machine learning algorithms for the Stellenbosch municipal water authorities, to enable them to accurately forecast and predict short- and medium-term water demand and supply for their jurisdiction. Issues identified in the Stellenbosch urban water system include water supply capacity, dilapidated infrastructure, and leakage in the supply system.

An initial EDA was performed using pivot tables, and river discharge was set as the target variable, since the external quantities of raw and treated water that would need to be purchased to adequately supply the intake system would be determined by the amount of RoRabs. This in turn would affect budget, water tariffs, and policy. In analysing the relationships between the independent variables, the study found that there is no correlation between total water use or total input to the system with increases in population or households, which should be the norm. This suggests that Stellenbosch Municipality's water management is reactive rather than proactive. Although the policy is reactive, it appears to be working, as the study showed that Stellenbosch Municipality provided sufficient clean water to its residents during the study period. In some ways, this can be seen as poor water management practice, as the risk is very high when one is more dependent on an external supply than an internal one that can be controlled. To mitigate this risk, the focus of the modelling was to predict and forecast water withdrawals from the rivers over which Stellenbosch Municipality has some influence. Once the Stellenbosch Municipality water authority can accurately predict and forecast water withdrawals from the RoRabs, it will be able to accurately budget and develop infrastructure, optimise its water supply system, and develop a strategy to proactively expand its water supply.

Once the target variable was established, modelling began with traditional ARIMA and SARIMA time series modelling of the target variable and models were successfully developed and their results were presented. Supervised machine learning models were then developed, where the dependent variable was RoRabs and the independent variables were mtmin, mtmax, spre, and month extracted from date.

EDA was also performed, and relationships were established between dependent and independent variables and between independent variables and independent variables.

One of the most striking results was the low withdrawal from rivers during the peak of the rainy season in 2013. There is a general trend towards less precipitation with warmer temperatures. This is indicative of the impact of climate change on precipitation and requires adequate preparation to ensure that the needs of water users are met as climate change becomes more evident. These observations make the machine learning methodology superior to traditional ARIMA and SARIMA modelling. This is because researchers or agencies gain deeper insight into what is happening in their water system. As a result, better-informed actions can be taken to improve water system management.

Due to the limited amount of data, ensemble machine learning techniques dominated the machine learning component of the study. The ensemble models developed were AdaBoost, GBM, SGB, and Random Forest. In addition to the ensemble models, ANNs were also developed. The results were presented and discussed. The best model was Random Forest, followed by SGB. ANNs performed poorly due to the limited data the researcher could obtain. Details of the modelling process can be found in the attached Jupyter notebook (see Appendix E2). The best models were saved and can be reloaded when new data become available, and predictions can be made about RoRabs. The researcher recommends the Random Forest and SGB models for production. However, once rich data become available, an ANN could also be a very good candidate.

CHAPTER 8:
CONCLUSIONS AND RECOMMENDATIONS

8.1 INTRODUCTION

This chapter presents a summary of the major contributions and findings of this study. The limitations of the study are discussed and suggestions to address these limitations are highlighted. Recommendations for improving the developed supervised machine learning models are also provided, with the focus on the importance of data collection in the water sector. Although the fourth objective took the centre stage of the study, namely to evaluate how a supervised machine learning model developed for predicting and forecasting water demand and supply in Stellenbosch Municipality can improve the municipality's water management system, it did not preclude the important role of transdisciplinary research methodology in water management in holistically addressing water management issues in Stellenbosch Municipality and in South Africa as a whole.

8.2 CONTRIBUTIONS

This study contributes to the importance of a human-centred design approach and the use of data-driven, supervised machine learning techniques in the management of urban water systems, which the researcher considers a human-centred, data-driven, technological triad (HC-T-DD) in the management of urban water systems. This is illustrated in Figure 8.1.

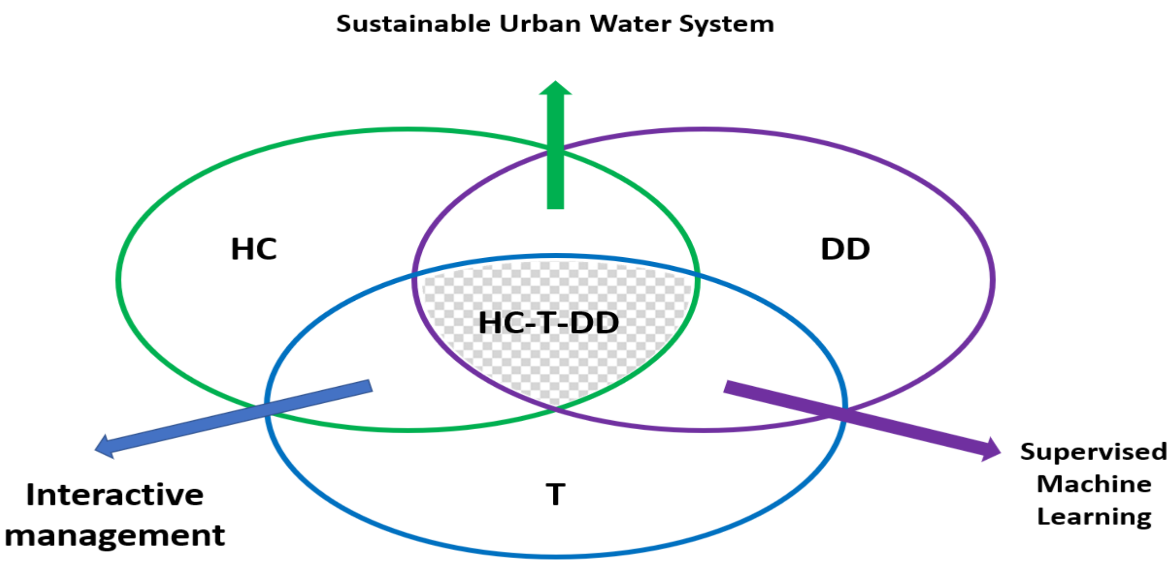


Figure 8.1: The human-centred, data-driven, technological triad (HC-T-DD) framework

The effective management of an urban water system requires the application of the four principles of the human-centred design approach and data-driven decision-making processes supported by technological tools. The interactive management methodology and supervised machine learning techniques are proposed approaches that can achieve an HC-T-DD approach to urban water system management. The HC-T-DD framework lies at the intersection of a human-centred design, data-driven approach, and technology; that is, a human-centred design approach provides the guidance necessary to understand the needs of the urban water system in the context of a particular community, which leads to appropriate data collection that provides a deeper understanding of the issues to be addressed. Through the use of technology, robust solutions can be provided that lead to effective improvements in the management of urban water systems.

In addition, this study provided a guiding process or procedure for the HC-T-DD framework, using Stellenbosch Municipality as a case study. The resulting supervised machine learning models that demonstrate the need for the reuse of treated municipal wastewater in the context of Stellenbosch Municipality represent a significant contribution, as the application of supervised machine learning to the management of urban water systems in South Africa is still in its infancy. The use of treated municipal wastewater as an alternative water source in South African municipalities is also still in its infancy. The study can be replicated in all parts of South Africa, as South Africa is a water-scarce country and its location in the sub-Saharan region lends it to being severely impacted by climate change in the coming years, which will result in altered rainfall cycles and reduced water volumes.

8.2.1 Discussion of research findings

The literature review highlighted the need for a paradigm shift in the management of urban water systems in Stellenbosch Municipality; i.e., from government water management to a governance approach. In South Africa, principles such as IUWM are still in their infancy. The IUWM principle was reviewed in the context of alternative water sources. To this end, treated municipal wastewater was explored and advocated as a possible alternative water source in Stellenbosch Municipality. Since agriculture is the largest consumer of freshwater, literature on the reuse of treated municipal wastewater in agriculture was reviewed. The literature shows that the reuse of treated municipal wastewater in irrigated agriculture is increasing in the Global North, but it is still an untapped resource in Africa. In South Africa, issues of policy, laws, and guidelines are prevalent. Countries that have successfully used treated municipal wastewater in irrigated agriculture have well-formulated policies that explicitly regulate urban wastewater reuse. The role of a supranational body and, in the case of the USA, the federal government, played an important role in the success of reusing treated municipal wastewater in irrigated agriculture in the Global North. In South Africa, however, there are still policy issues, conflicting laws, and a lack of guidelines that explicitly describe the processes and procedures to be followed when reusing treated municipal wastewater in irrigated agriculture. The study found that Stellenbosch Municipality, as the water authority, has not yet adopted the reuse of treated municipal wastewater as an alternative water source.

The literature review also revealed that the use of supervised machine learning in urban water management in South Africa has not been fully explored. To date, there are only sporadic research reports on a few municipalities, but none of the South African municipalities use machine learning models to manage their urban water systems. Research to date on the use of machine learning techniques in the management of urban water systems remains an academic exercise, which creates a divide between computer engineers, water engineers, and managers. In the Global North, however, the application of machine learning in predicting urban water demand is increasing and the benefits are being recognised.

During the drought and famous Day Zero-year of 2018, Stellenbosch Municipality presented plans to explore alternative water sources and introduced the reuse of treated municipal wastewater, but to date, the needle has not moved in this regard. The researcher therefore investigated the barriers to reusing treated municipal wastewater as an alternative water source. Since reusing treated municipal wastewater is a water governance issue, significant gains will be realised through a stakeholder-centred approach. However, the inclusion of all stakeholders in a basin in water management decision-making processes is one of the unresolved challenges of water governance, especially in South Africa. Accordingly, the researcher used the interactive management approach to investigate the barriers to reusing treated municipal wastewater in Stellenbosch Municipality, with the goal of involving everyone and realising meaningful engagement with a diverse group that represented the various communities in Stellenbosch Municipality.

The reasons for using interactive management in this study included the need to meaningfully engage all stakeholders in the decision-making process, as the Stellenbosch community is very diverse, and it is difficult to meaningfully engage such a diverse group. The researcher believes that the interactive management research methodology can overcome the above challenges. As far as the researcher is aware, the interactive management methodology has been used minimally, if at all, as a method to improve stakeholder engagement in the management of urban water systems in South Africa. Furthermore, in a diverse community such as Stellenbosch, the participation of the poor is seen as a formality for meeting the IDP. The views of the poor are not taken seriously or included in decision-making processes. In contrast, the interactive management method allows the input of all stakeholders to be captured and incorporated into the development of an interpretive structural model. The versatility of the interactive management method stems from its ability to draw out different perceptions of an issue in a meaningful way and to give a voice to the excluded within the group.

During the interactive management workshop, the participants identified 41 factors that impede the reuse of treated municipal wastewater. Some of the factors could be traced back to the literature. Although the researcher identified the three main topics for discussion as water laws, policies, and administration, people-related issues became the focus of the discussion. The result of the interpretive structural modelling showed that the main drivers for implementing treated municipal wastewater reuse in Stellenbosch Municipality revolve around societal issues, and that Stellenbosch Municipality must take a people-centred approach to implementing a novel water management concept such as treated municipal wastewater reuse in order to be successful. Once the societal issues are addressed, the political and legal issues would be more manageable. On the far right of the interpretive structural model were management issues that would be easier to address once the societal and legal issues are resolved. During the interactive management process, it was noted that an important sector that has a significant impact on the reuse of treated municipal wastewater in Stellenbosch is the agriculture and business sector. This means that

Stellenbosch Municipality needs to focus on these sectors as prime candidates for the reuse of treated municipal wastewater.

The interpretive structural model demonstrated the need for Stellenbosch Municipality to adopt a people-centred planning approach when implementing new methods of managing its water system. To achieve the stakeholder engagement and public participation called for in the IDP, Stellenbosch Municipality can use technology to ensure that input from a diverse group representing its diverse community is captured and modelled.

During the literature review, the researcher found that Stellenbosch Municipality has developed a strategic plan to address the challenges of the urban water system. The use of the interactive management methodology is proving useful in crafting a strategy to successfully implement its goals. The tools from interactive management can be used for stakeholder engagement, as well as at the management level. As far as the researcher is aware, the interactive management methodology is not being used by South African municipalities to support their planning and implementation processes in addressing the challenges in their urban water systems. This study thus succeeded in introducing the interactive management methodology in the planning and implementation of a novel approach to urban water system management that uses technology to improve operations, through meaningful collaboration with all stakeholders and a well-informed strategy as described by the interpretive structural model.

The study examined the technology- and data-driven (T-DD) hypothesis, a component of the HC-T-DD framework, which states that supervised machine learning algorithms can produce powerful models to predict water demand and supply for Stellenbosch Municipality compared to conventionally developed models. This will improve the management of the municipality's urban water system. This is because data-driven machine learning techniques are considered robust in creating powerful models to predict water demand due to their ability to handle large amounts of data and multiple variables and to quantify uncertainty. Despite the lack of large datasets in urban water systems that reflect daily and weekly water use, Stellenbosch Municipality was selected as a case study to demonstrate the advantages of using supervised machine learning techniques over conventional techniques in modelling the prediction and forecasting of water demand in an urban water system. Both conventional and supervised machine learning algorithms were used in the development of models and their performance was compared.

The researcher had planned to forecast all time horizons of water demand for Stellenbosch Municipality. Due to the lack of data, the study was limited to short- and medium-term forecasts. Short-term forecasts of days to weeks could not be made because the available water balance data were reported on a monthly basis. However, based on the available data, the researcher determined that the target variable in modelling the forecast model for water demand in Stellenbosch Municipality was RoRabs. This was justified by the EDA process. To the researcher's knowledge, there is no RoRabs model developed for Stellenbosch Municipality using supervised machine learning algorithms to assist water managers in planning and optimising the operation of their water system.

Using supervised machine learning, the study found the following:

- Supervised machine learning models outperforms conventional models.
- During the supervised machine learning modelling process (the EDA phase, where the relationships between independent and dependent features were analysed), more insights could be drawn from the dataset. For example, the discrepancy uncovered between peak precipitation and the volume of RoRabs, and how easy it would be to infer the decrease in precipitation as temperatures increase.
- It could easily demonstrate the disadvantages of the conventional ARIMA and SARIMA models to make predictions compared to supervised machine learning models, since predictions with the conventional models are only possible not too far into the future compared to the supervised machine learning models.
- The developed machine learning models performed better compared to the conventional models.
- The developed models were saved and can be re-run and improved when new datasets become available.
- The research showed that more detailed data collection is needed in Stellenbosch Municipality to build better models that have great implications for the management of the urban water system.

Using the above research findings, the researcher was able to successfully demonstrate how the use of technology, i.e., Concept Star's decision-making tools for professionals and supervised machine learning algorithms, can improve the management of Stellenbosch Municipality's water system. To this end, the researcher was able to convincingly adopt the null hypothesis (H_0): Supervised machine learning models can accurately predict and forecast urban water demand compared to conventional models.

The fourth research objective of the study, namely to develop, train and deploy a highly accurate water demand and supply prediction and forecasting model for Stellenbosch Municipality to assist water policy and decision makers in the sustainable management of its urban water system, was achieved; considering that powerful supervised machine learning models were developed, whose performance was superior to conventional models.

As presented in the research findings, a literature review on the application of machine learning techniques in urban water system management was successfully conducted. In addition, a literature review was conducted on the barriers in South African water legislation, policy, and administration to the reuse of municipal wastewater as a supplementary water source and for reuse in irrigated agriculture. Finally, the models developed proved to perform reasonably well. To this end, the researcher stored the models so that they could be retrieved when needed.

8.3 THEORETICAL AND PRACTICAL IMPLICATIONS OF THE RESEARCH

This study highlighted some important implications for the theory development and pragmatic application of interactive management as a human-centred design approach and of data-driven supervised machine learning techniques as an approach to implement the HC-T-DD framework. The developed HC-T-DD framework provides a general framework for understanding an appropriate methodology for engaging stakeholders in an urban setting on water issues and techniques for predicting the water demand of an urban water system. The framework can be extended to and tested on other areas of urban supply management, such as energy or waste. The transdisciplinary research methodology enabled the inclusion of interactive management in this research. This is an important contribution to water management research, as several researchers have pointed out the inadequacy of monodisciplinary research methods in finding holistic solutions to water management challenges. As far as the researcher is aware, transdisciplinary research in water resources management has not previously been applied in South Africa. It is therefore to the credit of the researcher that the transdisciplinary research methodology has been introduced in water resources management in South Africa.

An examination of the inherent characteristics of the main elements of the HC-T-DD framework, i.e., human-centred design, technology application, and data-driven methods, shows that the interactive management and supervised machine learning approach has the potential to improve urban water system management. This is because technology can replicate the characteristics of human-centred design and data-driven elements. Using interactive management in combination with supervised machine learning allowed accurate assessment of challenges to be addressed and solutions to be provided. While this framework may seem complex, it has become apparent that it is no longer possible to solve water problems using the government-centred water management approach. For researchers to provide appropriate water management solutions, water professionals need to understand the transdisciplinary nature of an urban water system. They need knowledge that transcends disciplines to holistically address the problems at hand. In addition, the participation of non-disciplines is critical, as local communities are indirectly or directly affected by the decisions that are made regarding the management of their urban water system. Overall, the transdisciplinary research methodology can be seen as one that is capable of facilitating key elements of the stakeholder-centred approach to water management. These include Principle 7, which requires that sound water governance frameworks be effectively implemented and enforced in the public interest, and Principle 10, which encourages stakeholder engagement to provide informed and results-oriented input into water policy and implementation.

The government-centred water management approach continues to be practised in South Africa. Research has identified the lack of a governance water management approach in an urban setting and confirmed that such a practice is needed. It is worth noting that meaningful stakeholder engagement and the use of technology in managing urban water systems are recognised globally; however, in South Africa, these aspects of water management have not been fully explored. As a result, novel water management approaches such as the reuse of treated municipal wastewater are immature and in their infancy. This was confirmed by the case study of Stellenbosch Municipality, where the reuse of treated municipal wastewater was brought to the table in 2018 but is not yet practised.

In summary, interactive management has the strength of integrating different stakeholder perceptions of the challenges that need to be addressed in an urban water system and developing a strategy to address these challenges. While supervised machine learning techniques are versatile in building powerful models that water agencies can use to gain insight into the urban water system and make data-driven decisions to address the challenges facing their water system, they are also powerful in providing water agencies with the means to understand the challenges facing their water system. When water agencies recognise the need to adopt the HC-T-DD framework, it can facilitate meaningful stakeholder engagement in their jurisdictions and lead to lasting decisions that drive change in the urban water system. In addition, data-driven solutions are central to the management of an urban water system because the optimisation of operations can be achieved in a sustainable manner. All water agencies therefore need to adopt the HC-T-DD framework for the better management of their urban water systems.

8.4 LIMITATIONS OF THE RESEARCH

The major limitation of this study was the availability of data. Some of the state agencies from which data were collected did not have the high-quality data needed to develop supervised machine learning models. There were discrepancies in data collected from different organisations for the same feature. Since the study involved a very diverse group of participants, getting them to agree on an appropriate day for the workshop was a major challenge. During the workshop, the power imbalance between the participants was a problem due to their different education levels. However, the purpose of convening a diverse group of participants is that sustainable solutions to a community problem are likely to emerge from such a group. It is therefore imperative that diversity is a requirement for water management focus groups.

8.5 RECOMMENDATIONS FOR FUTURE WORK

Opportunities for future research emerged from this study. First, the researcher explored the possibilities of applying the HC-T-DD framework in managing urban water systems, which began with the human-centred portion involving the interactive management methodology. Further research is needed to apply the interactive management methodology in a way that is most appropriate for managing urban water systems – given the diversity of groups involved and the perceptions of different communities under the same jurisdiction. The focus is on meaningful stakeholder engagement, capturing all perceptions, and integrating them into the interpretive structural modelling so that the resulting interpretive structural model reflects the reality on the ground and resonates with diverse communities across jurisdictions. Further research is needed on how to recruit participants and organise multiple focus groups with different participants to explore similarities and differences in the factors that emerge from the different focus groups in the Stellenbosch community. The results that emerge from the different interpretive structural models on the same topic must then be compared and contrasted. This will be quite interesting as the researcher only conducted one focus group discussion with only 11 participants. It can be concluded that the interpretive structural model is not a true reflection of the diverse Stellenbosch community, as only 11 participants gave their input on topics that affect an estimated population of 200 000 people. Multiple focus groups, consistent with the methodology of interactive management, would therefore provide a better picture of Stellenbosch as a whole.

In addition, the interactive management methodology can be applied to various water issues that Stellenbosch Municipality may need to address. These include infrastructure development and other novel approaches to water management to be introduced in an urban environment or watershed. This could be an interesting study to determine if the methodology developed by interpretive structural modelling will lead to what needs to be addressed and if the desired results can be achieved. Once the methodology has been tried and tested in Stellenbosch Municipality, it can be duplicated in other municipalities or catchments to discover and improve approaches to water management.

Secondly, the researched case study provided only monthly water use data, which limited the researcher to predicting water demand for a monthly and annual period. However, forecasting daily and weekly water demand is extremely important to improve the daily operation and management of an urban water system. The benefits of accurately forecasting daily and weekly water use include:

- establishing the pumping schedule for the next 24 hours for fresh water;
- ensuring adequate fresh water in the system while reducing the amount of unused fresh water pumped into the system;
- accurately estimating the water usage profile for the next day; and
- accurately calculating water treatment costs and energy consumption.

The above could not be investigated by the researcher. Ideally, to optimise the operation and management of an urban water system, water suppliers need to determine their pumping schedules each morning for the next 24 hours of potable water production. Optimised operations should ensure that potable water is always available to reduce the amount of unused potable water pumped into the system. To achieve this, operators must accurately estimate the water use profile for the next day. This is where powerful supervised machine learning models are needed to enable water managers to accurately plan and meet consumer demand while managing the water system sustainably. Stellenbosch Municipality also struggles with high water losses due to leakage. To mitigate this problem, daily and weekly forecasts of water consumption would help water managers to detect water leaks by identifying fluctuations in water demand early. In addition, accurate forecasting of water demand is necessary for long-term planning to avoid building oversized or undersized infrastructure that leads to water shortages. Further research is therefore needed to predict daily and weekly water withdrawals from the river, as well as long-term forecasting, i.e., for 10 years or longer.

The researcher therefore recommends the following to Stellenbosch Municipality:

- The establishment of an interdisciplinary research unit on wastewater reuse and the application of a transdisciplinary research methodology.
- Prioritise the reuse of treated municipal wastewater, as Stellenbosch will experience a significant decrease in rainfall by 2040, which will require freshwater to be used multiple times before it can be discharged into natural water bodies.
- Establish a water demand data science department with the following strategic team members:
    - An information technology manager with a data science background to lead and support the team in the data centre.

- A data scientist with cloud computing skills to lead research and model development activities.
- A water demand manager who will focus on identifying the requirements of the projects undertaken and act as an expert on the scope and evaluate the performance/interpretability of the artefacts developed by the data science members.
- Data engineers responsible for managing data acquisition and processing.

- The application of the HC-T-DD framework in the management of water resources for Stellenbosch Municipality.

8.6 CONCLUSION

From the study, it can be concluded that the transdisciplinary research methodology has the potential to enable researchers to find holistic solutions for managing urban water systems. In addition, the interactive management methodology has proven to be an effective approach to improve stakeholder engagement in solving water problems in a watershed by addressing key issues through modelling the contributions of all stakeholders. This approach solves a major problem that exists in implementing water governance management principles, which emphasise a bottom-up approach to water management. In addition, this study showed that the use of data-driven technological methods can improve the management of an urban water system. This is because water managers and agencies can gain deeper insights into water system activities by analysing data in conjunction with developed models that help them to predict water demand in their jurisdictions. In this way, water agencies will be enabled to develop appropriate and effective strategies and plans to maintain their water system and meet the water needs of their community.

REFERENCES

Abdel-Kader, A.M. & Abdel-Rassoul, S.M. 2010. Prospects of water conservation in Egypt (special reference to wastewater reuse). In *Proceedings of the Fourteenth International Water Technology Conference*, Cairo, Egypt, 25-27 March.

Abdel-Lateef, E.M., Hall, J.E., Farrag, M.A.A. & Farrag, A.A. 2011. Agroeconomic studies on wastewater reuse in developing marginal areas in West Delta, Egypt. *International Journal of Water Resources and Arid Environments*, 1(2):110-115.

Abdel-Shafy, H.I. & Mansour, M.S. 2013. Overview on water reuse in Egypt: Present and future. *Journal of Sustainable Sanitation Practice*, 14:17-25.

Abdel-Shafy, H.I. & Mansour, M.S. 2020. Rehabilitation and upgrading wastewater treatment plant for safe irrigation reuse in remote area. *Water Practice and Technology*, 15(4):1213-1227.

Abukila, A.F. 2015. Assessing the drain estuaries' water quality in response to pollution abatement. *Water Science*, 29(1):1-18.

Adamowski, J. 2008. Peak demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management*, 134(2):119-128.

Adamowski, J. & Karapataki, C. 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering*, 15(10):729-743.

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B. & Sliusarieva, A. 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, 48(1):1-14.

Adejo, O.W. & Connolly, T. 2018. Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, 10(1):61-75.

Adewumi, J.R., Ilemobade, A.A. & Van Zyl, J.E. 2010. Treated wastewater reuse in South Africa: Overview, potential and challenges. *Resources, Conservation and Recycling*, 55(2):221-231.

Adom, R.K. & Simatele, M.D. 2021. Analysis of public policies and programmes towards water security in post-apartheid South Africa. *Water Policy*, 23(3):503-520.

Ahmadalipour, A., Moradkhani, H., Castelletti, A. & Magliocca, N. 2019. Future drought risk in Africa: Integrating vulnerability, climate change, and population growth. *Science of the Total Environment*, 662:672-686.

Akhtar, N., Syakir Ishak, M.I., Bhawani, S.A. & Umar, K. 2021. Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water*, 13(19):2660.

Ako, A.A., Eyong, G.E.T. & Nkeng, G.E. 2010. Water resources management and integrated water resources management (IWRM) in Cameroon. *Water Resources Management*, 24(5):871-888.

Alexopoulos, E.C. 2010. Introduction to multivariate regression analysis. *Hippokratia*, 14(S1):23-28.

Al-Ghamdi, A.B., Kamel, S. & Khayyat, M. 2021. Evaluation of artificial neural networks performance using various normalization methods for water demand forecasting. In Institute of Electrical and Electronics Engineers (IEEE). *2021 National Computing Colleges Conference (NCCC)*. New Jersey: IEEE. pp. 1-6.

Alghamdi, A.H. & Li, L. 2013. Adapting design-based research as a research methodology in educational settings. *International Journal of Education and Research*, 1(10):1-12.

Almandoz, J., Cabrera, E., Gil, J. & Pellejero, I. 2003. Evaluation of leakage by means of night flow measurements and analytical discrimination: A comparative study. In *Proceedings of the IWA – IAHR International Conference PEDS 2003 (Pumps, Electromechanical Devices and Systems, Applied to Urban Water Management)*. Valencia: Balkema Publishers. pp. 327-342.

Almendarez-Hernández, M.A., Avilés Polanco, G., Hernández Trejo, V., Ortega-Rubio, A. & Beltrán Morales, L.F. 2016. Residential water demand in a Mexican biosphere reserve: Evidence of the effects of perceived price. *Water*, *8*(10):428.

Al-Saadi, H. 2014. *Demystifying Ontology and Epistemology in Research Methods.* Available at: https://www.academia.edu/26531411/Demystifying_Ontology_and_Epistemology_in_research_methods.

Al-Saidi, M. 2021. From acceptance snapshots to the social acceptability process: Structuring knowledge on attitudes towards water reuse. *Frontiers in Environmental Science*, 9:633841.

Altunkaynak, A. & Nigussie, T.A. 2017. Monthly water consumption prediction using season algorithm and wavelet transform-based models. *Journal of Water Resources Planning and Management*, 143(6):04017011.

Andrew Leung International Consultants and Investments Limited. 2015. *China Water Pollution Prevention and Control Action Plan ("Water Ten Directives")*. Available at: https://www.andrewleunginternationalconsultants.com/new/ 2015/05/water-pollution-prevention-and-control-action-plan-water-ten-directives.html.

Angelakis, A.N. & Gikas, P. 2014. Water reuse: Overview of current practices and trends in the world with emphasis on EU states. *Water Utility Journal*, 8(67):e78.

Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, A. & Oliveira, M.S. 2018. Short-term water demand forecasting using machine learning techniques. *Journal of Hydroinformatics*, 20(6):1343-1366.

Arandia, E., Ba, A., Eck, B. & McKenna, S. 2015. Tailoring seasonal time series models to forecast short-term water demand. *Journal of Water Resources Planning and Management*, 142(3):04015067.

Arena, C., Genco, M. & Mazzola, M.R. 2020. Environmental benefits and economical sustainability of urban wastewater reuse for irrigation – A cost-benefit analysis of an existing reuse project in Puglia, Italy. *Water*, 12(10):2926.

Asano, T. & Pettygrove, G. 1987. Using reclaimed municipal wastewater for irrigation. *California Agriculture*, 41(3):15-18.

Awad, M. & Khanna, R. 2015. Support vector regression. In M. Awad & R. Khanna. *Efficient Learning Machines*. Berkeley: Apress. pp. 67-80.

Ayers, R. & Wescott, D. 1985. *Water Quality for Agriculture*. Rome: FAO.

Bahri, A. 2012. *Integrated Urban Water Management*. Global Water Partnership (GWP) Technical Committee (TEC) Background Papers. Stockholm: GWP.

Bahri, A., Brikké, F. & Vairavamoorthy, K. 2016. Managing change to implement integrated urban water management in African cities. *Aquatic Procedia*, 6:3-14.

Bai, Y., Wang, P., Li, C., Xie, J. & Wang, Y. 2015. Dynamic forecast of daily urban water consumption using a variable-structure support vector regression model. *Journal of Water Resources Planning and Management*, 141(3):04014058.

Baker, B., Omer, A. & Aldridge, C. 2017. *Water: Sink to Sea*. Mississippi: Mississippi State University Extension Service.

Banda, P., Bhuiyan, M., Zhang, K. & Song, A. 2021. Multivariate monthly water demand prediction using ensemble and gradient boosting machine learning techniques. In *Proceedings of the International Conference on Evolving Cities*, 2021:29-36.

Bandyopadhyay, J. 2017. Restoration of ecological status of Himalayan rivers in China and India: The case of the two mother rivers – The Yellow and the Ganges. In S. Dong, J. Bandyopadhyay & S. Chaturvedi (Eds.). *Environmental Sustainability from the Himalayas to the Oceans: Struggles and Innovations in China and India*. Cham: Springer. pp. 69-98.

Bangdiwala, S.I. 2018. Regression: Simple linear. *International Journal of Injury Control and Safety Promotion*, 25(1):113-115.

Barona, J. & Mestre, J. 2008. *La Salud y el Estado: El Movimiento Sanitario Internacional y la Administración Española, 1815–1945*. València: Universitat de València.

Bata, M.T., Carriveau, R. & Ting, D.S.K. 2020. Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water*, 5:1-18.

Batie, S.S. 2008. Wicked problems and applied economics. *American Journal of Agricultural Economics*, 90(5):1176-1191.

Bayliss, K., Fine, B., Saad-Filho, A. & Robertson, M. 2016. 13 things you need to know about neoliberalism: Political economy. *New Agenda: South African Journal of Social and Economic Policy*, 2016(63):24-33.

Benson, D., Gain, A. & Rouillard, J. 2015. Water governance in a comparative perspective: From IWRM to a “nexus” approach? *Water Alternatives*, 8(1):756-773.

Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F., Burgmann, H., Sorum, H., Norstrom, M. & Pons, M.N. 2015. Tackling antibiotic resistance: The environmental framework. *Nature Reviews Microbiology*, 13(5):310-317.

Bernstein, J.H. 2015. Transdisciplinarity: A review of its origins, development, and current issues. *Journal of Research Practice*, 11(1):Article R1. Available at: http://jrp.icaap.org/index.php/jrp/article/view/510/412.

Bichai, F., Ryan, H., Fitzgerald, C., Williams, K., Abdelmoteleb, A., Brotchie, R. & Komatsu, R. 2015. Understanding the role of alternative water supply in an urban water security strategy: An analytical framework for decision-making. *Urban Water Journal*, 12(3):175-189.

Biswas, A.K. 2004. Integrated water resources management: A reassessment: A water forum contribution. *Water International*, 29(2):248-256.

Biswas, A.K. 2008. Integrated water resources management: Is it working? *International Journal of Water Resources Development*, 24(1):5-22.

Bixio, D., De Heyder, B., Cikurel, H., Muston, M., Miska, V., Joksimovic, D., Schäfer, A.I., Ravazzini, A., Aharoni, A., Savic, D. & Thoeye, C. 2005. Municipal wastewater reclamation: Where do we stand? An overview of treatment technology and management practice. *Water Science and Technology: Water Supply*, 5(1):77-85.

Bixio, D., Thoeye, C., De Koning, J., Joksimovic, D., Savic, D., Wintgens, T. & Melin, T. 2006. Wastewater reuse in Europe. *Desalination*, 187(1-3):89-101.

Blättel-Mink, B. & Kastenholz, H. 2005. Transdisciplinarity in sustainability research: Diffusion conditions of an institutional innovation. *The International Journal of Sustainable Development & World Ecology*, 12(1):1-12.

Bougadis, J., Adamowski, K. & Diduch, R. 2005. Short-term municipal water demand forecasting. *Hydrological Processes: An International Journal*, 19(1):137-148.

Bourblanc, M. & Blanchon, D. 2014. The challenges of rescaling South African water resources management: Catchment Management Agencies and interbasin transfers. *Journal of Hydrology*, 519:2381-2391.

Box, G.E., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M. 2015. *Time Series Analysis: Forecasting and Control*. New Jersey: John Wiley & Sons.

Bradley, R.S., Diaz, H.F., Eischeid, J.K., Jones, P.D., Kelly, P.M. & Goodess, C.M. 1987. Precipitation fluctuations over Northern Hemisphere land areas since the mid-19th century. *Science*, 237:171-175.

Braun, M., Bernard, T., Piller, O. & Sedehizade, F. 2014. 24-hours demand forecasting based on SARIMA and support vector machines. *Procedia Engineering*, 89:926-933.

Brentan, B.M., Luvizotto Jr, E., Herrera, M., Izquierdo, J. & Pérez-García, R. 2017. Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, 309:532-541.

Brissaud, F. 2008. Criteria for water recycling and reuse in the Mediterranean countries. *Desalination*, 218:24-33.

Bromley, D.B. 1986. *The Case-Study Method in Psychology and Related Disciplines*. Chichester: John Wiley and Sons.

Brown, E.C. & Weinstock, N. 1980. Legal issues in implementing water reuse in California. *Ecology LQ*, 9:243.

Brown, L.E., Mitchell, G., Holden, J., Folkard, A., Wright, N., Beharry-Borg, N., Berry, G., Brierley, B., Chapman, P., Clarke, S.J. & Cotton, L. 2010. Priority water research questions as determined by UK practitioners and policy makers. *Science of the Total Environment*, 409(2):256-266.

Burges, C.J. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121-167.

Burhanuddin, M.A., Mohammed, A.A.J., Ismail, R., Hameed, M.E., Kareem, A.N. & Basiron, H. 2018. A review on security challenges and features in wireless sensor networks: IoT perspective. *Journal of Telecommunication, Electronic and Computer Engineering*, 10(1-7):17-21.

Cai, B., Hubacek, K., Feng, K., Zhang, W., Wang, F. & Liu, Y. 2020. Tension of agricultural land and water use in China’s trade: Tele-connections, hidden drivers and potential solutions. *Environmental Science & Technology*, 54(9):5365-5375.

Caiado, J. 2010. Performance of combined double seasonal univariate time series models for forecasting water demand. *Journal of Hydrologic Engineering*, 15(3):215-222.

California Department of Public Health. 2014. *Regulations Related to Recycled Water*. Available at: https://dot.ca.gov/-/media/dot-media/programs/design/documents/rwregulations-20140618-a11y.pdf.

California State Board of Health. 1918. *Regulations Governing Use of Sewage for Irrigation Purposes*. Sacramento: California State Board of Health.

California State Water Resources Control Board and Department of Water Resources. 2015. *State Water Resources Control Board Resolution No. 2015-0032: To Adopt an Emergency Regulation for Statewide Urban Water Conservation*. Available at: https://www.waterboards.ca.gov/board_decisions/adopted_orders/resolutions/2015/rs2015_0032.pdf.

California Water Boards. 2000. *Water Recycling Criteria Update (Title 22, Division 4, Chapter 3)*. Available at: https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/water-recycling-criteria.html.

California Water Boards. 2020. *California State Water Resources Control Board Recycled Water Funding for Fiscal Year 2020/21*. Available at: https://watereuse.org/wp-content/uploads/2020/07/WateReuse-Chapter-Meeting-Presentation-on-WRFP-Funding_20200716.pdf.

Campisi-Pinto, S., Adamowski, J. & Oron, G. 2012. Forecasting urban water demand via wavelet-denoising and neural network models – Case study: City of Syracuse, Italy. *Water Resources Management*, 26(12):3539-3558.

Candelieri, A. 2017. Clustering and support vector regression for water demand forecasting and anomaly detection. *Water*, 9(3):224.

Candelieri, A. & Archetti, F. 2014. Identifying typical urban water demand patterns for a reliable short-term forecasting: The icewater project approach. *Procedia Engineering*, 89:1004-1012.

Candelieri, A., Giordani, I., Archetti, F., Barkalov, K., Meyerov, I., Polovinkin, A., Sysoyev, A. & Zolotykh, N. 2019. Tuning hyperparameters of a SVM-based water demand forecasting system through parallel global optimization. *Computers & Operations Research*, 106:202-209.

Carr, R. 2005. WHO guidelines for safe wastewater use: More than just numbers. *Irrigation and Drainage*, 2005(54):103-111.

Carvalho, T.M.N., Souza Filho, F.A. & Porto, V.C. 2021. Urban water demand modelling using machine learning techniques: Case study of Fortaleza, Brazil. *Journal of Water Resources Planning and Management*, 147(1):05020026.

Castro, J.E. 2007. Water governance in the twentieth-first century. *Ambiente & Sociedade*, 10(2):97-118.

Chan, S.J., Nutting, V.I., Natterson, T.A. & Horowitz, B.N. 2021. Impacts of psychopharmaceuticals on the neurodevelopment of aquatic wildlife: A call for increased knowledge exchange across disciplines to highlight implications for human health. *International Journal of Environmental Research and Public Health*, 18(10):5094.

Checkland, P. 1985. From optimizing to learning: A development of systems thinking for the 1990s. *The Journal of the Operational Research Society*, 36(9):757-767.

Chen, B., Han, M.Y., Peng, K., Zhou, S.L., Shao, L., Wu, X.F., Wei, W.D., Liu, S.Y., Li, Z., Li, J.S. & Chen, G.Q. 2018. Global land-water nexus: Agricultural land and freshwater use embodied in worldwide supply chains. *Science of the Total Environment*, 613:931-943.

Chen, T. & Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August, pp. 785-794. Available at: https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H. & Chen, K. 2015. Xgboost: Extreme gradient boosting. *R Package Version 0.4-2*, *1*(4):1-4.

Chen, X.Y., Li, P., Yuan, Y. & Shi, X. 2005. Forecast of water using improved Chebyshev neural network. *Journal of Petrochemical Universities*, 18(1):70-72.

Chikozho, C. 2008. Globalizing integrated water resources management: A complicated option in Southern Africa. *Water Resource Management*, 22:1241-1257.

Chikozho, C., Managa, R. & Dabata, T. 2020. Ensuring access to water for food production by emerging farmers in South Africa: What are the missing ingredients? *Water SA*, 46(2):225-233.

Chini, C.M. & Stillwell, A.S. 2018. The state of US urban water: Data and the energy-water nexus. *Water Resources Research*, 54(3):1796-1811.

Chu, J., Chen, J., Wang, C. & Fu, P. 2004. Wastewater reuse potential analysis: Implications for China's water resources management. *Water Research*, *38*(11):2746-2756.

Cicovacki, P. 2004. Transdisciplinarity as an interactive method: A critical reflection on the three pillars of transdisciplinarity. *TRANS: Internet Journal for Cultural Sciences*, 15(1). Available at: http://www.inst.at/trans/15Nr/01_6/cicovacki15.htm.

Cicovacki, P. 2009. Transdisciplinarity as an interactive method. *Integral Leadership Review*, 9(5). Available at: http://www.integralleadershipreview.com/archives/2009-10/2009-10-06-article-cicovaki.php.

Closas, A., Schuring, M. & Rodriguez, D. 2012. *Integrated Urban Water Management: Lessons and Recommendations from Regional Experiences in Latin America, Central Asia, and Africa (No. 75043)*. Washington, D.C.: World Bank.

Coelho, A.C., Labadie, J.W. & Fontane, D.G. 2012. Multicriteria decision support system for regionalization of integrated water resources management. *Water Resources Management*, 26(5):1325-1346.

Cohen, L., Manion, L. & Morrison, K. 2007. *Research Methods in Education.* 6th edition. London: Routledge.

Comisión Nacional del Agua (CONAGUA). 2007. *Programa Nacional Hídrico [National Water Programme] 2007-2012.* Available at: http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/PNH_Ingles.pdf.

Comisión Nacional del Agua (CONAGUA). 2010. *Financing Water Resources Management in Mexico*. Available at: www.conagua.gob.mx/english07/publications/OECD.pdf.

Comisión Nacional del Agua (CONAGUA). 2018. *Programa Nacional Hídrico [National Water Programme] 2014-2018*. Available at: https://www.gob.mx/cms/uploads/attachment/file/411191/PNH_Reporte_Transversalidad_061118_s_a.pdf.

Commonwealth Scientific and Industrial Research Organisation. 2012. *The Impacts of Climate Change and Urban Development on Future Water Security and the Adaptation Options for Makassar City, Indonesia*. Available at: https://publications.csiro.au/rpr/pub?pid=csiro:EP126372.

Cook, T.D. & Reichardt, C.S. 1979. *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills: Sage Publications.

Cooper, P. 2001. Historical aspect of wastewater treatment. In P. Lens, G. Zeemann & G. Lettinga (Eds.). *Decentralised Sanitation Reuse: Concepts, System, and Implementation*. London: IWA Publishing. pp. 11-38.

Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine Learning*, 20(3):273-297.

Cvitanovic, C., Hobday, A.J., Van Kerkhoff, L., Wilson, S.K., Dobbs, K. & Marshall, N.A. 2015. Improving knowledge exchange among scientists and decision-makers to facilitate the adaptive governance of marine resources: A review of knowledge and research needs. *Ocean & Coastal Management*, 112:25-35.

Decker, M. & Fleischer, T. 2010. When should there be which kind of technology assessment? A plea for a strictly problem-oriented approach from the very outset. *Poiesis & Praxis*, 7(1-2):117-133.

De Myttenaere, A., Golden, B., Le Grand, B. & Rossi, F. 2016. Mean absolute percentage error for regression models. *Neurocomputing*, 192:38-48.

Department of Water Affairs (DWA). 2011. *Green Drop Report, 2011*. Available at: https://www.dws.gov.za/Documents/GD/GDIntro.pdf.

Department of Water Affairs (DWA). 2013. *National Water Resource Strategy: Second Edition*. Available at: https://www.dws.gov.za/documents/Other/Strategic%20Plan/NWRS2-Final-email-version.pdf.

Department of Water Affairs and Forestry (DWAF). 1996. *South African Water Quality Guidelines – Volume 7: Aquatic Ecosystems*. Available at: https://www.dws.gov.za/iwqs/wq_guide/edited/Pol_saWQguideFRESH_vol7_Aquaticecosystems.pdf.

Department of Water Affairs and Forestry (DWAF). 1997. *White Paper on a National Water Policy for South Africa*. Available at: https://www.gov.za/sites/default/files/gcis_document/201409/nwpwp.pdf.

Department of Water Affairs and Forestry (DWAF). 1999. GN 1191, 8 October 1999: General authorizations in terms of section 39 of the National Water Act, 1998 (Act No. 36 of 1998). *Government Gazette*, 20526. Available at: http://ward2forum.org/wp-content/uploads/2017/03/GENERAL-AUTHORISATIONS-IN-TERMS-OF-SECTION-39-OF-THE-NATIONAL-WATER-ACT-1998-ACT-NO.-36-OF-1998.pdf.

Department of Water Affairs and Forestry (DWAF). 2001a. *Annual Report 2000-2001*. Available at: https://www.gov.za/sites/default/files/gcis_document/201409/dwaf-annualreport2000-2001.pdf.

Department of Water Affairs and Forestry (DWAF). 2001b. GN R509, 8 June 2001: Regulations: Compulsory national standards and measures to conserve water. *Government Gazette*, 22355. Available at: https://www.gov.za/sites/default/files/gcis_document/201409/223550.pdf.

Department of Water Affairs and Forestry (DWAF). 2004. *National Water Resource Strategy: First Edition*. Available at: https://cer.org.za/wp-content/uploads/2017/10/NWRS-2004.pdf.

Department of Water Affairs and Forestry (DWAF). 2009. *Annual Report: 1 April – 31 March 2009*. Available at: https://www.gov.za/sites/default/files/gcis_document/201409/dwafannualreport200809.pdf.

Department of Water and Sanitation (DWS). n.d. *Overview of the South African Water Sector*. Available at: https://www.dws.gov.za/IO/Docs/CMA/CMA%20GB%20Training%20Manuals/gbtrainingmanualchapter1.pdf.

Dieter, C.A., Maupin, M.A., Caldwell, R.R., Harris, M.A., Ivahnenko, T.I., Lovelace, J.K., Barber, N.L. & Linsey, K.S. 2018. Estimated use of water in the United States in 2015. *US Geological Survey Circular*, 1441:65. Available at: https://pubs.er.usgs.gov/publication/cir1441.

Dogo, E.M., Nwulu, N.I., Twala, B. & Aigbavboa, C. 2019. A survey of machine learning methods applied to anomaly detection on drinking-water quality data. *Urban Water Journal*, 16(3):235-248.

Domènech, L., March, H. & Saurí, D. 2013. Degrowth initiatives in the urban water sector? A social multi-criteria evaluation of non-conventional water alternatives in Metropolitan Barcelona. *Journal of Cleaner Production*, 38:44-55.

Donnenfeld, Z., Hedden, S. & Crookes, C. 2018. *A Delicate Balance: Water Scarcity in South Africa*. Pretoria: Institute for Security Studies.

Dooley, G. & Lenihan, H. 2005. An assessment of time series methods in metal price forecasting. *Resources Policy*, 30(3):208-217.

Dooley, J., Alkhaddar, R.M., Abdellatif, M., Zubaidi, S.L. & Gharghan, S.K. 2018. Short-term urban water demand prediction considering weather factors. *Water Resources Management*, 14(32):4527-4542.

Drechsel, P., Scott, A., Sally, R., Redwood, M. & Bachir, A. 2010. *Wastewater Irrigation and Health: Assessing and Mitigating Risk in Low-Income Countries*. London: Earthscan.

Dyer Jr, W.G. & Wilkins, A.L. 1991. Better stories, not better constructs, to generate better theory: A rejoinder to Eisenhardt. *Academy of Management Review*, 16(3):613-619.

Edokpayi, J.N., Enitan-Folami, A.M., Adeeyo, A.O., Durowoju, O.S., Jegede, A.O. & Odiyo, J.O. 2020. Recent trends and national policies for water provision and wastewater treatment in South Africa. In P. Singh (Ed.). *Water Conservation and Wastewater Treatment in BRICS Nations: Technologies, Challenges, Strategies and Policies*. Amsterdam: Elsevier. pp. 187-211.

Eftelioglu, E., Jiang, Z., Tang, X. & Shekhar, S. 2017. The nexus of food, energy, and water resources: Visions and challenges in spatial computing. In D.A. Griffith, Y. Chun & D.J. Dean (Eds.). *Advances in Geocomputation*. Cham: Springer. pp. 5-20.

Eisner, E.W. 1998. Does experience in the arts boost academic achievement? *Arts Education Policy Review*, 100(1):32-40.

Elbana, T.A., Bakr, N. & Elbana, M. 2017. Reuse of treated wastewater in Egypt: Challenges and opportunities. In A.M. Negm (Ed.). *Unconventional Water Resources and Agriculture in Egypt*. Cham: Springer. pp. 429-453.

Elmenoufy, H.M., Morsy, M., Eid, M.M., El Ganzoury, A., El-Hussainy, F.M. & Wahab, M.A. 2017. Towards enhancing rainfall projection using bias correction method: Case study Egypt. *IJSRSET*, 6(3):187-194.

El-Zanfaly, H.T. 2015. Wastewater reuse in agriculture: A way to develop the economies of arid regions of the developing countries. *Journal of Environment Protection and Sustainable Development*, 1:144-158.

Environmental Protection Agency (EPA). 2004. *Guidelines for Water Reuse*. Washington, D.C.: EPA.

Environmental Protection Agency (EPA). 2012. *Guidelines for Water Reuse*. Washington, D.C.: EPA.

Erasmus, Z. 2010. Contact theory: Too timid for “race” and racism. *Journal of Social Issues*, 66(2):387-400.

Eslamian, S. (Ed.). 2016. *Urban Water Reuse Handbook*. Boca Raton: CRC Press.

European Commission. 2007. *Communication from the Commission to the Council and the European Parliament, Addressing the Challenge of Water Scarcity and Droughts in the European Union. Brussels*. Available at: https://eur-ex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0414:FIN:en:PDF.

European Commission. 2012. *A Blueprint to Safeguard Europe’s Water Resources*. Available at: https://www.riob.org/sites/default/files/IMG/pdf/Presentation_synthetique_du_BP.pdf.

European Commission. 2014. *Regulation on Minimum Requirements for Water Reuse Enters Into Force*. Available at: https://ec.europa.eu/environment/water/reuse.htm.

European Commission. 2015. *Optimising Water Reuse in the EU: Final Report – Part I*. Available at: https://ec.europa.eu/environment/water/blueprint/pdf/BIO_IA%20on%20water%20reuse_Final%20Part%20I.pdf.

European Commission. 2020. *Water Reuse: Regulation on Minimum Requirements for Water Reuse*. Available at: https://ec.europa.eu/environment/water/reuse.htm.

European Environment Agency. 2015. *Use of Freshwater Resources*. Available at: https://www.eea.europa.eu/data-and-maps/indicators/use-of-freshwater-resources-2/assessment-3.

European Environment Agency. 2017. Available at: https://www.eea.europa.eu/data-and-maps/indicators/use-of-freshwater-resources-3/assessment-4.

European Parliament and Council. 2000. *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy.* Available at: https://eur-lex.europa.eu/eli/dir/2000/60/oj.

European Parliament and Council. 2010. *Directive 2010/75/EU of the European Parliament and of the Council on Industrial Emissions (Integrated Pollution Prevention and Control)*. Available at: https://www.ecolex.org/details/legislation/directive-201075eu-of-the-european-parliament-and-of-the-council-on-industrial-emissions-integrated-pollution-prevention-and-control-lex-faoc109066/.

Fairbridge, D. 1922. *Historic Houses of South Africa*. London: Oxford University Press.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X. & Xiang, Y. 2018. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, 164:102-111.

Farooqui, T.A., Renouf, M.A. & Kenway, S.J. 2016. A metabolism perspective on alternative urban water servicing options using water mass balance. *Water Research*, 106:415-428.

Feagin, J.R., Orum, A.M. & Sjoberg, G. (Eds.). 1991. *A Case for the Case Study*. Chapel Hill: UNC Press.

Felizatto, M. 2001. Projeto integrado de tratamento avançado e reúso direto de águas residuárias. In *Proceedings of the 21st Congresso Brasileiro de Engenharia Sanitária e Ambiental*, João Pessoa, Brazil, 16-21 September, pp. 1-17.

Fetters, M.D., Curry, L.A. & Creswell, J.W. 2013. Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, 48(6.2):2134-2156.

Fletcher, T.D., Mitchell, V., Deletic, A., Ladson, T.R. & Seven, A. 2007. Is stormwater harvesting beneficial to urban waterway environmental flows? *Water Science and Technology*, 55(4):265-272.

Flinterman, J.F., Teclemariam-Mesbah, R., Broerse, J.E.W. & Bunders, J.F.G. 2001. Transdisciplinarity: The new challenge for biomedical research*. Bulletin of Science, Technology & Society*, 21:253-266.

Flörke, M., Schneider, C. & McDonald, R.I. 2018. Water competition between cities and agriculture driven by climate change and urban growth. *Nature Sustainability*, 1(1):51-58.

Flyvbjerg, B. 2006. Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2):219-245.

Folifac, F.A. 2007. National water policies and water services at the extremes: What challenges must be faced in bridging the gap? Learning from the South Africa experience. *African Water Journal*, 1(1):5-22.

Food and Agriculture Organization (FAO). 1999. *Wastewater Treatment and Use in Agriculture*. Available at: https://www.fao.org/3/t0551e/t0551e00.htm.

Food and Agriculture Organization (FAO)-AQUASTAT. 2016. *Global Water Information System Website*. Available at: http://www.fao.org/nr/water/aquastat/data/query/results.html.

Frantzeskaki, N. & Loorbach, D. 2010. Towards governing infrasystem transitions: Reinforcing lock-in or facilitating change? *Technological Forecasting and Social Change,* 77(8):1292-1301.

Freedman, J., Tseng, J., Meeker, M. & Vallero, M. 2015. *Addressing Water Scarcity Through Recycling and Reuse: A Menu for Policymakers – Perspective on Latin America, Brazil, and Mexico*. Available at: https://watereuse.org/wp-content/uploads/2015/01/Addressing_Water_Scarcity_in_LAM.pdf.

Funtowicz, S. & Ravetz, J. 2008. Values and uncertainties. In G.H. Hadorn, H. Hoffman-Riem, S. Biber-Klemm, W. Grossenbacher-Mansuy, D. Joye, C. Phol, U. Wiesmann & E. Zemp (Eds.). *Handbook of Transdisciplinary Research*. Dordrecht: Springer. pp. 361-368.

Furlong, C., Gan, K. & De Silva, S. 2016. Governance of integrated urban water management in Melbourne, Australia. *Utilities Policy*, 43:48-58.

Furlong, C., Guthrie, L., De Silva, S. & Considine, R. 2015. Analysing the terminology of integration in the water management field. *Water Policy*, 17(1):46-60.

Gabr, M.E. 2018. *Wastewater Reuse Standards for Agriculture Irrigation in Egypt*. Paper Presented at the 21st International Water Technology Conference, Egypt. Available at: https://www.researchgate.net/publication/333676905_WASTEWATER_REUSE_STANDARDS_FOR_AGRICULTURE_IRRIGATION_IN_EGYPT.

Gagliardi, F., Alvisi, S., Kapelan, Z. & Franchini, M. 2017. A probabilistic short-term water demand forecasting model based on the Markov Chain. *Water*, 9(7):507.

Gallego-Ayala, J. & Juízo, D. 2014. Integrating stakeholders' preferences into water resources management planning in the Incomati river basin. *Water Resources Management*, 28(2):527-540.

Garcia, X., Barceló, D., Comas, J., Corominas, L., Hadjimichael, A., Page, T.J. & Acuña, V. 2016. Placing ecosystem services at the heart of urban water systems management. *Science of the Total Environment*, 1(563-564):1078-1085.

Georgakakos, A.P., Yao, H., Kistenmacher, M., Georgakakos, K.P., Graham, N.E., Cheng, F.Y., Spencer, C. & Shamir, E. 2012. Value of adaptive water resources management in Northern California under climatic variability and change: Reservoir management. *Journal of Hydrology*, 412:34-46.

Gernaey, K.V., Van Loosdrecht, M.C., Henze, M., Lind, M. & Jørgensen, S.B. 2004. Activated sludge wastewater treatment plant modelling and simulation: State of the art. *Environmental Modelling & Software*, 19(9):763-783.

Ghalehkhondabi, I., Ardjmand, E., Young, W.A. & Weckman, G.R. 2017. Water demand forecasting: Review of soft computing methods. *Environmental Monitoring and Assessment*, 189(7):1-13.

Ghernaout, D. & Ibn-Elkhattab, R.O. 2020. On the treatment trains for municipal wastewater reuse for irrigation. *Open Access Library Journal*, 7(2):1-15.

Ghiassi, M., Zimbra, D.K. & Saidane, H. 2008. Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management*, 134(2):138-146.

Gilabert-Alarcón, C., Salgado-Méndez, S.O., Daesslé, L.W., Mendoza-Espinosa, L.G. & Villada-Canela, M. 2018. Regulatory challenges for the use of reclaimed water in Mexico: A case study in Baja California. *Water*, 10(10):1432.

Gillham, B. 2000. *Case Study Research Methods*. London: Continuum.

Glasser, M.J. 2006. Cutting through the jargon. *Event Horizon Blog*, 21 June. Available at: http://technoeventhorizon.blogspot.com/2006_06_01_archive.html.

Glazewski, J. 1999. *Environmental Law in South Africa*. Durban: Butterworth-Heinemann.

Gleick, P.H. 2009. China and water. In P.H. Gleick, H. Cooley, M.J. Cohen, M. Morikawa, J. Morrison & M. Palaniappan (Eds.). *The World's Water 2008–2009: The Biennial Report on Freshwater Resources*. Washington, D.C.: Island Press. pp. 79-100.

Global Water Intelligence. 2012. *Global Water and Wastewater Quality Regulations 2012*. Available at: https://www.globalwaterintel.com/client_media/uploaded/Global%20Regulation/GWI_Regulations_2012_Prepub_Table_of_Contents.pdf.

Grigg, N.S. 2008. Integrated water resources management: Balancing views and improving practice. *Water International*, 33(3):279-292.

Gummesson, E. 1988. Service quality and product quality combined. *Review of Business*, 9(3):14-19.

Guo, G. & Liu, S. 2018. Short-term water demand forecast based on deep neural network. In *WDSA/CCWI Joint Conference Proceedings*, Vol. 1. Available at: https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12039.

Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R. & Zhu, X. 2018. Short-term water demand forecast based on deep learning method. *Journal of Water Resources Planning and Management*, 144(12):04018076.

Guo, L., Fang, W., Zhao, Q. & Wang, X. 2021. The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. *Computers & Industrial Engineering*, 161:107598.

Gutmann, B. & Kersting, K. 2007. Stratified conjugate gradient boosting for fast training of conditional random fields. In *Proceedings of the 5th International Workshop on Mining and Learning with Graphs*, pp. 131-134. Available at: https://www.researchgate.net/publication/268325634_Stratified_Gradient_Boosting_for_Fast_Training_of_Conditional_Random_Fields.

Hadjimichael, A., Comas, J. & Corominas, L. 2016. Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector. *AI Communications*, 29(6):747-756.

Hadorn, G.H., Biber-Klemm, S., Grossenbacher-Mansuy, W., Hoffmann-Riem, H., Joye, D., Pohl, C., Wiesmann, U. & Zemp, E. 2008. The emergence of transdisciplinarity as a form of research. In G.H. Hadorn, H. Hoffman-Riem, S. Biber-Klemm, W. Grossenbacher-Mansuy, D. Joye, C. Phol, U. Wiesmann & E. Zemp (Eds.). *Handbook of Transdisciplinary Research*. Dordrecht: Springer. pp. 19-39.

Handelman, G.S., Kok, H.K., Chandra, R.V., Razavi, A.H., Huang, S., Brooks, M., Lee, M.J. & Asadi, H. 2019. Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods. *American Journal of Roentgenology*, 212(1):38-43.

Hanjra, M.A., Blackwell, J., Carr, G., Zhang, F. & Jackson, T.M. 2012. Wastewater irrigation and environmental health: Implications for water governance and public policy. *International Journal of Hygiene and Environmental Health*, 215(3):255-269.

Harding, S. 1987. Is there a feminist method? In S. Harding (Ed.). *Feminism and Methodology*. Bloomington: Indiana University Press. pp. 1-14.

Hartini, S., Hadi, M.P., Sudibyakto, S. & Poniman, A. 2015. Application of vector auto regression model for rainfall-river discharge analysis. *Forum Geografi*, 29(1):1-10.

Hartman, R.S. 1967. *The Structure of Value: Foundations of Scientific Axiology*. Carbondale: Southern Illinois University Press.

Hashem, M.S. & Qi, X. 2021. Treated wastewater irrigation: A review. *Water*, 13(11):1527.

Hassanzadeh, E., Elshorbagy, A., Wheater, H. & Gober, P. 2016. A risk-based framework for water resource management under changing water availability, policy options, and irrigation expansion. *Advances in Water Resources*, 94:291-306.

Hattingh, J.L. 1983. Naamgewing aan slawe, vryswartes en ander gekleurdes. *Kronos*, 6:5-20.

He, G., Zhang, L., Mol, A.P., Wang, T. & Lu, Y. 2014. Why small and medium chemical companies continue to pose severe environmental risks in rural China. *Environmental Pollution*, 2014(185):158-167.

Healy, R.W., Alley, W.M., Engle, M.A., McMahon, P.B. & Bales, J.D. 2015. The water-energy nexus – An earth science perspective. *U.S. Geological Survey Circular*, 1407(107). Available at: https://pubs.usgs.gov/circ/1407/.

Heaney, J.P., Pitt, R. & Field, R. 1999. *Innovative Urban Wet-Weather Flow Management Systems*. Cincinnati: National Risk Management Research Laboratory, U.S. Environmental Protection Agency.

Heller, M. & Thind, H. 1994. Forecasting with cascade correlation: An application to potable water demand. *Artificial Neural Networks in Engineering: Proceedings*, 4:1155-1160.

Hernández-Mora, N. & Del Moral, L. 2015. Developing markets for water reallocation: Revisiting the experience of Spanish water mercantilización. *Geoforum*, 62:143-155.

Herrera, M., Izquierdo, J., Pérez- García, R. & Ayala-Cabrera, D. 2014. On-line learning of predictive kernel models for urban water demand in a smart city. *Procedia Engineering*, 70:791-799.

Herrera, M., Torgo, L., Izquierdo, J. & Pérez-García, R. 2010. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, 387(1-2):141-150.

Herrfahrdt-Pähle, E. 2013. Integrated and adaptive governance of water resources: The case of South Africa. *Regional Environmental Change*, 13(3):551-561.

House-Peters, L.A. & Chang, H. 2011. Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, 47(5):1-15.

Huang, L., Zhang, C., Peng, Y. & Zhou, H. 2014. Application of a combination model based on wavelet transform and KPLS-ARMA for urban annual water demand forecasting. *Journal of Water Resources Planning and Management*, 140(8):04014013.

Hyndman, R.J. & Khandakar, Y. 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(2008):1-22.

International Organization for Standardization (ISO). 2013. *ISO/TC 282: Water Reuse*. Available at: https://www.iso.org/committee/4856734.html.

International Organization for Standardization (ISO). 2015. *ISO 16075-2:2015: Guidelines for Treated Wastewater Use for Irrigation Projects – Part 2: Development of the Project*. Available at: https://www.iso.org/standard/62758.html.

Ivanko, D., Sørensen, Å.L. & Nord, N. 2020. Selecting the model and influencing variables for DHW heat use prediction in hotels in Norway. *Energy and Buildings*, 228:110441.

Jackson, D.L. 2003. Revisiting sample size and number of parameter estimates: Some support for the N: q hypothesis. *Structural Equation Modeling*, 10(1):128-141.

Jackson, M.C. 2007. *Systems Approaches to Management*. Cham: Springer Science & Business Media.

Jacobs, I.M. & Nienaber, S. 2011. Waters without borders: Transboundary water governance and the role of the 'transdisciplinary individual' in Southern Africa. *Water SA*, 37(5):665-678.

Jacobsen, M., Webster, M. & Vairavamoorthy, K. (Eds.). 2012. *The Future of Water in African Cities: Why Waste Water?* Washington D.C.: World Bank.

Jain, A. & Kumar, A.M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, 7(2):585-592.

Jain, A. & Ormsbee, L.E. 2002. Short-term water demand forecast modeling techniques – Conventional methods versus AI. *American Water Works Association Journal*, 94(7):64-72.

Jain, A., Varshney, A.K. & Joshi, U.C. 2001. Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resources Management*, 15(5):299-321.

Janes, F.R. 1988. Interpretive structural modelling: A methodology for structuring complex issues. *Transactions of the Institute of Measurement and Control*, 10(3):145-154.

Jaramillo, M.F. & Restrepo, I. 2017. Wastewater reuse in agriculture: A review about its limitations and benefits. *Sustainability*, 9(10):1734.

Jia, P., An, S., Chen, G., Jeon, J.Y. & Jee, H.K. 2007. Urban water demand forecasting using artificial neural network model: Case study of Daegu City. In Korea Water Resources Association. *Proceedings of the Korea Water Resources Association Conference*, pp. 1910-1914. Available at: https://koreascience.kr/article/CFKO200724737423898.pdf.

Jiang, T., Gradus, J.L. & Rosellini, A.J. 2020. Supervised machine learning: A brief primer. *Behavior Therapy*, 51(5):675-687.

Jiménez, B. 2006. Irrigation in developing countries using wastewater. *International Review for Environmental Strategies*, 6(2):229-250.

Jiménez, B. & Asano, T. 2008. *Water Reuse: An International Survey of Current Practice, Issues and Needs*. London: IWA Publishing.

Jódar-Abellán, A., Ruiz-Álvarez, M.A.R.C.O.S. & Valdes-Abellan, J. 2019. Calibration and validation of et 0 through an r-cran code in agricultural lands of south-east Spain. *WIT Transactions on Ecology and the Environment*, 234:167-179.

Jussah, O., Orabi, M.O., Sušnik, J., Bichai, F. & Zevenbergen, C. 2020. Assessment of the potential contribution of alternative water supply systems in two contrasting locations: Lilongwe, Malawi and Sharm El-Sheikh, Egypt. *Journal of Water and Climate Change*, 11(1):130-149.

Kahinda, J.M.M. & Boroto, J.R. 2009. *IWRM Survey and Status Report: South Africa*. Stockholm: Global Water Partnerships Southern Africa.

Kahinda, J.M.M., Meissner, R. & Engelbrecht, F.A. 2016. Implementing integrated catchment management in the upper Limpopo River basin: A situational assessment. *Physics and Chemistry of the Earth, Parts A/B/C*, 93:104-118.

Kamizoulis, G. 2008. Setting health-based targets for water reuse (in agriculture). *Desalination*, 2008(218):154-163.

Kang, H.S., Kim, H., Lee, J., Lee, I., Kwak, B.Y. & Im, H. 2015. Optimization of pumping schedule based on water demand forecasting using a combined model of autoregressive integrated moving average and exponential smoothing. *Water Science and Technology: Water Supply*, 15(1):188-195.

Kapfudzaruwa, F. & Sowman, M. 2009. Is there a role for traditional governance systems in South Africa's new water management regime? *Water SA*, 35(5):683-692.

Karodia, H. & Weston, D.R. 2001. South Africa's new water policy and law. In *Intersectoral Management of River Basins. Pretoria, DWAF/IWMI.* pp. 13-22. Available at: https://publications.iwmi.org/pdf/H029111.pdf.

Kellis, M., Kalavrouziotis, I.K. & Gikas, P. 2013. Review of wastewater reuse in the Mediterranean countries, focusing on regulations and policies for municipal and industrial applications. *Global NEST Journal*, 15(3):333-350.

Keng, C.Y., Shan, F.P., Shimizu, K., Imoto, T., Lateh, H. & Peng, K.S. 2017. *Application of Vector Autoregressive Model for Rainfall and Groundwater Level Analysis*. Available at: https://ui.adsabs.harvard.edu/abs/2017AIPC.1870f0013K/abstract.

Kessel, F. & Rosenfield, P.L. 2008. Toward transdisciplinary research: Historical and contemporary perspectives. *American Journal of Preventive Medicine*, 35(2):S225-S234.

Khalid, R.M. 2018. Review of the water supply management and reforms needed to ensure water security in Malaysia. *International Journal of Business and Society*, 19(S3):472-483.

Khanum, M., Mahboob, T., Imtiaz, W., Ghafoor, H.A. & Sehar, R. 2015. A survey on unsupervised machine learning algorithms for automation, classification and maintenance. *International Journal of Computer Applications*, 119(13):34-39.

Kidd, M. 2009. South Africa: The development of water law. In J.W. Dellapenna & J. Gupta (Eds.). *The Evolution of the Law and Politics of Water.* Dordrecht: Springer. pp. 87-104.

Kim, B., Lee, D.E., Hu, G., Natarajan, Y., Preethaa, S. & Rathinakumar, A.P. 2022. Ensemble machine learning-based approach for predicting of FRP-concrete interfacial bonding. *Mathematics*, 10(2):231.

Kirshen, P., Aytur, S., Hecht, J., Walker, A., Burdick, D., Jones, S., Fennessey, N., Bourdeau, R. & Mather, L. 2018. Integrated urban water management applied to adaptation to climate change. *Urban Climate*, 24:247-263.

Klein, J.T. 2001. The discourse of transdisciplinarity: An expanding global field. In J.T. Klein. *Transdisciplinarity: Joint Problem Solving Among Science, Technology, and Society.* Basel: Birkhäuser. pp. 35-44.

Klein, J.T. 2004. Prospects for transdisciplinarity. *Futures*, 36(4):515-526.

Kockelmans, J.J. 1979. Why interdisciplinarity? In J.J. Kockelmans (Ed.). *Interdisciplinarity and Higher Education*. Pennsylvania: Pennsylvania State University Press. pp. 123-160.

Kofinas, D., Mellios, N., Papageorgiou, E. & Laspidou, C. 2014. Urban water demand forecasting for the island of Skiathos. *Procedia Engineering*, 89:1023-1030.

Koop, S.H. & Van Leeuwen, C.J. 2017. The challenges of water, waste and climate change in cities. *Environment, Development and Sustainability*, 19(2):385-418.

Koutroulis, A.G., Papadimitriou, L.V., Grillakis, M.G., Tsanis, I.K., Wyser, K. & Betts, R.A. 2018. Freshwater vulnerability under high end climate change: A pan-European assessment. *Science of the Total Environment*, 613:271-286.

Kranz, N., Interwies, E. & Vidaurre, R. 2005. *Transboundary River Basin Management Regimes: The Orange Basin Case Study*. Berlin: Ecologic Institute for International and European Environmental Policy.

Kühnert, C., Gonuguntla, N.M., Krieg, H., Nowak, D. & Thomas, J.A. 2021. Application of LSTM networks for water demand prediction in optimal pump control. *Water*, 13(5):644.

Kumar, I. & Singh, S.P. 2022. Machine learning in bioinformatics. In D.B. Singh & R.K. Pathak (Eds.). *Bioinformatics*. Cambridge: Academic Press. pp. 443-456.

Kumarasamy, M.V. & Dube, V.N. 2016. Study on recycling urban wastewater for non-potable uses for water conservation. *Polish Journal of Environmental Studies*, 25(1):167-171.

Lahiri, S.K. & Ghanta, K.C. 2008. Prediction of pressure drop of slurry flow in pipeline by hybrid support vector regression and genetic algorithm model. *Chinese Journal of Chemical Engineering*, 16(6):841-848.

Leach, W.D. & Pelkey, N.W. 2001. Making watershed partnerships work: A review of the empirical literature. *Journal of Water Resources Planning and Management*, 127(6):378-385.

Lee, K. & Jepson, W. 2020. Drivers and barriers to urban water reuse: A systematic review. *Water Security*, 11:100073.

Leonard-Barton, D. 1990. A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science*, 1(3):248-266.

Li, Y., Ge, Y. & Zhang, Y. 2021. Tutorial on fairness of machine learning in recommender systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2654-2657. Available at: https://dl.acm.org/doi/abs/10.1145/3404835.3462814.

Liu, B. & Speed, R. 2009. Water resources management in the People's Republic of China. *Water Resources Development*, 25(2):193-208.

Liu, S. & Persson, K.M. 2013. Situations of water reuse in China. *Water Policy*, 15(5):705-727.

Liu, Y., Roberts, M.C. & Sioshansi, R. 2018. A vector autoregression weather model for electricity supply and demand modeling. *Journal of Modern Power Systems and Clean Energy*, 6(4):763-776.

López-Huertas, M. 2013. Reflexions on multidimensional knowledge: Its influence on the foundation of knowledge organization. *Knowledge Organization*, 40(6):400-407.

Lovarelli, D., Ingrao, C., Fiala, M. & Bacenetti, J. 2018. Beyond the water footprint: A new framework proposal to assess freshwater environmental impact and consumption. *Journal of Cleaner Production*, 172:4189-4199.

Lu, H. & Ma, X. 2020. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 249:126169.

Lyu, S., Chen, W., Zhang, W., Fan, Y. & Jiao, W. 2016. Wastewater reclamation and reuse in China: Opportunities and challenges. *Journal of Environmental Sciences*, 39:86-96.

Ma, Y., Liu, K., Guan, Z., Xu, X., Qian, X. & Bao, H. 2018. Background augmentation generative adversarial networks (BAGANs): Effective data generation based on GAN-augmented 3D synthesizing. *Symmetry*, 10(12):734.

MacKay, H.M., Rogers, K.H. & Roux, D.J. 2003. Implementing the South African water policy: Holding the vision while exploring an uncharted mountain. *Water SA*, 29(4):353-358.

Mahan, J.L., Jr. 1970. Toward Transdisciplinary Inquiry in the Humane Sciences. Doctoral Dissertation. Kenya: United States International University.

Malisa, R., Schwella, E. & Kidd, M. 2019. From ‘government’ to ‘governance’: A quantitative transition analysis of urban wastewater management principles in Stellenbosch Municipality. *Science of the Total Environment*, 674:494-511.

Malisa-Van der Walt, R. & Taigbenu, A. 2022. Policy, laws, and guidelines of wastewater reuse for agricultural purposes in developing countries. In M. Nasr & A.M. Negm (Eds.). *Cost-Efficient Wastewater Treatment Technologies*. Cham: Springer International Publishing. pp. 1-24.

Mapedza, E., Manzungu, E., Rosen, T., Ncube, P. & Van Koppen, B. 2016. Decentralised water governance in Zimbabwe: Disorder within order. *Water Resources and Rural Development*, 8:1-11.

Mara, D.D., Sleigh, P.A., Blumenthal, U.J. & Carr, R.M. 2007. Health risks in wastewater irrigation: Comparing estimates from quantitative microbial risk analyses and epidemiological studies. *Journal of Water and Health*, 5(1):39-50.

Martin Jr, D., Prabhakaran, V., Kuhlberg, J., Smart, A. & Isaac, W.S. 2020. Participatory problem formulation for fairer machine learning through community-based system dynamics. *arXiv* preprint arXiv:2005.07572. Available at: https://arxiv.org/abs/2005.07572.

Maryam, B. & Büyükgüngör, H. 2019. Wastewater reclamation and reuse trends in Turkey: Opportunities and challenges. *Journal of Water Process Engineering*, 30:100501.

McGregor, S.L.T. 2004. *The Nature of Transdisciplinary Research and Practice*. Available at: https://www.academia.edu/26721302/The_Nature_of_Transdisciplinary_Research_and_Practice.

McGregor, S.L.T. 2011. Transdisciplinary axiology: To be or not to be. *Integral Leadership* Review, 11(3):1-9.

McGregor, S.L.T. 2012. Place and transdisciplinarity. In B. Nicolescu (Ed.). *Transdisciplinarity and Sustainability*. Lubbock: The Atlas. pp. 8-12.

McGregor, S.L.T. & Volckmann, R. 2013. Transversity: Transdisciplinarity in higher education. In G. Hampson & M. Rich-Tolsma (Eds.). *Leading Transformative Higher Education*. Olomouc: Palacky University Press. pp. 58-81.

Meiring, R. 2017. A Case Study of Women's Households, Sanitation and Care in Zwelitsha, An Informal Settlement Section in Stellenbosch Municipality. Doctoral Dissertation. Stellenbosch: Stellenbosch University.

Melgarejo, J., Prats, D., Molina, A. & Trapote, A. 2016. A case study of urban wastewater reclamation in Spain: Comparison of water quality produced by using alternative processes and related costs. *Journal of Water Reuse and Desalination*, 2016(6):72-81.

Melgarejo-Moreno, J. 2019. Agua y economía circular: Ponencia Marco. In *Proceedings of the Congreso Nacional del Agua 2019: Innovación y Sostenibilidad*, Alicante, Spain, 7-9 May. pp. 27-52.

Mendoza-Espinosa, L., Orozco-Borbón, M.V. & Silva-Nava, P. 2004. Quality assessment of reclaimed water for its possible use for crop irrigation and aquifer recharge in Ensenada, Baja California, Mexico. *Water Science and Technology*, 50(2):285-291.

Merriam, S.B. 1998. *Qualitative Research and Case Study Applications in Education*. San Francisco: Jossey-Bass.

Messerli, B. & Messerli, P. 2008. From local projects in the Alps to global change programmes in the mountains of the world: Milestones in transdisciplinary research. In G.H. Hadorn, H. Hoffman-Riem, S. Biber-Klemm, W. Grossenbacher-Mansuy, D. Joye, C. Phol, U. Wiesmann & E. Zemp (Eds.). *Handbook of Transdisciplinary Research*. Dordrecht: Springer. pp. 43-62.

Meyer, C.B. 2001. A case in case study methodology. *Field Methods*, 13(4):329-352.

Mexico Now. 2013. *Mexico's National Development Plan 2013-2018*. Available at: https://mexico-now.com/mexico-s-national-development-plan-2013-2018/.

Michalski, R.S., Carbonell, J.G. & Mitchell, T.M. (Eds.). 2013. *Machine Learning: An Artificial Intelligence Approach*. Cham: Springer Science & Business Media.

Ministry of Water Resources and Irrigation. 2005. *Water for the Future: National Water Resources Plan for Egypt – 2017*. Available at: https://faolex.fao.org/docs/pdf/egy147082.pdf.

Ministry of Water Resources and Irrigation. 2014. *Water Scarcity in Egypt: The Urgent Need for Regional Cooperation Among the Nile Basin Countries*. Technical Report. Available at: http://www.mfa.gov.eg/SiteCollection Documents/Egypt%20Water%20Resources%20Paper_2014.pdf.

Mitchell, T.M. 1997. *Machine Learning*. 7th edition. New York: McGraw-Hill.

Mittelstraß, J. 1992. Auf dem wege zur transdisziplinarität. *GAIA*, 1(5):250.

Mobjörk, M. 2010. Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. *Futures*, 42(8):866-873.

Mohammed, J.R. & Ibrahim, H.M. 2012. Hybrid wavelet artificial neural network model for municipal water demand forecasting. *ARPN Journal of Engineering and Applied Sciences*, 7(8):1047-1065.

Molinos-Senante, M., Hernández-Sancho, F., Mocholí-Aarce, M. & Sala-Garrido, R. 2014. A management and optimisation model for water supply planning in water deficit areas. *Journal of Hydrology*, 515:139-146.

Mombeni, H.A., Rezaei, S., Nadarajah, S. & Emami, M. 2013. Estimation of water demand in Iran based on SARIMA models. *Environmental Modeling & Assessment*, 18(5):559-565.

Moran-Ellis, J., Alexander, V.D., Cronin, A., Dickinson, M., Fielding, J., Sleney, J. & Thomas, H. 2006. Triangulation and integration: Processes, claims and implications. *Qualitative Research*, 6(1):45-59.

Mouatadid, S. & Adamowski, J. 2017. Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal*, 14(6):630-638.

Mount, J. & Hanak, E. 2019. *Water Use in California*. California: Public Policy Institute of California.

Msiza, I.S., Nelwamondo, F.V. & Marwala, T. 2008. Artificial neural networks and support vector machines for water demand time series forecasting. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pp. 638-643. Available at: https://arxiv.org/ftp/arxiv/papers/0705/0705.0969.pdf.

Mu, L., Zheng, F., Tao, R., Zhang, Q. & Kapelan, Z. 2020. Hourly and daily urban water demand predictions using a long short-term memory-based model. *Journal of Water Resources Planning and Management*, 146(9):05020017.

Mukhtarov, F.G. 2008. Intellectual history and current status of integrated water resources management: A global perspective. In C. Pahl-Wostl, P. Kabat & J. Möltgen (Eds.). *Adaptive and Integrated Water Management*. Berlin: Springer. pp. 167-185.

Munroe, R., Crawford, T. & Curtis, S. 2014. Geospatial analysis of space-time patterning of ENSO forced daily precipitation distributions in the Gulf of Mexico. *The Professional Geographer*, 66(1):91-101.

Nasr, F., El-Shafai, S.A. & Abdelfadil, A.S. 2022. Decentralized domestic wastewater management as unconventional water resource for agricultural purposes. *Egyptian Journal of Chemistry*, 65(5):1-2.

Nasteski, V. 2017. An overview of the supervised machine learning methods. *Horizons*, 4:51-62.

National Institute of Statistics and Geography (INEGI). 2013. *Datos Estadísticos.* Aguascalientes: INEGI.

National Planning Commission (NPC). 2011. *National Development Plan 2030*. Pretoria: NPC.

National Water Research Institute (NWRI). 2013. *Examining the Criteria for Direct Potable Reuse: Recommendations of an NWRI Independent Advisory Panel, Project 11-02*. Fountain Valley: NWRI.

Navarro-Caballero, M.T. 2018. Water reuse and desalination in Spain: Challenges and opportunities. *Journal of Water Reuse and Desalination*, 2018(8):153-168.

Nicolescu, B. 1997. *The Transdisciplinary Evolution of the University Condition for Sustainable Development*. Paper Presented at the International Congress of the International Association of Universities, Bangkok, Thailand, Chulalongkorn University. Available at: https://ciret-transdisciplinarity.org/bulletin/b12c8.php.

Nicolescu, B. 2002. *Manifesto of Transdisciplinarity* (Karen-Claire Voss, Trans.). Albany: State University of New York Press.

Nicolescu, B. 2004. Gurdjieff's philosophy of nature. In J. Needleman & G. Baker (Eds.). *Gurdjieff*. New York: The Continuum International Publishing Group. pp. 37-69.

Nicolescu, B. 2005. *Towards Transdisciplinary Education and Learning*. Paper Presented at the Science and Religion: Global Perspectives Metanexus Institute Conference. Available at: http://www.metanexus.net/archive/conference2005/pdf/nicolescu.pdf.

Nicolescu, B. 2006. Transdisciplinarity – Past, present and future. In B. Haverkott & C. Reijntjes (Eds.). *Moving Worldviews Conference Proceedings*. Leusden: ETC/Compas. pp. 142-165. Available at: http://www.movingworldviews.net/Downloads/Papers/Nicolescu.pdf.

Nicolescu, B. 2007. Transdisciplinarity as methodological framework for going beyond the science-religion debate. *The Global Spiral*, 8(3). Available at: https://www.fernandosantiago.com.br/transdisx.pdf.

Nicolescu, B. (Ed.). 2008. *Transdisciplinarity: Theory and Practice*. Creskill: Hampton Press.

Nicolescu, B. 2010. Methodology of transdisciplinarity: Levels of reality, logic of the included middle and complexity. *Transdisciplinary Journal of Engineering & Science*, 1:17-32.

Niquice, C., Marques Arsenio, A., Brito, R.M.C.L. & Van Lier, J.B. 2020. Use of (partially) treated municipal wastewater in irrigated agriculture: Potentials and constraints for sub-Saharan Africa. *Physics and Chemistry of the Earth*, 118-119. Available at: https://pure.tudelft.nl/ws/portalfiles/portal/83627561/1_s2.0_S1474706519301056_main.pdf.

Norman, D.A. & Draper, S.W. (Eds.). 1986. *User-Centred System Design: New Perspectives on Human-Computer Interaction*. New Jersey: Lawrence Erlbaum Associates.

Ntombela, C., Funke, N., Meissner, R., Steyn, M. & Masangane, W. 2016. A critical look at South Africa’s Green Drop programme. *Water SA*, 42(4):703-710.

Null, S.E. & Prudencio, L. 2016. Climate change effects on water allocations with season dependent water rights. *Science of the Total Environment*, 571(2016):943-954.

Oertlé, E., Mueller, S.R., Choukr-Allah, R. & Jaouani, A. 2020. Decision support tool for water reclamation beyond technical considerations – Egyptian, Moroccan, and Tunisian case studies. *Integrated Environmental Assessment and Management*, 16(6):885-897.

Oliveira, P.J., Steffen, J.L. & Cheung, P. 2017. Parameter estimation of seasonal ARIMA models for water demand forecasting using the Harmony Search Algorithm. *Procedia Engineering*, 186:177-185.

Olivieri, A.W., Pecson, B., Crook, J. & Hultquist, R. 2020. California water reuse – Past, present and future perspectives. *Advances in Chemical Pollution, Environmental Management and Protection*, 5:65-111.

O’Neill, A. 2022. *Egypt: Total Population From 2017 to 2027 (In Million Inhabitants)*. Available at: https://www.statista.com/statistics/377302/total-population-of-egypt/.

Opher, T., Friedler, E. & Shapira, A. 2019. Comparative life cycle sustainability assessment of urban water reuse at various centralization scales. *The International Journal of Life Cycle Assessment*, 24(7):1319-1332.

Organization for Economic Co-operation and Development. 2013. *Making Water Reform Happen in Mexico*. Available at: http://www.oecd-ilibrary.org/fr/governance/making-water-reform-happen-in-mexico_9789264187894-en.

Orlando, L. 2015. Fighting water wars: Regional environmental cooperation as a roadmap for peace. *Fletcher Forum of World Affairs*, 39(2):101-110.

Osman, A.I.A., Ahmed, A.N., Chow, M.F., Huang, Y.F. & El-Shafie, A. 2021. Extreme gradient boosting (XGBoost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 12(2):1545-1556.

Ostertagová, E. 2012. Modelling using polynomial regression. *Procedia Engineering*, 48:500-506.

Oyebode, O. & Ighravwe, D.E. 2019. Urban water demand forecasting: A comparative evaluation of conventional and soft computing techniques. *Resources*, 8(3):156.

Oyebode, O.K., Otieno, F.A.O. & Adeyemo, J. 2014. Review of three data-driven modelling techniques for hydrological modelling and forecasting. *Fresenius Environmental Bulletin*, 23(7):1443-1454.

Pacchin, E., Gagliardi, F., Alvisi, S. & Franchini, M. 2019. A comparison of short-term water demand forecasting models. *Water Resources Management*, 33(4):1481-1497.

Pahl-Wostl, C. 2002. Towards sustainability in the water sector – The importance of human actors and processes of social learning. *Aquatic Sciences*, 64(4):394-411.

Pahl-Wostl, C. 2007. Transitions towards adaptive management of water facing climate and global change. *Water Resources Management*, 21(1):49-62.

Pahl-Wostl, C., Sendzimir, J., Jeffrey, P., Aerts, J., Berkamp, G. & Cross, K. 2007. Managing change toward adaptive water management through social learning. *Ecology and Society*, 12(2):30.

Pandey, P., Bokde, N.D., Dongre, S. & Gupta, R. 2021. Hybrid models for water demand forecasting. *Journal of Water Resources Planning and Management*, 147(2):04020106.

Paneque, P. 2015. Drought management strategies in Spain. *Water*, 7(12):6689-6701.

Passi, S. & Barocas, S. 2019. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 39-48. Available at: https://dl.acm.org/doi/10.1145/3287560.3287567.

Patton, M.Q. 1990. *Qualitative Evaluation and Research Methods*. Thousand Oaks: Sage Publications.

Peasey, A., Blumenthal, U., Mara, D. & Ruiz-Palacios, G. 2000. *A Review of Policy and Standards for Wastewater Reuse in Agriculture: A Latin American Perspective*. Available at: https://www.ehproject.org/well/resources/well-studies/full-reports-pdf/task0068ii.pdf.

Pelaccia, T., Forestier, G. & Wemmert, C. 2019. Deconstructing the diagnostic reasoning of human versus artificial intelligence. *CMAJ*, 191(48):E1332-E1335.

Pesantez, J.E., Berglund, E.Z. & Kaza, N. 2020. Smart meters data for modeling and forecasting water demand at the user-level. *Environmental Modelling & Software*, 125:104633.

Piaget, J. 1972. The epistemology of interdisciplinary relationships. In Centre Pour La Recherche & l'Innovation dans l'Enseignement (Eds.). *L'interdisciplinarité – Problèmes D'enseignement et de Recherche dans les Universités.* Paris: OECD. pp. 127-139.

Plecher, H. 2020a. *Distribution of the Workforce Across Economic Sectors in Mexico 2020*. Available at: https://www.statista.com/statistics/275428/distribution-of-the-workforce-across-economic-sectors-in-mexico/.

Plecher, H. 2020b. *Distribution of Gross Domestic Product (GDP) Across Economic Sectors in Mexico 2019*. Available at: https://www.statista.com/statistics/275420/distribution-of-gross-domestic-product-gdp-across-economic-sectors-in-mexico/.

Plecher, H. 2020c. *Distribution of Gross Domestic Product (GDP) Across Economic Sectors Egypt 2019*. Available at: https://www.statista.com/statistics/377309/egypt-gdp-distribution-across-economic-sectors/.

Plecher, H. 2020d. *Employment by Economic Sector in Egypt 2020*. Available at: https://www.statista.com/statistics/377950/employment-by-economic-sector-in-egypt/.

Pohl, C. & Hardon, G.H. 2007. *Principles of Designing Transdisciplinary Research Proposed by the Swiss Academies of Arts and Sciences*. Munich: Oekom Verlag.

Pohl, C. & Hadorn, G.H. 2008. Core terms in transdisciplinary research. In G.H. Hadorn, H. Hoffman-Riem, S. Biber-Klemm, W. Grossenbacher-Mansuy, D. Joye, C. Phol, U. Wiesmann & E. Zemp (Eds.). *Handbook of Transdisciplinary Research*. Dordrecht: Springer. pp. 427-432.

Pradhan, A. 2012. Support vector machine: A survey. *International Journal of Emerging Technology and Advanced Engineering*, 2(8):82-85.

Pradhan, S. 2017. Water war thesis: A myth or a reality? *International Journal of Arts, Humanities and Social Science*, 2(1):12-15.

Pu, G., Wang, L., Shen, J. & Dong, F. 2020. A hybrid unsupervised clustering-based anomaly detection method. *Tsinghua Science and Technology*, 26(2):146-153.

Pulido-Calvo, I., Montesinos, P., Roldán, J. & Ruiz-Navarro, F. 2007. Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems. *Biosystems Engineering*, 97(2):283-293.

Pulido-Calvo, I., Roldán, J., López-Luque, R. & Gutiérrez-Estrada, J.C. 2003. Demand forecasting for irrigation water distribution systems. *Journal of Irrigation and Drainage Engineering*, 129(6):422-431.

Qadir, M., Wichelns, D., Raschid-Sally, L., McCornick, P.G., Drechsel, P., Bahri, A. & Minhas, P.S. 2010. The challenges of wastewater irrigation in developing countries. *Agricultural Water Management*, 97:561-568.

Radingoana, M.P., Dube, T. & Mazvimavi, D. 2020. Progress in greywater reuse for home gardening: Opportunities, perceptions and challenges. *Physics and Chemistry of the Earth, Parts A/B/C*, 116:102853.

Raj, K.S. & Kumar, P. 2022. To analyse the impact of water scarcity in developing countries using machine learning. In V. Bhateja, J. Tang, S.C. Satapathy, P.

Peer & R. Das (Eds.). *Evolution in Computational Intelligence*. Singapore: Springer. pp. 53-63.

Ramulongo, L., Nethengwe, N.S. & Musyoki, A. 2017. The nature of urban household water demand and consumption in Makhado Local Municipality: A case study of Makhado Newtown. *Procedia Environmental Sciences*, 37(2017):182-194.

Rassoul, E.M.A. 2006. Prospects of water reuse in Egypt. In *Proceedings of the 10th International Water Technology Conference (IWTC10 2006)*, Alexandria, Egypt, pp. 561-567.

Ray, S. 2019. A quick review of machine learning algorithms. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 35-39. Available at: https://translateyar.ir/wp-content/uploads/2021/12/A-Quick-Review-of-Machine.pdf.

Republic of South Africa (RSA). 1912. *Water Act (Act No. 8 of 1912)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 1956. *Water Act (Act No. 54 of 1956)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 1996. *The Constitution of the Republic of South Africa (Act No. 108 of 1996)*. Pretoria Government Printers.

Republic of South Africa (RSA). 1997. *Water Services Act (Act No. 108 of 1997)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 1998a. *National Water Act (Act No. 36 of 1998).* Pretoria: Government Printers.

Republic of South Africa (RSA). 1998b. *National Environmental Act (Act No. 107 of 1998)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 1998c. *Municipal Structures Act (Act No. 117 of 1998)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 2000. *Municipal Systems Act (Act No. 32 of 2000).* Pretoria: Government Printers.

Republic of South Africa (RSA). 2003. *Municipal Finance Management Act (Act No. 56 of 2003)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 2004. *Water Services Amendment Act (Act No. 30 of 2004)*. Pretoria: Government Printers.

Republic of South Africa (RSA). 2008. *National Environmental Management: Waste Act (Act No. 59 of 2008)*. Pretoria: Government Printers.

Reyers, B., Roux, D.J., Cowling, R.M., Ginsburg, A.E., Nel, J.L. & Farrell, P.O. 2010. Conservation planning as a transdisciplinary process. *Conservation Biology*, 24(4):957-965.

Rezaali, M., Quilty, J. & Karimi, A. 2021. Probabilistic urban water demand forecasting using wavelet-based machine learning models. *Journal of Hydrology*, 600:126358.

Ricart, S. & Rico, A.M. 2019. Assessing technical and social driving factors of water reuse in agriculture: A review on risks, regulation and the yuck factor. *Agricultural Water Management*, 217:426-439.

Rinaudo, J.D. 2015. Long-term water demand forecasting. In Q. Grafton, K.A. Daniell, C. Nauges, J.D. Rinaudo & N. Wai Wah Cha (Eds.). *Understanding and Managing Urban Water in Transition*. Cham: Springer. pp. 239-268.

Ritter, W. 2021. State regulations and guidelines for wastewater reuse for irrigation in the US. *Water*, 13(20):2818.

Safavi, H.R., Golmohammadi, M.H. & Sandoval-Solis, S. 2015. Expert knowledge-based modelling for integrated water resources planning and management in the Zayandehrud River Basin. *Journal of Hydrology*, 528:773-789.

Saleth, R.M. & Dinar, A. 2004. *The Institutional Economics of Water: A Cross-Country Analysis of Institutions and Performance*. London: Edward Elgar Publishing.

Salloom, T., Kaynak, O., Yu, X. & He, W. 2022. Proportional integral derivative booster for neural networks-based time-series prediction: Case of water demand prediction. *Engineering Applications of Artificial Intelligence*, 108:104570.

Samuel, A.L. 1959. Some studies in machine learning using the game of checkers. *IBM Journal of R&D*, 3(3):210-229.

Santillán-Fernández, A., García-Chávez, L., Vásquez-Bautista, N., Santoyo Cortés, V.H., Melgar Morales, M., Pereira, W., Aguilar, J.E.L. & García, A.M. 2018.

Impact of the Substitution of Cane Sugar for High Intensity Sweeteners in Mexico. Available at: https://ciestaam.edu.mx/publicaciones2018/libros/edulcorantes.pdf.

Santillán-Fernández, A., Salinas-Moreno, Y., Valdez-Lazalde, J.R., Carmona-Arellano, M.A., Vera-López, J.E. & Pereira-Lorenzo, S. 2021. Relationship between maize seed productivity in Mexico between 1983 and 2018 with the adoption of genetically modified maize and the resilience of local races. *Agriculture*, 11(8):737.

Saravanan, V.S., McDonald, G.T. & Mollinga, P.P. 2009. Critical review of integrated water resources management: Moving beyond polarised discourse. *Natural Resources Forum*, 33(1):76-86.

Schapire, R.E. 2013. Explaining AdaBoost. In Z. Luo, B. Schölkopf, V. Vovk & V.N. Vapnik (Eds.). *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. Berlin: Springer. pp. 37-52.

Schleich, J. & Hillenbrand, T. 2009. Determinants of residential water demand in Germany. *Ecological Economics*, 68(6):1756-1769.

Schmidt, R., Lyytinen, K., Keil, M. & Cule, P. 2001. Identifying software project risks: An international Delphi study. *Journal of Management Information Systems*, 17(4):5-36.

Scholz, R.W., Lang, D.J., Wiek, A., Walter, A.I. & Stauffacher, M. 2006. Transdisciplinary case studies as a means of sustainability learning: Historical framework and theory. *International Journal of Sustainability in Higher Education*, 7(3):226-251.

Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., Von Gunten, U. & Wehrli, B. 2006. The challenge of micropollutants in aquatic systems. *Science*, 313(5790):1072-1077.

Scoullos, M. 2012. Transboundary IWRM attempts in the Mediterranean emphasis on the Drin River case and the involvement of stakeholders. In R. Choukr-Allah, R. Ragab & R. Rodriquez-Clemente (Eds.). *Integrated Water Resources Management in the Mediterranean Region: Dialogue Towards New Strategy.* Dordrecht: Springer. pp. 3-23.

Seekings, J. 2008. Deserving individuals and groups: The post-apartheid state's justification of the shape of South Africa's system of social assistance. *Transformation: Critical Perspectives on Southern Africa*, 68(1):28-52.

Seeliger, L. & Turok, I. 2014. Averting a downward spiral: Building resilience in informal urban settlements through adaptive governance. *Environment and Urbanization*, 26(1):184-199.

Seguí, A. 2004. Sistemas de Regeneración y Reutilización de Aguas Residuales: Metodología Para el Análisis Técnico-Económico y Casos. Doctoral Dissertation. Barcelona: Universidad Politécnica de Cataluña Espana.

Senge, P.M. 1997. The fifth discipline. *Measuring Business Excellence*, 1(3):46-51.

Shabani, S., Yousefi, P., Adamowski, J., Naser, G. & Rahman, I.M.M. 2016. Intelligent soft computing models in water demand forecasting. In I.M.M. Rahman, Z.A. Begum & H. Hasegawa (Eds.). *Water Stress in Plants*. InTech Open. pp. 99-117.

Shabani, S., Yousefi, P. & Naser, G. 2017. Support vector machines in urban water demand forecasting using phase space reconstruction. *Procedia Engineering*, 186:537-543.

Shanthamallu, U.S., Spanias, A., Tepedelenlioglu, C. & Stanley, M. 2017. A brief survey of machine learning methods and their sensor and IoT applications. In *8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pp. 1-8. Available at: https://sensip.engineering.asu.edu/wp-content/uploads/2020/03/1_Machine-Learning-Survey-Paper.pdf.

Shao, W., Yang, D., Hu, H. & Sanbongi, K. 2009. Water resources allocation considering the water use flexible limit to water shortage: A case study in the Yellow River Basin of China. *Water Resources Management*, 23(5):869-880.

Shoushtarian, F. & Negahban-Azar, M. 2020. Worldwide regulations and guidelines for agricultural water reuse: A critical review. *Water*, 12(4):971.

Shrestha, M., Manandhar, S. & Shrestha, S. 2020. Forecasting water demand under climate change using artificial neural network: A case study of Kathmandu Valley, Nepal. *Water Supply*, 20(5):1823-1833.

Sithole, S.L. & Mathonsi, N.S. 2015. Local governance service delivery issues during apartheid and post-apartheid South Africa. *Africa's Public Service Delivery & Performance Review*, 3(3):5-30.

Slobodiuk, S., Niven, C., Arthur, G., Thakur, S. & Ercumen, A. 2021. Does irrigation with treated and untreated wastewater increase antimicrobial resistance in soil and water: A systematic review. *International Journal of Environmental Research and Public Health*, 18(21):11046.

Smolak, K., Kasieczka, B., Fialkiewicz, W., Rohm, W., Siła-Nowicka, K. & Kopańczyk, K. 2020. Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. *Urban Water Journal*, 17(1):32-42.

Song, P., Wang, C., Zhang, W., Liu, W., Sun, J., Wang, X., Lei, X. & Wang, H. 2020. Urban multi-source water supply in China: Variation tendency, modeling methods and challenges. *Water*, 12(4):1199.

Stake, R.E. 1995. *The Art of Case Study Research*. Thousand Oaks: Sage Publications.

Stellenbosch Municipality. 2010. *Stellenbosch Municipality Integrated Development Plan, 2nd Generation – Revision 3*. Stellenbosch: Stellenbosch Municipality.

Stellenbosch Municipality. 2012. *Annual Report: Compiled in Terms of Section 121 of the Municipal Finance Management Act (Act 56 of 2003), 01 July 2010 – 30 June 2011*. Available at: https://www.westerncape.gov.za/sites/www.westerncape.gov.za/files/documents/2012/11/stellenbosch-annual_report_final_march_2012.pdf.

Stellenbosch Municipality. 2017. *Annual Report 2015/16*. Available at: https://stellenbosch.gov.za/download/final-annual-report-2015-16/?ind=1627384863886&filename=Final-Annual-Report-2015-16.pdf&wpdmdl=14270&refresh=6437d2b6875be1681380022.

Stellenbosch Municipality. 2018a. *Stellenbosch Municipality Medium Term Revenue and Expenditure Framework 2018/2019 – 2020/2021*. Available at: https://stellenbosch.gov.za/download/medium-term-revenue-and-expenditure-framework-2018_2019-2020_2021-decision/.

Stellenbosch Municipality. 2018b. *Tabling the Final Budget to Council for Final Approval – 28 May 2018: Budget of Opportunities: For Better Living and Economic Opportunities and Job Creation*. Available at: https://stellenbosch.gov.za/download/budget-speech-by-the-executive-mayor-2018-2019/?wpdmdl=10503&refresh=6437d3fe8b0451681380350.

Stellenbosch Municipality. 2011. *Water Services Development Plan for 2011/2012*. Available at: https://silo.tips/download/stellenbosch-municipality.

Stijnen, T. & Mulder, P.G.H. 1999. *Classical Methods for Data Analyses*. Rotterdam: NIHES Program.

Suh, D., Kim, H. & Kim, J. 2015. Estimation of water demand in residential building using machine learning approach. In *IT Convergence and Security (ICITCS), 5th International Conference*. Kuala Lumpur, Malaysia: IEEE. pp. 1-2. Available at: https://www.semanticscholar.org/paper/Estimation-of-Water-Demand-in-Residential-Building-Suh-Kim/f0cb6f0d9ae4c56e9b21fdb55fffd5d8a2405ffd.

Sundui, B., Calderon, O.A.R., Abdeldayem, O.M., Lázaro-Gil, J., Rene, E.R. & Sambuu, U. 2021. Applications of machine learning algorithms for biological wastewater treatment: Updates and perspectives. *Clean Technologies and Environmental Policy*, 2021:1-17.

Stellenbosch Municipality. 2012. *Spatial Development Framework*. Available at: http://www.stellenboschheritage.co.za/wp-content/uploads/Stellenbosch_Municipality_SDF_Part_1-1.pdf.

Swatuk, L.A. 2005. Political challenges to implementing IWRM in Southern Africa. *Physics and Chemistry of the Earth*, 30(11-16):872-880.

Swatuk, L.A. 2010. The state and water resources development through the lens of history: A South African case study. *Water Alternatives*, 3(3):521-536.

Tamminen, K.A. & Poucher, Z.A. 2020. Research philosophies. In D. Hackfort & R. Schinke (Eds.). *The Routledge International Encyclopedia of Sport and Exercise Psychology*. London: Routledge. pp. 535-549.

Tempelhoff, J. 2004. Rand water and the transition to a multiracial democratic South Africa 1989–94. *African Historical Review*, 36(1):79-106.

Tempelhoff, J. 2017. The Water Act, No. 54 of 1956 and the first phase of apartheid in South Africa (1948–1960). *Water History*, 9(2):189-213.

Textor, C. 2020a. *Distribution of GDP Across Economic Sectors in China 2009-2019*. Available at: https://www.statista.com/statistics/270325/distribution-of-gross-domestic-product-gdp-across-economic-sectors-in-china/.

Textor, C. 2020b. *Distribution of the Workforce Across Economic Sectors in China 2019*. Available at: https://www.statista.com/statistics/270327/distribution-of-the-workforce-across-economic-sectors-in-china/#:~:text=In%202019%2C%20around%2025.1%20percent,percent%20in%20the%20service%20sector.&text=In%202012%.

The Global Economy.com. 2019. *Spain: GDP Share of Agriculture*. Available at: https://www.theglobaleconomy.com/Spain/share_of_agriculture/.

Thompson, H. 2006. *Water Law: A Practical Approach to Resource Management and the Provision of Services*. Cape Town: Juta.

Thompson, H., Stimie, C., Richters, E. & Perret, S. 2001. *Policies, Legislation, and Organizations Related to Water in South Africa With Special Reference to the Olifants River Basin*. Working Paper 18. Sri Lanka: International Water Management Institute.

Tian, D., Martinez, C.J. & Asefa, T. 2016. Improving short-term urban water demand forecasts with reforecast analog ensembles. *Journal of Water Resources Planning and Management*, 142(6):04016008.

Tiwari, M.K. & Adamowski, J. 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resources Research*, 49(10):6486-6507.

Tiwari, M.K. & Adamowski, J.F. 2015. Medium-term urban water demand forecasting with limited data using an ensemble wavelet-bootstrap machine-learning approach. *Journal of Water Resources Planning and Management*, 141(2):04014053.

Tiwari, M.K. & Adamowski, J.F. 2017. An ensemble wavelet bootstrap machine learning approach to water demand forecasting: A case study in the city of Calgary, Canada. *Urban Water Journal*, 14(2):185-201.

Tong, S. & Koller, D. 2001. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2(Nov.):45-66.

Trading Economics. 2020. *Spain – Employment in Agriculture (% of Total Employment)*. Available at: https://tradingeconomics.com/spain/employment-in-agriculture-percent-of-total-employment-wb-data.html.

Tsegaye, S., Eckart, J. & Vairavamoorthy, K. 2012. Urban water management in cities of the future – Emerging areas in developing countries. In J. Lundqvist (Ed.). *On the Waterfront Selections from the 2011 World Water Week in Stockholm*. Stockholm: SIWI. pp. 42-48.

Turner, J.L. 2006. New ripples and responses to China's water woes. *China Brief*, 6(25). Available at: https://jamestown.org/program/new-ripples-and-responses-to-chinas-water-woes-3/.

TYPSA Consulting Engineers & Architects. 2013. *Updated Report on Wastewater Reuse in the European Union*. Available at: https://ec.europa.eu/environment/water/blueprint/pdf/Final%20Report_Water%20Reuse_April%202013.pdf.

Tzanakakis, V., Koo-Oshima, S., Haddad, M., Apostolidis, N. & Angelakis, A. 2014. The history of land application and hydroponic systems for wastewater treatment and reuse. In A.N. Angelakis & J.B. Rose (Eds). *Evolution of Sanitation and Wastewater Technologies Through the Centuries*. London: IWA Publishing. pp. 457-479.

Tzanakakis, V., Paranychianaki, N. & Angelakis, A. 2007. Soil as a wastewater treatment system: Historical development. *Water Science and Technology: Water Supplement,* 2007(7):67-75.

Udimal, T.B., Jincai, Z., Ayamba, E.C. & Owusu, S.M. 2017. China's water situation: The supply of water and the pattern of its usage. *International Journal of Sustainable Built Environment*, 6(2):491-500.

Ungureanu, N., Vlăduț, V. & Voicu, G. 2020. Water scarcity and wastewater reuse in crop irrigation. *Sustainability*, 12(21):9055.

United Nations Environment Programme (UNEP). 1991. *Council Directive 91/271/EEC Concerning Urban Waste Water Treatment*. Available at:

https://leap.unep.org/countries/eu/national-legislation/council-directive-91271eec-concerning-urban-waste-water-treatment.

United Nations Water (UN Water). 2015. *UN World Water Development Report 2015*. Available at: https://www.unwater.org/publications/un-world-water-development-report-2015.

United Nations Water (UN Water). 2017. *UN World Water Development Report 2017*. Available at: https://www.unwater.org/publications/world-water-development-report-2017/.

Valdes Ramos, A., Aguilera Gonzalez, E.N., Tobón Echeverri, G., Samaniego Moreno, L., Díaz Jiménez, L. & Carlos Hernández, S. 2019. Potential uses of treated municipal wastewater in a semiarid region of Mexico. *Sustainability*, 11(8):2217.

VanBerlo, B., Ross, M.A. & Hsia, D. 2021. Univariate long-term municipal water demand forecasting. *arXiv* preprint arXiv:2105.08486. Available at: https://arxiv.org/abs/2105.08486.

Van Niekerk, A.M. & Schneider, B. 2013. *Implementation Plan for Direct and Indirect Water Reuse for Domestic Purposes: Sector Discussion Document*. Research Report No. KV 320/13. Pretoria: Water Research Commission.

Vijai, P. & Sivakumar, P.B. 2018. Performance comparison of techniques for water demand forecasting. *Procedia Computer Science*, 143:258-266.

Viljoen, G. & Van der Walt, K. 2018. South Africa's water crisis-an interdisciplinary approach. *Tydskrif vir Geesteswetenskappe*, 58(3):483-500.

Villarin, M.C. & Rodriguez-Galiano, V.F. 2019. Machine learning for modeling water demand. *Journal of Water Resources Planning and Management*, 145(5):04019017.

Vozhehova, R.A., Lykhovyd, P.V., Lavrenko, S.O., Kokovikhin, S.V., Lavrenko, N.M., Marchenko, T.Y., Sydyakina, O.V., Hlushko, T.V. & Nesterchuk, V.V. 2019. Artificial neural network use for sweet corn water consumption prediction depending on cultivation technology peculiarities. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 10(1):354-358.

Walker, G. 2014. Water scarcity in England and Wales as a failure of (meta) governance. *Water Alternatives*, 7(2):388-413.

Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B., Janssen, P. & Krayer Von Krauss, M.P. 2003. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5-17.

Warfield, J.N. & Cárdenas, A.R. 1994. *A Handbook of Interactive Management.* Ames: Iowa State University Press.

Warner, J., Wester, P. & Bolding, A. 2008. Going with the flow: River basins as the natural units for water management? *Water Policy*, 10(S2):121-138.

Water Quality Australia. 2006. *Australian Guidelines for Water Recycling*. Available at: https://www.waterquality.gov.au/guidelines/recycled-water#managing-health-and-environmental-risks-phase-1.

Weaver, M.J.T., Hamer, N., O'Keeffe, J. & Palmer, C.G. 2017. Water service delivery challenges in a small South African municipality: Identifying and exploring key elements and relationships in a complex social-ecological system. *Water SA*, 43(3):398-408.

Western Cape Government. 2017. *Stellenbosch Local Municipality: Overview*. Available at: https://www.westerncape.gov.za/your_gov/8.

Westley, F., Olsson, P., Folke, C., Homer-Dixon, T., Vredenburg, H., Loorbach, D., Thompson, J., Nilsson, M., Lambin, E., Sendzimir, J. & Banerjee, B. 2011. Tipping toward sustainability: Emerging pathways of transformation. *Ambio*, 40(7):762-780.

Wintgens, T., Bixio, D., Thoeye, C., Jeffrey, P., Hochstrat, R. & Melin, T. 2006. *Integrated Concepts for Reuse of Upgraded Wastewater*. Aachen: AQUAREC.

Wirth, R. & Hipp, J. 2000. CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Vol. 1*. London: Springer-Verlag. pp. 29-39. Available at: http://cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf.

World Bank. 2016. *Mainstreaming Water Resources Management in Urban Projects: Taking an Integrated Urban Water Management Approach – A Guidance Note*. Available at: https://documents1.worldbank.org/curated/fr/6337315218

18888260/pdf/Mainstreaming-water-resources-management-in-urban-projects-taking-an-integrated-urban-water-management-approach-a-guidance-note.pdf.

World Bank. 2020. *Employment in agriculture (% of Total Employment) (Modeled ILO Estimate)*. Available at: https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS.

World Health Organization (WHO). 1973. *Reuse of Effluents: Methods of Wastewater Treatment and Health Safeguards*. Available at: https://apps.who.int/iris/bitstream/handle/10665/41032/WHO_TRS_517.pdf?sequence=1&isAllowed=y.

World Health Organization (WHO). 1989. *Health Guidelines for the Use of Wastewater in Agriculture and Aquaculture: Report of a WHO Scientific Group*. Technical Report Series No. 778. Available at: https://apps.who.int/iris/bitstream/handle/10665/39401/WHO_TRS_778.pdf?sequence=1&isAllowed=y.

World Health Organization (WHO). 2006. *WHO Guidelines for the Safe Use of Wastewater Excreta and Greywater, Vol. 1: Policy and Regulatory Aspects*. Available at: https://apps.who.int/iris/bitstream/handle/10665/78265/9241546824_eng.pdf?sequence=1.

Worthington, E.B. 1977. United Nations water conference, held in Mar del Plata, Argentina, 14-25 March 1977. *Environmental Conservation*, 4(2):153-154.

Xenochristou, M. & Kapelan, Z. 2020. An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water Journal*, 17(3):212-223.

Xie, M. 2006. *Integrated Water Resources Management (IWRM) – Introduction to Principles and Practices*. Presentation at the Africa Regional Workshop on IWRM, Nairobi, 29 October – 2 November. Available at: https://iwlearn.net/resolveuid/aef0e14aa7070ede8a0f1af8c619c6fd.

Xu, Y., Zhang, J., Long, Z., Tang, H. & Zhang, X. 2019. Hourly urban water demand forecasting using the continuous deep belief echo state network. *Water*, 11(2):351.

Yadav, S.K., Singh, S. & Gupta, R. 2019. Univariate logistic regression: Theoretical aspects. In S.K. Yadav, S. Singh & R. Gupta. *Biomedical Statistics*. Singapore: Springer. pp. 219-222. Available at: https://doi.org/10.1007/978-981-32-9294-9_28.

Yan, K. & Yang, M.Z. 2018. Water demand forecast model of least squares support vector machine based on particle swarm optimization. *MATEC Web of Conferences*, 246:01029. Available at: https://www.matec-conferences.org/articles/matecconf/abs/2018/105/matecconf_iswso2018_01029/matecconf_iswso2018_01029.html.

Yassine, A., Mohamed, L. & Al Achhab, M. 2021. Intelligent recommender system based on unsupervised machine learning and demographic attributes. *Simulation Modelling Practice and Theory*, 107:102198.

Yi, L., Jiao, W., Chen, X. & Chen, W. 2011. An overview of reclaimed water reuse in China. *Journal of Environmental Sciences*, 23(10):1585-1593.

Yin, R.K. 1993. *Applications of Case Study Research*. Newbury Park: Sage Publications.

Yin, R.K. 1994. Discovering the future of the case study: Method in evaluation research. *Evaluation Practice*, 15(3):283-290.

Yin, R.K. 2003. *Case Study Research: Designs and Method*. 3rd edition. Thousand Oaks: Sage Publications.

Yin, R.K. 2009. *Case Study Research: Design and Methods – Essential Guide to Qualitative Methods in Organizational Research*. Singapore: Sage Publications.

Yue, L., Zhang, H. & Wang, L. 2007. Application of particle swarm optimization in urban water demand prediction. *Journal of Tianjin University Science and Technology*, 40(6):742-746.

Zhang, J., Fu, D., Urich, C. & Singh, R.P. 2018. Accelerated exploration for long-term urban water infrastructure planning through machine learning. *Sustainability*, 10(12):4600.

Zhang, Y., Deng, J., Qin, B., Zhu, G., Zhang, Y., Jeppesen, E. & Tong, Y. 2022. Importance and vulnerability of lakes and reservoirs supporting drinking water in China. *Fundamental Research*, 3(2):265-273.

Zhang, Y. & Haghani, A. 2015. A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58:308-324.

Zhou, S.L., McMahon, T.A., Walton, A. & Lewis, J. 2002. Forecasting operational demand for an urban water supply zone. *Journal of Hydrology*, 259(1-4):189-202.

Zhu, Z. & Dou, J. 2018. Current status of reclaimed water in China: an overview. *Journal of Water Reuse and Desalination*, 8(3):293-307.

Zubaidi, S.L., Dooley, J., Alkhaddar, R.M., Abdellatif, M., Al-Bugharbee, H. & Ortega-Martorell, S. 2018. A novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. *Journal of Hydrology*, 561:136-145.

Zubaidi, S.L., Gharghan, S.K., Dooley, J., Alkhaddar, R.M. & Abdellatif, M. 2018. Short-term urban water demand prediction considering weather factors. *Water Resources Management*, 32(14):4527-4542.

Zubaidi, S.L., Ortega-Martorell, S., Al-Bugharbee, H., Olier, I., Hashim, K.S., Gharghan, S.K., Kot, P. & Al-Khaddar, R. 2020. Urban water demand prediction for a city that suffers from climate change and population growth: Gauteng province case study. *Water*, 12(7):1885.

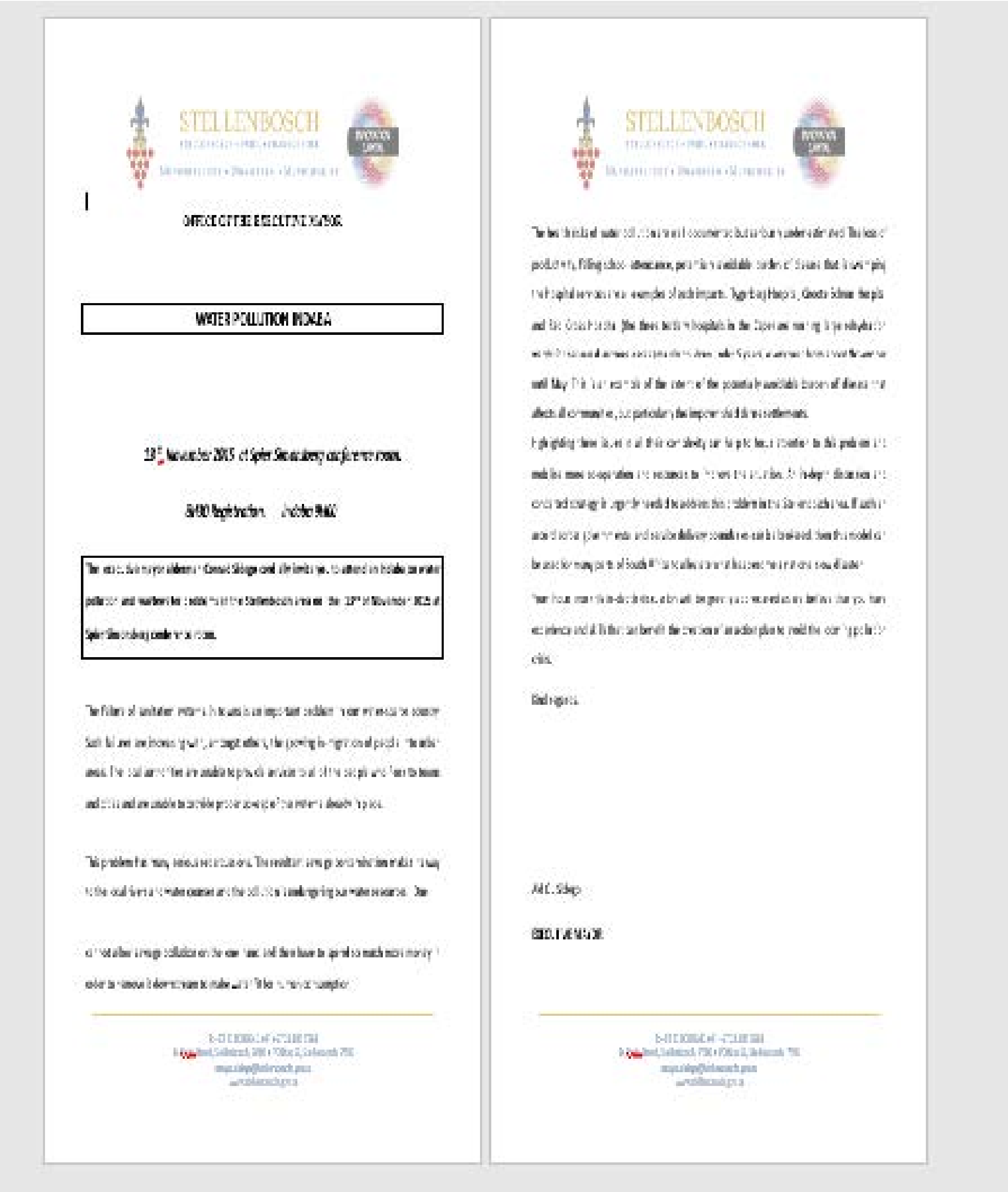
Zubaidi, S.L., Ortega-Martorell, S., Kot, P., Alkhaddar, R.M., Abdellatif, M., Gharghan, S.K., Ahmed, M.S. & Hashim, K. 2020. A method for predicting long-term municipal water demands under climate change. *Water Resources Management*, 34(3):1265-1279.

Žukauskas, P., Vveinhardt, J. & Andriukaitienė, R. 2018. *Management Culture and Corporate Social Responsibility*. London: InTech.

APPENDICES

Appendix A:
Water pollution indaba

A1. Invitation to the water pollution indaba



A2. Wastewater indaba November 2015 programme

08:00 – 09:00	Registration
09:00 – 09:15	Opening (Executive Mayor Cllr Conrad Sidego)
09:15 – 10:00	Keynote speaker (Prof. Eugene Cloete Rector Research & Innovation, Stellenbosch University)
10:00 – 10:30	Tea break
10:30 – 11:15	Dr Jo Barnes (senior lecturer at Stellenbosch University)
11:15 – 11:30	Quinton Brynard (CFO Stellenbosch Water Board)
11:30 – 11:45	Director of Engineering services (Stellenbosch Municipality)
11:45 – 12:30	Dr Jeremy R Biddle, technical director Bluewater Bio Ltd UK (Technology)
12:30 – 13:00	Reflections
13:00 – 13:10	Heidi Newton-King: Director sustainability and human resources Spier (closing remarks)
13:00 – 14:00	Lunch & Networking

A3. Guest list for the indaba

1. Anton Bredell (Min of Environment WC)

2. Deputy Mayor of Stellenbosch
3. Andricus van der Westhuizen (MP):
4. Quiton Brynard (Stellenbosch water board):
5. Jan Boland (Stellenbosch Water Board):
6. Angelica van der Merwe (CEO of Stellenbosch Farmers Association):
7. Prof. Eugene Cloete (Deputy Vice Chancellor of Stellenbosch University):
8. Diamond student representative chairperson. (SRC)
9. Lillies Ratshidi: STEM
10. Duncan Michelle:
11. Mbatha LS:
12. Moletsane (Stellenbosch University)
13. Barnes (Stellenbosch University):
14. André Pelser (chairperson Stellenbosch Ratepayers Association):
15. Danie Keet (Editor) Eikestad News:
16. Thatha Madiba: By hand
17. Mama Madiba: By hand
18. Dann Ngece: By hand
19. Wanana Maidas
20. Heide (Spier):

21. Cllr van der Walt:
22. Charon Marais
23. Farmer’s representative
24. Rejoice Malisa (organiser):
25. Justine Moore:
26. Carla Tenzer BGU University Israel
27. Stefan de Villiers:
28. Jeremy Biddle:
29. Nina Rivers:
30. Karabo Chigwiza:
31. Lindre:
32. Thumakele:
33. Derrick Hendricks:
34. Bernard Pieters Stellenbosch Ratepayers Association
35. Jan Dryer:
36. Dr Johann van Wyk:
37. Christian Wolf Mahncke:
38. Elke Watson:
39. Boet Grobler:
40. Jacques Rossouw: Distell

A4. Motion: Water indaba held at Spier, 13 November 2015

We, the attendees of this public meeting on the state of the water pollution of Stellenbosch rivers and the main sewerage treatment plant and the upgrade thereof, resolve as follows:

1.The pollution of rivers flowing through Stellenbosch are at critical levels and pose health and economic problems that needs to be addressed as a matter of grave concern;

2.The pollution problem should be treated holistically. Three main source areas are identified by the attendees that contribute to pollution:

a) The informal settlements of Kayamandi and Nkaneni and the total lack of infrastructure to treat water-based affluent;
b) The industrial enterprises along the Plankenberg and Eerste Rivers; and
c) The total lack of capacity of the main sewerage treatment plant the past ten years and the lack of sufficient capacity of the present upgrade;

3.The convenors of the indaba (with the right of co-option of stakeholders) are mandated to consult and engage with the private sector, the informal settlements, organised agriculture associations, a broad spectrum of professional consultants, and departments on all three tiers of government with the view to conduct feasibility studies and if need be a business plan to solve the pollution problems in a sustainable manner;

That alternative technologies and plant construction for the treatment of sewerage affluent be investigated;

That the option of decentralised plant construction is supported by the delegates; and

That the convenors release a press statement no later than 30 June 2016 as to the progress or the problems encountered to dispose of this mandate.

A5. Press statement

The dysfunctional in-house sanitation infrastructure of Kayamandi and Inkanini, as well as the bulk sanitation infrastructure of Stellenbosch, has resulted in raw sewage seeping into the Plankenburg River, which in turn drains into Eerste River. A collective of stakeholders who are concerned about the aforementioned scenario had deliberations. The collective comprised the following: Spier's Director of Human Resources and Sustainability, Heide Newton-King, who co-birthed the ideology of the indaba together with Ms Rejoice Malisa, a PhD candidate at Stellenbosch University. In addition to that, Heide is the main sponsor of the upcoming indaba of 13 November 2015. Others include André Pelser, the chairperson of the Stellenbosch Ratepayers Association; Cllr André van der Walt; Angelika van der Merwe, general manager of the Stellenbosch Farmers Association; Quiton Brynard, CEO of the Water Board; Justin Moore, CEO of Headstream; the business sector; the farming community; Kayamandi community headed by Tata Madiba Mpemnyama; Dr Jo Barnes, senior lecturer at Stellenbosch University; Carla Tenzer, the executive director of South African Associates of Ben-Gurion University of the Negev; and Prof. Eugene Cloete, the Rector of Innovation and Research from Stellenbosch University, who is the mentor of Ms Rejoice Malisa and provides guidance to this matter. With the support of the executive mayor, Alderman Conrad Sidego, and Stellenbosch Municipality Engineering Services, the collective is convening at Spier on the 13th of November 2015 for an indaba together with invited delegates.

The Deputy Mayor of Stellenbosch, Cllr Martin Smuts, will deliver the opening address, while the keynote speaker is Prof. Eugene Cloete, Vice-Rector: Research and Innovation at Stellenbosch University.

"We want to bring together stakeholders from diverse fields and political spectrums to find solutions to the dysfunctional sanitation infrastructure of Kayamandi and Inkanini, as well as the main wastewater treatment plant of Stellenbosch, which has contributed to river pollution in general and specifically the severe pollution in the Plankenburg River in Kayamandi," says Ms Rejoice Malisa, a PhD student who is doing her thesis on **The disparity between the sanitation needs of two low-income areas in Stellenbosch, South Africa and the prospect to provide services – A systems analysis** and she is also the convener of the indaba.

"Inadequate sanitation leads to water pollution and affects all of us and together we need to find solutions."

Other speakers include Dr Jeremy R Biddle, Technical Director of the UK-based company Bluewater Bio; Carla Tenzer, executive director of South Africa Associates of Ben-Gurion University of the Negev (SAABGU); Dr Jo Barnes, lecturer in community health at Stellenbosch University; Quinton Brynard, CEO of the Stellenbosch Water Board; E.J. Wentzel, Acting Director of Engineering Services at Stellenbosch Municipality; and Heidi Newton-King, Director of Sustainability at Spier.

For more information and to attend, please contact Rejoice Malisa on 078 6447 726 or Zelda Loos on 012 808 8941.

Appendix B:

Consent letter

Water workshop on Stellenbosch University Main Campus

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

STELLENBOSCH UNIVERSITY CONSENT TO PARTICIPATE IN RESEARCH

You are invited to take part in a research project. Please take some time to read the information below, which will explain the details of this research project. Please feel free to contact the researcher about any part of this project that you do not fully understand. It is very important that you are completely satisfied that you clearly understand what this research is about and how you could be involved.

Your participation is completely voluntary, and you are free to decline to participate. In other words, you may choose to take part, or not. Saying no will not affect you negatively in any way whatsoever.

You are also free to withdraw from the study at any point, even if you agreed to take part initially. None of the information that you will contribute to the study will be linked to your identity. I will not be able to retrieve any information during and after the modelling.

The Research Ethics Committee: Social, Behavioural and Education Research at Stellenbosch University has approved this study (Project ID #: 24924). We commit to conduct the study according to the ethical guidelines and principles of the South African Department of Health Ethics in Health Research: Principles, Processes and Studies (2015) and global ethics code.

WHO IS CONDUCTING THIS STUDY?

The study will be conducted by Rejoice van der Walt, from the Faculty of Military Science at Stellenbosch University.

WHY DO WE INVITE YOU TO PARTICIPATE?

You will be invited as a possible participant because of your vast knowledge and expertise in water management or because you are aware and understand the challenges facing the management of Stellenbosch's urban water system.

WHAT IS THIS RESEARCH PROJECT ABOUT?

The purpose of the study is to construct a model that can improve the management of Stellenbosch's urban water system through improved forecasting and predicting of water demands of Stellenbosch Municipality.

WHAT WILL BE ASKED OF ME?

If you consent to taking part in this study, the researcher will kindly ask you to participate in an online virtual interactive management workshop, during which you will be expected to make contributions to a list of issues that impact on the effective management of the Stellenbosch urban water system. The main areas of discussion will be around water laws, policy, and administration. Once the list is compiled, there will be a voting session to determine which issues are interdependent and interconnected. From this exercise, the modelling will be conducted and the resultant model will inform the water managers on how to address the issues that emerged from the group. This exercise can take up to four hours. If need be, there might be a follow-up, which will take less than four hours on a specific date that suits the participants.

ARE THERE ANY RISKS IN MY TAKING PART IN THIS RESEARCH?

According to the researcher, there is no potential risk; however, since water management is highly politicised and also considering the diversity of the participants, there might be issues that may arise during the brainstorming session, which might be a bone of contention to particular groupings.

Therefore, in the eventuality of such happening, the researcher will be highly alert to navigate through amicably respecting all participates according to their world views and values of the issues discussed.

WILL I BENEFIT FROM TAKING PART IN THIS RESEARCH?

There will be direct benefits to the participants in the form of knowledge exchange. Additionally, the resultant model will be presented to Stellenbosch Municipality water authorities, and if they deem it fit, the model can provide insights into how to improve the management of their urban water system.

WILL I BE PAID TO TAKE PART IN THIS STUDY AND ARE THERE ANY COSTS INVOLVED?

There will not be any form of compensation to the participants.

WHO WILL HAVE ACCESS TO MY INFORMATION?

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. Confidentiality will be maintained by means of keeping the identity of the participants anonymous. The data compiled and processed will be published in peer-reviewed journals and as dissertation. You can choose whether to be in this study or not. If you consent to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer and still remain in the study. The investigator may withdraw you from this research if circumstances arise that warrant doing so.

HOW DO I MAKE CONTACT WITH THE RESEARCHERS?

If you have any questions or concerns about this study, please feel free to contact Rejoice van der Walt at cell: xxxxxxxx; email: xxxxxxxxxxxx and/or the supervisor, Prof. K.I. Theletsane, at email: xxxxxxxxxx.

RIGHTS OF RESEARCH PARTICIPANTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights, or remedies because of your participation in this research study. If you have questions regarding your rights as a research subject, contact Mrs Clarissa Robertson (cgraham@sun.ac.za; 021 808 9183) at the Division for Research Development.

DECLARATION OF CONSENT BY THE PARTICIPANT

As the participant, I declare that:

- I have read this information and consent form, or it was read to me, and it is written in a language in which I am fluent and with which I am comfortable.
- I have had a chance to ask questions and I am satisfied that all my questions have been answered.
- I understand that taking part in this study is voluntary, and I have not been pressured to take part.
- I may choose to leave the study at any time and nothing bad will come of it – I will not be penalised or prejudiced in any way.
- I agree that the interview with me can be [video-recorded / audio-recorded].

By signing below, I ____ (name of participant) agree to take part in this research study, as conducted by Rejoice van der Walt.

____________________________________ ____________________________

Signature of Participant **Date**

DECLARATION BY THE RESEARCHER

As the researcher, I hereby declare that the information contained in this document has been thoroughly explained to the participant. I also declare that the participant has been encouraged (and has been given ample time) to ask any questions. In addition, I would like to select the following option:

	The conversation with the participant was conducted in a language in which the participant is fluent.
	I did/did not use an interpreter. (If an interpreter was used, then the interpreter must sign the declaration below.)

______________________________ ______________________

Signature of Principal Investigator **Date**

______________________________ ______________________

Signature of Interpreter (if applicable) **Date**

Permission to have all anonymous data shared with journals:

When this study is finished, we would like to publish results of the study in journals. Most journals require us to share your anonymous data with them before they publish the results. Therefore, we would like to obtain your permission to have your anonymous data shared with journals.

Tick the option you choose for anonymous data sharing with journals:

I agree to have my anonymous data shared with journals during publication of results of this study ☐

Signature______________________________

OR

I do not agree to have my anonymous data shared with journals during publication of results of this study ☐

Signature___________________________________

Permission for sharing data/information with other investigators:

In order to do the research we have discussed, we must collect and store [*describe the raw data that will be collected and stored*] from people like you. Once we have done the research that we are planning for this research project, we would like to store your information for further research to be done in the future. Other investigators from all over the world can ask to use your data in future research [*please indicate if the data will be transferred from South Africa, where the data will be stored and who will have access to the data*]. To protect your privacy, we will replace your name with a unique study number. We will only use this code for data/information about you. We will do our best to keep the code private. It is, however, always possible that someone could find out your name, but this is very unlikely to happen. Therefore, we would like to ask for your permission to share your data/information with other investigators for future, related research.

Tick the option you choose for sharing your data/information with other investigators:

I do not want my data to be shared with other investigators ☐

Signature ___________________________________

OR

I want my data to be shared with other investigators ☐

Signature ___________________________________

Appendix C:
Requests for institutional permission

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

REQUEST LETTER FOR INSTITUTIONAL PERMISSION

INSTITUTION NAME & ADDRESS: Stellenbosch Municipality
INSTITUTION CONTACT PERSON: Mr Deon Louw
INSTITUTION CONTACT NUMBER: 021 808 8111
INSTITUTION EMAIL ADDRESS: engineering.services@stellenbosch.gov.za

---

TITLE OF RESEARCH PROJECT: Development of a model to enhance sustainable urban water system management and assess the impact of municipal wastewater reuse: A case study of Stellenbosch.
RESEARCHER: Rejoice van der Walt
DEPT NAME & ADDRESS: Faculty of Military Sciences
CONTACT NUMBER: xxxxxxxxxx
EMAIL ADDRESS: xxxxxxxxxx

Dear Mr Louw

Kindly note that I am a PhD researcher at the Department of Military Sciences at Stellenbosch University, and I would appreciate your assistance with one facet of my research project. Please take some time to read the information presented in the following five points, which will explain the purpose of this letter, as well as the purpose of my research project, and then feel free to contact me if you require any additional information.

1. A short introduction to the project:

The main goal of the project is to develop a supervised machine learning model for prediction and forecasting of urban water supply and demand of Stellenbosch Municipality. Historical data on water supply and demand, demographics (age distribution in the case study), and weather statistics influencing precipitation will be utilised.

2. The purpose of the project:

The purpose of the study is to construct a supervised machine learning model that can improve the management of Stellenbosch's urban water system through improved forecasting and predicting water demands of the municipality.

3. Your assistance would be appreciated in the following regard:

I will need your assistance with access to historical achieved data that are no longer available on the public domain: daily, monthly, and annual water demand and supply for Stellenbosch Municipality over the period of 2006 to 2021.

4. Confidentiality:

The data compiled and processed will be published in peer-reviewed journals and as a dissertation.

5. Timeframe of research project:

The research will be conducted over the period of 2021 to 2023. If you have any further questions or concerns about the research, please feel free to contact me via email xxxxxxxxxxx or telephonically xxxxxxxxxx. Alternatively, feel free to contact my supervisor, K.I. Theletsane, via email xxxxxxxxx or telephonically xxxxxxxxxx. Thank you in advance for your assistance in this regard.

Kind regards,

Rejoice van der Walt
Principal Investigator

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

REQUEST LETTER FOR INSTITUTIONAL PERMISSION

INSTITUTION NAME & ADDRESS: South African Weather Service
INSTITUTION CONTACT PERSON: Prof. Joel Botai
INSTITUTION CONTACT NUMBER: +27 12 367 6000
INSTITUTION EMAIL ADDRESS: joel.botai@weathersa.co.za

TITLE OF RESEARCH PROJECT: Development of a model to enhance sustainable urban water system management and assess the impact of municipal wastewater reuse: A case study of Stellenbosch.
RESEARCHER: Rejoice van der Walt
DEPT NAME & ADDRESS: Faculty of military sciences
CONTACT NUMBER: xxxxxxxxxx
EMAIL ADDRESS: xxxxxxxxxx

Dear

Kindly note that I am a PhD researcher at the Department of Military Sciences at Stellenbosch University, and I would appreciate your assistance with one facet of my research project.

Please take some time to read the information presented in the following five points, which will explain the purpose of this letter as well as the purpose of my research project, and then feel free to contact me if you require any additional information.

1. A short introduction to the project:

The main goal of the project is to develop a supervised machine learning model for prediction and forecasting of urban water supply and demand of Stellenbosch Municipality. Historical data on water supply and demand, demographics (age distribution in the case study), and weather statistics influencing precipitation will be utilised.

2. The purpose of the project:

The purpose of the study is to construct a supervised machine learning model that can improve the management of Stellenbosch's urban water system through improved forecasting and predicting water demands of the municipality.

3. Your assistance would be appreciated in the following regard:

I will need your assistance with access to the following historical achieved data that are no longer available on the public domain: weather statistics that influence precipitation cycle and levels in the demarcated research area covering the period 2006 to 2021.

4. Confidentiality:

The data compiled and processed will be published in peer-reviewed journals and as a dissertation.

5. Timeframe of research project:

The research will be conducted over the period of 2021 to 2023.

If you have any further questions or concerns about the research, please feel free to contact me via email xxxxxxxxxxx or telephonically xxxxxxxxxx Alternatively, feel free to contact my supervisor, K.I. Theletsane, via email xxxxxxxxxx or telephonically xxxxxxxxxx.

Thank you in advance for your assistance in this regard.

Kind regards,

Rejoice van der Walt
Principal Investigator

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

REQUEST LETTER FOR INSTITUTIONAL PERMISSION

INSTITUTION NAME & ADDRESS: Department of Environmental Affairs Western Cape Government
INSTITUTION CONTACT PERSON: Mr M. Mukanya
INSTITUTION CONTACT NUMBER: 021 483 4091
INSTITUTION EMAIL ADDRESS: ronald.mukanya@westerncape.gov.za

TITLE OF RESEARCH PROJECT: Development of a model to enhance sustainable urban water system management and assess the impact of municipal wastewater reuse: a case study of Stellenbosch.
RESEARCHER: Rejoice van der Walt
DEPT NAME & ADDRESS: Faculty of Military Sciences
CONTACT NUMBER: xxxxxxxxxx
EMAIL ADDRESS: xxxxxxxxxx

Dear Mr Mukanya

Kindly note that I am a PhD researcher at the Department of Military Sciences at Stellenbosch University, and I would appreciate your assistance with one facet of my research project.

Please take some time to read the information presented in the following five points, which will explain the purpose of this letter, as well as the purpose of my research project, and then feel free to contact me if you require any additional information.

1. A short introduction to the project:

The main goal of the project is to develop a supervised machine learning model for prediction and forecasting of urban water supply and demand of Stellenbosch Municipality. Historical data on water supply and demand, demographics (age distribution in the case study), and weather statistics influencing precipitation will be utilised.

2. The purpose of the project:

The purpose of the study is to construct a supervised machine learning model that can improve the management of Stellenbosch's urban water system through improved forecasting and predicting water demands of the municipality.

3. Your assistance would be appreciated in the following regard:

I will need your assistance with access to the following historical achieved data that are no longer available on the public domain: population statistics covering age distribution in Stellenbosch Municipality for the 2006 to 2021 period.

4. Confidentiality:

The data compiled and processed will be published in peer-reviewed journals and as a dissertation.

5. Timeframe of research project:

The research will be conducted over the period of 2021 to 2023.

If you have any further questions or concerns about the research, please feel free to contact me via email xxxxxxxxx or telephonically xxxxxxxxxx. Alternatively, feel free to contact my supervisor, K.I. Theletsane, via email xxxxxxxxxxx or telephonically xxxxxxxxxx. Thank you in advance for your assistance in this regard.

Kind regards,

Rejoice van der Walt
Principal Investigator

Appendix D:
Dataset CSV file

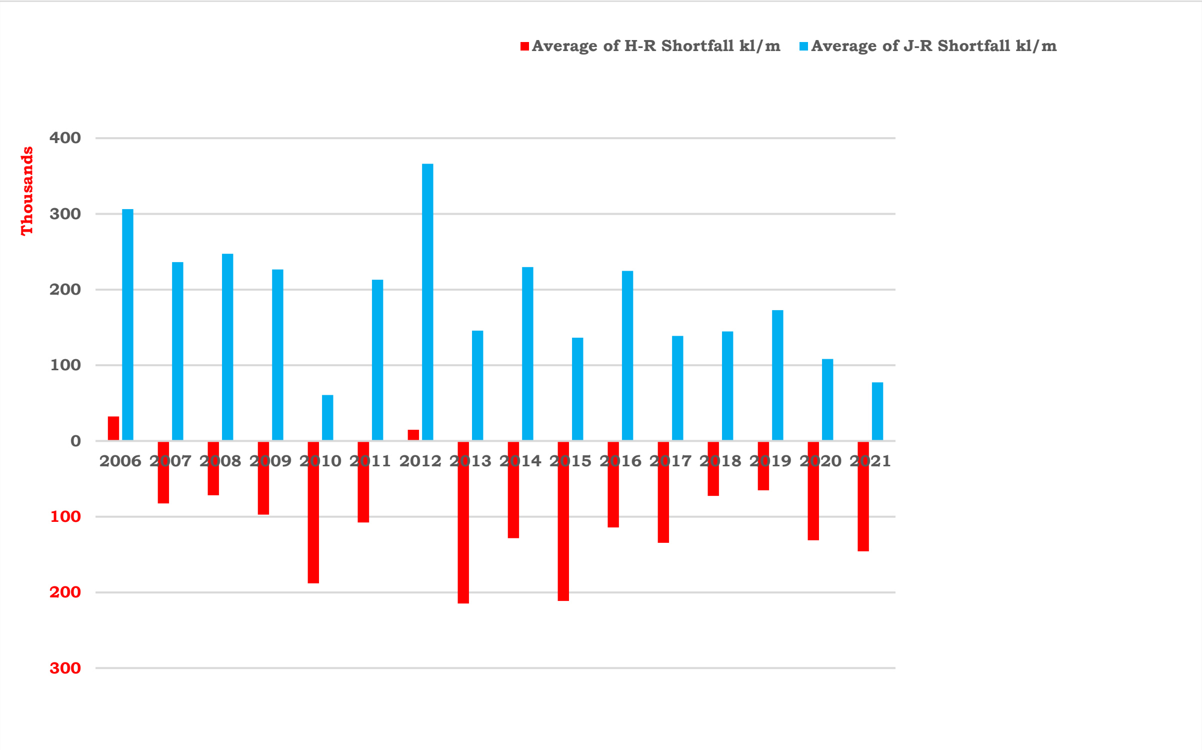
D1. Dictionary for the dataset

1. Combined monthly _merged_data_long: CMMDL
2. Run-of-river abstraction-monthly: RoRa
3. Other Raw Water Resource/Purchase: OrWr/P
4. Total Raw Water Abstraction: TRWA
5. Total Raw Water input to all WTWs: TRWI
6. Total bulk water prior to treatment: Tbwp
7. Treated water(after all WTWs): TWA
8. Bulk treated water purchased: BTWP
9. System Input Volume: SIV
10. Total expenditure on raw water-Rands: Texp
11. Total water revenue collected-Rands: Twrc
12. Billed Metered Consumption: BMC
13. Billed Unmetered Consumption: BuMC
14. Billed Consumption: BC
15. Unbilled Metered Consumption: UnBMC
16. Unbilled Unmetered Consumption: UnB&MC
17. Total water consumption: TWC
18. Proposed total wastewater treated: Ptwwt
19. Possible Water reuse for irrigation: Pwr
20. Treatment losses (12month): TLs
21. Reticulation water loss (12month): Rwls
22. Non-Revenue Water (12month): Nrw
23. Total non-revenue water(12month): TNrw
24. Run-of-river abstraction (12month): RoRo/y
25. Groundwater abstraction (12month): GA/y
26. Other water resources/purchased(12month): Owr/P/y
27. Total allocation(12month):Ta/y
28. Population: POPU
29. Student Population: SPopu

30. Population under 15 years: Popu<15
31. Population 15 to 65 years: Popu< 65
32. Population over 65: Popu>65
33. Number of Households: HNo
34. Formal dwellings: FD
35. Informal dwellings: IfH
36. Households with flush Toilets: Hft
37. Households with Piped water inside: Hpwl
38. Households using public tap: Hpt
39. Monthly minimum temperature: mtmin
40. Monthly maximum temperature: mtmax
41. Sum precipitation: Spre
42. Station name: sname
43. Water Demand: WD

D2. Exploratory data analysis for Stellenbosch Municipality

Period	Values Average of H-R Shortfall kl/m	Average of J-R Shortfall kl/m
2006	32 470	306 242
2007	82 518	236 311
2008	71 680	247 290
2009	97 221	226 558
2010	188 051	60 648
2011	107 692	213 099
2012	14 908	366 121
2013	214 654	145 698
2014	128 432	229 730
2015	211 356	136 424
2016	114 245	224 622
2017	134 441	138 705
2018	72 642	144 705
2019	65 100	172 919
2020	131 192	108 237
2021	145 813	77 418
Grand Total	**110 732**	**189 526**



Month	Period	Run-of-River abstraction kl/m	Other Raw Water Resource / Purchased kl/m	Total Raw Water Abstraction kl/m	Total Raw water input to all WTWs kl/m	Total bulk water prior to treatment kl/m	Treated water (after all WTWs) kl/m	Bulk treated water purchased kl/m	System Input Volume kl/m	Total expenditure on raw water-Rands kl/m	Total water revenue collected-Rands kl/m	Billed Metered Consumption kl/m	Billed Un-Metered Consumption kl/m	Billed Consumption kl/m	Un-Billed Metered Consumption kl/m	Unbilled Unmetered Consumption kl/m	Total water consumption kl/m	Proposed total wastewater treated kl/m	possible Water re-use for irrigation kl/m	Treatment losses (12 month) kl/y	Reticulation Water Loss (12 Month) kl/y	Non-Revenue Water (12 Month) kl/y	Total non-revenue water (12 month) kl/y	Run-of-River abstraction kl/y	Groundwater abstraction kl/y	Other water resource / purchased kl/y	Total allocation kl/y
Jul-06	2006	404,000	160,560	564,560	564,560	595,800	599,155	238,452	837,607	3,688,000	778,000	672,901	0	672,901	11,000	1,675	685,576	514,182	154,255	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Aug-06	2006	455,000	140,800	595,800	595,800	702,310	628,989	265,985	894,974	3,688,000	788,000	604,058	0	604,058	11,000	1,790	616,848	462,636	138,791	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Sep-06	2006	697,000	5,310	702,310	702,310	759,055	730,706	252,111	982,817	3,688,000	778,000	569,111	0	569,111	11,000	1,966	582,077	436,557	130,967	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Oct-06	2006	529,664	229,391	759,055	759,055	661,833	784,898	296,975	1,081,873	3,688,000	778,000	609,118	0	609,118	11,000	2,164	622,282	466,711	140,013	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Nov-06	2006	458,241	203,592	661,833	661,833	750,659	692,051	292,503	984,554	3,688,000	778,000	644,201	0	644,201	11,000	1,969	657,170	492,878	147,863	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Dec-06	2006	447,574	303,085	750,659	750,659	870,000	776,879	296,604	1,073,483	3,688,000	778,000	840,756	0	840,756	11,000	2,147	853,903	640,427	192,128	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Jan-07	2007	572,560	297,440	870,000	870,000	833,365	890,850	385,944	1,276,794	3,688,000	778,000	794,773	0	794,773	11,000	2,554	808,327	606,245	181,873	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Feb-07	2007	548,431	284,934	833,365	833,365	844,794	855,864	365,657	1,221,521	3,688,000	778,000	899,637	0	899,637	11,000	2,443	913,080	684,810	205,443	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Mar-07	2007	548,851	295,943	844,794	844,794	629,196	866,778	313,921	1,180,699	3,688,000	778,000	1,139,000	0	1,139,000	11,000	2,361	1,152,361	864,271	259,281	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Apr-07	2007	396,150	233,046	629,196	629,196	540,729	660,882	462,633	1,123,515	3,688,000	778,000	999,851	0	999,851	11,000	2,247	1,013,098	759,823	227,947	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
May-07	2007	362,503	178,226	540,729	540,729	465,001	576,396	318,518	894,914	3,688,000	778,000	981,925	0	981,925	11,000	1,790	994,715	746,036	223,811	-4,143,619	2,834,634	2,664,436	-1,241,374	7,957,000	251,000	4,525,000	12,733,000
Jun-07	2007	339,402	125,599	465,001	465,001	597,540	504,076	301,340	805,416	3,688,000	778,000	794,900	0	794,900	11,000	1,611	807,511	605,633	181,690	-4,140,865	2,783,219	2,807,936	-1,332,929	7,957,000	251,000	4,525,000	12,733,000
Jul-07	2007	592,850	4,690	597,540	597,540	657,730	630,651	272,782	903,433	3,727,000	788,000	652,510	0	652,510	11,000	1,807	665,317	498,988	149,696	-4,173,710	2,869,304	2,894,152	-1,279,558	7,957,000	251,000	4,525,000	12,733,000
Aug-07	2007	657,730	0	657,730	657,730	683,164	688,132	250,602	938,734	3,727,000	788,000	635,776	0	635,776	11,000	1,877	648,653	486,490	145,947	-4,155,541	2,881,259	2,906,195	-1,249,346	7,957,000	251,000	4,525,000	12,733,000
Sep-07	2007	627,000	56,164	683,164	683,164	603,479	712,422	220,006	932,428	3,727,000	788,000	659,268	0	659,268	11,000	1,865	672,133	504,100	151,230	-4,124,297	2,740,813	2,765,648	-1,358,649	7,957,000	251,000	4,525,000	12,733,000
Oct-07	2007	404,000	199,479	603,479	603,479	648,231	636,322	306,791	943,113	3,727,000	788,000	599,934	0	599,934	11,000	1,886	612,820	459,615	137,885	-4,141,114	2,611,516	2,636,073	-1,505,041	7,957,000	251,000	4,525,000	12,733,000
Nov-07	2007	455,000	193,231	648,231	648,231	936,293	679,061	277,722	956,783	3,727,000	788,000	623,439	0	623,439	11,000	1,914	636,353	477,264	143,179	-4,126,945	2,604,562	2,629,064	-1,497,881	7,957,000	251,000	4,525,000	12,733,000
Dec-07	2007	697,000	239,293	936,293	936,293	795,440	954,160	350,035	1,304,195	3,727,000	788,000	707,831	0	707,831	11,000	2,608	721,439	541,080	162,324	-4,172,023	2,967,738	2,992,701	-1,179,322	7,957,000	251,000	4,525,000	12,733,000
Jan-08	2008	498,000	297,440	795,440	795,440	773,073	819,645	377,893	1,197,538	3,727,000	788,000	758,646	0	758,646	11,000	2,395	772,041	579,031	173,709	-4,167,327	2,924,767	2,949,572	-1,217,755	7,957,000	251,000	4,525,000	12,733,000
Feb-08	2008	530,000	243,073	773,073	773,073	775,834	798,285	389,013	1,187,298	3,727,000	788,000	1,006,418	0	1,006,418	11,000	2,375	1,019,793	764,844	229,453	-4,193,396	2,783,832	2,808,568	-1,384,828	7,957,000	251,000	4,525,000	12,733,000
Mar-08	2008	575,000	200,834	775,834	775,834	690,624	800,921	343,815	1,144,736	3,727,000	788,000	962,926	0	962,926	11,000	2,289	976,215	732,162	219,648	-4,226,393	2,924,015	2,948,679	-1,277,714	7,957,000	251,000	4,525,000	12,733,000
Apr-08	2008	506,000	184,624	690,624	690,624	681,367	719,546	326,851	1,046,397	3,727,000	788,000	935,478	0	935,478	11,000	2,093	948,571	711,428	213,428	-4,087,847	2,911,423	2,935,933	-1,151,913	7,957,000	251,000	4,525,000	12,733,000
May-08	2008	490,000	191,367	681,367	681,367	587,350	710,705	250,966	961,671	3,727,000	788,000	946,806	0	946,806	11,000	1,923	959,729	719,797	215,939	-4,013,966	3,013,167	3,037,810	-976,156	7,957,000	251,000	4,525,000	12,733,000
Jun-08	2008	419,000	168,350	587,350	587,350	571,430	620,919	310,055	930,974	3,727,000	788,000	759,422	0	759,422	11,000	1,862	772,284	579,213	173,764	-4,017,175	3,173,952	3,198,846	-818,329	7,957,000	251,000	4,525,000	12,733,000
Jul-08	2008	455,657	115,773	571,430	571,430	596,689	605,716	235,688	841,404	5,001,000	839,000	678,959	0	678,959	11,000	1,683	691,642	518,731	155,619	-3,981,256	3,085,598	3,110,368	-870,888	7,957,000	251,000	4,525,000	12,733,000
Aug-08	2008	433,826	162,863	596,689	596,689	566,290	629,838	279,209	909,047	5,001,000	839,000	501,189	0	501,189	11,000	1,818	514,007	385,505	115,652	-4,012,610	3,190,557	3,215,268	-797,342	7,957,000	251,000	4,525,000	12,733,000
Sep-08	2008	426,518	139,772	566,290	566,290	758,176	600,807	300,632	901,439	5,001,000	839,000	565,820	0	565,820	11,000	1,803	578,623	433,967	130,190	-4,098,496	3,253,078	3,277,728	-820,768	7,957,000	251,000	4,525,000	12,733,000
Oct-08	2008	529,664	228,512	758,176	758,176	682,126	784,058	298,295	1,082,353	5,001,000	839,000	638,325	0	638,325	11,000	2,165	651,490	488,617	146,585	-4,083,038	3,353,648	3,378,576	-704,462	7,957,000	251,000	4,525,000	12,733,000
Nov-08	2008	458,241	223,885	682,126	682,126	728,856	711,430	334,768	1,046,198	5,001,000	839,000	668,336	0	668,336	11,000	2,092	681,428	511,071	153,321	-4,138,559	3,397,988	3,423,095	-715,464	7,957,000	251,000	4,525,000	12,733,000
Dec-08	2008	447,574	281,282	728,856	728,856	808,565	756,057	380,456	1,136,513	5,001,000	839,000	838,988	0	838,988	11,000	2,273	852,261	639,196	191,759	-4,178,315	3,099,485	3,124,257	-1,054,058	7,957,000	251,000	4,525,000	12,733,000
Jan-09	2009	544,000	264,565	808,565	808,565	892,677	814,579	356,686	1,171,265	5,001,000	839,000	1,031,448	0	1,031,448	11,000	2,343	1,044,791	783,593	235,078	-4,138,916	2,800,462	2,825,181	-1,313,735	7,957,000	251,000	4,525,000	12,733,000
Feb-09	2009	557,000	335,677	892,677	892,677	907,013	871,749	419,123	1,290,872	5,001,000	839,000	1,054,902	0	1,054,902	11,000	2,582	1,068,484	801,363	240,409	-4,122,886	2,855,345	2,880,271	-1,242,615	7,957,000	251,000	4,525,000	12,733,000
Mar-09	2009	564,000	343,013	907,013	907,013	743,068	905,819	418,906	1,324,725	5,001,000	839,000	1,059,714	0	1,059,714	11,000	2,649	1,073,364	805,023	241,507	-4,171,696	2,938,186	2,963,471	-1,208,224	7,957,000	251,000	4,525,000	12,733,000
Apr-09	2009	396,150	346,918	743,068	743,068	674,610	769,630	457,513	1,227,143	5,001,000	839,000	1,006,108	0	1,006,108	11,000	2,454	1,019,562	764,672	229,401	-4,299,998	3,047,940	3,073,588	-1,226,410	7,957,000	251,000	4,525,000	12,733,000
May-09	2009	362,503	312,107	674,610	674,610	529,931	704,253	338,405	1,042,658	5,001,000	839,000	1,036,821	0	1,036,821	11,000	2,085	1,049,906	787,430	236,229	-4,387,741	3,038,750	3,064,559	-1,323,182	7,957,000	251,000	4,525,000	12,733,000
Jun-09	2009	339,402	190,529	529,931	529,931	571,535	566,084	285,222	851,306	5,001,000	839,000	761,374	0	761,374	11,000	1,703	774,077	580,558	174,167	-4,365,492	2,957,289	2,982,939	-1,382,553	7,957,000	251,000	4,525,000	12,733,000
Jul-09	2009	477,887	93,648	571,535	571,535	613,552	575,656	258,864	834,520	5,848,000	932,000	537,682	0	537,682	11,000	1,669	550,351	412,763	123,829	-4,358,503	3,091,696	3,117,332	-1,241,171	7,957,000	251,000	4,525,000	12,733,000
Aug-09	2009	418,792	194,760	613,552	613,552	600,123	629,882	260,434	890,316	5,848,000	932,000	512,526	0	512,526	11,000	1,781	525,307	393,980	118,194	-4,322,909	3,061,665	3,087,264	-1,235,645	7,957,000	251,000	4,525,000	12,733,000
Sep-09	2009	460,830	139,293	600,123	600,123	707,682	607,697	258,769	866,466	5,848,000	932,000	638,994	0	638,994	11,000	1,733	651,727	488,795	146,639	-4,254,104	2,953,589	2,979,118	-1,274,986	7,957,000	251,000	4,525,000	12,733,000
Oct-09	2009	529,664	178,018	707,682	707,682	641,539	729,766	246,659	976,425	5,848,000	932,000	618,153	0	618,153	11,000	1,953	631,106	473,329	141,999	-4,198,670	2,868,045	2,893,362	-1,305,308	7,957,000	251,000	4,525,000	12,733,000

Month	Period	Run-of-River abstraction kl/m	Other Raw Water Resource / Purchased kl/m	Total Raw Water Abstraction kl/m	Total Raw water input to all WTWs kl/m	Total bulk water prior to treatment kl/m	Treated water (after all WTWs) kl/m	Bulk treated water purchased kl/m	System Input Volume kl/m	Total expenditure on raw water-Rands kl/m	Total water revenue collected-Rands kl/m	Billed Metered Consumption kl/m	Billed Un-Metered Consumption kl/m	Billed Consumption kl/m	Un-Billed Metered Consumption kl/m	Unbilled Unmetered Consumption kl/m	Total water consumption kl/m	Proposed total wastewater treated kl/m	possible Water re-use for irrigation kl/m	Treatment losses (12 month) kl/y	Reticulation Water Loss (12 Month) kl/y	Non-Revenue Water (12 Month) kl/y	Total non-revenue water (12 month) kl/y	Run-of-River abstraction kl/y	Groundwater abstraction kl/y	Other water resource / purchased kl/y	Total allocation kl/y
Nov-09	2009	458,241	183,298	641,539	641,539	772,461	696,080	278,540	974,620	5,848,000	932,000	674,915	0	674,915	11,000	1,949	687,864	515,898	154,769	-4,167,678	2,790,031	2,815,204	-1,352,474	7,957,000	251,000	4,525,000	12,733,000
Dec-09	2009	447,574	324,887	772,461	772,461	926,659	791,399	306,220	1,097,619	5,848,000	932,000	739,511	0	739,511	11,000	2,195	752,706	564,530	169,359	-4,085,179	2,850,691	2,875,787	-1,209,392	7,957,000	251,000	4,525,000	12,733,000
Jan-10	2010	583,239	343,420	926,659	926,659	749,872	954,502	319,162	1,273,664	5,848,000	932,000	853,540	0	853,540	11,000	2,547	867,087	650,315	195,095	-4,069,485	3,130,794	3,156,095	-913,390	7,957,000	251,000	4,525,000	12,733,000
Feb-10	2010	475,722	274,150	749,872	749,872	753,160	774,365	281,658	1,056,023	5,848,000	932,000	1,126,005	0	1,126,005	11,000	2,112	1,139,117	854,338	256,301	-3,977,441	2,825,312	2,850,143	-1,127,298	7,957,000	251,000	4,525,000	12,733,000
Mar-10	2010	480,403	272,757	753,160	753,160	413,059	774,592	326,609	1,101,201	5,848,000	932,000	1,053,201	0	1,053,201	11,000	2,202	1,066,403	799,803	239,941	-3,907,769	2,608,748	2,633,132	-1,274,638	7,957,000	251,000	4,525,000	12,733,000
Apr-10	2010	185,450	227,609	413,059	413,059	457,531	438,963	267,733	706,696	5,848,000	932,000	1,110,079	0	1,110,079	11,000	1,413	1,122,492	841,869	252,561	-3,717,332	1,985,371	2,008,714	-1,708,618	7,957,000	251,000	4,525,000	12,733,000
May-10	2010	193,508	264,023	457,531	457,531	420,277	464,553	242,537	707,090	5,848,000	932,000	983,122	0	983,122	11,000	1,414	995,536	746,652	223,996	-3,598,843	1,704,173	1,726,845	-1,871,998	7,957,000	251,000	4,525,000	12,733,000
Jun-10	2010	234,205	186,072	420,277	420,277	571,535	438,265	0	438,265	5,848,000	932,000	887,895	0	887,895	11,000	877	899,772	674,829	202,449	-3,295,456	1,165,437	1,187,283	-2,108,173	7,957,000	251,000	4,525,000	12,733,000
Jul-10	2010	477,887	93,648	571,535	571,535	613,552	622,194	218,073	840,267	6,853,000	1,054,000	692,017	0	692,017	11,000	1,681	704,698	528,523	158,557	-3,301,203	1,016,838	1,038,695	-2,262,508	7,957,000	251,000	4,525,000	12,733,000
Aug-10	2010	418,792	194,760	613,552	613,552	600,123	679,652	270,467	950,119	6,853,000	1,054,000	637,565	0	637,565	11,000	1,900	650,465	487,849	146,355	-3,361,006	951,482	973,459	-2,387,547	7,957,000	251,000	4,525,000	12,733,000
Sep-10	2010	460,830	139,293	600,123	600,123	707,682	657,461	213,555	871,016	6,853,000	1,054,000	680,382	0	680,382	11,000	1,742	693,124	519,843	155,953	-3,365,556	914,635	936,621	-2,428,935	7,957,000	251,000	4,525,000	12,733,000
Oct-10	2010	529,664	178,018	707,682	707,682	641,539	788,679	217,380	1,006,059	6,853,000	1,054,000	685,499	0	685,499	11,000	2,012	698,511	523,883	157,165	-3,395,190	876,864	898,909	-2,496,281	7,957,000	251,000	4,525,000	12,733,000
Nov-10	2010	458,241	183,298	641,539	641,539	772,461	734,023	298,939	1,032,962	6,853,000	1,054,000	804,746	0	804,746	11,000	2,066	817,812	613,359	184,008	-3,453,532	805,258	827,420	-2,626,112	7,957,000	251,000	4,525,000	12,733,000
Dec-10	2010	447,574	324,887	772,461	772,461	926,659	873,586	328,273	1,201,859	6,853,000	1,054,000	789,028	0	789,028	11,000	2,404	802,432	601,824	180,547	-3,557,772	859,773	882,143	-2,675,629	7,957,000	251,000	4,525,000	12,733,000
Jan-11	2011	583,239	343,420	926,659	926,659	749,872	1,085,385	390,413	1,475,798	6,853,000	1,054,000	891,004	0	891,004	11,000	2,952	904,956	678,717	203,615	-3,759,906	1,024,038	1,046,813	-2,713,093	7,957,000	251,000	4,525,000	12,733,000
Feb-11	2011	475,722	274,150	749,872	749,872	753,160	872,253	351,051	1,223,304	6,853,000	1,054,000	998,275	0	998,275	11,000	2,447	1,011,721	758,791	227,637	-3,927,187	1,318,715	1,341,824	-2,585,363	7,957,000	251,000	4,525,000	12,733,000
Mar-11	2011	480,403	272,757	753,160	753,160	413,059	872,047	405,920	1,277,967	6,853,000	1,054,000	1,090,723	0	1,090,723	11,000	2,556	1,104,279	828,209	248,463	-4,103,953	1,457,605	1,481,068	-2,622,885	7,957,000	251,000	4,525,000	12,733,000
Apr-11	2011	185,450	227,609	413,059	413,059	457,531	528,550	282,995	811,545	6,853,000	1,054,000	1,275,189	0	1,275,189	11,000	1,623	1,287,812	965,859	289,758	-4,208,802	1,397,134	1,420,807	-2,787,995	7,957,000	251,000	4,525,000	12,733,000
May-11	2011	193,508	264,023	457,531	457,531	420,277	559,196	275,897	835,093	6,853,000	1,054,000	983,512	0	983,512	11,000	1,670	996,182	747,137	224,141	-4,336,805	1,524,491	1,548,420	-2,788,385	7,957,000	251,000	4,525,000	12,733,000
Jun-11	2011	234,205	186,072	420,277	420,277	588,852	483,463	226,851	710,314	6,853,000	1,054,000	913,068	0	913,068	11,000	1,421	925,489	694,116	208,235	-4,608,854	1,770,823	1,795,296	-2,813,558	7,957,000	251,000	4,525,000	12,733,000
Jul-11	2011	530,373	58,479	588,852	588,852	591,728	693,206	353,336	1,046,542	7,806,000	1,023,000	512,254	10,465	522,719	10,465	2,093	535,278	401,458	120,438	-4,797,811	2,135,517	2,170,868	-2,626,943	7,957,000	251,000	4,525,000	12,733,000
Aug-11	2011	469,064	122,664	591,728	591,728	734,267	680,461	346,819	1,027,280	7,806,000	1,023,000	525,429	10,273	535,702	10,273	2,055	548,029	411,022	123,307	-4,896,797	2,304,115	2,349,893	-2,546,904	7,957,000	251,000	4,525,000	12,733,000
Sep-11	2011	595,505	138,762	734,267	734,267	599,606	789,718	304,055	1,093,773	7,806,000	1,023,000	588,776	10,938	599,714	10,938	2,188	612,839	459,629	137,889	-4,985,409	2,596,156	2,653,317	-2,332,092	7,957,000	251,000	4,525,000	12,733,000
Oct-11	2011	480,806	118,800	599,606	599,606	598,251	713,849	352,568	1,066,417	7,806,000	1,023,000	638,442	10,664	649,106	10,664	2,133	661,903	496,427	148,928	-5,153,843	2,682,122	2,750,067	-2,403,775	7,957,000	251,000	4,525,000	12,733,000
Nov-11	2011	470,591	127,660	598,251	598,251	712,898	671,114	298,240	969,354	7,806,000	1,023,000	671,784	9,694	681,478	9,694	1,939	693,110	519,832	155,950	-5,133,522	2,732,216	2,809,728	-2,323,795	7,957,000	251,000	4,525,000	12,733,000
Dec-11	2011	466,005	246,893	712,898	712,898	785,011	798,441	261,348	1,059,789	7,806,000	1,023,000	735,068	10,598	745,666	10,598	2,120	758,383	568,788	170,636	-5,051,015	2,623,193	2,711,019	-2,339,996	7,957,000	251,000	4,525,000	12,733,000
Jan-12	2012	460,440	324,571	785,011	785,011	707,827	954,394	518,998	1,473,392	7,806,000	1,023,000	860,828	14,734	875,562	14,734	2,947	893,243	669,932	200,980	-5,190,256	2,621,500	2,724,055	-2,466,201	7,957,000	251,000	4,525,000	12,733,000
Feb-12	2012	477,297	230,530	707,827	707,827	778,147	808,139	410,202	1,218,341	7,806,000	1,023,000	828,721	12,183	840,904	12,183	2,437	855,524	641,643	192,493	-5,227,338	2,761,734	2,876,462	-2,350,876	7,957,000	251,000	4,525,000	12,733,000
Mar-12	2012	424,469	353,678	778,147	778,147	731,009	836,225	400,869	1,237,094	7,806,000	1,023,000	853,023	12,371	865,394	12,371	2,474	880,239	660,179	198,054	-5,161,479	2,933,901	3,060,919	-2,100,560	7,957,000	251,000	4,525,000	12,733,000
Apr-12	2012	424,469	306,540	731,009	731,009	489,276	801,257	322,852	1,124,109	7,806,000	1,023,000	762,749	11,241	773,990	11,241	2,248	787,479	590,610	177,183	-5,156,092	3,735,797	3,874,681	-1,281,411	7,957,000	251,000	4,525,000	12,733,000
May-12	2012	340,477	148,799	489,276	489,276	610,867	575,585	367,074	942,659	7,806,000	1,023,000	631,894	9,427	641,321	9,427	1,885	652,632	489,474	146,842	-5,231,912	4,175,912	4,324,438	-907,474	7,957,000	251,000	4,525,000	12,733,000
Jun-12	2012	466,005	144,862	610,867	610,867	463,705	689,656	251,777	941,433	7,806,000	1,023,000	576,791	9,414	586,205	9,414	1,883	597,503	448,127	134,438	-5,272,442	4,724,018	4,882,420	-390,022	7,957,000	251,000	4,525,000	12,733,000
Jul-12	2012	333,995	129,710	463,705	463,705	401,480	578,110	388,334	966,444	7,955,000	1,354,000	516,607	9,664	526,271	9,664	1,933	537,869	403,402	121,020	-5,317,491	4,641,330	4,798,771	-518,721	7,957,000	251,000	4,525,000	12,733,000
Aug-12	2012	353,563	47,917	401,480	401,480	441,785	464,138	195,227	659,365	7,955,000	1,354,000	519,134	6,594	525,728	6,594	1,319	533,640	400,230	120,069	-5,139,825	4,287,804	4,440,830	-698,995	7,957,000	251,000	4,525,000	12,733,000
Sep-12	2012	353,623	88,162	441,785	441,785	647,873	518,505	318,487	836,992	7,955,000	1,354,000	559,349	8,370	567,719	8,370	1,674	577,763	433,322	129,997	-5,175,525	4,066,099	4,216,043	-959,482	7,957,000	251,000	4,525,000	12,733,000
Oct-12	2012	360,184	287,689	647,873	647,873	790,589	717,109	311,606	1,028,715	7,955,000	1,354,000	605,804	10,287	616,091	10,287	2,057	628,436	471,327	141,398	-5,089,556	4,061,864	4,211,356	-878,200	7,957,000	251,000	4,525,000	12,733,000
Nov-12	2012	412,607	377,982	790,589	790,589	844,222	883,251	347,636	1,230,887	7,955,000	1,354,000	742,450	12,309	754,759	12,309	2,462	769,530	577,147	173,144	-5,158,752	4,246,978	4,399,609	-759,143	7,957,000	251,000	4,525,000	12,733,000
Dec-12	2012	433,259	410,963	844,222	844,222	842,904	894,160	381,495	1,275,655	7,955,000	1,354,000	799,712	12,757	812,469	12,757	2,551	827,776	620,832	186,250	-5,243,294	4,393,452	4,548,673	-694,622	7,957,000	251,000	4,525,000	12,733,000
Jan-13	2013	430,623	412,281	842,904	842,904	636,693	899,297	396,302	1,295,599	7,955,000	1,354,000	891,649	12,956	904,605	12,956	2,591	920,152	690,114	207,034	-5,007,609	4,188,750	4,341,837	-665,772	7,957,000	251,000	4,525,000	12,733,000
Feb-13	2013	310,192	326,501	636,693	636,693	590,327	703,942	377,612	1,081,554	7,955,000	1,354,000	910,968	10,816	921,784	10,816	2,163	934,762	701,072	210,321	-4,941,956	3,972,725	4,124,171	-817,785	7,957,000	251,000	4,525,000	12,733,000

Month	Period	Run-of-River abstraction kl/m	Other Raw Water Resource / Purchased kl/m	Total Raw Water Abstraction kl/m	Total Raw water input to all WTWs kl/m	Total bulk water prior to treatment kl/m	Treated water (after all WTWs) kl/m	Bulk treated water purchased kl/m	System Input Volume kl/m	Total expenditure on raw water-Rands kl/m	Total water revenue collected-Rands kl/m	Billed Metered Consumption kl/m	Billed Un-Metered Consumption kl/m	Billed Consumption kl/m	Un-Billed Metered Consumption kl/m	Unbilled Unmetered Consumption kl/m	Total water consumption kl/m	Proposed total wastewater treated kl/m	possible Water re-use for irrigation kl/m	Treatment losses (12 month) kl/y	Reticulation Water Loss (12 Month) kl/y	Non-Revenue Water (12 Month) kl/y	Total non-revenue water (12 month) kl/y	Run-of-River abstraction kl/y	Groundwater abstraction kl/y	Other water resource / purchased kl/y	Total allocation kl/y
Mar-13	2013	231,640	358,687	590,327	590,327	541,115	631,762	270,127	901,889	7,955,000	1,354,000	801,776	9,019	810,795	9,019	1,804	821,618	616,213	184,864	-4,794,571	3,696,142	3,843,565	-951,006	7,957,000	251,000	4,525,000	12,733,000
Apr-13	2013	216,303	324,812	541,115	541,115	521,941	584,545	332,961	917,506	7,955,000	1,354,000	946,830	9,175	956,005	9,175	1,835	967,015	725,261	217,578	-4,777,862	3,310,003	3,454,947	-1,322,915	7,957,000	251,000	4,525,000	12,733,000
May-13	2013	221,631	300,310	521,941	521,941	457,117	546,279	378,447	924,726	7,955,000	1,354,000	953,306	9,247	962,553	9,247	1,849	973,650	730,237	219,071	-4,727,264	2,971,053	3,115,782	-1,611,483	7,957,000	251,000	4,525,000	12,733,000
Jun-13	2013	205,980	251,137	457,117	457,117	482,199	462,847	269,595	732,442	7,955,000	1,354,000	757,085	7,324	764,409	7,324	1,465	773,199	579,899	173,970	-4,672,023	2,586,365	2,728,586	-1,943,437	7,957,000	251,000	4,525,000	12,733,000
Jul-13	2013	163,722	318,477	482,199	482,199	474,247	486,741	431,031	917,772	8,585,000	1,516,000	767,560	9,178	776,738	9,178	1,836	787,751	590,813	177,244	-4,604,857	2,287,811	2,429,448	-2,175,409	7,957,000	251,000	4,525,000	12,733,000
Aug-13	2013	146,477	327,770	474,247	474,247	481,103	494,363	460,400	954,763	8,585,000	1,516,000	604,112	9,548	613,660	9,548	1,910	625,117	468,838	140,651	-4,827,487	2,491,731	2,636,913	-2,190,574	7,957,000	251,000	4,525,000	12,733,000
Sep-13	2013	169,776	311,327	481,103	481,103	656,199	473,664	281,697	755,361	8,585,000	1,516,000	686,286	7,554	693,840	7,554	1,511	702,904	527,178	158,153	-4,706,539	2,284,960	2,429,162	-2,277,377	7,957,000	251,000	4,525,000	12,733,000
Oct-13	2013	291,819	364,380	656,199	656,199	673,814	664,612	329,303	993,915	8,585,000	1,516,000	750,577	9,939	760,516	9,939	1,988	772,443	579,332	173,800	-4,663,413	2,106,153	2,249,938	-2,413,476	7,957,000	251,000	4,525,000	12,733,000
Nov-13	2013	305,899	367,915	673,814	673,814	670,228	679,294	375,669	1,054,963	8,585,000	1,516,000	750,807	10,550	761,357	10,550	2,110	774,016	580,512	174,154	-4,604,264	1,925,742	2,067,416	-2,536,848	7,957,000	251,000	4,525,000	12,733,000
Dec-13	2013	369,033	301,195	670,228	670,228	666,213	703,747	421,074	1,124,821	8,585,000	1,516,000	829,564	11,248	840,812	11,248	2,250	854,310	640,733	192,220	-4,627,424	1,748,374	1,888,238	-2,739,186	7,957,000	251,000	4,525,000	12,733,000
Jan-14	2014	376,982	289,231	666,213	666,213	834,697	684,796	363,821	1,048,617	8,585,000	1,516,000	901,166	10,486	911,652	10,486	2,097	924,236	693,177	207,953	-4,557,133	1,497,309	1,634,209	-2,922,924	7,957,000	251,000	4,525,000	12,733,000
Feb-14	2014	475,626	359,071	834,697	834,697	769,004	853,671	442,139	1,295,810	8,585,000	1,516,000	1,134,683	12,958	1,147,641	12,958	2,592	1,163,191	872,393	261,718	-4,573,385	1,483,136	1,622,607	-2,950,778	7,957,000	251,000	4,525,000	12,733,000
Mar-14	2014	425,117	343,887	769,004	769,004	726,340	782,913	415,308	1,198,221	8,585,000	1,516,000	1,018,419	11,982	1,030,401	11,982	2,396	1,044,780	783,585	235,075	-4,691,040	1,556,305	1,699,332	-2,991,707	7,957,000	251,000	4,525,000	12,733,000
Apr-14	2014	386,594	339,746	726,340	726,340	749,857	758,909	364,989	1,123,898	8,585,000	1,516,000	1,164,850	11,239	1,176,089	11,239	2,248	1,189,576	892,182	267,655	-4,712,206	1,540,137	1,685,640	-3,026,566	7,957,000	251,000	4,525,000	12,733,000
May-14	2014	416,427	333,430	749,857	749,857	740,260	752,058	239,971	992,029	8,585,000	1,516,000	958,829	9,920	968,749	9,920	1,984	980,654	735,490	220,647	-4,551,594	1,600,437	1,746,748	-2,804,846	7,957,000	251,000	4,525,000	12,733,000
Jun-14	2014	382,637	357,623	740,260	740,260	754,297	734,082	255,394	989,476	8,585,000	1,516,000	888,894	9,895	898,789	9,895	1,979	910,662	682,997	204,899	-4,525,486	1,720,008	1,869,403	-2,656,082	7,957,000	251,000	4,525,000	12,733,000
Jul-14	2014	409,230	345,067	754,297	754,297	799,208	752,238	302,776	1,055,014	10,171,000	1,607,000	758,956	10,550	769,506	10,550	2,110	782,166	586,625	175,987	-4,390,629	1,862,834	2,013,876	-2,376,753	7,957,000	251,000	4,525,000	12,733,000
Aug-14	2014	434,402	364,806	799,208	799,208	753,395	783,806	397,132	1,180,938	10,171,000	1,607,000	701,126	11,809	712,935	11,809	2,362	727,107	545,330	163,599	-4,291,843	1,987,019	2,140,775	-2,151,068	7,957,000	251,000	4,525,000	12,733,000
Sep-14	2014	442,068	311,327	753,395	753,395	830,862	739,568	368,583	1,108,151	10,171,000	1,607,000	694,313	11,082	705,395	11,082	2,216	718,692	539,019	161,706	-4,372,341	2,324,020	2,482,010	-1,890,331	7,957,000	251,000	4,525,000	12,733,000
Oct-14	2014	535,226	295,636	830,862	830,862	866,232	844,637	346,457	1,191,094	10,171,000	1,607,000	852,912	11,911	864,823	11,911	2,382	879,116	659,337	197,801	-4,394,857	2,414,526	2,574,883	-1,819,974	7,957,000	251,000	4,525,000	12,733,000
Nov-14	2014	508,372	357,860	866,232	866,232	917,860	879,562	398,489	1,278,051	10,171,000	1,607,000	793,102	12,781	805,883	12,781	2,556	821,219	615,914	184,774	-4,425,526	2,590,411	2,753,444	-1,672,082	7,957,000	251,000	4,525,000	12,733,000
Dec-14	2014	545,614	372,246	917,860	917,860	936,916	941,376	402,874	1,344,250	10,171,000	1,607,000	877,822	13,443	891,265	13,443	2,689	907,396	680,547	204,164	-4,397,324	2,756,755	2,922,421	-1,474,902	7,957,000	251,000	4,525,000	12,733,000
Jan-15	2015	551,367	385,549	936,916	936,916	823,614	948,936	404,349	1,353,285	10,171,000	1,607,000	997,856	13,533	1,011,389	13,533	2,707	1,027,628	770,721	231,216	-4,431,288	2,958,029	3,127,352	-1,303,936	7,957,000	251,000	4,525,000	12,733,000
Feb-15	2015	477,499	346,115	823,614	823,614	1,056,726	808,496	460,519	1,269,015	10,171,000	1,607,000	1,143,053	12,690	1,155,743	12,690	2,538	1,170,971	878,229	263,469	-4,415,577	2,923,455	3,092,456	-1,323,121	7,957,000	251,000	4,525,000	12,733,000
Mar-15	2015	628,306	428,420	1,056,726	1,056,726	823,332	1,059,083	391,694	1,450,777	10,171,000	1,607,000	1,210,252	14,508	1,224,760	14,508	2,902	1,242,169	931,627	279,488	-4,380,411	2,978,622	3,150,654	-1,229,758	7,957,000	251,000	4,525,000	12,733,000
Apr-15	2015	416,735	406,597	823,332	823,332	699,309	832,260	403,087	1,235,347	10,171,000	1,607,000	1,198,585	12,353	1,210,938	12,353	2,471	1,225,763	919,322	275,797	-4,394,869	3,053,884	3,227,254	-1,167,615	7,957,000	251,000	4,525,000	12,733,000
May-15	2015	296,454	402,855	699,309	699,309	645,869	708,900	316,927	1,025,827	10,171,000	1,607,000	1,143,051	10,258	1,153,309	10,258	2,052	1,165,619	874,214	262,264	-4,479,215	2,902,717	3,076,491	-1,402,723	7,957,000	251,000	4,525,000	12,733,000
Jun-15	2015	292,937	352,932	645,869	645,869	641,381	649,488	289,077	938,565	10,171,000	1,607,000	1,046,326	9,386	1,055,712	9,386	1,877	1,066,974	800,231	240,069	-4,522,694	2,695,493	2,868,657	-1,654,037	7,957,000	251,000	4,525,000	12,733,000
Jul-15	2015	359,857	281,524	641,381	641,381	675,653	652,599	292,854	945,453	11,833,000	1,700,000	761,789	17,502	779,291	17,502	1,891	798,684	599,013	179,704	-4,526,049	2,569,415	2,749,311	-1,776,738	7,957,000	251,000	4,525,000	12,733,000
Aug-15	2015	374,783	300,870	675,653	675,653	648,851	677,302	206,603	883,905	11,833,000	1,700,000	722,178	16,923	739,101	16,923	1,768	757,793	568,345	170,503	-4,352,571	2,241,696	2,426,112	-1,926,459	7,957,000	251,000	4,525,000	12,733,000
Sep-15	2015	340,140	308,711	648,851	648,851	820,131	637,935	285,129	923,064	11,833,000	1,700,000	732,224	17,419	749,643	17,419	1,846	768,908	576,681	173,004	-4,272,028	2,006,393	2,196,776	-2,075,251	7,957,000	251,000	4,525,000	12,733,000
Oct-15	2015	437,181	382,950	820,131	820,131	857,036	812,635	346,297	1,158,932	11,833,000	1,700,000	756,517	21,877	778,394	21,877	2,318	802,589	601,942	180,583	-4,250,597	2,050,758	2,251,043	-1,999,554	7,957,000	251,000	4,525,000	12,733,000
Nov-15	2015	453,707	403,329	857,036	857,036	819,970	844,826	396,189	1,241,015	11,833,000	1,700,000	818,995	23,331	842,326	23,331	2,482	868,140	651,105	195,331	-4,222,758	1,966,802	2,177,564	-2,045,194	7,957,000	251,000	4,525,000	12,733,000
Dec-15	2015	435,176	384,794	819,970	819,970	887,090	810,738	380,631	1,191,369	11,833,000	1,700,000	1,036,815	22,515	1,059,330	22,515	2,383	1,084,228	813,171	243,951	-4,167,767	1,637,088	1,856,617	-2,311,149	7,957,000	251,000	4,525,000	12,733,000
Jan-16	2016	480,257	406,833	887,090	887,090	839,017	871,588	454,573	1,326,161	11,833,000	1,700,000	1,083,339	24,942	1,108,281	24,942	2,652	1,135,875	851,906	255,572	-4,190,469	1,501,717	1,732,601	-2,457,867	7,957,000	251,000	4,525,000	12,733,000
Feb-16	2016	491,587	347,430	839,017	839,017	823,078	819,839	447,032	1,266,871	11,833,000	1,700,000	1,042,165	23,871	1,066,036	23,871	2,534	1,092,441	819,331	245,799	-4,172,922	1,578,104	1,820,164	-2,352,757	7,957,000	251,000	4,525,000	12,733,000
Mar-16	2016	462,244	360,834	823,078	823,078	850,831	804,531	481,163	1,285,694	11,833,000	1,700,000	960,290	24,210	984,500	24,210	2,571	1,011,281	758,461	227,538	-4,241,487	1,643,909	1,895,342	-2,346,145	7,957,000	251,000	4,525,000	12,733,000
Apr-16	2016	405,251	445,580	850,831	850,831	832,086	826,717	408,012	1,234,729	11,833,000	1,700,000	857,345	23,426	880,771	23,426	2,469	906,667	680,000	204,000	-4,213,369	1,962,386	2,224,890	-1,988,479	7,957,000	251,000	4,525,000	12,733,000
May-16	2016	366,784	465,302	832,086	832,086	714,811	808,051	364,648	1,172,699	11,833,000	1,700,000	796,611	22,268	818,879	22,268	2,345	843,493	632,619	189,786	-4,227,464	2,431,385	2,706,193	-1,521,272	7,957,000	251,000	4,525,000	12,733,000
Jun-16	2016	334,330	380,481	714,811	714,811	697,001	688,415	147,233	835,648	11,833,000	1,700,000	666,770	16,225	682,995	16,225	1,671	700,891	525,668	157,700	-4,055,605	2,694,551	2,975,992	-1,079,613	7,957,000	251,000	4,525,000	12,733,000

Month	Period	Run-of-River abstraction kl/m	Other Raw Water Resource / Purchased kl/m	Total Raw Water Abstraction kl/m	Total Raw water input to all WTWs kl/m	Total bulk water prior to treatment kl/m	Treated water (after all WTWs) kl/m	Bulk treated water purchased kl/m	System Input Volume kl/m	Total expenditure on raw water-Rands kl/m	Total water revenue collected-Rands kl/m	Billed Metered Consumption kl/m	Billed Un-Metered Consumption kl/m	Billed Consumption kl/m	Un-Billed Metered Consumption kl/m	Unbilled Unmetered Consumption kl/m	Total water consumption kl/m	Proposed total wastewater treated kl/m	possible Water re-use for irrigation kl/m	Treatment losses (12 month) kl/y	Reticulation Water Loss (12 Month) kl/y	Non-Revenue Water (12 Month) kl/y	Total non-revenue water (12 month) kl/y	Run-of-River abstraction kl/y	Groundwater abstraction kl/y	Other water resource / purchased kl/y	Total allocation kl/y
Jul-16	2016	333,920	363,081	697,001	697,001	722,333	677,039	271,219	948,258	13,295,000	2,012,000	655,070	28,448	683,518	28,448	1,897	713,862	535,396	160,619	-4,002,790	2,782,178	3,074,570	-928,220	7,957,000	251,000	4,525,000	12,733,000
Aug-16	2016	351,082	371,251	722,333	722,333	713,467	704,349	268,915	973,264	13,295,000	2,012,000	707,052	29,198	736,250	29,198	1,947	767,394	575,546	172,664	-4,045,469	2,861,936	3,166,781	-878,688	7,957,000	251,000	4,525,000	12,733,000
Sep-16	2016	335,936	377,531	713,467	713,467	823,887	727,534	351,925	1,079,459	13,295,000	2,012,000	744,350	32,384	776,734	32,384	2,159	811,276	608,457	182,537	-4,137,249	2,975,962	3,296,086	-841,163	7,957,000	251,000	4,525,000	12,733,000
Oct-16	2016	368,484	455,403	823,887	823,887	878,704	808,385	201,237	1,009,622	13,295,000	2,012,000	785,029	30,289	815,318	30,289	2,019	847,626	635,719	190,716	-3,984,183	2,781,616	3,109,852	-874,331	7,957,000	251,000	4,525,000	12,733,000
Nov-16	2016	414,899	463,805	878,704	878,704	758,600	865,744	356,413	1,222,157	13,295,000	2,012,000	895,190	36,665	931,855	36,665	2,444	970,964	728,223	218,467	-3,943,657	2,659,934	3,001,466	-942,191	7,957,000	251,000	4,525,000	12,733,000
Dec-16	2016	377,209	381,391	758,600	758,600	791,861	748,407	314,030	1,062,437	13,295,000	2,012,000	853,895	31,873	885,768	31,873	2,125	919,766	689,825	206,947	-3,876,094	2,695,464	3,046,095	-829,999	7,957,000	251,000	4,525,000	12,733,000
Jan-17	2017	389,055	402,806	791,861	791,861	680,742	778,727	401,077	1,179,804	13,295,000	2,012,000	876,190	35,394	911,584	35,394	2,360	949,338	712,003	213,601	-3,824,967	2,735,644	3,096,436	-728,531	7,957,000	251,000	4,525,000	12,733,000
Feb-17	2017	371,083	309,659	680,742	680,742	747,775	675,105	304,622	979,727	13,295,000	2,012,000	859,114	29,392	888,506	29,392	1,959	919,857	689,893	206,968	-3,696,097	2,621,083	2,986,821	-709,276	7,957,000	251,000	4,525,000	12,733,000
Mar-17	2017	373,214	374,561	747,775	747,775	681,992	733,035	380,052	1,113,087	13,295,000	2,012,000	742,245	33,393	775,638	33,393	2,226	811,256	608,442	182,533	-3,598,793	2,648,501	3,023,076	-575,717	7,957,000	251,000	4,525,000	12,733,000
Apr-17	2017	322,059	359,933	681,992	681,992	671,589	665,810	245,339	911,149	13,295,000	2,012,000	727,048	43,707	770,755	27,334	1,822	799,912	599,934	179,980	-3,444,052	2,431,676	2,809,513	-634,540	7,957,000	251,000	4,525,000	12,733,000
May-17	2017	313,212	358,377	671,589	671,589	549,935	663,688	260,892	924,580	13,295,000	2,012,000	669,301	44,732	714,033	27,737	1,849	743,620	557,715	167,314	-3,356,431	2,283,430	2,666,240	-690,191	7,957,000	251,000	4,525,000	12,733,000
Jun-17	2017	267,188	282,747	549,935	549,935	564,757	549,347	267,414	816,761	13,295,000	2,012,000	642,309	39,657	681,966	24,503	1,634	708,102	531,077	159,323	-3,502,420	2,257,333	2,648,383	-854,038	7,957,000	251,000	4,525,000	12,733,000
Jul-17	2017	284,047	280,710	564,757	564,757	551,163	560,727	236,341	797,068	16,442,000	1,344,000	565,209	38,369	603,578	23,912	1,594	629,085	471,813	141,544	-3,483,474	2,190,920	2,577,132	-906,342	7,957,000	251,000	4,525,000	12,733,000
Aug-17	2017	267,853	283,310	551,163	551,163	501,197	539,837	238,877	778,714	16,442,000	1,344,000	578,546	37,395	615,941	23,361	1,557	640,860	480,645	144,193	-3,460,094	2,122,905	2,502,891	-957,203	7,957,000	251,000	4,525,000	12,733,000
Sep-17	2017	253,737	247,460	501,197	501,197	522,832	496,652	235,252	731,904	16,442,000	1,344,000	575,276	35,274	610,550	21,957	1,464	633,971	475,478	142,643	-3,324,809	1,952,656	2,321,520	-1,003,289	7,957,000	251,000	4,525,000	12,733,000
Oct-17	2017	358,092	164,740	522,832	522,832	480,958	522,387	245,896	768,282	16,442,000	1,344,000	545,833	37,119	582,952	23,048	1,537	607,537	455,652	136,696	-3,384,524	1,951,405	2,312,546	-1,071,978	7,957,000	251,000	4,525,000	12,733,000
Nov-17	2017	477,648	3,310	480,958	480,958	500,809	478,118	227,567	705,685	16,442,000	1,344,000	572,507	33,900	606,407	21,171	1,411	628,989	471,742	141,522	-3,265,798	1,776,907	2,121,522	-1,144,276	7,957,000	251,000	4,525,000	12,733,000
Dec-17	2017	418,149	82,660	500,809	500,809	540,380	494,805	234,431	729,236	16,442,000	1,344,000	640,587	35,087	675,674	21,877	1,458	699,010	524,257	157,277	-3,190,388	1,664,462	1,998,414	-1,191,973	7,957,000	251,000	4,525,000	12,733,000
Jan-18	2018	230,275	310,105	540,380	540,380	448,726	537,864	277,361	815,225	16,442,000	1,344,000	525,142	39,520	564,662	24,457	1,630	590,749	443,062	132,919	-3,077,289	1,658,471	1,980,757	-1,096,532	7,957,000	251,000	4,525,000	12,733,000
Feb-18	2018	158,741	289,985	448,726	448,726	499,748	446,572	244,483	691,055	16,442,000	1,344,000	456,252	33,297	489,549	20,732	1,382	511,663	383,747	115,124	-3,020,634	1,777,994	2,091,042	-929,592	7,957,000	251,000	4,525,000	12,733,000
Mar-18	2018	176,378	323,370	499,748	499,748	465,318	493,563	230,396	723,959	16,442,000	1,344,000	481,140	34,729	515,869	21,719	1,448	539,036	404,277	121,283	-2,879,533	1,661,087	1,961,683	-917,850	7,957,000	251,000	4,525,000	12,733,000
Apr-18	2018	174,088	291,230	465,318	465,318	450,645	457,518	217,344	674,861	16,442,000	1,344,000	464,861	32,251	497,112	20,246	1,350	518,708	389,031	116,709	-2,859,919	1,706,003	1,999,038	-860,881	7,957,000	251,000	4,525,000	12,733,000
May-18	2018	213,768	231,890	450,645	450,645	392,440	444,653	153,353	598,006	16,442,000	1,344,000	465,524	28,701	494,225	17,940	1,196	513,362	385,021	115,506	-2,754,289	1,609,687	1,892,271	-862,017	7,957,000	251,000	4,525,000	12,733,000
Jun-18	2018	352,491	37,670	392,440	392,440	410,061	386,655	145,265	531,920	16,442,000	1,344,000	419,906	25,410	445,316	15,958	1,064	462,338	346,753	104,026	-2,626,942	1,570,610	1,844,080	-782,862	7,957,000	251,000	4,525,000	12,733,000
Jul-18	2018	390,254	310	410,061	410,061	428,963	404,976	213,283	618,259	12,273,000	2,555,000	429,945	30,025	459,970	18,548	1,237	479,754	359,816	107,945	-2,602,829	1,541,131	1,808,879	-793,950	7,957,000	251,000	4,525,000	12,733,000
Aug-18	2018	409,348	1,410	428,963	428,963	459,597	423,669	204,083	627,752	12,273,000	2,555,000	463,637	30,293	493,930	18,833	1,256	514,018	385,514	115,654	-2,574,066	1,517,010	1,779,927	-794,139	7,957,000	251,000	4,525,000	12,733,000
Sep-18	2018	437,751	0	459,597	459,597	535,888	453,712	201,912	655,624	12,273,000	2,555,000	455,708	31,635	487,343	19,669	1,311	508,323	381,242	114,373	-2,539,387	1,566,378	1,826,854	-712,533	7,957,000	251,000	4,525,000	12,733,000
Oct-18	2018	513,451	0	535,888	535,888	492,573	527,633	228,785	756,418	12,273,000	2,555,000	531,947	36,658	568,605	22,693	1,513	592,811	444,608	133,382	-2,514,466	1,569,239	1,829,335	-685,131	7,957,000	251,000	4,525,000	12,733,000
Nov-18	2018	472,902	170	492,573	492,573	460,643	489,724	240,631	730,354	12,273,000	2,555,000	578,260	35,649	613,909	21,911	1,461	637,280	477,960	143,388	-2,527,521	1,585,617	1,846,503	-681,018	7,957,000	251,000	4,525,000	12,733,000
Dec-18	2018	312,332	119,330	460,643	460,643	435,034	460,040	251,269	711,309	12,273,000	2,555,000	472,775	34,709	507,484	21,339	1,423	530,246	397,684	119,305	-2,549,760	1,736,455	1,996,766	-552,994	7,957,000	251,000	4,525,000	12,733,000
Jan-19	2019	354,563	48,080	435,034	435,034	492,446	434,314	239,396	673,710	12,273,000	2,555,000	573,711	32,629	606,340	20,211	1,347	627,899	470,924	141,277	-2,513,591	1,557,790	1,813,573	-700,018	7,957,000	251,000	4,525,000	12,733,000
Feb-19	2019	274,550	186,200	492,446	492,446	433,244	489,633	259,606	749,239	12,273,000	2,555,000	580,978	36,521	617,499	22,477	1,498	641,474	481,106	144,332	-2,528,054	1,486,162	1,743,807	-784,247	7,957,000	251,000	4,525,000	12,733,000
Mar-19	2019	175,148	226,400	433,244	433,244	445,541	432,528	253,072	685,600	12,273,000	2,555,000	491,720	33,656	525,376	20,568	1,371	547,315	410,487	123,146	-2,556,199	1,439,523	1,695,941	-860,258	7,957,000	251,000	4,525,000	12,733,000
Apr-19	2019	208,665	205,180	445,541	445,541	487,184	442,609	253,916	696,525	12,273,000	2,555,000	472,730	34,090	506,819	20,896	1,393	529,108	396,831	119,049	-2,597,639	1,450,786	1,707,897	-889,742	7,957,000	251,000	4,525,000	12,733,000
May-19	2019	193,928	261,560	487,184	487,184	449,606	478,974	181,432	660,406	12,273,000	2,555,000	465,228	31,594	496,822	19,812	1,321	517,955	388,466	116,540	-2,623,499	1,508,593	1,767,700	-855,799	7,957,000	251,000	4,525,000	12,733,000
Jun-19	2019	283,130	134,780	449,606	449,606	738,377	442,454	146,163	588,617	12,273,000	2,555,000	443,252	28,152	471,403	17,659	1,177	490,239	367,679	110,304	-2,623,031	1,537,389	1,798,311	-824,720	7,957,000	251,000	4,525,000	12,733,000
Jul-19	2019	343,600	363,081	738,377	738,377	776,377	716,553	238,107	954,660	14,303,000	2,571,000	575,249	45,480	620,729	28,640	1,909	651,278	488,459	146,538	-2,631,116	1,702,266	1,973,953	-657,163	7,957,000	251,000	4,525,000	12,733,000
Aug-19	2019	373,430	371,251	776,377	776,377	745,163	755,961	231,584	987,545	14,303,000	2,571,000	709,159	47,160	756,319	29,626	1,975	787,921	590,941	177,282	-2,643,496	1,788,157	2,071,357	-572,138	7,957,000	251,000	4,525,000	12,733,000
Sep-19	2019	335,936	377,531	745,163	745,163	855,583	757,804	290,229	1,048,033	14,303,000	2,571,000	846,833	51,175	898,008	31,441	2,096	931,545	698,659	209,598	-2,750,338	1,757,344	2,053,101	-697,237	7,957,000	251,000	4,525,000	12,733,000
Oct-19	2019	368,484	455,403	855,583	855,583	587,293	838,655	228,265	1,066,920	14,303,000	2,571,000	457,248	51,012	508,260	29,800	2,134	540,194	405,146	121,544	-2,741,145	2,120,463	2,423,949	-317,196	7,957,000	251,000	4,525,000	12,733,000

Month	Period	Run-of-River abstraction kl/m	Other Raw Water Resource / Purchased kl/m	Total Raw Water Abstraction kl/m	Total Raw water input to all WTWs kl/m	Total bulk water prior to treatment kl/m	Treated water (after all WTWs) kl/m	Bulk treated water purchased kl/m	System Input Volume kl/m	Total expenditure on raw water-Rands kl/m	Total water revenue collected-Rands kl/m	Billed Metered Consumption kl/m	Billed Un-Metered Consumption kl/m
Nov-19	2019	321,484	234,113	587,293	587,293	554,362	574,209	274,111	848,320	14,303,000	2,571,000	643,137	41,412
Dec-19	2019	278,396	244,270	554,362	554,362	526,915	542,760	260,352	803,112	14,303,000	2,571,000	646,138	39,194
Jan-20	2020	278,888	248,027	526,915	526,915	452,106	516,548	277,494	794,042	14,303,000	2,571,000	536,957	38,861
Feb-20	2020	250,556	201,550	452,106	452,106	618,751	449,089	241,663	690,752	14,303,000	2,571,000	556,676	33,967
Mar-20	2020	279,179	339,572	618,751	618,751	274,123	619,421	266,602	886,023	14,303,000	2,571,000	588,641	42,793
Apr-20	2020	274,123	0	274,123	274,123	274,150	276,956	203,038	479,994	14,303,000	2,571,000	365,519	23,891
May-20	2020	274,150	0	274,150	274,150	519,982	272,999	242,751	515,750	14,303,000	2,571,000	384,081	25,392
Jun-20	2020	259,991	259,991	519,982	519,982	413,282	519,157	302,219	821,376	14,303,000	2,571,000	583,219	39,585
Jul-20	2020	331,178	82,104	413,282	413,282	490,273	427,428	195,344	622,772	11,202,000	2,585,000	497,451	30,740
Aug-20	2020	346,610	143,663	490,273	490,273	466,960	508,165	207,041	715,206	11,202,000	2,585,000	615,062	35,062
Sep-20	2020	344,550	122,410	466,960	466,960	466,960	465,201	267,307	732,508	11,202,000	2,585,000	558,374	35,294
Oct-20	2020	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Nov-20	2020	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Dec-20	2020	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Jan-21	2021	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Feb-21	2021	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Mar-21	2021	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Apr-21	2021	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
May-21	2021	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698
Jun-21	2021	340,779	116,059	456,838	456,838	474,731	466,931	223,230	690,162	11,202,000	2,585,000	556,962	33,698

Month	Billed Consumption kl/m	Un-Billed Metered Consumption kl/m	Unbilled Unmetered Consumption kl/m	Total water consumption kl/m	Proposed total wastewater treated kl/m	possible Water re-use for irrigation kl/m	Treatment losses (12 month) kl/y	Reticulation Water Loss (12 Month) kl/y	Non-Revenue Water (12 Month) kl/y	Total non-revenue water (12 month) kl/y	Run-of-River abstraction kl/y	Groundwater abstraction kl/y	Other water resource / purchased kl/y	Total allocation kl/y
Nov-19	684,549	25,450	1,697	711,695	533,771	160,131	-2,764,390	2,164,014	2,471,275	-293,115	7,957,000	251,000	4,525,000	12,733,000
Dec-19	685,332	24,093	1,606	711,031	533,273	159,982	-2,762,474	2,075,032	2,385,230	-377,244	7,957,000	251,000	4,525,000	12,733,000
Jan-20	575,818	23,821	1,588	601,227	450,921	135,276	-2,790,925	2,222,035	2,536,085	-254,841	7,957,000	251,000	4,525,000	12,733,000
Feb-20	590,643	20,723	1,382	612,747	459,560	137,868	-2,772,779	2,192,276	2,504,454	-268,325	7,957,000	251,000	4,525,000	12,733,000
Mar-20	631,434	26,581	1,772	659,787	494,840	148,452	-2,787,696	2,280,227	2,598,819	-188,877	7,957,000	251,000	4,525,000	12,733,000
Apr-20	389,410	14,400	960	404,770	303,577	91,073	-2,742,583	2,188,035	2,499,698	-242,886	7,957,000	251,000	4,525,000	12,733,000
May-20	409,473	15,473	1,032	425,977	319,483	95,845	-2,810,962	2,135,357	2,442,391	-368,571	7,957,000	251,000	4,525,000	12,733,000
Jun-20	622,804	24,641	1,643	649,088	486,816	146,045	-2,973,345	2,209,267	2,523,749	-449,596	7,957,000	251,000	4,525,000	12,733,000
Jul-20	528,190	18,683	1,246	548,119	411,089	123,327	-2,966,552	1,980,539	2,284,400	-682,152	7,957,000	251,000	4,525,000	12,733,000
Aug-20	650,124	21,456	1,430	673,011	504,758	151,427	-2,980,317	1,823,110	2,118,256	-862,060	7,957,000	251,000	4,525,000	12,733,000
Sep-20	593,668	21,975	1,465	617,108	462,831	138,849	-2,942,995	1,822,022	2,107,071	-835,924	7,957,000	251,000	4,525,000	12,733,000
Oct-20	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Nov-20	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Dec-20	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Jan-21	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Feb-21	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Mar-21	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Apr-21	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
May-21	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000
Jun-21	590,660	20,704	1,380	612,744	459,558	137,867	-2,963,288	1,875,223	2,169,909	-793,378	7,957,000	251,000	4,525,000	12,733,000

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
135,874	17 ,407	17407	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-86,421	152,031
135,874	17 ,407	17407	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	12,141	278,126
135,874	17 ,407	17407	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	148,629	400,740
135,874	17 ,407	17407	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	162,616	459,591
135,874	17 ,407	17407	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	34,880	327,383
135,874	17 ,407	17407	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-77,024	219,580
135,874	18, 054	18054	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	82,523	468,467
135,874	18, 054	18054	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-57,216	308,441
135,874	18, 054	18054	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-285,583	28,338
135,874	18, 054	18054	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-352,215	110,418
135,874	18, 054	18054	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-418,319	-99,801
135,874	18, 054	18054	35,190	95,112	5,570	29,200	19,692	9,508	18,885	16,864	4,500	0,62	-303,435	-2,095
200,524	18, 054	18054	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	-34,666	238,116
200,524	18, 054	18054	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	39,479	290,081
200,524	18, 054	18054	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	40,289	260,295
200,524	18, 054	18054	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	23,502	330,293
200,524	18, 054	18054	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	42,708	320,430
200,524	18, 054	18054	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	232,720	582,755
200,524	18, 930	18930	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	47,604	425,497
200,524	18, 930	18930	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	-221,508	167,505
200,524	18, 930	18930	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	-175,294	168,521
200,524	18, 930	18930	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	-229,025	97,826
200,524	18, 930	18930	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	-249,024	1,942
200,524	18, 930	18930	33,200	97,775	5,726	36,413	32,918	3,495	15,750	16,864	4,500	0,62	-151,365	158,690
141,095	18, 930	18930	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-85,926	149,762
141,095	18, 930	18930	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	115,831	395,040

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
141,095	18, 930	18930	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	22,184	322,816
141,095	18, 930	18930	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	132,568	430,863
141,095	18, 930	18930	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	30,002	364,770
141,095	18, 930	18930	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-96,204	284,252
141,095	19, 793	19793	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-230,212	126,474
141,095	19, 793	19793	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-196,735	222,388
141,095	19, 793	19793	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-167,544	251,362
141,095	19, 793	19793	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-249,932	207,581
141,095	19, 793	19793	34,129	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-345,654	-7,249
141,095	19, 793	19793	39,308	101,704	5,110	42,756	35,724	7,032	18,000	18,101	4,500	0,62	-207,993	77,229
142,250	19, 793	19793	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	25,305	284,169
142,250	19, 793	19793	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	104,576	365,010
142,250	19, 793	19793	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	-44,029	214,740
142,250	19, 793	19793	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	98,660	345,319
142,250	19, 793	19793	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	8,216	286,756
142,250	19, 793	19793	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	38,693	344,913
142,250	20, 865	20865	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	87,415	406,577
142,250	20, 865	20865	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	-364,752	-83,094
142,250	20, 865	20865	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	-291,812	34,797
142,250	20, 865	20865	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	-683,529	-415,796
142,250	20, 865	20865	39,308	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	-530,983	-288,446
142,250	20, 865	20865	35,544	101,704	5,110	43,010	32,620	10,390	16,651	43,358	7,819	0,62	-461,507	-461,507
146,126	20, 865	20865	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-82,504	135,569
146,126	20, 865	20865	35,544	112,583	7,552	39,000	29,000	10,000	16,651	17,828	10,135	0,59	29,187	299,654
146,126	20, 865	20865	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-35,662	177,893
146,126	20, 865	20865	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	90,168	307,548

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
146,126	20, 865	20865	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-83,789	215,150
146,126	20, 865	20865	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	71,155	399,428
146,126	21, 408	21408	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	180,430	570,843
146,126	21, 408	21408	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-139,469	211,582
146,126	21, 408	21408	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-232,232	173,688
146,126	21, 408	21408	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-759,262	-476,267
146,126	21, 408	21408	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-436,986	-161,089
146,126	21, 408	21408	35,544	112,583	7,652	39,000	29,000	10,000	16,651	17,828	10,135	0,59	-442,026	-215,175
155,728	21, 408	21408	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	157,928	511,264
155,728	21, 408	21408	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	132,432	479,251
155,728	21, 408	21408	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	176,879	480,934
155,728	21, 408	21408	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	51,946	404,514
155,728	21, 408	21408	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	-21,996	276,244
155,728	21, 408	21408	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	40,057	301,405
155,728	21, 175	21175	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	61,151	580,149
155,728	21, 175	21175	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	-47,386	362,816
155,728	21, 175	21175	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	-44,014	356,855
155,728	21, 175	21175	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	13,777	336,629
155,728	21, 175	21175	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	-77,048	290,026
155,728	21, 175	21175	35,545	112,533	7,654	44,953	32,624	6,400	16,888	18,477	10,135	0,625	92,153	343,930
153,088	21, 175	21175	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	40,242	428,576
153,088	21, 175	21175	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-69,502	125,725
153,088	21, 175	21175	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-59,258	259,229
153,088	21, 175	21175	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	88,673	400,279
153,088	21, 175	21175	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	113,722	461,358
153,088	21, 175	21175	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	66,384	447,879

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
153,088	21, 492	21492	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-20,855	375,447
153,088	21, 492	21492	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-230,820	146,792
153,088	21, 492	21492	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-189,855	80,272
153,088	21, 492	21492	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-382,470	-49,509
153,088	21, 492	21492	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-427,371	-48,924
153,088	21, 492	21492	35,545	112,533	7,654	45,695	24,416	2,444	17,608	18,477	13,500	0,625	-310,352	-40,757
156,263	21, 492	21492	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-301,010	130,021
156,263	21, 492	21492	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-130,754	329,646
156,263	21, 492	21492	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-229,240	52,457
156,263	21, 492	21492	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-107,831	221,472
156,263	21, 492	21492	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-94,722	280,947
156,263	21, 492	21492	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-150,563	270,511
156,263	22, 541	22541	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-239,439	124,382
156,263	22, 541	22541	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-309,520	132,619
156,263	22, 541	22541	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-261,867	153,441
156,263	22, 541	22541	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-430,667	-65,678
156,263	22, 541	22541	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-228,595	11,376
156,263	22, 541	22541	35,545	112,533	7,654	46,446	24,554	13,000	18,408	18,946	14,903	0,62	-176,580	78,814
160,078	22, 541	22541	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-29,929	272,847
160,078	22, 541	22541	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	56,699	453,831
160,078	22, 541	22541	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	20,876	389,459
160,078	22, 541	22541	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-34,479	311,978
160,078	22, 541	22541	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	58,342	456,831
160,078	22, 541	22541	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	33,981	436,855
160,078	22, 892	22892	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-78,692	325,657
160,078	22, 892	22892	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-362,475	98,044

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
160,078	22, 892	22892	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-183,086	208,608
160,078	22, 892	22892	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-393,503	9,584
160,078	22, 892	22892	44,280	112,720	5,190	47,219	15,000	15,000	37,804	34,954	6,231	0,62	-456,719	-139,792
160,078	22, 892	22892	44,280	112,720	5,190	47,219	15,000	16,112	37,804	34,954	6,231	0,62	-417,487	-128,410
162,198	22, 892	22892	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-146,085	146,769
162,198	22, 892	22892	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-80,491	126,112
162,198	22, 892	22892	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-130,973	154,156
162,198	22, 892	22892	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	10,046	356,343
162,198	22, 892	22892	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-23,314	372,875
162,198	22, 892	22892	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-273,490	107,141
162,198	23, 334	23334	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-264,287	190,286
162,198	23, 334	23334	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-272,602	174,430
162,198	23, 334	23334	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-206,750	274,413
162,198	23, 334	23334	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-79,950	328,062
162,198	23, 334	23334	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-35,441	329,207
162,198	23, 334	23334	41,195	124,861	7,140	48,008	16,112	16,112	47,594	51,581	8,491	0,62	-12,476	134,757
179,523	23, 334	23334	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-36,823	234,396
178,523	23, 334	23334	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-63,045	205,870
179,523	23, 334	23334	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-83,742	268,183
179,523	23, 334	23334	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-39,240	161,997
179,523	23, 334	23334	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-105,219	251,194
179,523	23, 334	23334	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-171,359	142,671
179,523	24, 384	24384	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-170,611	230,466
179,523	24, 384	24384	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-244,752	59,870
179,523	24, 384	24384	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-78,221	301,831
179,523	24, 384	24384	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-134,102	111,237

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
179,523	24, 384	24384	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-79,931	180,961
179,523	24, 384	24384	40,500	123,338	9,719	52,374	34,071	13,000	37,939	41,585	9,238	0,63	-158,755	108,659
169,441	24, 384	24384	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-68,358	167,984
169,441	24, 384	24384	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-101,023	137,854
169,441	24, 384	24384	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-137,318	97,933
169,441	24, 384	24384	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-85,150	160,746
169,441	24, 384	24384	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-150,871	76,696
169,441	24, 384	24384	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-204,205	30,226
169,441	24 ,623	24623	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-52,885	224,476
169,441	24 ,623	24623	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-65,090	179,393
169,441	24 ,623	24623	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-45,473	184,923
169,441	24 ,623	24623	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-61,190	156,154
169,441	24 ,623	24623	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-68,708	84,645
169,441	24 ,623	24623	41,354	125,042	10,123	52,374	26,506	25,868	37,939	41,623	9,699	0,64	-75,683	69,582
186,730	24 ,623	24623	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-74,778	138,504
186,730	24 ,623	24623	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-90,349	113,733
186,730	24 ,623	24623	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-54,611	147,301
186,730	24 ,623	24623	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-65,178	163,607
186,730	24 ,623	24623	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-147,556	93,074
186,730	24 ,623	24623	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-70,206	181,063
186,730	24, 612	24612	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-193,585	45,811
186,730	24, 612	24612	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-151,841	107,765
186,730	24, 612	24612	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-114,787	138,284
186,730	24, 612	24612	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-86,499	167,417
186,730	24, 612	24612	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-38,981	142,451
186,730	24, 612	24612	43,478	133,357	14,376	52,374	34,071	18,303	39,224	47,018	11,327	0,65	-47,785	98,378

Population size	Population size	Student population	Population under 15	Population 15 to 65	Population over 65	Total no. of households	Formal dwellings	Informal dwellings	Flush toilet connected to sewerage	Piped water inside dwelling	Using public tap	Gini coefficient	H-R shortfall kl/m	J-R shortfall kl/m
189,746	24, 612	24612	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	65,275	303,382
189,746	24, 612	24612	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-31,960	199,624
189,746	24, 612	24612	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-173,741	116,488
189,746	24, 612	24612	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	298,461	526,726
189,746	24, 612	24612	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-137,486	136,625
189,746	24, 612	24612	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-168,271	92,081
189,746	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-84,680	192,814
189,746	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-163,658	78,005
189,746	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-40,366	226,236
189,746	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-127,813	75,225
189,746	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-152,978	89,773
189,746	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,884	47,660	49,493	12,884	0,61	-129,931	172,288
192,879	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-120,690	74,654
192,879	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-164,846	42,195
192,879	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-151,907	115,400
192,879	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	24, 597	24597	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	25 ,725	25725	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	25 ,725	25725	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	25 ,725	25725	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	25 ,725	25725	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	25 ,725	25725	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418
192,879	25 ,725	25725	43,095	139,312	10,472	52,374	36,214	12,135	44,991	46,619	12,135	0,65	-145,813	77,418

D3. StellRRA.csv

```
Date,RiverRunOffAbsklm
7/1/2006,404000
8/1/2006,455000
9/1/2006,697000
10/1/2006,529664
11/1/2006,458241
12/1/2006,447574
1/1/2007,572560
2/1/2007,548431
3/1/2007,548851
4/1/2007,396150
5/1/2007,362503
6/1/2007,339402
7/1/2007,592850
8/1/2007,657730
9/1/2007,627000
10/1/2007,404000
11/1/2007,455000
12/1/2007,697000
1/1/2008,498000
2/1/2008,530000
3/1/2008,575000
4/1/2008,506000
5/1/2008,490000
6/1/2008,419000
7/1/2008,455657
8/1/2008,433826
9/1/2008,426518
10/1/2008,529664
11/1/2008,458241
12/1/2008,447574
1/1/2009,544000
2/1/2009,557000
3/1/2009,564000
4/1/2009,396150
5/1/2009,362503
6/1/2009,339402
7/1/2009,477887
8/1/2009,418792
9/1/2009,460830
10/1/2009,529664
11/1/2009,458241
12/1/2009,447574
1/1/2010,583239
```

2/1/2010,475722
3/1/2010,480403
4/1/2010,185450
5/1/2010,193508
6/1/2010,234205
7/1/2010,477887
8/1/2010,418792
9/1/2010,460830
10/1/2010,529664
11/1/2010,458241
12/1/2010,447574
1/1/2011,583239
2/1/2011,475722
3/1/2011,480403
4/1/2011,185450
5/1/2011,193508
6/1/2011,234205
7/1/2011,530373
8/1/2011,469064
9/1/2011,595505
10/1/2011,480806
11/1/2011,470591
12/1/2011,466005
1/1/2012,460440
2/1/2012,477297
3/1/2012,424469
4/1/2012,424469
5/1/2012,340477
6/1/2012,466005
7/1/2012,333995
8/1/2012,353563
9/1/2012,353623
10/1/2012,360184
11/1/2012,412607
12/1/2012,433259
1/1/2013,430623
2/1/2013,310192
3/1/2013,231640
4/1/2013,216303
5/1/2013,221631
6/1/2013,205980
7/1/2013,163722
8/1/2013,146477
9/1/2013,169776
10/1/2013,291819
11/1/2013,305899

12/1/2013,369033
1/1/2014,376982
2/1/2014,475626
3/1/2014,425117
4/1/2014,386594
5/1/2014,416427
6/1/2014,382637
7/1/2014,409230
8/1/2014,434402
9/1/2014,442068
10/1/2014,535226
11/1/2014,508372
12/1/2014,545614
1/1/2015,551367
2/1/2015,477499
3/1/2015,628306
4/1/2015,416735
5/1/2015,296454
6/1/2015,292937
7/1/2015,359857
8/1/2015,374783
9/1/2015,340140
10/1/2015,437181
11/1/2015,453707
12/1/2015,435176
1/1/2016,480257
2/1/2016,491587
3/1/2016,462244
4/1/2016,405251
5/1/2016,366784
6/1/2016,334330
7/1/2016,333920
8/1/2016,351082
9/1/2016,335936
10/1/2016,368484
11/1/2016,414899
12/1/2016,377209
1/1/2017,389055
2/1/2017,371083
3/1/2017,373214
4/1/2017,322059
5/1/2017,313212
6/1/2017,267188
7/1/2017,284047
8/1/2017,267853
9/1/2017,253737

10/1/2017,358092
11/1/2017,477648
12/1/2017,418149
1/1/2018,230275
2/1/2018,158741
3/1/2018,176378
4/1/2018,174088
5/1/2018,213768
6/1/2018,352491
7/1/2018,390254
8/1/2018,409348
9/1/2018,437751
10/1/2018,513451
11/1/2018,472902
12/1/2018,312332
1/1/2019,354563
2/1/2019,274550
3/1/2019,175148
4/1/2019,208665
5/1/2019,193928
6/1/2019,283130
7/1/2019,343600
8/1/2019,373430
9/1/2019,335936
10/1/2019,368484
11/1/2019,321484
12/1/2019,278396
1/1/2020,278888
2/1/2020,250556
3/1/2020,279179
4/1/2020,274123
5/1/2020,274150
6/1/2020,259991
7/1/2020,331178
8/1/2020,346610
9/1/2020,344550
10/1/2020,340779
11/1/2020,340779
12/1/2020,340779
1/1/2021,340779
2/1/2021,340779
3/1/2021,340779
4/1/2021,340779
5/1/2021,340779
6/1/2021,340779

D4. StellWaterClimate2.csv

```
Date,RoRabs,mtmin,mtmax, spre
7/1/2006,404000,8.658064516,16.92258065,12.79032258,71.4
8/1/2006,455000,7.932258065,17.73870968,12.83548387,56.2
9/1/2006,697000,10.32333333,20.87333333,15.59833333,20
10/1/2006,529664,11.27419355,22.38064516,16.82741935,37.2
11/1/2006,458241,13.90666667,24.55333333,19.23,37.7
12/1/2006,447574,15.42580645,25.00322581,20.21451613,10
1/1/2007,572560,17.60322581,28.15483871,22.87903226,0.5
2/1/2007,548431,16.19285714,26.36071429,21.27678571,27.3
3/1/2007,548851,14.17741935,26.46129032,20.31935484,18.6
4/1/2007,396150,12.83,24.02333333,18.42666667,65.6
5/1/2007,362503,9.7,21.08064516,15.39032258,96
6/1/2007,339402,8.1,17.84,12.97,123.4
7/1/2007,592850,6.909677419,17.62903226,12.26935484,151.5
8/1/2007,657730,8.196774194,17.7483871,12.97258065,101.5
9/1/2007,627000,9.153333333,19.82,14.48666667,18.2
10/1/2007,404000,12.13225806,23.32903226,17.73064516,18.7
11/1/2007,455000,12.78666667,22.20333333,17.495,40.8
12/1/2007,697000,15.70967742,26.33870968,21.02419355,18.5
1/1/2008,498000,16.78709677,26.46451613,21.62580645,6.8
2/1/2008,530000,16.86896552,26.55862069,21.7137931,13.9
3/1/2008,575000,14.7516129,26.59032258,20.67096774,5.2
4/1/2008,506000,12.23666667,24.09333333,18.165,15.2
5/1/2008,490000,12.9,21.43870968,17.16935484,51.4
6/1/2008,419000,9.806666667,17.73,13.76833333,63.2
7/1/2008,455657,7.5,16.71935484,12.10967742,182.4
8/1/2008,433826,7.690322581,18.3,12.99516129,#N/A
9/1/2008,426518,#N/A,#N/A,#N/A,137.8
10/1/2008,529664,10.96774194,21.97741935,16.47258065,12.4
11/1/2008,458241,13.98333333,23.17666667,18.58,53.1
12/1/2008,447574,16.00967742,25.49677419,20.75322581,7.8
1/1/2009,544000,16.38709677,26.2483871,21.31774194,1.4
2/1/2009,557000,16.98214286,28.075,22.52857143,3.6
3/1/2009,564000,15.73548387,26.90322581,21.31935484,0.8
4/1/2009,396150,13.23333333,23.95,18.59166667,24
5/1/2009,362503,10.60967742,20.33548387,15.47258065,64.4
6/1/2009,339402,9.473333333,18.58,14.02666667,108.4
7/1/2009,477887,7.748387097,19.66774194,13.70806452,88.4
8/1/2009,418792,8.651612903,18.7483871,13.7,52
9/1/2009,460830,9.753333333,19.09,14.42166667,60.2
10/1/2009,529664,12.54516129,23.03870968,17.79193548,31.6
```

11/1/2009,458241,14.08666667,24.11,19.09833333,86.2
12/1/2009,447574,15.1516129,24.92258065,20.03709677,4.4
1/1/2010,583239,17.15483871,26.65806452,21.90645161,3.4
2/1/2010,475722,17.125,27.46785714,22.29642857,7.9
3/1/2010,480403,15.90967742,26.79032258,21.35,6.4
4/1/2010,185450,12.02666667,23.04,17.53333333,12
5/1/2010,193508,10.25806452,19.77096774,15.01451613,95.4
6/1/2010,234205,7.52,18.56,13.04,70.2
7/1/2010,477887,6.090322581,18.23225806,12.16129032,40.3
8/1/2010,418792,7.541935484,19.28709677,13.41451613,32.2
9/1/2010,460830,9.45,20.03333333,14.74166667,24.4
10/1/2010,529664,11.09032258,21.76774194,16.42903226,31
11/1/2010,458241,13.22,23.62333333,18.42166667,27.8
12/1/2010,447574,16.26451613,26.91935484,21.59193548,18
1/1/2011,583239,16.36129032,27.83870968,22.1,6.2
2/1/2011,475722,17.86785714,28.61428571,23.24107143,3.2
3/1/2011,480403,15.69354839,26.82580645,21.25967742,6
4/1/2011,185450,11.67,23.39333333,17.53166667,27.8
5/1/2011,193508,10.79032258,20.32903226,15.55967742,59.6
6/1/2011,234205,8.27,17.68666667,12.97833333,84.8
7/1/2011,530373,6.95483871,19.09677419,13.02580645,25
8/1/2011,469064,6.403225806,19.00645161,12.70483871,53.4
9/1/2011,595505,9.02,19.21666667,14.11833333,23.8
10/1/2011,480806,10.93870968,21.75483871,16.34677419,12.4
11/1/2011,470591,12.51666667,22.29333333,17.405,27.6
12/1/2011,466005,14.98709677,24.28387097,19.63548387,18.6
1/1/2012,460440,17.89354839,28.29032258,23.09193548,3
2/1/2012,477297,16.73103448,26.69655172,21.7137931,5.6
3/1/2012,424469,16.09677419,26.13870968,21.11774194,23
4/1/2012,424469,12.51,23.00666667,17.75833333,44
5/1/2012,340477,8.938709677,19.51290323,14.22580645,39.8
6/1/2012,466005,8.063333333,17.9,12.98166667,78.2
7/1/2012,333995,7.470967742,17.3,12.38548387,92
8/1/2012,353563,7.393548387,16.40645161,11.9,82
9/1/2012,353623,8.466666667,19.04333333,13.755,55.2
10/1/2012,360184,11.91935484,21.12258065,16.52096774,34.2
11/1/2012,412607,13.25333333,23.74666667,18.5,8.2
12/1/2012,433259,17.22903226,27.35483871,22.29193548,1
1/1/2013,430623,17.06451613,26.53870968,21.8016129,12.6
2/1/2013,310192,17.225,26.36071429,21.79285714,37.4
3/1/2013,231640,16.17096774,26.15806452,21.16451613,14.4
4/1/2013,216303,10.71,23.18,16.945,36.4
5/1/2013,221631,10.12258065,21.06451613,15.59354839,53.6

6/1/2013,205980,8.01,17.56,12.785,115.4
7/1/2013,163722,8.503225806,18.22903226,13.36612903,43.6
8/1/2013,146477,8.035483871,17.47419355,12.75483871,168
9/1/2013,169776,8.863333333,17.25,13.05666667,68
10/1/2013,291819,12.26451613,21.0516129,16.65806452,16
11/1/2013,305899,14.39,23.74,19.065,85.2
12/1/2013,369033,16.60967742,27.03870968,21.82419355,4.8
1/1/2014,376982,17.88709677,27.03548387,22.46129032,23.2
2/1/2014,475626,18.06428571,28.20357143,23.13392857,2.2
3/1/2014,425117,14.94516129,24.00967742,19.47741935,43.8
4/1/2014,386594,13.00666667,25.52333333,19.265,24
5/1/2014,416427,10.53870968,20.2483871,15.39354839,61
6/1/2014,382637,7.793333333,17.97333333,12.88333333,109.4
7/1/2014,409230,7.219354839,17.34193548,12.28064516,105.6
8/1/2014,434402,9.661290323,19.18387097,14.42258065,91.4
9/1/2014,442068,#N/A,#N/A,#N/A,27.6
10/1/2014,535226,12.11612903,24.98064516,18.5483871,4.8
11/1/2014,508372,14.06,24.28,19.17,21.2
12/1/2014,545614,16.21290323,25.81612903,21.01451613,2.6
1/1/2015,551367,17.16451613,27.47741935,22.32096774,13.6
2/1/2015,477499,#N/A,26.27857143,#N/A,3
3/1/2015,628306,15.5,26.86129032,21.18064516,1.6
4/1/2015,416735,12.05,23.94666667,17.99833333,4.4
5/1/2015,296454,10.94193548,21.08709677,16.01451613,27
6/1/2015,292937,7.34,17.07666667,12.20833333,106.6
7/1/2015,359857,7.316129032,16.50967742,11.91290323,87.8
8/1/2015,374783,9.190322581,18.41290323,13.8016129,35.6
9/1/2015,340140,10.14666667,21.01666667,15.58166667,21.4
10/1/2015,437181,12.54516129,#N/A,#N/A,5.4
11/1/2015,453707,13.45666667,24.47,18.96333333,25.6
12/1/2015,435176,16.67419355,26.96774194,21.82096774,16.2
1/1/2016,480257,18.63548387,29.72580645,24.18064516,9.4
2/1/2016,491587,16.97931034,27.57241379,22.27586207,3.2
3/1/2016,462244,14.85483871,25.5516129,20.20322581,35.6
4/1/2016,405251,12.59,23.64666667,18.11833333,48
5/1/2016,366784,10.03870968,#N/A,#N/A,17
6/1/2016,334330,7.653333333,18.22666667,12.94,58
7/1/2016,333920,6.996774194,17.98064516,12.48870968,89
8/1/2016,351082,8.603225806,19.99032258,14.29677419,60.8
9/1/2016,335936,9.29,19.99666667,14.64333333,29.6
10/1/2016,368484,11.46774194,23.08387097,17.27580645,9.2
11/1/2016,414899,14.63666667,25.12,19.87833333,3.6
12/1/2016,377209,15.91935484,26.46774194,21.19354839,19

1/1/2017,389055,16.41290323,26.6516129,21.53225806,3.6
2/1/2017,371083,17.05,27.83928571,22.44464286,0.6
3/1/2017,373214,14.80967742,26.64516129,20.72741935,7.6
4/1/2017,322059,13.02333333,26.60333333,19.81333333,19.4
5/1/2017,313212,9.729032258,22.44516129,16.08709677,5.6
6/1/2017,267188,7.48,17.70333333,12.59166667,86
7/1/2017,284047,6.238709677,18.28064516,12.25967742,36
8/1/2017,267853,8.222580645,18.1483871,13.18548387,48
9/1/2017,253737,9.45,21.18666667,15.31833333,16.8
10/1/2017,358092,10.3,22.13870968,16.21935484,19
11/1/2017,477648,13.13,24.12333333,18.62666667,33.2
12/1/2017,418149,15.4483871,26.46451613,20.95645161,7.8
1/1/2018,230275,17.28387097,27.69677419,22.49032258,5.4
2/1/2018,158741,16.24285714,27.92857143,22.08571429,8.8
3/1/2018,176378.063,14.81935484,25.1,19.95967742,9.2
4/1/2018,174088.063,12.32666667,23.74,18.03333333,54.4
5/1/2018,213768,11.20967742,21.37419355,16.29193548,66
6/1/2018,352491,9.333333333,18.61666667,13.975,77.4
7/1/2018,390254,8.083870968,20.28709677,14.18548387,49
8/1/2018,409348,6.922580645,17.25806452,12.09032258,55
9/1/2018,437751,8.816666667,18.79666667,13.80666667,65.4
10/1/2018,513451,13.3516129,25.71290323,19.53225806,9
11/1/2018,472902,12.94666667,24.76,18.85333333,8.8
12/1/2018,312332,15.83548387,25.97096774,20.90322581,14
1/1/2019,354563,15.83548387,26.74516129,21.29032258,10.6
2/1/2019,274550,#N/A,#N/A,#N/A,5.2
3/1/2019,175148,15.81290323,24.8483871,20.33064516,26
4/1/2019,208665,12.77,23.02666667,17.89833333,13.9
5/1/2019,193928.0315,10.68387097,21.59354839,16.13870968,29.3
6/1/2019,283129.5,#N/A,19.06,#N/A,58
7/1/2019,343600,9.061290323,17.25483871,13.15806452,89.8
8/1/2019,373430,8.164516129,18.72258065,13.44354839,26.7
9/1/2019,335936,11.39333333,22.51666667,16.955,11.3
10/1/2019,368484,11.4516129,21.67741935,16.56451613,62.9
11/1/2019,321484,14.57333333,23.83666667,19.205,0.5
12/1/2019,278396,14.79032258,24.9,19.84516129,13.5
1/1/2020,278888,17.01612903,26.40645161,21.71129032,6
2/1/2020,250556,16.75862069,27.42413793,22.09137931,4
3/1/2020,279179,14.96451613,25.67741935,20.32096774,1.1
4/1/2020,274123,11.63333333,23.68666667,17.66,23.3
5/1/2020,274150,10.29032258,21.47419355,15.88225806,40.2
6/1/2020,259991,9.28,19.40333333,14.34166667,70.6
7/1/2020,331178,7.122580645,19.08709677,13.10483871,68.4

8/1/2020,346610,7.1,17.11935484,12.10967742,77.2
9/1/2020,344550,8.7,19.04333333,13.87166667,50.6
10/1/2020,340779,11.25806452,21.53548387,16.39677419,7.7
11/1/2020,340779,13.58,23.2,18.39,29.1
12/1/2020,340779,15.67419355,25.1483871,20.41129032,3.4
1/1/2021,340779,16.81935484,27.08064516,21.95,3.4
2/1/2021,340779,16.91785714,26.36428571,21.64107143,1.4
3/1/2021,340779,15.00645161,25.37419355,20.19032258,27.4
4/1/2021,340779,13.25333333,24.47333333,18.86333333,1
5/1/2021,340779,10.1516129,20.36774194,15.25967742,70.4
6/1/2021,340779,9.55,20.13666667,14.84333333,84.4

Appendix E:
The Jupyter notebook

E1. Conventional models

Stellenbosch River Runoff Abstraction in kl/m Forecasting using Time Series Modeling

In [1]:
```
import pandas as pd
```

In [2]:
```
import numpy as np
```

In [3]:
```
import matplotlib.pyplot as plt
```

In [4]:
```
import seaborn as sns
```

In [5]:
```
import pmdarima as pm
```

In [6]:
```
import sklearn as sk
```

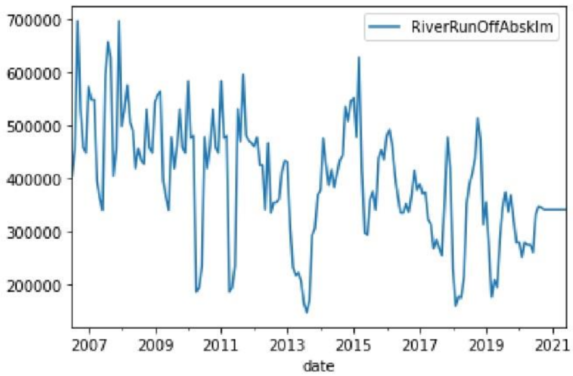
In [7]:
```
rrabs=pd.read_csv('StellRRA.csv',index_col='date',parse_dates=True)
```

In [8]:
```
rrabs.head()
```

Out[8]:

date	RiverRunOffAbsklm
2006-07-01	404000
2006-08-01	455000
2006-09-01	697000
2006-10-01	529664
2006-11-01	458241

In [9]:
```
fig, ax = plt.subplots()
rrabs.plot(ax=ax)
plt.show()
```

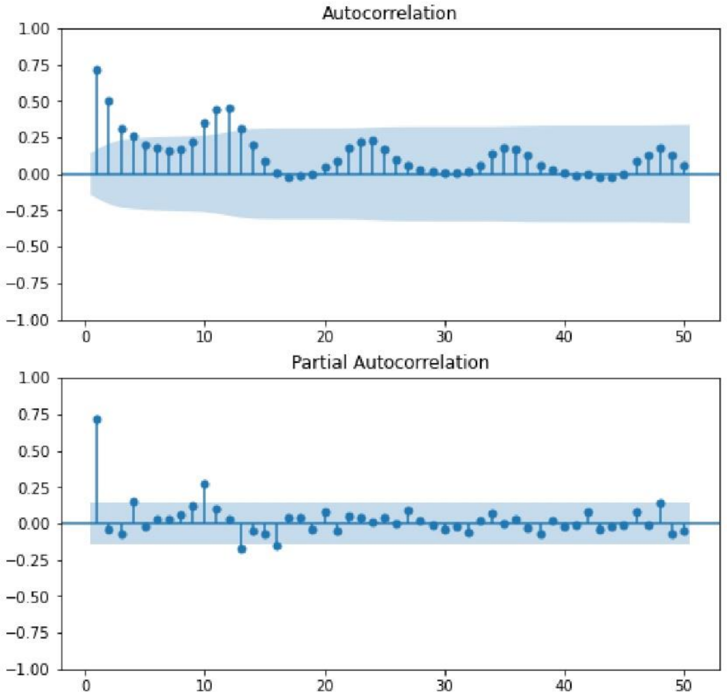


One important feature of a time series is its trend: there is a subtle negative trend where the values are decreasing over time

Another important feature is seasonality. A seasonal time series has patterns that repeat at regular intervals.

In [10]:
```
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
```

In [11]:
```
fig, (ax1, ax2) = plt.subplots(2,1, figsize=(8,8))
plot_acf(rrabs, lags=50, zero=False, ax=ax1)
plot_pacf(rrabs, method = 'ywm', lags=50, zero=False, ax=ax2)
plt.show()
```

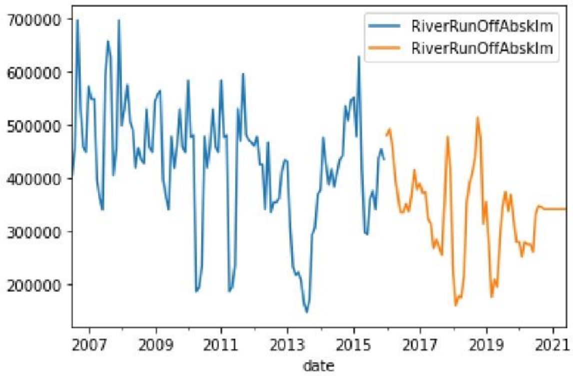


In [12]:
```
rrabs_train = rrabs.loc[:'2015']
```

In [13]:
```
rrabs_test = rrabs.loc['2016':]
```

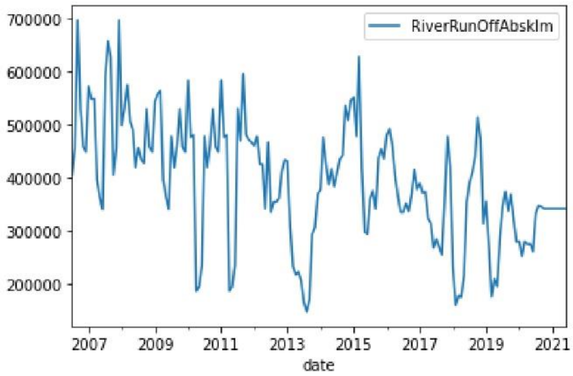
In time series modeling we use the past values to predict the future and so we need to split the data in time. We train on the data earlier in the time series (from 2006 to end of 2019) and test in the data that comes latter (from 2020 to mid 2021).

In [14]:
```
fig, ax = plt.subplots()
rrabs_train.plot(ax=ax)
rrabs_test.plot(ax=ax)
plt.show()
```



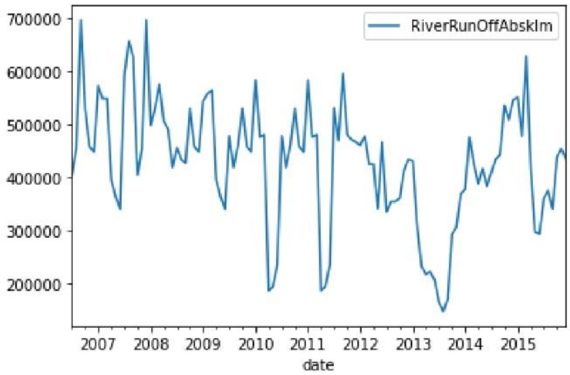
The series in blue is the training set and the amber series is the testing set against which the predictions are compared.

In [16]:
```
rrabs.plot()
plt.show()
```



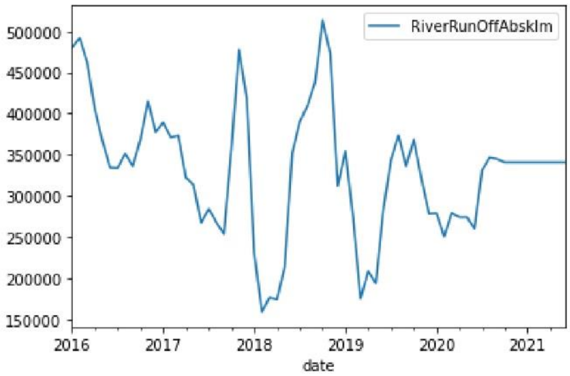
To model a time series, it must be stationary. Stationary means the distribution of the data does not change with time. The plot indicates a negative trend and nonconstant variance - so this time series is not stationary.

In [17]:
```
rrabs_train.plot()
plt.show()
```



In [18]:

```
rrabs_test.plot()
plt.show()
```



In [19]:

```
rrabs_test
```

Out[19]:

	RiverRunOffAbsklm
date	
2016-01-01	480257
2016-02-01	491587
2016-03-01	462244
2016-04-01	405251
2016-05-01	366784
...	...
2021-02-01	340779
2021-03-01	340779
2021-04-01	340779
2021-05-01	340779
2021-06-01	340779

66 rows × 1 columns

The most common test for identifying whether a time series is non-stationary is the augmented Dicky-Fuller test. This is a statistical test with a null hypothesis stating that your time series is non-stationary due to trend.

In [20]:
```
from statsmodels.tsa.stattools import adfuller
```

In [21]:
```
adfresults = adfuller(rrabs['RiverRunOffAbsklm'])
```

In [22]:
```
print(adfresults)
```

```
(-2.177207163726718, 0.21462168731149478, 12, 167, {'1%': -3.470126426071447, '5%': -2.879
0075987120027, '10%': -2.5760826967621644}, 4175.227469270265)
```

• The zeroth element (0th element) is the test statistic (-2.177) o The more negative this number is the more likely that the data is stationary. The next item in the results tuple is the test p-value.

• 1st element is p-value: (0.2146) o If p-value is smaller than 0.05 (p-value < 0.05) → reject null hypothesis and assume that our time series is stationary. Reject non-stationarity. So here we cannot reject the null hypothesis as the p-value > 0.05 - the time series is non-stationary. The last element in the tuple is a dictionary that stores the critical values of the test statistics which equate to different p-values. In this case, if we wanted a p-value of 0.05 or below, our test statistics needed to be below -2.879.

• 4th element is the critical test statistics We will ignore the rest of the test statistics

One very common way to make a time series stationary is to take it difference. This is where from each value in your time series, we subtract the previous value.

In [23]:
```
rrabs_stationary = rrabs.diff().dropna()
```

In [24]:
```
rrabs_stationary.head()
```

Out[24]:

date	RiverRunOffAbsklm
2006-08-01	51000.0
2006-09-01	242000.0
2006-10-01	-167336.0
2006-11-01	-71423.0
2006-12-01	-10667.0

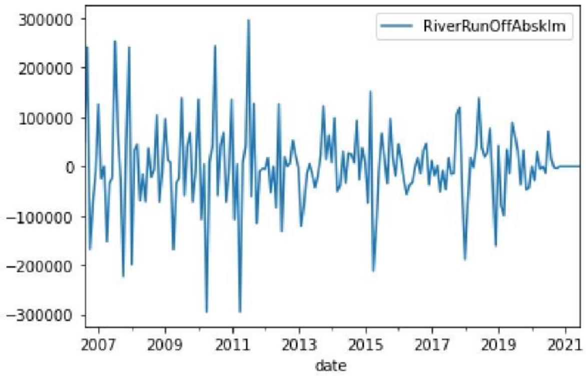
In [25]:
```
adfresults = adfuller(rrabs_stationary['RiverRunOffAbsklm'])
```

In [26]:
```
print(adfresults)
```

```
(-5.340814595569276, 4.504520893142619e-06, 11, 167, {'1%': -3.470126426071447, '5%': -2.8
790075987120027, '10%': -2.5760826967621644}, 4150.461464465352)
```

In [28]:
```
fig, ax = plt.subplots()
```

```
rrabs_stationary.plot(ax=ax)
plt.show()
```



The differenced time series rrabs_stationary is indeed stationary as confimed by the Augmented Dicky-Fuller test as well as the plot above

In [29]:
```
log_return = np.log(rrabs/rrabs.shift(1)).dropna()
```

In [30]:
```
log_return.head()
```

Out[30]:

	RiverRunOffAbsklm
date	
2006-08-01	0.118883
2006-09-01	0.426488
2006-10-01	-0.274543
2006-11-01	-0.144848
2006-12-01	-0.023553

In [31]:
```
result_log = adfuller(log_return['RiverRunOffAbsklm'])
```

In [32]:
```
print(result_log)
```

```
(-5.1427908931902895, 1.1491816647489372e-05, 11, 167, {'1%': -3.470126426071447, '5%': -
2.8790075987120027, '10%': -2.5760826967621644}, -52.18078702350789)
```

Notice that both the differenced and the log-return transformed time series have a small p-value, but the differenced time series has a much more negative test statistic. This means the differenced tranformation is better.

AR, MA and ARMA models

In an autoregressive model (AR(p)), we regress the values of the time series against previous values of this same time series.

AR(p) model: yt = a(1) y_(t-1) + a2 y(t-2) + ⋯ + ap y(t-p) + ϵ_t

This means we have p autoregressive coefficients and p independent variables of the series at each lag.

In a moving average (MA(q)) model, we regress the values of the time series against the previous shock values of this same time series.

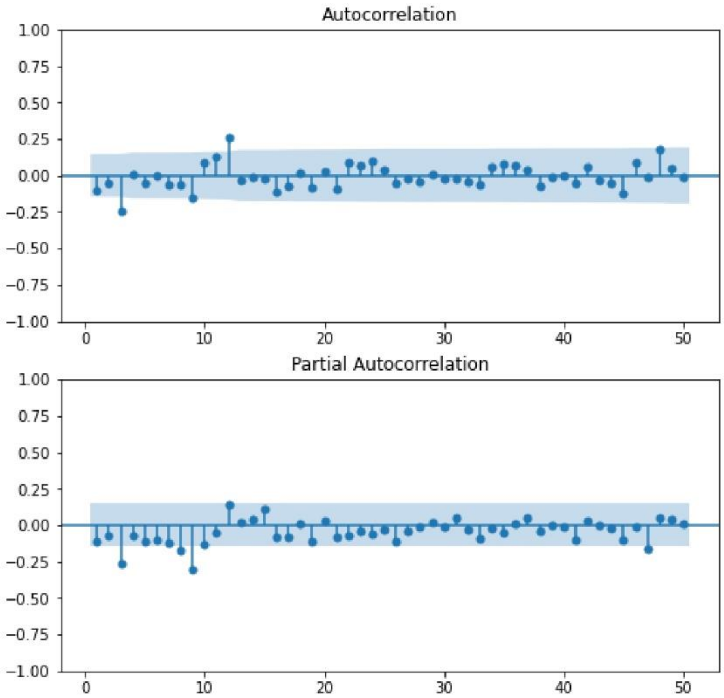
MA(q) model: $yt = m(1)\ \epsilon(t-1) + m\_2\ \epsilon(t-2) + \cdots + mq\ \epsilon(t-q) + \epsilon\_t$

This means we have q moving average coefficients and q shock variables, from previous q steps.

Autoregressive moving average (ARMA) model: An ARMA model is a combination of the AR and MA models. ARMA = AR + MA

$yt = a(1)\ y\_(t-1) + a2\ y(t-2) + \cdots + ap\ y(t-p) + m\_(1)\ \epsilon(t-1) + m\_2\ \epsilon(t-2) + \cdots + mq\ \epsilon(t-q) + \epsilon\_t$

In [33]:
```
fig, (ax1, ax2) = plt.subplots(2,1, figsize=(8,8))
plot_acf(rrabs_stationary, lags=50, zero=False, ax=ax1)
plot_pacf(rrabs_stationary, method = 'ywm', lags=50, zero=False, ax=ax2)
plt.show()
```



What is the ACF

One of the many ways to identify the correct order is by using the autocorrelation function the ACF and the partial autocorrelation function the PACF. The plots for the time series is shown above. The ACF or autocorrelation function at lag 1 is the correlation between the time series and the same time series off set by one step corr(*yt* ,*y*(t-1)). The autocorrelation at lag 2 is the correlation between the time series and itself off set by two steps corr(*yt* ,*y*(t-2)) and so on.

When we talk of the autocorrelation function, we talk of the set of autocorrelation values for different lags. The bars show the ACF values at increasing lags. If these values are small and fall within the blue shaded region then they are not statistically significant.

What is the PACF

The Partial autocorrelation function is the correlation between the time series and the lag version of itself after we subtract the effect of correlation at smaller lags. So it is the correlation associated with just that particular lag.

Using ACF and PACF to choose model order

By comparing the ACF and PACF for time series we can deduce the model order.

If the amplitude of the ACF tails off with increasing lag and the PACF cuts off after some lag p, then we have an AR(p) model.

If the amplitude of the ACF cuts off after some lag q and the amplitude of the PACF tails off, then we have an MA(q) model.

If both the ACF and PACF tail off then we have an ARMA(p, q) model. In this case, we cannot deduce the model orders of p and q from the plot.

This is the case here, both tail off and we can not deduce the values of p and q from the plots of the ACF and PACF!

In [34]:
```
from statsmodels.tsa.statespace.sarimax import SARIMAX
```

In [35]:
```
order_aic_bic =[]
for p in range(5):
    for q in range(5):
        model = SARIMAX(rrabs_train, order=(p,1,q))
        results = model.fit()
        order_aic_bic.append((p, 1, q, results.aic, results.bic))
        order_df = pd.DataFrame(order_aic_bic, columns=['p', '1', 'q', 'aic', 'bic'])
        print(order_df.sort_values('aic'))
        print(order_df.sort_values('bic'))
        print(p, 1, q, None, None)
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
0  0  1  0  2923.001971  2925.729359
   p  1  q          aic          bic
0  0  1  0  2923.001971  2925.729359
0 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
0  0  1  0  2923.001971  2925.729359
```

```
1  0  1  1  2923.044906  2928.499681
   p  1  q          aic          bic
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
0 1 1 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
0 1 2 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
0 1 3 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
4  0  1  4  2907.694787  2921.331726
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
4  0  1  4  2907.694787  2921.331726
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
0 1 4 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
```

```
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
4  0  1  4  2907.694787  2921.331726
5  1  1  0  2922.832803  2928.287579
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
4  0  1  4  2907.694787  2921.331726
0  0  1  0  2923.001971  2925.729359
5  1  1  0  2922.832803  2928.287579
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
1 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
4  0  1  4  2907.694787  2921.331726
6  1  1  1  2911.557952  2919.740115
5  1  1  0  2922.832803  2928.287579
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
6  1  1  1  2911.557952  2919.740115
4  0  1  4  2907.694787  2921.331726
0  0  1  0  2923.001971  2925.729359
5  1  1  0  2922.832803  2928.287579
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
1 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
4  0  1  4  2907.694787  2921.331726
6  1  1  1  2911.557952  2919.740115
7  1  1  2  2913.632934  2924.542485
5  1  1  0  2922.832803  2928.287579
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
```

```
6  1  1  1  2911.557952  2919.740115
4  0  1  4  2907.694787  2921.331726
7  1  1  2  2913.632934  2924.542485
0  0  1  0  2923.001971  2925.729359
5  1  1  0  2922.832803  2928.287579
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
1 1 2 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
8  1  1  3  2907.302587  2920.939526
4  0  1  4  2907.694787  2921.331726
6  1  1  1  2911.557952  2919.740115
7  1  1  2  2913.632934  2924.542485
5  1  1  0  2922.832803  2928.287579
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
6  1  1  1  2911.557952  2919.740115
8  1  1  3  2907.302587  2920.939526
4  0  1  4  2907.694787  2921.331726
7  1  1  2  2913.632934  2924.542485
0  0  1  0  2923.001971  2925.729359
5  1  1  0  2922.832803  2928.287579
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
1 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\base\model.py:604: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
8  1  1  3  2907.302587  2920.939526
4  0  1  4  2907.694787  2921.331726
9  1  1  4  2908.786543  2925.150870
6  1  1  1  2911.557952  2919.740115
7  1  1  2  2913.632934  2924.542485
5  1  1  0  2922.832803  2928.287579
0  0  1  0  2923.001971  2925.729359
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
   p  1  q          aic          bic
3  0  1  3  2905.943935  2916.853486
```

```
6  1  1  1  2911.557952  2919.740115
8  1  1  3  2907.302587  2920.939526
4  0  1  4  2907.694787  2921.331726
7  1  1  2  2913.632934  2924.542485
9  1  1  4  2908.786543  2925.150870
0  0  1  0  2923.001971  2925.729359
5  1  1  0  2922.832803  2928.287579
1  0  1  1  2923.044906  2928.499681
2  0  1  2  2924.151590  2932.333753
1 1 4 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
6   1  1  1  2911.557952  2919.740115
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
0   0  1  0  2923.001971  2925.729359
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
2 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
6   1  1  1  2911.557952  2919.740115
```

```
11   2  1  1  2913.380890  2924.290441
7    1  1  2  2913.632934  2924.542485
5    1  1  0  2922.832803  2928.287579
0    0  1  0  2923.001971  2925.729359
1    0  1  1  2923.044906  2928.499681
2    0  1  2  2924.151590  2932.333753
10   2  1  0  2924.998547  2933.180710
     p  1  q          aic          bic
3    0  1  3  2905.943935  2916.853486
6    1  1  1  2911.557952  2919.740115
8    1  1  3  2907.302587  2920.939526
4    0  1  4  2907.694787  2921.331726
11   2  1  1  2913.380890  2924.290441
7    1  1  2  2913.632934  2924.542485
9    1  1  4  2908.786543  2925.150870
0    0  1  0  2923.001971  2925.729359
5    1  1  0  2922.832803  2928.287579
1    0  1  1  2923.044906  2928.499681
2    0  1  2  2924.151590  2932.333753
10   2  1  0  2924.998547  2933.180710
2 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
     p  1  q          aic          bic
3    0  1  3  2905.943935  2916.853486
8    1  1  3  2907.302587  2920.939526
4    0  1  4  2907.694787  2921.331726
9    1  1  4  2908.786543  2925.150870
6    1  1  1  2911.557952  2919.740115
12   2  1  2  2911.763148  2925.400087
11   2  1  1  2913.380890  2924.290441
7    1  1  2  2913.632934  2924.542485
5    1  1  0  2922.832803  2928.287579
0    0  1  0  2923.001971  2925.729359
1    0  1  1  2923.044906  2928.499681
2    0  1  2  2924.151590  2932.333753
10   2  1  0  2924.998547  2933.180710
     p  1  q          aic          bic
3    0  1  3  2905.943935  2916.853486
6    1  1  1  2911.557952  2919.740115
8    1  1  3  2907.302587  2920.939526
4    0  1  4  2907.694787  2921.331726
11   2  1  1  2913.380890  2924.290441
7    1  1  2  2913.632934  2924.542485
9    1  1  4  2908.786543  2925.150870
12   2  1  2  2911.763148  2925.400087
0    0  1  0  2923.001971  2925.729359
5    1  1  0  2922.832803  2928.287579
1    0  1  1  2923.044906  2928.499681
2    0  1  2  2924.151590  2932.333753
10   2  1  0  2924.998547  2933.180710
2 1 2 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
```

```
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
2 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
14  2  1  4  2910.070054  2929.161769
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
```

```
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
2 1 4 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
14  2  1  4  2910.070054  2929.161769
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
3 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
```

```
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
14  2  1  4  2910.070054  2929.161769
16  3  1  1  2910.831552  2924.468491
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
3 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
16  3  1  1  2910.831552  2924.468491
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
```

```
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
3 1 2 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
5   1  1  0  2922.832803  2928.287579
```

```
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
3 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
19  3  1  4  2911.991523  2933.810625
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
19  3  1  4  2911.991523  2933.810625
3 1 4 None None
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
19  3  1  4  2911.991523  2933.810625
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
19  3  1  4  2911.991523  2933.810625
4 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
```

```
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
19  3  1  4  2911.991523  2933.810625
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
19  3  1  4  2911.991523  2933.810625
4 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\base\model.py:604: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
```

```
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
22  4  1  2  2910.482981  2929.574696
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
19  3  1  4  2911.991523  2933.810625
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
22  4  1  2  2910.482981  2929.574696
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
19  3  1  4  2911.991523  2933.810625
4 1 2 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\base\model.py:604: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
```

```
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
22  4  1  2  2910.482981  2929.574696
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
19  3  1  4  2911.991523  2933.810625
23  4  1  3  2912.112933  2933.932036
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
22  4  1  2  2910.482981  2929.574696
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
19  3  1  4  2911.991523  2933.810625
23  4  1  3  2912.112933  2933.932036
4 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
```

```
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
13  2  1  3  2909.072805  2925.437132
17  3  1  2  2910.025070  2926.389397
14  2  1  4  2910.070054  2929.161769
22  4  1  2  2910.482981  2929.574696
16  3  1  1  2910.831552  2924.468491
18  3  1  3  2911.081757  2930.173472
6   1  1  1  2911.557952  2919.740115
12  2  1  2  2911.763148  2925.400087
19  3  1  4  2911.991523  2933.810625
23  4  1  3  2912.112933  2933.932036
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
7   1  1  2  2913.632934  2924.542485
24  4  1  4  2913.720737  2938.267228
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
0   0  1  0  2923.001971  2925.729359
1   0  1  1  2923.044906  2928.499681
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
    p  1  q          aic          bic
3   0  1  3  2905.943935  2916.853486
6   1  1  1  2911.557952  2919.740115
8   1  1  3  2907.302587  2920.939526
4   0  1  4  2907.694787  2921.331726
15  3  1  0  2913.358238  2924.267790
11  2  1  1  2913.380890  2924.290441
16  3  1  1  2910.831552  2924.468491
7   1  1  2  2913.632934  2924.542485
21  4  1  1  2908.445897  2924.810223
9   1  1  4  2908.786543  2925.150870
12  2  1  2  2911.763148  2925.400087
13  2  1  3  2909.072805  2925.437132
0   0  1  0  2923.001971  2925.729359
17  3  1  2  2910.025070  2926.389397
20  4  1  0  2914.074914  2927.711853
5   1  1  0  2922.832803  2928.287579
1   0  1  1  2923.044906  2928.499681
14  2  1  4  2910.070054  2929.161769
22  4  1  2  2910.482981  2929.574696
18  3  1  3  2911.081757  2930.173472
2   0  1  2  2924.151590  2932.333753
10  2  1  0  2924.998547  2933.180710
19  3  1  4  2911.991523  2933.810625
23  4  1  3  2912.112933  2933.932036
24  4  1  4  2913.720737  2938.267228
4 1 4 None None
```

In [36]:

```
order_aic_bic =[]
for p in range(5):
    for q in range(5):
        model = SARIMAX(rrabs_train, order=(p,2,q))
        results = model.fit()
        order_aic_bic.append((p, 2, q, results.aic, results.bic))
        order_df = pd.DataFrame(order_aic_bic, columns=['p', '2', 'q', 'aic', 'bic'])
        print(order_df.sort_values('aic'))
```

```
        print(order_df.sort_values('bic'))
        print(p, 1, q, None, None)
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
0  0  2  0  2992.865256  2995.583755
0 1 0 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
1  0  2  1  2913.757778  2919.194775
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
1  0  2  1  2913.757778  2919.194775
0  0  2  0  2992.865256  2995.583755
0 1 1 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
2  0  2  2  2912.617182  2920.772679
1  0  2  1  2913.757778  2919.194775
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
0  0  2  0  2992.865256  2995.583755
0 1 2 None None

C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
   p  2  q          aic          bic
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
```

```
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
0  0  2  0  2992.865256  2995.583755
0 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
0  0  2  0  2992.865256  2995.583755
0 1 4 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
1 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
6  1  2  1  2914.805552  2922.961048
```

```
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
6  1  2  1  2914.805552  2922.961048
3  0  2  3  2913.251230  2924.125226
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
1 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
7  1  2  2  2910.169695  2921.043690
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
6  1  2  1  2914.805552  2922.961048
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
7  1  2  2  2910.169695  2921.043690
6  1  2  1  2914.805552  2922.961048
3  0  2  3  2913.251230  2924.125226
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
1 1 2 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
7  1  2  2  2910.169695  2921.043690
8  1  2  3  2911.663202  2925.255697
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
6  1  2  1  2914.805552  2922.961048
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
7  1  2  2  2910.169695  2921.043690
```

```
6  1  2  1  2914.805552  2922.961048
3  0  2  3  2913.251230  2924.125226
8  1  2  3  2911.663202  2925.255697
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
1 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
9  1  2  4  2902.249763  2918.560757
7  1  2  2  2910.169695  2921.043690
8  1  2  3  2911.663202  2925.255697
2  0  2  2  2912.617182  2920.772679
3  0  2  3  2913.251230  2924.125226
1  0  2  1  2913.757778  2919.194775
6  1  2  1  2914.805552  2922.961048
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
   p  2  q          aic          bic
4  0  2  4  2901.407824  2915.000318
9  1  2  4  2902.249763  2918.560757
1  0  2  1  2913.757778  2919.194775
2  0  2  2  2912.617182  2920.772679
7  1  2  2  2910.169695  2921.043690
6  1  2  1  2914.805552  2922.961048
3  0  2  3  2913.251230  2924.125226
8  1  2  3  2911.663202  2925.255697
5  1  2  0  2951.124644  2956.561642
0  0  2  0  2992.865256  2995.583755
1 1 4 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
7   1  2  2  2910.169695  2921.043690
8   1  2  3  2911.663202  2925.255697
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
```

```
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
6   1  2  1  2914.805552  2922.961048
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
2 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
7   1  2  2  2910.169695  2921.043690
8   1  2  3  2911.663202  2925.255697
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
6   1  2  1  2914.805552  2922.961048
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
11  2  2  1  2914.849328  2925.723323
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
2 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  2  q          aic          bic
```

```
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
7    1  2  2  2910.169695  2921.043690
8    1  2  3  2911.663202  2925.255697
12   2  2  2  2911.848244  2925.440739
2    0  2  2  2912.617182  2920.772679
3    0  2  3  2913.251230  2924.125226
1    0  2  1  2913.757778  2919.194775
6    1  2  1  2914.805552  2922.961048
11   2  2  1  2914.849328  2925.723323
10   2  2  0  2950.611833  2958.767330
5    1  2  0  2951.124644  2956.561642
0    0  2  0  2992.865256  2995.583755
     p  2  q          aic          bic
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
1    0  2  1  2913.757778  2919.194775
2    0  2  2  2912.617182  2920.772679
7    1  2  2  2910.169695  2921.043690
6    1  2  1  2914.805552  2922.961048
3    0  2  3  2913.251230  2924.125226
8    1  2  3  2911.663202  2925.255697
12   2  2  2  2911.848244  2925.440739
11   2  2  1  2914.849328  2925.723323
5    1  2  0  2951.124644  2956.561642
10   2  2  0  2950.611833  2958.767330
0    0  2  0  2992.865256  2995.583755
2 1 2 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
     p  2  q          aic          bic
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
7    1  2  2  2910.169695  2921.043690
13   2  2  3  2910.914616  2927.225609
8    1  2  3  2911.663202  2925.255697
12   2  2  2  2911.848244  2925.440739
2    0  2  2  2912.617182  2920.772679
3    0  2  3  2913.251230  2924.125226
1    0  2  1  2913.757778  2919.194775
6    1  2  1  2914.805552  2922.961048
11   2  2  1  2914.849328  2925.723323
10   2  2  0  2950.611833  2958.767330
5    1  2  0  2951.124644  2956.561642
0    0  2  0  2992.865256  2995.583755
     p  2  q          aic          bic
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
1    0  2  1  2913.757778  2919.194775
2    0  2  2  2912.617182  2920.772679
7    1  2  2  2910.169695  2921.043690
6    1  2  1  2914.805552  2922.961048
3    0  2  3  2913.251230  2924.125226
8    1  2  3  2911.663202  2925.255697
12   2  2  2  2911.848244  2925.440739
```

```
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
2 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
14  2  2  4  2904.105127  2923.134619
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
2 1 4 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
14  2  2  4  2904.105127  2923.134619
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
```

```
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
3 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
14  2  2  4  2904.105127  2923.134619
16  3  2  1  2907.768013  2921.360507
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
```

```
15   3  2  0  2936.380009  2947.254004
5    1  2  0  2951.124644  2956.561642
10   2  2  0  2950.611833  2958.767330
0    0  2  0  2992.865256  2995.583755
3 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
14   2  2  4  2904.105127  2923.134619
17   3  2  2  2904.999698  2921.310691
16   3  2  1  2907.768013  2921.360507
7    1  2  2  2910.169695  2921.043690
13   2  2  3  2910.914616  2927.225609
8    1  2  3  2911.663202  2925.255697
12   2  2  2  2911.848244  2925.440739
2    0  2  2  2912.617182  2920.772679
3    0  2  3  2913.251230  2924.125226
1    0  2  1  2913.757778  2919.194775
6    1  2  1  2914.805552  2922.961048
11   2  2  1  2914.849328  2925.723323
15   3  2  0  2936.380009  2947.254004
10   2  2  0  2950.611833  2958.767330
5    1  2  0  2951.124644  2956.561642
0    0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
1    0  2  1  2913.757778  2919.194775
2    0  2  2  2912.617182  2920.772679
7    1  2  2  2910.169695  2921.043690
17   3  2  2  2904.999698  2921.310691
16   3  2  1  2907.768013  2921.360507
6    1  2  1  2914.805552  2922.961048
14   2  2  4  2904.105127  2923.134619
3    0  2  3  2913.251230  2924.125226
8    1  2  3  2911.663202  2925.255697
12   2  2  2  2911.848244  2925.440739
11   2  2  1  2914.849328  2925.723323
13   2  2  3  2910.914616  2927.225609
15   3  2  0  2936.380009  2947.254004
5    1  2  0  2951.124644  2956.561642
10   2  2  0  2950.611833  2958.767330
0    0  2  0  2992.865256  2995.583755
3 1 2 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4    0  2  4  2901.407824  2915.000318
9    1  2  4  2902.249763  2918.560757
14   2  2  4  2904.105127  2923.134619
18   3  2  3  2904.774082  2923.803574
```

```
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
3 1 3 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
19  3  2  4  2903.898643  2925.646634
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
```

```
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
19  3  2  4  2903.898643  2925.646634
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
3 1 4 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
19  3  2  4  2903.898643  2925.646634
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
```

```
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
19  3  2  4  2903.898643  2925.646634
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
4 1 0 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
19  3  2  4  2903.898643  2925.646634
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
17  3  2  2  2904.999698  2921.310691
21  4  2  1  2907.036053  2923.347046
16  3  2  1  2907.768013  2921.360507
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
21  4  2  1  2907.036053  2923.347046
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
19  3  2  4  2903.898643  2925.646634
```

```
11  2  2  1  2914.849328  2925.723323
13  2  2  3  2910.914616  2927.225609
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
4 1 1 None None
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
19  3  2  4  2903.898643  2925.646634
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
17  3  2  2  2904.999698  2921.310691
21  4  2  1  2907.036053  2923.347046
16  3  2  1  2907.768013  2921.360507
22  4  2  2  2907.869039  2926.898532
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
21  4  2  1  2907.036053  2923.347046
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
19  3  2  4  2903.898643  2925.646634
11  2  2  1  2914.849328  2925.723323
22  4  2  2  2907.869039  2926.898532
13  2  2  3  2910.914616  2927.225609
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
4 1 2 None None
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\base\model.py:604: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
19  3  2  4  2903.898643  2925.646634
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
17  3  2  2  2904.999698  2921.310691
23  4  2  3  2906.386831  2928.134822
21  4  2  1  2907.036053  2923.347046
16  3  2  1  2907.768013  2921.360507
22  4  2  2  2907.869039  2926.898532
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
21  4  2  1  2907.036053  2923.347046
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
19  3  2  4  2903.898643  2925.646634
11  2  2  1  2914.849328  2925.723323
22  4  2  2  2907.869039  2926.898532
13  2  2  3  2910.914616  2927.225609
23  4  2  3  2906.386831  2928.134822
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
4 1 3 None None
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
19  3  2  4  2903.898643  2925.646634
14  2  2  4  2904.105127  2923.134619
18  3  2  3  2904.774082  2923.803574
17  3  2  2  2904.999698  2921.310691
24  4  2  4  2906.190925  2930.657415
23  4  2  3  2906.386831  2928.134822
21  4  2  1  2907.036053  2923.347046
16  3  2  1  2907.768013  2921.360507
22  4  2  2  2907.869039  2926.898532
7   1  2  2  2910.169695  2921.043690
13  2  2  3  2910.914616  2927.225609
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
2   0  2  2  2912.617182  2920.772679
3   0  2  3  2913.251230  2924.125226
1   0  2  1  2913.757778  2919.194775
6   1  2  1  2914.805552  2922.961048
11  2  2  1  2914.849328  2925.723323
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
10  2  2  0  2950.611833  2958.767330
5   1  2  0  2951.124644  2956.561642
0   0  2  0  2992.865256  2995.583755
    p  2  q          aic          bic
4   0  2  4  2901.407824  2915.000318
9   1  2  4  2902.249763  2918.560757
1   0  2  1  2913.757778  2919.194775
2   0  2  2  2912.617182  2920.772679
7   1  2  2  2910.169695  2921.043690
17  3  2  2  2904.999698  2921.310691
16  3  2  1  2907.768013  2921.360507
6   1  2  1  2914.805552  2922.961048
14  2  2  4  2904.105127  2923.134619
21  4  2  1  2907.036053  2923.347046
18  3  2  3  2904.774082  2923.803574
3   0  2  3  2913.251230  2924.125226
8   1  2  3  2911.663202  2925.255697
12  2  2  2  2911.848244  2925.440739
19  3  2  4  2903.898643  2925.646634
11  2  2  1  2914.849328  2925.723323
22  4  2  2  2907.869039  2926.898532
13  2  2  3  2910.914616  2927.225609
23  4  2  3  2906.386831  2928.134822
24  4  2  4  2906.190925  2930.657415
20  4  2  0  2928.238365  2941.830859
15  3  2  0  2936.380009  2947.254004
```

```
5   1  2  0  2951.124644  2956.561642
10  2  2  0  2950.611833  2958.767330
0   0  2  0  2992.865256  2995.583755
4 1 4 None None
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\base\model.py:604: C
onvergenceWarning: Maximum Likelihood optimization failed to converge. Check mle_retvals
  warnings.warn("Maximum Likelihood optimization failed to "
```

Above wrote loops to fit multiple ARIMA models to time series rrabs to find the best model order using the AIC and BIC.

Here we loop over AR and MA orders from zero through 5 and fit each model using the first difference and then the second difference.

Then we print the model along with the AIC and BIC scores.

The Akaike information criterion or AIC is a metric that tells us how good a model is.

A model that makes better predictions is given a smaller AIC score.

The AIC also penalises models which have lots of parameters.

This means if we set the order too high compared to the data, we will get a high AIC value.

This stops us overfitting to the training data.

The Bayesian information criterion or BIC is very similar to the AIC.

The BIC penalises over complex models. For both of these, a lower value is a better model.

The difference between these two metrics is in how much they penalize model complexity! The BIC penalizes more additional model orders than the AIC and so the BIC will sometimes suggest a simpler model. The AIC and BIC will often choose the same model but when they don't we have to make a choice. Since our goal is to identify good predictive models, we should use the AIC.

The best models were shown to be the ARIMA(0,2,4), ARIMA(1,2,4), ARIMA(3,2,4), ARIMA(2,2,4), ARIMA(3,2,3), ARIMA(3,2,2), ARIMA(0,1,3)

In [37]:

```
model_024 = SARIMAX(rrabs_train, order=(0,2,4))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [38]:

```
results_024 = model_024.fit()
```

In [39]:

```
print(results_024.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:      RiverRunOffAbsklm   No. Observations:                  114
Model:               SARIMAX(0, 2, 4)   Log Likelihood               -1445.704
```

```
Date:                  Sun, 05 Feb 2023   AIC                          2901.408
Time:                          09:47:28   BIC                          2915.000
Sample:                      07-01-2006   HQIC                         2906.923
                           - 12-01-2015
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -1.1775      0.133     -8.826      0.000      -1.439      -0.916
ma.L2          0.0546      0.159      0.344      0.731      -0.256       0.365
ma.L3         -0.2498      0.126     -1.983      0.047      -0.497      -0.003
ma.L4          0.3752      0.077      4.855      0.000       0.224       0.527
sigma2      9.891e+09    1.3e-11   7.61e+20      0.000    9.89e+09    9.89e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 0.50
Prob(Q):                              0.88   Prob(JB):                         0.78
Heteroskedasticity (H):               0.43   Skew:                            -0.14
Prob(H) (two-sided):                  0.01   Kurtosis:                         3.17
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 7.21e+36. Standa
rd errors may be unstable.
```

The top section of the results summary includes useful information such as the order of the model we fit, the number of observations or data points, and the name of the time series.

The next section of this summary shows the fitted model parameters. We fitted an ARIMA(1,2,4) model. So the model has ar lag 1 and ma lags 1 to 4 coefficients. In the table these are the ar.L1 and ma.L1 to ma.L4 rows row. The first column shows the model coefficients while the second column shows the errors in these coefficients. This is the uncertainty on the fitted coefficient values.

Ljung-Box Prob(Q) - p-value for null hypothesis that residuals are uncorrelated.

Jarque-Bera Prob(JB) - p-value for null hypothesis that residuals are normal i. e. that the residuals are Gaussian normally distributed.

If either p-value is less than 0.05 we reject the null hypothesis.

Here we can conclude that the residuals are not correlated and are normally distributed.

In [40]:

```
residuals_024 = results_024.resid
```

In [41]:

```
mae_024 = np.mean(np.abs(residuals_024))
```

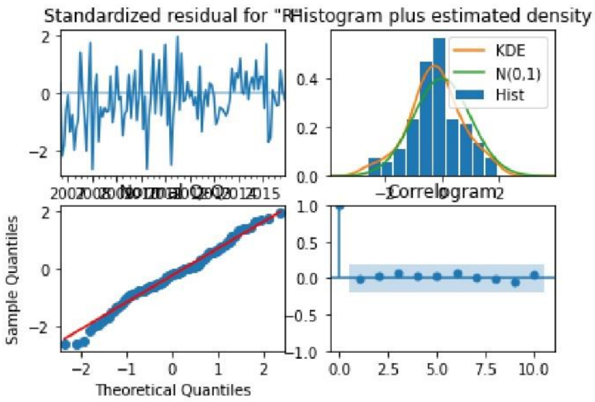
In [42]:

```
print(mae_024)
```

```
79102.61322010116
```

In [43]:

```
results_024.plot_diagnostics()
plt.show()
```



The next step is using common model diagnostics to confirm our model is behaving well. To diagnose our models we focus on the residuals to the training data. The residuals are the difference between our models one step predictions and the real values of the time series. To answer how far our predictions are from the true values, we can calculate the mean absolute error of the residuals. For the above model, this was found to be 69164.48226456114.

Plot diagnostics

If the model fits well the residuals will be white Gaussian noise. For an ideal model the residuals should be uncorrelated white Gaussian noise centred on 0. The rest of our diagnostics will help us to see if this is true.

One of the plots shows the standardised one step ahead residuals if our model is working correctly, there should be no structure in residuals. The plot should not have a pattern at all.

Another of the four plots is the Histogram plus estimated density shows us the distribution of the residuals. The histogram shows us the measured distribution and there is also a smooth version of the histogram and the normal distribution and if our model is good, these two lines should almost the same.

The Normal QQ plot is another way to show how the distribution of the model residuals compares to a normal distribution. If our residuals re normally distributed, then all the points should lie along the major 45 degree line (the red line) except perhaps some values at either end.

The last plot is the Correlogram which is just an acf plot of the residuals rather than the data. 95% of the correlations for the lag > 0 should not be significant. If there is significant correlation in the residuals, it means that there is information in the data that our model has not captured.

The Q-Q plot suggests that the model can be improved on

In [44]:

```
model_124 = SARIMAX(rrabs_train, order=(1,2,4))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [45]:

```
results_124 = model_124.fit()
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
```

In [46]:
```
residuals_124 = results_124.resid
```

In [47]:
```
print(results_124.summary())
```

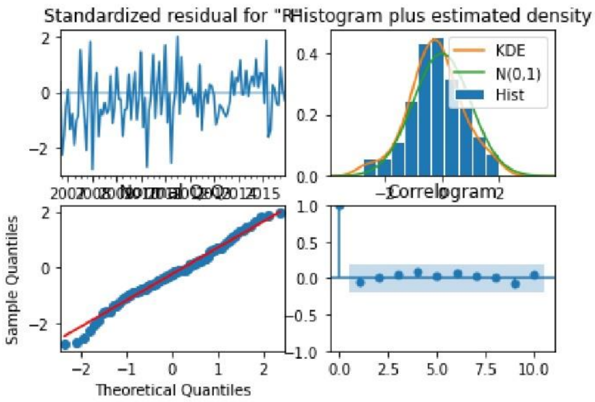
```
                               SARIMAX Results
==============================================================================
Dep. Variable:      RiverRunOffAbsklm   No. Observations:                  114
Model:               SARIMAX(1, 2, 4)   Log Likelihood               -1445.125
Date:                Sun, 05 Feb 2023   AIC                           2902.250
Time:                        09:53:49   BIC                           2918.561
Sample:                    07-01-2006   HQIC                          2908.868
                         - 12-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2384      0.257      0.929      0.353      -0.264       0.741
ma.L1         -1.3993      0.275     -5.080      0.000      -1.939      -0.859
ma.L2          0.3220      0.371      0.867      0.386      -0.406       1.050
ma.L3         -0.2761      0.173     -1.592      0.111      -0.616       0.064
ma.L4          0.3546      0.084      4.201      0.000       0.189       0.520
sigma2      9.499e+09   1.08e-11    8.8e+20      0.000     9.5e+09     9.5e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                 0.68
Prob(Q):                              0.59   Prob(JB):                         0.71
Heteroskedasticity (H):               0.43   Skew:                            -0.12
Prob(H) (two-sided):                  0.01   Kurtosis:                         3.30
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 4.22e+36. Standa
rd errors may be unstable.
```

In [48]:
```
mae_124 = np.mean(np.abs(residuals_124))
```

In [49]:
```
print(mae_124)
```

```
79212.05115388677
```

In [50]:
```
results_124.plot_diagnostics()
plt.show()
```



In [51]:
```
model_324 = SARIMAX(rrabs_train, order=(3,2,4))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [52]:
```
results_324 = model_324.fit()
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:966: UserWarning: Non-stationary starting autoregressive parameters found. Using zer
os as starting parameters.
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
```

In [53]:
```
residuals_324 = results_324.resid
```

In [54]:
```
print(results_324.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:      RiverRunOffAbsklm   No. Observations:                  114
Model:               SARIMAX(3, 2, 4)   Log Likelihood               -1443.949
Date:                Sun, 05 Feb 2023   AIC                           2903.899
Time:                        09:54:17   BIC                           2925.647
Sample:                    07-01-2006   HQIC                          2912.722
                         - 12-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.6376      0.067      9.458      0.000       0.505       0.770
ar.L2          0.7665      0.045     16.936      0.000       0.678       0.855
ar.L3         -0.8709      0.064    -13.704      0.000      -0.995      -0.746
ma.L1         -1.8315      0.369     -4.963      0.000      -2.555      -1.108
ma.L2         -0.0329      0.656     -0.050      0.960      -1.318       1.253
```

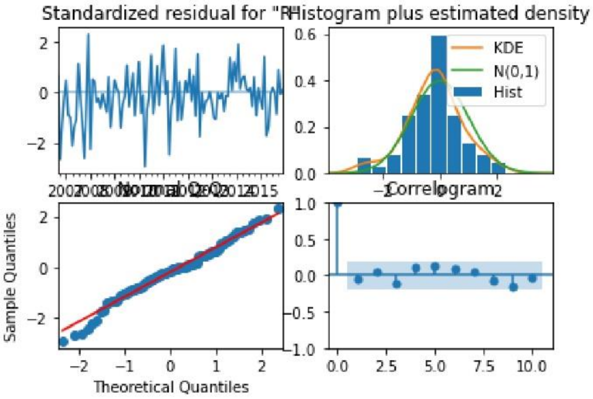
```
ma.L3          1.8245      0.389      4.689      0.000       1.062       2.587
ma.L4         -0.9601      0.139     -6.912      0.000      -1.232      -0.688
sigma2      8.658e+09   7.66e-11   1.13e+20      0.000    8.66e+09    8.66e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.37   Jarque-Bera (JB):                 2.17
Prob(Q):                              0.54   Prob(JB):                         0.34
Heteroskedasticity (H):               0.45   Skew:                            -0.23
Prob(H) (two-sided):                  0.02   Kurtosis:                         3.51
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 6.72e+35. Standa
rd errors may be unstable.
```

In [55]:
```
mae_324 = np.mean(np.abs(residuals_324))
```

In [56]:
```
print(mae_324)
```

```
76508.99114655265
```

In [57]:
```
results_324.plot_diagnostics()
plt.show()
```



In [58]:
```
model_224 = SARIMAX(rrabs_train, order=(2,2,4))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [59]:
```
results_224 = model_224.fit()
```

In [60]:
```
residuals_224 = results_224.resid
```

In [61]:
```
print(results_224.summary())
```

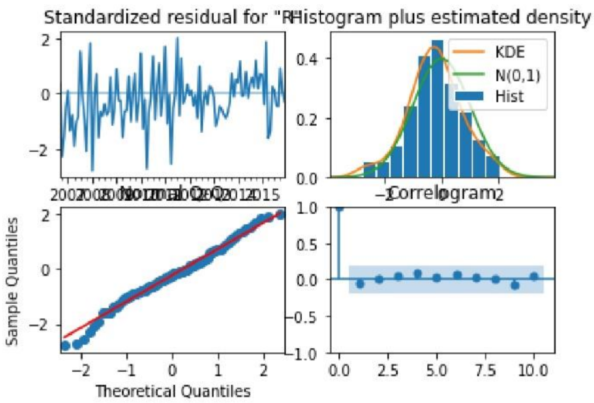
```
                               SARIMAX Results
==============================================================================
Dep. Variable:      RiverRunOffAbsklm   No. Observations:                  114
Model:               SARIMAX(2, 2, 4)   Log Likelihood               -1445.053
Date:                Sun, 05 Feb 2023   AIC                           2904.105
Time:                        09:54:50   BIC                           2923.135
Sample:                    07-01-2006   HQIC                          2911.826
                         - 12-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2512      0.353      0.712      0.476      -0.440       0.943
ar.L2         -0.0128      0.340     -0.038      0.970      -0.679       0.654
ma.L1         -1.4145      0.388     -3.649      0.000      -2.174      -0.655
ma.L2          0.3490      0.720      0.485      0.628      -1.062       1.760
ma.L3         -0.2907      0.407     -0.714      0.475      -1.089       0.507
ma.L4          0.3570      0.085      4.183      0.000       0.190       0.524
sigma2      9.276e+09   1.05e-10   8.79e+19      0.000    9.28e+09    9.28e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.31   Jarque-Bera (JB):                 0.66
Prob(Q):                              0.58   Prob(JB):                         0.72
Heteroskedasticity (H):               0.43   Skew:                            -0.12
Prob(H) (two-sided):                  0.01   Kurtosis:                         3.29
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.05e+36. Standa
rd errors may be unstable.
```

In [62]:
```
mae_224 = np.mean(np.abs(residuals_224))
```

In [63]:
```
print(mae_224)
```

```
79289.22768937329
```

In [64]:
```
results_224.plot_diagnostics()
plt.show()
```



In [65]:
```
model_323 = SARIMAX(rrabs_train, order=(3,2,3))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [66]:
```
results_323 = model_323.fit()
```

In [67]:
```
residuals_323 = results_323.resid
```

In [68]:
```
print(results_323.summary())
```

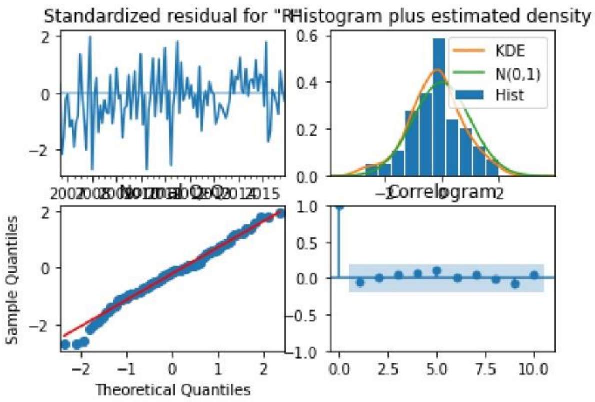
```
                               SARIMAX Results
==============================================================================
Dep. Variable:     RiverRunOffAbsklm   No. Observations:                  114
Model:              SARIMAX(3, 2, 3)   Log Likelihood               -1445.387
Date:               Sun, 05 Feb 2023   AIC                           2904.774
Time:                       09:55:24   BIC                           2923.804
Sample:                   07-01-2006   HQIC                          2912.495
                        - 12-01-2015
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2649      0.359      0.737      0.461      -0.439       0.969
ar.L2          0.2721      0.278      0.979      0.328      -0.273       0.817
ar.L3         -0.2955      0.080     -3.685      0.000      -0.453      -0.138
ma.L1         -1.4156      0.384     -3.684      0.000      -2.169      -0.662
ma.L2          0.0634      0.701      0.090      0.928      -1.310       1.436
ma.L3          0.3541      0.345      1.027      0.304      -0.321       1.030
sigma2      9.957e+09   1.49e-10    6.7e+19      0.000    9.96e+09    9.96e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.22   Jarque-Bera (JB):                 1.35
Prob(Q):                              0.64   Prob(JB):                         0.51
Heteroskedasticity (H):               0.45   Skew:                            -0.21
Prob(H) (two-sided):                  0.02   Kurtosis:                         3.35
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 3.68e+35. Standa
rd errors may be unstable.
```

In [69]:
```
mae_323 = np.mean(np.abs(residuals_323))
```

In [70]:
```
print(mae_323)
```

```
78017.01539697628
```

In [71]:
```
results_323.plot_diagnostics()
plt.show()
```



In [72]:
```
model_322 = SARIMAX(rrabs_train, order=(3,2,2))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [73]:
```
results_322 = model_322.fit()
```

In [74]:
```
residuals_322 = results_322.resid
```

In [75]:
```
print(results_322.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:      RiverRunOffAbsklm   No. Observations:                  114
Model:               SARIMAX(3, 2, 2)   Log Likelihood               -1446.500
Date:                Sun, 05 Feb 2023   AIC                           2905.000
Time:                        09:55:52   BIC                           2921.311
Sample:                    07-01-2006   HQIC                          2911.618
                         - 12-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.5845      0.188      3.103      0.002       0.215       0.954
ar.L2         -0.0054      0.131     -0.041      0.967      -0.263       0.252
ar.L3         -0.2669      0.098     -2.721      0.007      -0.459      -0.075
ma.L1         -1.7516      0.174    -10.085      0.000      -2.092      -1.411
ma.L2          0.7542      0.170      4.428      0.000       0.420       1.088
sigma2       1.07e+10   4.53e-12   2.37e+21      0.000    1.07e+10    1.07e+10
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                 1.48
Prob(Q):                              0.82   Prob(JB):                         0.48
Heteroskedasticity (H):               0.44   Skew:                            -0.19
Prob(H) (two-sided):                  0.01   Kurtosis:                         3.41
===================================================================================
```

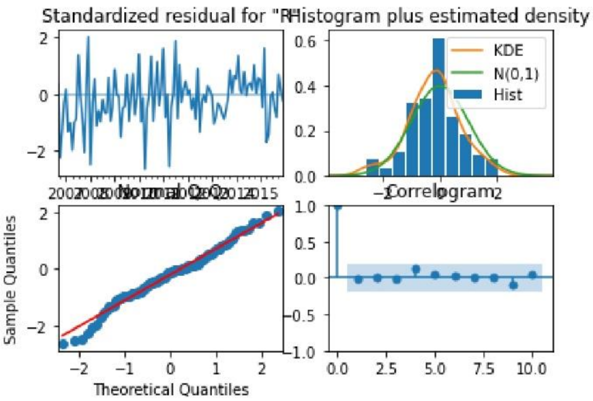
```
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 2.09e+37. Standa
rd errors may be unstable.
```

In [76]:
```
mae_322 = np.mean(np.abs(residuals_322))
```

In [77]:
```
print(mae_322)
```

```
78025.20048591417
```

In [78]:
```
results_322.plot_diagnostics()
plt.show()
```



In [79]:
```
model_013 = SARIMAX(rrabs_train, order=(0,1,3))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [80]:
```
results_013 = model_013.fit()
```

In [81]:
```
residuals_013 = results_013.resid
```

In [82]:
```
print(results_013.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:       RiverRunOffAbsklm   No. Observations:                  114
Model:                SARIMAX(0, 1, 3)   Log Likelihood               -1448.972
Date:                 Sun, 05 Feb 2023   AIC                           2905.944
Time:                         09:57:35   BIC                           2916.853
Sample:                     07-01-2006   HQIC                          2910.371
                          - 12-01-2015
Covariance Type:                   opg
```

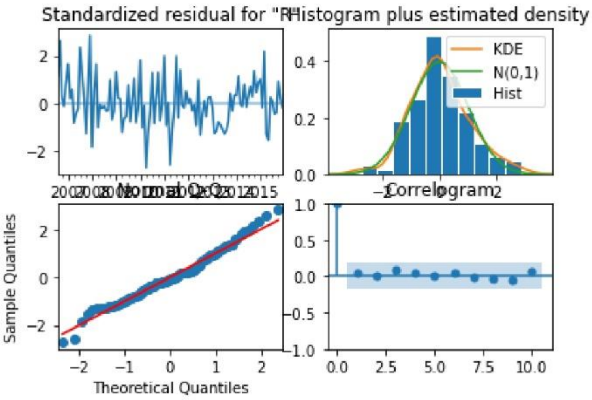
```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.3117      0.082     -3.794      0.000      -0.473      -0.151
ma.L2         -0.1410      0.076     -1.851      0.064      -0.290       0.008
ma.L3         -0.4393      0.067     -6.525      0.000      -0.571      -0.307
sigma2      7.725e+09   2.02e-12   3.82e+21      0.000    7.73e+09    7.73e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.21   Jarque-Bera (JB):                 2.27
Prob(Q):                              0.65   Prob(JB):                         0.32
Heteroskedasticity (H):               0.60   Skew:                             0.28
Prob(H) (two-sided):                  0.12   Kurtosis:                         3.41
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.33e+37. Standa
rd errors may be unstable.
```

In [83]:
```
mae_013 = np.mean(np.abs(residuals_013))
```

In [84]:
```
print(mae_013)
```

```
71766.71871195053
```

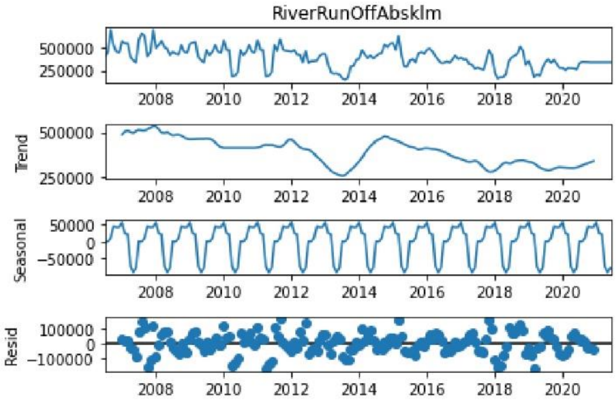
In [85]:
```
results_013.plot_diagnostics()
plt.show()
```



In [86]:
```
from statsmodels.tsa.seasonal import seasonal_decompose
```

In [87]:
```
rrabs_decomp = seasonal_decompose(rrabs['RiverRunOffAbsklm'], period=12)
```

In [88]:
```
rrabs_decomp.plot()
plt.show()
```



A seasonal time series has predictable patterns that repeat regularly. Although we call this feature seasonality it can repeat after any length of time.

The seasonal time series can be thought of as being made of three things: • the trend • seasonal component and • the residual. The full time series is these three parts added together.

The above decomposition confirms this. Rain is seasonal so a period of 12 months was used to decompose this series.

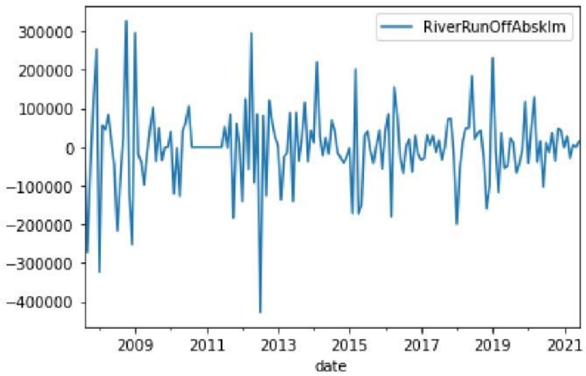
In [89]:
```
rrabs_sdiff = rrabs_stationary.diff(12).dropna()
```

In [90]:
```
rrabs_sdiff.head()
```

Out[90]:

	RiverRunOffAbsklm
date	
2007-08-01	13880.0
2007-09-01	-272730.0
2007-10-01	-55664.0
2007-11-01	122423.0
2007-12-01	252667.0

In [91]:
```
fig, ax = plt.subplots()
rrabs_sdiff.plot(ax=ax)
plt.show()
```



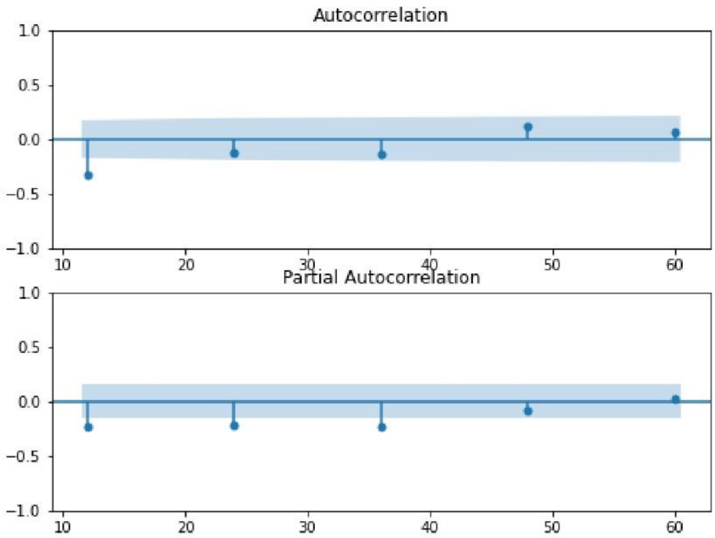
In [92]:
```
adfresults = adfuller(rrabs_sdiff['RiverRunOffAbsklm'])
```

In [93]:
```
print(adfresults)
```

```
(-4.4982077874259865, 0.00019811137651967137, 11, 155, {'1%': -3.4732590518613002, '5%': -
2.880374082105334, '10%': -2.5768120811654525}, 3890.712265213651)
```

The Augmented Dicky Fuller test still seems to suggest that the single differencing is slightly superior than the double differencing (differencing + seadonal differencing)

In [94]:
```
lags = [12, 24, 36, 48, 60]
fig, (ax1, ax2) = plt.subplots(2,1,figsize=(8,6))
plot_acf(rrabs_sdiff, lags = [12, 24, 36, 48, 60], ax = ax1)
plot_pacf(rrabs_sdiff, method = 'ywm', lags = [12, 24, 36, 48, 60], ax = ax2)
plt.show()
```



The acf shows a periodic correlation pattern! Look for lag greater 1 which is the peak for acf, here there is a peak at 12 lags, so what this means is that the seasonal component peaks every time after every 12 time steps.

In [95]:
```
rrabs_st = rrabs - rrabs.rolling(15).mean()
```

In [96]:
```
rrabs_st = rrabs_st.dropna()
```

In [97]:
```
rrabs_st.head()
```

Out[97]:

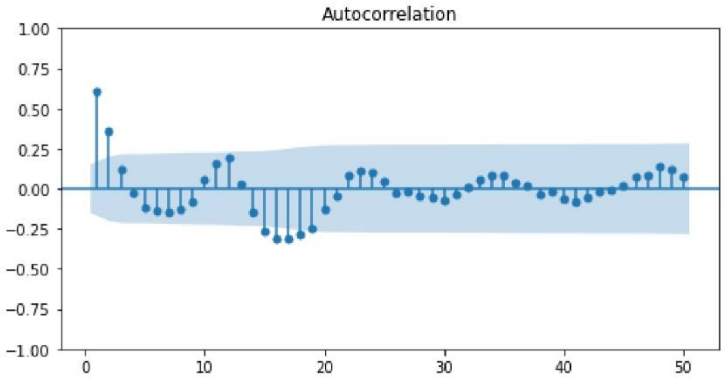
	RiverRunOffAbsklm
date	
2007-09-01	117869.600000
2007-10-01	-105130.400000
2007-11-01	-54130.400000
2007-12-01	187869.600000
2008-01-01	-9019.466667

In [98]:
```
adfresults = adfuller(rrabs_st['RiverRunOffAbsklm'])
```

In [99]:
```
print(adfresults)
```

```
(-4.080995484144464, 0.0010410004809968205, 13, 152, {'1%': -3.474120870218417, '5%': -2.8
80749791423677, '10%': -2.5770126333102494}, 3784.6147300568327)
```

In [100…
```
fig, ax = plt.subplots(1,1, figsize=(8,4))
plot_acf(rrabs_st, ax=ax, lags=50, zero=False)
plt.show()
```



In [107…
```
pip install pmdarima
```

```
Requirement already satisfied: pmdarima in c:\users\user\programs\python\python310\lib\sit
e-packages (2.0.2)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in c:\users\user\programs\pytho
n\python310\lib\site-packages (from pmdarima) (63.2.0)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in c:\users\user\programs
\python\python310\lib\site-packages (from pmdarima) (0.29.33)
Requirement already satisfied: pandas>=0.19 in c:\users\user\programs\python\python310\lib
\site-packages (from pmdarima) (1.5.2)
Requirement already satisfied: scipy>=1.3.2 in c:\users\user\programs\python\python310\lib
\site-packages (from pmdarima) (1.9.3)
```

```
Requirement already satisfied: joblib>=0.11 in c:\users\user\programs\python\python310\lib
\site-packages (from pmdarima) (1.2.0)
Requirement already satisfied: scikit-learn>=0.22 in c:\users\user\programs\python\python3
10\lib\site-packages (from pmdarima) (1.1.3)
Requirement already satisfied: urllib3 in c:\users\user\programs\python\python310\lib\site
-packages (from pmdarima) (1.26.13)
Requirement already satisfied: numpy>=1.21.2 in c:\users\user\programs\python\python310\li
b\site-packages (from pmdarima) (1.23.5)
Requirement already satisfied: statsmodels>=0.13.2 in c:\users\user\programs\python\python
310\lib\site-packages (from pmdarima) (0.13.5)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\programs\python\python310\lib
\site-packages (from pandas>=0.19->pmdarima) (2022.6)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\user\programs\python\pyt
hon310\lib\site-packages (from pandas>=0.19->pmdarima) (2.8.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\user\programs\python\pytho
n310\lib\site-packages (from scikit-learn>=0.22->pmdarima) (3.1.0)
Requirement already satisfied: packaging>=21.3 in c:\users\user\programs\python\python310
\lib\site-packages (from statsmodels>=0.13.2->pmdarima) (21.3)
Requirement already satisfied: patsy>=0.5.2 in c:\users\user\programs\python\python310\lib
\site-packages (from statsmodels>=0.13.2->pmdarima) (0.5.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\user\programs\python\p
ython310\lib\site-packages (from packaging>=21.3->statsmodels>=0.13.2->pmdarima) (3.0.9)
Requirement already satisfied: six in c:\users\user\programs\python\python310\lib\site-pac
kages (from patsy>=0.5.2->statsmodels>=0.13.2->pmdarima) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip available: 22.3.1 -> 23.0
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [101…

```
results_auto = pm.auto_arima(rrabs_train)
```

In [102…

```
results_auto
```

Out[102…

```
ARIMA(order=(4, 1, 1), scoring_args={}, suppress_warnings=True)
```

In [103…

```
print(results_auto.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  114
Model:               SARIMAX(4, 1, 1)   Log Likelihood               -1449.609
Date:                Sun, 05 Feb 2023   AIC                           2913.217
Time:                        10:01:43   BIC                           2932.309
Sample:                    07-01-2006   HQIC                          2920.965
                         - 12-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept  -1024.8157   1306.763     -0.784      0.433   -3586.024    1536.393
ar.L1          0.6417      0.143      4.484      0.000       0.361       0.922
ar.L2          0.0565      0.096      0.591      0.555      -0.131       0.244
ar.L3         -0.3155      0.075     -4.223      0.000      -0.462      -0.169
ar.L4          0.1536      0.107      1.429      0.153      -0.057       0.364
ma.L1         -0.9124      0.100     -9.158      0.000      -1.108      -0.717
sigma2      8.153e+09      0.000   3.16e+13      0.000    8.15e+09    8.15e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 2.08
Prob(Q):                              0.98   Prob(JB):                         0.35
Heteroskedasticity (H):               0.58   Skew:                             0.13
Prob(H) (two-sided):                  0.10   Kurtosis:                         3.61
===================================================================================
```

```
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 6.06e+28. Standa
rd errors may be unstable.
```

In [104…

```
model_312 = SARIMAX(rrabs_train, order=(3,1,2))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [105…

```
results_312 = model_312.fit()
```

In [106…

```
residuals_312 = results_312.resid
```

In [107…

```
print(results_312.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:      RiverRunOffAbsklm   No. Observations:                  114
Model:               SARIMAX(3, 1, 2)   Log Likelihood               -1449.013
Date:                Sun, 05 Feb 2023   AIC                           2910.025
Time:                        10:02:33   BIC                           2926.389
Sample:                    07-01-2006   HQIC                          2916.666
                         - 12-01-2015
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1477      0.300      0.492      0.623      -0.441       0.736
ar.L2          0.4291      0.213      2.011      0.044       0.011       0.847
ar.L3         -0.2304      0.085     -2.717      0.007      -0.397      -0.064
ma.L1         -0.4253      0.299     -1.424      0.155      -1.011       0.160
ma.L2         -0.5227      0.302     -1.730      0.084      -1.115       0.069
sigma2      8.249e+09   7.28e-12   1.13e+21      0.000    8.25e+09    8.25e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 4.70
Prob(Q):                              0.93   Prob(JB):                         0.10
Heteroskedasticity (H):               0.56   Skew:                             0.20
Prob(H) (two-sided):                  0.08   Kurtosis:                         3.92
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.02e+38. Standa
rd errors may be unstable.
```

In [108…

```
mae_312 = np.mean(np.abs(residuals_312))
```

In [109…

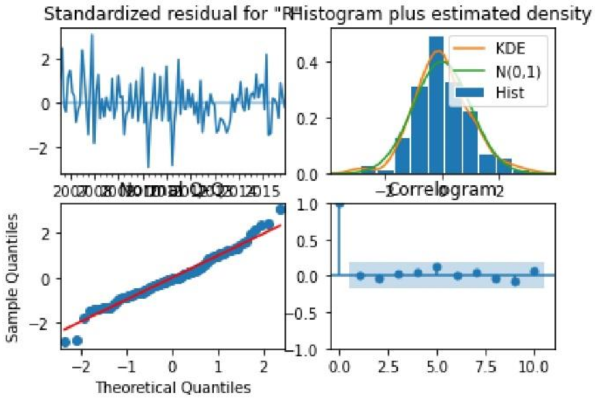
```
print(mae_312)
```

```
70630.27919913779
```

In [110…

```
results_312.plot_diagnostics()
```

```
plt.show()
```



In [111...
```
from sklearn.metrics import mean_absolute_error
```

In [112...
```
from sklearn.metrics import mean_squared_error
```

In [113...
```
rrabs_testdf = pd.DataFrame(rrabs_test)
print(rrabs_testdf.head())
```

```
            RiverRunOffAbsklm
date
2016-01-01             480257
2016-02-01             491587
2016-03-01             462244
2016-04-01             405251
2016-05-01             366784
```

In [114...
```
arima_pred024 = results_024.get_forecast(steps=66)
```

In [115...
```
arima_mean024 = arima_pred024.predicted_mean
```

In [116...
```
print(arima_mean024)
```

```
2016-01-01    409767.841797
2016-02-01    410423.598785
2016-03-01    420065.485296
2016-04-01    420917.473752
2016-05-01    421769.462208
                  ...
2021-02-01    470332.804202
2021-03-01    471184.792659
2021-04-01    472036.781115
2021-05-01    472888.769571
2021-06-01    473740.758027
Freq: MS, Name: predicted_mean, Length: 66, dtype: float64
```

In [117...
```
print(arima_mean024.shape)
```

```
(66,)
```

In [118…
```
arima_mean024df = pd.DataFrame(arima_mean024, columns = ['predicted_mean'])
arima_mean024df["predicted_mean"] = arima_mean024df["predicted_mean"].astype("float")
arima024_out = pd.merge(arima_mean024df, rrabs_testdf, left_index = True, right_index = Tr
arima024_out['errors'] = abs(arima024_out['predicted_mean'] - arima024_out['RiverRunOffAbs
arima024_out['percent_error'] = (arima024_out['errors']/arima024_out['RiverRunOffAbsklm'])
print(arima024_out.head())
mape_test024 = arima024_out['percent_error'].mean()
```

```
            predicted_mean  RiverRunOffAbsklm        errors  percent_error
2016-01-01   409767.841797             480257  70489.158203      14.677383
2016-02-01   410423.598785             491587  81163.401215      16.510486
2016-03-01   420065.485296             462244  42178.514704       9.124730
2016-04-01   420917.473752             405251  15666.473752       3.865869
2016-05-01   421769.462208             366784  54985.462208      14.991238
```

In [119…
```
mae_test024 = mean_absolute_error(rrabs_test, arima_mean024)
print('Mean_Absolute_Error_024: %f' % mae_test024)
mse_test024 = mean_squared_error(rrabs_test, arima_mean024)
print('Mean_Squared_Error_024: %f' % mse_test024)
rmse_test024 = mse_test024**(1/2)
print('Root_Mean_Squared_Error_024: %f' % rmse_test024)
rmspe_test024 = (rmse_test024 / arima024_out['RiverRunOffAbsklm'].mean())*100
print('Root_Mean_Squared_Percentage_Error_024: %f' % rmspe_test024)
print('Mean_Absolute_Percentage_Error_024: %f' % mape_test024)
```

```
Mean_Absolute_Error_024: 124230.051207
Mean_Squared_Error_024: 20276893159.240288
Root_Mean_Squared_Error_024: 142396.956285
Root_Mean_Squared_Percentage_Error_024: 42.959958
Mean_Absolute_Percentage_Error_024: 46.013626
```

In [120…
```
arima_pred124 = results_124.get_forecast(steps=66)
```

In [121…
```
arima_mean124 = arima_pred124.predicted_mean
```

In [122…
```
print(arima_mean124)
```

```
2016-01-01    409949.351658
2016-02-01    402648.142255
2016-03-01    410935.735525
2016-04-01    413141.547543
2016-05-01    413897.352463
                  ...
2021-02-01    431246.091921
2021-03-01    431547.962665
2021-04-01    431849.833409
2021-05-01    432151.704153
2021-06-01    432453.574897
Freq: MS, Name: predicted_mean, Length: 66, dtype: float64
```

In [123…
```
arima_mean124df = pd.DataFrame(arima_mean124, columns = ['predicted_mean'])
arima_mean124df["predicted_mean"] = arima_mean124df["predicted_mean"].astype("float")
arima124_out = pd.merge(arima_mean124df, rrabs_testdf, left_index = True, right_index = Tr
arima124_out['errors'] = abs(arima124_out['predicted_mean'] - arima124_out['RiverRunOffAbs
arima124_out['percent_error'] = (arima124_out['errors']/arima124_out['RiverRunOffAbsklm'])
print(arima124_out.head())
mape_test124 = arima124_out['percent_error'].mean()
```

```
            predicted_mean  RiverRunOffAbsklm        errors  percent_error
2016-01-01   409949.351658             480257  70307.648342      14.639588
```

```
2016-02-01   402648.142255            491587  88938.857745      18.092191
2016-03-01   410935.735525            462244  51308.264475      11.099823
2016-04-01   413141.547543            405251   7890.547543       1.947077
2016-05-01   413897.352463            366784  47113.352463      12.844986
```

In [124…

```
mae_test124 = mean_absolute_error(rrabs_test, arima_mean124)
print('Mean_Absolute_Error_124: %f' % mae_test124)
mse_test124 = mean_squared_error(rrabs_test, arima_mean124)
print('Mean_Squared_Error_124: %f' % mse_test124)
rmse_test124 = mse_test124**(1/2)
print('Root_Mean_Squared_Error_124: %f' % rmse_test124)
rmspe_test124 = (rmse_test124 / arima124_out['RiverRunOffAbsklm'].mean())*100
print('Root_Mean_Squared_Percentage_Error_124: %f' % rmspe_test124)
print('Mean_Absolute_Percentage_Error_124: %f' % mape_test124)
```

```
Mean_Absolute_Error_124: 103788.306939
Mean_Squared_Error_124: 14806243829.369423
Root_Mean_Squared_Error_124: 121680.909881
Root_Mean_Squared_Percentage_Error_124: 36.710101
Mean_Absolute_Percentage_Error_124: 38.947657
```

In [125…

```
arima_pred324 = results_324.get_forecast(steps=66)
```

In [126…

```
arima_mean324 = arima_pred324.predicted_mean
```

In [127…

```
print(arima_mean324)
```

```
2016-01-01    430183.852943
2016-02-01    385613.715692
2016-03-01    378233.244090
2016-04-01    344336.892998
2016-05-01    356506.363660
                  ...
2021-02-01    461357.630563
2021-03-01    479793.571642
2021-04-01    463497.786196
2021-05-01    481755.738359
2021-06-01    465474.356306
Freq: MS, Name: predicted_mean, Length: 66, dtype: float64
```

In [128…

```
arima_mean324df = pd.DataFrame(arima_mean324, columns = ['predicted_mean'])
arima_mean324df["predicted_mean"] = arima_mean324df["predicted_mean"].astype("float")
arima324_out = pd.merge(arima_mean324df, rrabs_testdf, left_index = True, right_index = Tr
arima324_out['errors'] = abs(arima324_out['predicted_mean'] - arima324_out['RiverRunOffAbs
arima324_out['percent_error'] = (arima324_out['errors']/arima324_out['RiverRunOffAbsklm'])
print(arima324_out.head())
mape_test324 = arima324_out['percent_error'].mean()
```

```
             predicted_mean  RiverRunOffAbsklm         errors  percent_error
2016-01-01    430183.852943             480257   50073.147057      10.426323
2016-02-01    385613.715692             491587  105973.284308      21.557381
2016-03-01    378233.244090             462244   84010.755910      18.174548
2016-04-01    344336.892998             405251   60914.107002      15.031205
2016-05-01    356506.363660             366784   10277.636340       2.802095
```

In [129…

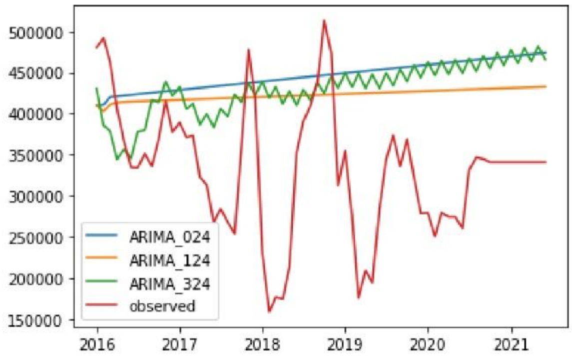
```
mae_test324 = mean_absolute_error(rrabs_test, arima_mean324)
print('Mean_Absolute_Error_324: %f' % mae_test324)
mse_test324 = mean_squared_error(rrabs_test, arima_mean324)
print('Mean_Squared_Error_324: %f' % mse_test324)
rmse_test324 = mse_test324**(1/2)
```

```
print('Root_Mean_Squared_Error_324: %f' % rmse_test324)
rmspe_test324 = (rmse_test324 / arima324_out['RiverRunOffAbsklm'].mean())*100
print('Root_Mean_Squared_Percentage_Error_324: %f' % rmspe_test324)
print('Mean_Absolute_Percentage_Error_324: %f' % mape_test324)
```

```
Mean_Absolute_Error_324: 112779.427517
Mean_Squared_Error_324: 17642058245.075390
Root_Mean_Squared_Error_324: 132823.410004
Root_Mean_Squared_Percentage_Error_324: 40.071700
Mean_Absolute_Percentage_Error_324: 41.952999
```

In [130…

```
plt.plot(arima_mean024, label='ARIMA_024')
plt.plot(arima_mean124, label='ARIMA_124')
plt.plot(arima_mean324, label='ARIMA_324')
plt.plot(rrabs_test, label='observed')
plt.legend()
plt.show()
```



The above are a comparison of the out of sample dynamic predictions of 3 ARIMA models against the test series. Predictions were made 66 months intyo the future after the training series' end period. it is clear that the better model would have been the ARIMA(3,2,4) model in terms of trying to follow the series however the approach of a 60:40 split for time series suffers from extrapolating too far from the data.

A SARIMA or seasonal ARIMA model is the tool choice for a seasonal time series. We can split up our time series into a seasonal and some non-seasonal components.

Fitting a SARIMA model is like fitting two different ARIMA models at once, one to the seasonal part and another to the non-seasonal part. Since we have these two models, we will have two sets of orders as shown below. Seasonal ARIMA = SARIMA

SARIMA(p,d,q)(P,D,Q)_S

Non-seasonal orders: p: autoregressive order, d: differencing order, q: moving average order

Seasonal Orders: P: seasonal autoregressive order, D: seasonal differencing order, Q: seasonal moving average order, S: number of time steps per cycle.

If the time series shows a trend, then we take the normal difference. First difference of time series.

If there is a strong seasonal cycle, we also take a seasonal difference. First difference and first seasonal difference of time series.

Once we have found the two orders of differencing, and made the time series stationary. We need to find the other orders. AR(p) MA(q) ARMA(p,q)

ACF Tails off Cuts off after lag q Tails off

PACF Cuts off after lag p Tails off Tails off

In [131…

```
results_sarima1 = pm.auto_arima(rrabs_train,
                        d=1,
                        start_p=0, # initial guess for p
                        start_q=0, # initial guess for q
                        max_p=4, # max value of p to test
                        max_q=4, # max value of q to test
                        seasonal=True, # is the time series seasonal
                        m=11, # the seasonal period
                        D=1, # seasonal difference order
                        start_P=0, # initial guess for P
                        start_Q=0, # initial guess for Q
                        max_P=4, # max value of P to test
                        max_Q=4, # max value of Q to test
                        information_criterion='aic', # used to select best model
                        trace=True, # print results whilst training
                        error_action='ignore' , # ignore orders that don't work
                        stepwise=True, # apply intelligent order search
)
```

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,1,0)[11]             : AIC=2697.244, Time=0.67 sec
 ARIMA(1,1,0)(1,1,0)[11]             : AIC=2668.895, Time=1.05 sec
 ARIMA(0,1,1)(0,1,1)[11]             : AIC=2662.130, Time=0.60 sec
 ARIMA(0,1,1)(0,1,0)[11]             : AIC=2676.785, Time=0.24 sec
 ARIMA(0,1,1)(1,1,1)[11]             : AIC=2659.300, Time=1.70 sec
 ARIMA(0,1,1)(1,1,0)[11]             : AIC=2667.011, Time=0.80 sec
 ARIMA(0,1,1)(2,1,1)[11]             : AIC=2661.100, Time=5.34 sec
 ARIMA(0,1,1)(1,1,2)[11]             : AIC=2661.197, Time=2.80 sec
 ARIMA(0,1,1)(0,1,2)[11]             : AIC=2660.938, Time=1.91 sec
 ARIMA(0,1,1)(2,1,0)[11]             : AIC=2667.854, Time=1.54 sec
 ARIMA(0,1,1)(2,1,2)[11]             : AIC=2662.480, Time=4.36 sec
 ARIMA(0,1,0)(1,1,1)[11]             : AIC=inf, Time=0.78 sec
 ARIMA(1,1,1)(1,1,1)[11]             : AIC=2659.276, Time=1.81 sec
 ARIMA(1,1,1)(0,1,1)[11]             : AIC=2659.788, Time=1.43 sec
 ARIMA(1,1,1)(1,1,0)[11]             : AIC=2666.133, Time=0.97 sec
 ARIMA(1,1,1)(2,1,1)[11]             : AIC=2660.747, Time=6.15 sec
 ARIMA(1,1,1)(1,1,2)[11]             : AIC=2661.014, Time=5.38 sec
 ARIMA(1,1,1)(0,1,0)[11]             : AIC=2678.628, Time=0.44 sec
 ARIMA(1,1,1)(0,1,2)[11]             : AIC=2660.132, Time=3.45 sec
 ARIMA(1,1,1)(2,1,0)[11]             : AIC=2666.494, Time=2.52 sec
 ARIMA(1,1,1)(2,1,2)[11]             : AIC=2662.086, Time=6.02 sec
 ARIMA(1,1,0)(1,1,1)[11]             : AIC=2660.107, Time=1.25 sec
 ARIMA(2,1,1)(1,1,1)[11]             : AIC=2658.531, Time=2.81 sec
 ARIMA(2,1,1)(0,1,1)[11]             : AIC=2659.980, Time=1.79 sec
 ARIMA(2,1,1)(1,1,0)[11]             : AIC=2666.891, Time=1.93 sec
 ARIMA(2,1,1)(2,1,1)[11]             : AIC=2660.208, Time=11.06 sec
 ARIMA(2,1,1)(1,1,2)[11]             : AIC=2660.374, Time=5.87 sec
 ARIMA(2,1,1)(0,1,0)[11]             : AIC=2677.412, Time=1.10 sec
 ARIMA(2,1,1)(0,1,2)[11]             : AIC=2659.720, Time=4.29 sec
 ARIMA(2,1,1)(2,1,0)[11]             : AIC=2667.134, Time=3.62 sec
 ARIMA(2,1,1)(2,1,2)[11]             : AIC=2661.537, Time=7.06 sec
 ARIMA(2,1,0)(1,1,1)[11]             : AIC=2662.121, Time=2.16 sec
 ARIMA(3,1,1)(1,1,1)[11]             : AIC=2657.471, Time=3.29 sec
 ARIMA(3,1,1)(0,1,1)[11]             : AIC=2658.765, Time=1.84 sec
 ARIMA(3,1,1)(1,1,0)[11]             : AIC=2664.590, Time=1.64 sec
 ARIMA(3,1,1)(2,1,1)[11]             : AIC=2659.059, Time=8.71 sec
```

```
 ARIMA(3,1,1)(1,1,2)[11]             : AIC=2659.255, Time=6.68 sec
 ARIMA(3,1,1)(0,1,0)[11]             : AIC=2675.720, Time=0.65 sec
 ARIMA(3,1,1)(0,1,2)[11]             : AIC=2658.562, Time=5.28 sec
 ARIMA(3,1,1)(2,1,0)[11]             : AIC=2664.970, Time=3.38 sec
 ARIMA(3,1,1)(2,1,2)[11]             : AIC=2660.598, Time=9.00 sec
 ARIMA(3,1,0)(1,1,1)[11]             : AIC=2656.044, Time=2.44 sec
 ARIMA(3,1,0)(0,1,1)[11]             : AIC=2657.691, Time=0.95 sec
 ARIMA(3,1,0)(1,1,0)[11]             : AIC=2663.193, Time=0.96 sec
 ARIMA(3,1,0)(2,1,1)[11]             : AIC=2657.558, Time=7.01 sec
 ARIMA(3,1,0)(1,1,2)[11]             : AIC=2657.797, Time=7.92 sec
 ARIMA(3,1,0)(0,1,0)[11]             : AIC=2674.662, Time=0.44 sec
 ARIMA(3,1,0)(0,1,2)[11]             : AIC=2657.320, Time=4.22 sec
 ARIMA(3,1,0)(2,1,0)[11]             : AIC=2663.757, Time=4.23 sec
 ARIMA(3,1,0)(2,1,2)[11]             : AIC=2659.075, Time=7.74 sec
 ARIMA(4,1,0)(1,1,1)[11]             : AIC=2656.759, Time=2.98 sec
 ARIMA(4,1,1)(1,1,1)[11]             : AIC=2658.784, Time=5.42 sec
 ARIMA(3,1,0)(1,1,1)[11] intercept   : AIC=2657.686, Time=5.13 sec

Best model:  ARIMA(3,1,0)(1,1,1)[11]
Total fit time: 183.522 seconds
```

In [132…

```
results_sarima2 = pm.auto_arima(rrabs_train,
                        d=2,
                        start_p=0, # initial guess for p
                        start_q=0, # initial guess for q
                        max_p=4, # max value of p to test
                        max_q=4, # max value of q to test
                        seasonal=True, # is the time series seasonal
                        m=11, # the seasonal period
                        D=1, # seasonal difference order
                        start_P=0, # initial guess for P
                        start_Q=0, # initial guess for Q
                        max_P=4, # max value of P to test
                        max_Q=4, # max value of Q to test
                        information_criterion='aic', # used to select best model
                        trace=True, # print results whilst training
                        error_action='ignore' , # ignore orders that don't work
                        stepwise=True, # apply intelligent order search
)
```

```
Performing stepwise search to minimize aic
 ARIMA(0,2,0)(0,1,0)[11]             : AIC=2774.723, Time=0.26 sec
 ARIMA(1,2,0)(1,1,0)[11]             : AIC=2694.767, Time=0.80 sec
 ARIMA(0,2,1)(0,1,1)[11]             : AIC=2663.215, Time=1.56 sec
 ARIMA(0,2,1)(0,1,0)[11]             : AIC=inf, Time=0.50 sec
 ARIMA(0,2,1)(1,1,1)[11]             : AIC=2665.069, Time=2.12 sec
 ARIMA(0,2,1)(0,1,2)[11]             : AIC=2665.117, Time=3.97 sec
 ARIMA(0,2,1)(1,1,0)[11]             : AIC=2666.847, Time=0.85 sec
 ARIMA(0,2,1)(1,1,2)[11]             : AIC=2666.980, Time=5.91 sec
 ARIMA(0,2,0)(0,1,1)[11]             : AIC=2726.978, Time=0.49 sec
 ARIMA(1,2,1)(0,1,1)[11]             : AIC=2656.141, Time=3.21 sec
 ARIMA(1,2,1)(0,1,0)[11]             : AIC=inf, Time=0.63 sec
 ARIMA(1,2,1)(1,1,1)[11]             : AIC=2656.649, Time=1.97 sec
 ARIMA(1,2,1)(0,1,2)[11]             : AIC=2657.231, Time=3.18 sec
 ARIMA(1,2,1)(1,1,0)[11]             : AIC=2659.436, Time=1.36 sec
 ARIMA(1,2,1)(1,1,2)[11]             : AIC=2658.522, Time=8.18 sec
 ARIMA(1,2,0)(0,1,1)[11]             : AIC=2685.926, Time=0.98 sec
 ARIMA(2,2,1)(0,1,1)[11]             : AIC=2655.841, Time=2.47 sec
 ARIMA(2,2,1)(0,1,0)[11]             : AIC=inf, Time=1.16 sec
 ARIMA(2,2,1)(1,1,1)[11]             : AIC=2656.209, Time=3.02 sec
 ARIMA(2,2,1)(0,1,2)[11]             : AIC=2656.852, Time=7.37 sec
 ARIMA(2,2,1)(1,1,0)[11]             : AIC=2659.331, Time=2.49 sec
 ARIMA(2,2,1)(1,1,2)[11]             : AIC=2658.075, Time=22.42 sec
 ARIMA(2,2,0)(0,1,1)[11]             : AIC=2683.301, Time=1.09 sec
 ARIMA(3,2,1)(0,1,1)[11]             : AIC=2653.205, Time=2.05 sec
```

```
 ARIMA(3,2,1)(0,1,0)[11]             : AIC=2662.722, Time=1.18 sec
 ARIMA(3,2,1)(1,1,1)[11]             : AIC=2654.071, Time=2.81 sec
 ARIMA(3,2,1)(0,1,2)[11]             : AIC=2654.576, Time=6.00 sec
 ARIMA(3,2,1)(1,1,0)[11]             : AIC=2656.142, Time=2.65 sec
 ARIMA(3,2,1)(1,1,2)[11]             : AIC=2655.819, Time=16.28 sec
 ARIMA(3,2,0)(0,1,1)[11]             : AIC=2672.597, Time=1.09 sec
 ARIMA(4,2,1)(0,1,1)[11]             : AIC=2648.553, Time=2.46 sec
 ARIMA(4,2,1)(0,1,0)[11]             : AIC=inf, Time=1.47 sec
 ARIMA(4,2,1)(1,1,1)[11]             : AIC=2648.872, Time=3.22 sec
 ARIMA(4,2,1)(0,1,2)[11]             : AIC=2649.558, Time=5.48 sec
 ARIMA(4,2,1)(1,1,0)[11]             : AIC=2652.160, Time=2.57 sec
 ARIMA(4,2,1)(1,1,2)[11]             : AIC=2650.636, Time=13.38 sec
 ARIMA(4,2,0)(0,1,1)[11]             : AIC=2664.353, Time=1.46 sec
 ARIMA(4,2,2)(0,1,1)[11]             : AIC=2643.918, Time=5.46 sec
 ARIMA(4,2,2)(0,1,0)[11]             : AIC=inf, Time=2.37 sec
 ARIMA(4,2,2)(1,1,1)[11]             : AIC=2644.136, Time=7.90 sec
 ARIMA(4,2,2)(0,1,2)[11]             : AIC=2644.690, Time=14.13 sec
 ARIMA(4,2,2)(1,1,0)[11]             : AIC=2649.358, Time=6.56 sec
 ARIMA(4,2,2)(1,1,2)[11]             : AIC=2645.953, Time=17.36 sec
 ARIMA(3,2,2)(0,1,1)[11]             : AIC=2643.026, Time=4.41 sec
 ARIMA(3,2,2)(0,1,0)[11]             : AIC=inf, Time=1.05 sec
 ARIMA(3,2,2)(1,1,1)[11]             : AIC=2643.175, Time=7.20 sec
 ARIMA(3,2,2)(0,1,2)[11]             : AIC=2643.764, Time=10.45 sec
 ARIMA(3,2,2)(1,1,0)[11]             : AIC=2647.816, Time=6.09 sec
 ARIMA(3,2,2)(1,1,2)[11]             : AIC=2644.980, Time=16.28 sec
 ARIMA(2,2,2)(0,1,1)[11]             : AIC=2646.636, Time=3.09 sec
 ARIMA(3,2,3)(0,1,1)[11]             : AIC=2639.349, Time=6.24 sec
 ARIMA(3,2,3)(0,1,0)[11]             : AIC=inf, Time=2.77 sec
 ARIMA(3,2,3)(1,1,1)[11]             : AIC=inf, Time=7.81 sec
 ARIMA(3,2,3)(0,1,2)[11]             : AIC=2639.719, Time=25.80 sec
 ARIMA(3,2,3)(1,1,0)[11]             : AIC=inf, Time=6.23 sec
 ARIMA(3,2,3)(1,1,2)[11]             : AIC=2641.076, Time=23.41 sec
 ARIMA(2,2,3)(0,1,1)[11]             : AIC=inf, Time=3.83 sec
 ARIMA(4,2,3)(0,1,1)[11]             : AIC=2641.065, Time=7.46 sec
 ARIMA(3,2,4)(0,1,1)[11]             : AIC=2641.507, Time=11.53 sec
 ARIMA(2,2,4)(0,1,1)[11]             : AIC=inf, Time=7.52 sec
 ARIMA(4,2,4)(0,1,1)[11]             : AIC=2640.038, Time=11.67 sec
 ARIMA(3,2,3)(0,1,1)[11] intercept   : AIC=2655.951, Time=6.52 sec

Best model:  ARIMA(3,2,3)(0,1,1)[11]
Total fit time: 361.414 seconds
```

Fitting SARIMA models is the beginning of the end of this journey into time series modeling. In the above two exercises, we applied the pmdarima package which is a powerful tool to search over model orders. In both cases we apply both the first differencing and then the seasonal differencing.

In the first run we applied a single first differencing followed by a single seasonal differencing and in the second, we apply a single first differencing followed by 2 seasonal differencing.

The .auto_arima() function of this package looks over model orders to find the best one. The advice is not to compare models with different differences, however the models with 2 seasonal differences had a lower AIC compared to the ones with a single seasonal differencing.

For each run, the object returned by this function is the result object of the best model find by the search. The only required input to the function is data (df), optionally, we can also set the order of non-seasonal differencing, initial estimates of the non-seasonal orders, and the maximum values of non-seasonal orders to test.

If the time series is seasonal as is the case here, then we set the seasonal parameter to true. We also need to specify the length of the period, and the order of seasonal differencing, as with non-seasonal parameters, we can specify initial guesses and maximum values for the seasonal orders.

Finally, there are a few non-order parameters that we may want to set. We select whether to choose the best model based on AIC or BIC. If trace is set to True, then this function prints the AIC and BIC for each model it fits as we did here. To ignore bad models set the error_action to ignore.

Some rules of thumb are that: you should never use more than one seasonal differencing and never more than two orders of differencing in total.

• D should be 0 or 1

• d + D should be 0-2

Sometimes you will be able to make the time series stationary by using either one seasonal difference or one no-seasonal difference. You might build models for each in this case and see which one makes better predictions.

In our case here we used more than one seasonal differencing resulting in 3 orders of differencing in total, 1 more than the recommended.

In [133…
```
model_sarima310 = SARIMAX(rrabs_train, order = (3,1,0), seasonal_order = (1,1,1,11))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [134…
```
results_sarima310 = model_sarima310.fit()
```

In [135…
```
residuals_sarima310 = results_sarima310.resid
```

In [136…
```
mae_sarima310 = np.mean(np.abs(residuals_sarima310))
```

In [137…
```
print(mae_sarima310)
```

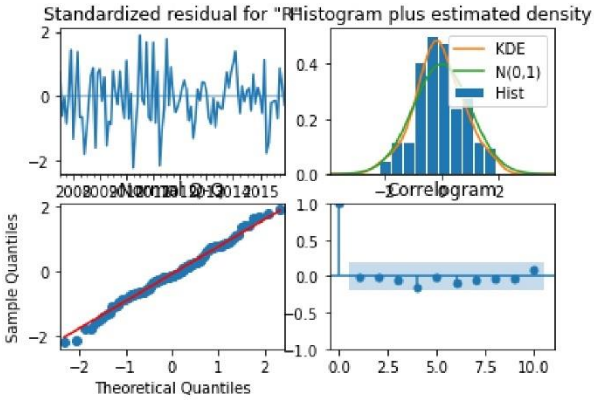
```
85220.53179724926
```

In [138…
```
print(results_sarima310.summary())
```

```
                                      SARIMAX Results

==========================================================================================
==
Dep. Variable:                   RiverRunOffAbsklm   No. Observations:                   1
14
Model:            SARIMAX(3, 1, 0)x(1, 1, [1], 11)   Log Likelihood                -1322.0
22
Date:                             Sun, 05 Feb 2023   AIC                            2656.0
44
Time:                                     10:25:43   BIC                            2671.7
94
Sample:                                 07-01-2006   HQIC                           2662.4
22
```

```
                                         - 12-01-2015

Covariance Type:                                       opg

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.2883      0.134     -2.158      0.031      -0.550      -0.026
ar.L2         -0.0826      0.153     -0.538      0.590      -0.383       0.218
ar.L3         -0.2570      0.102     -2.511      0.012      -0.458      -0.056
ar.S.L11       0.3093      0.212      1.459      0.145      -0.106       0.725
ma.S.L11      -0.8718      0.302     -2.882      0.004      -1.465      -0.279
sigma2      1.336e+10   2.55e-11   5.25e+20      0.000    1.34e+10    1.34e+10
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                 0.00
Prob(Q):                              0.83   Prob(JB):                         1.00
Heteroskedasticity (H):               0.52   Skew:                            -0.00
Prob(H) (two-sided):                  0.06   Kurtosis:                         3.01
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.02e+36. Standa
rd errors may be unstable.
```

In [139…
```
results_sarima310.plot_diagnostics()
plt.show()
```



In [140…
```
sarima310_pred = results_sarima310.get_forecast(steps=66)
```

In [141…
```
sarima310_mean = sarima310_pred.predicted_mean
```

In [142…
```
print(sarima310_mean)
```

```
2016-01-01    368976.192666
2016-02-01    468295.264025
2016-03-01    386936.173945
2016-04-01    328422.342891
2016-05-01    321532.629447
                  ...
2021-02-01    469174.780274
2021-03-01    476504.881172
2021-04-01    490965.348090
```

```
2021-05-01    480012.304690
2021-06-01    479942.450454
Freq: MS, Name: predicted_mean, Length: 66, dtype: float64
```

In [143...

```
sarima310_meandf = pd.DataFrame(sarima310_mean, columns = ['predicted_mean'])
sarima310_meandf["predicted_mean"] = sarima310_meandf["predicted_mean"].astype("float")
sarima310_out = pd.merge(sarima310_meandf, rrabs_testdf, left_index = True, right_index =
sarima310_out['errors'] = abs(sarima310_out['predicted_mean'] - sarima310_out['RiverRunOff
sarima310_out['percent_error'] = (sarima310_out['errors']/sarima310_out['RiverRunOffAbsklm
print(sarima310_out.head())
mape_testsarima310 = sarima310_out['percent_error'].mean()
```

```
            predicted_mean  RiverRunOffAbsklm         errors  percent_error
2016-01-01   368976.192666             480257  111280.807334      23.171095
2016-02-01   468295.264025             491587   23291.735975       4.738070
2016-03-01   386936.173945             462244   75307.826055      16.291791
2016-04-01   328422.342891             405251   76828.657109      18.958289
2016-05-01   321532.629447             366784   45251.370553      12.337335
```

In [144...

```
mae_testsarima310 = mean_absolute_error(rrabs_test, sarima310_mean)
print('Mean_Absolute_Error_310: %f' % mae_testsarima310)
mse_testsarima310 = mean_squared_error(rrabs_test, sarima310_mean)
print('Mean_Squared_Error_310: %f' % mse_testsarima310)
rmse_testsarima310 = mse_testsarima310**(1/2)
print('Root_Mean_Squared_Error_310: %f' % rmse_testsarima310)
rmspe_testsarima310 = (rmse_testsarima310 / sarima310_out['RiverRunOffAbsklm'].mean())*10(
print('Root_Mean_Squared_Percentage_Error_310: %f' % rmspe_testsarima310)
print('Mean_Absolute_Percentage_Error_sarima310: %f' % mape_testsarima310)
```

```
Mean_Absolute_Error_310: 105129.515790
Mean_Squared_Error_310: 15142869559.643194
Root_Mean_Squared_Error_310: 123056.367408
Root_Mean_Squared_Percentage_Error_310: 37.125065
Mean_Absolute_Percentage_Error_sarima310: 38.449217
```

In [145...

```
model_sarima323 = SARIMAX(rrabs_train, order = (3,2,3), seasonal_order = (0,1,1,11))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [146...

```
results_sarima323 = model_sarima323.fit()
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
```

In [147...

```
residuals_sarima323 = results_sarima323.resid
```

In [148...

```
mae_sarima323 = np.mean(np.abs(residuals_sarima323))
```

In [149...

```
print(mae_sarima323)
```

```
91916.76732885827
```

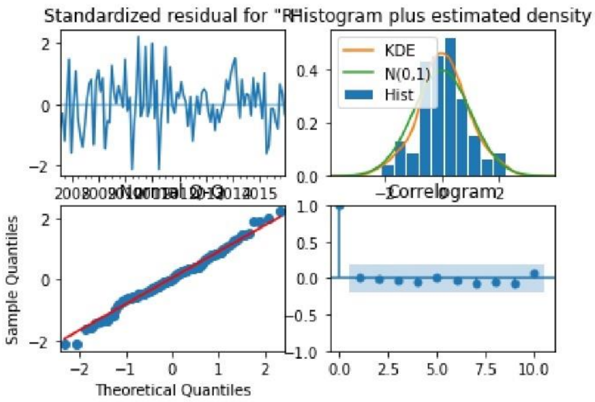
In [150…

```
print(results_sarima323.summary())
```

```
                                     SARIMAX Results
==========================================================================================
==
Dep. Variable:                  RiverRunOffAbsklm   No. Observations:                    1
14
Model:             SARIMAX(3, 2, 3)x(0, 1, [1], 11)   Log Likelihood                 -1311.6
75
Date:                            Sun, 05 Feb 2023   AIC                            2639.3
49
Time:                                    10:26:30   BIC                            2660.2
70
Sample:                                07-01-2006   HQIC                           2647.8
19
                                     - 12-01-2015
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1502      0.661      0.227      0.820      -1.146       1.446
ar.L2          0.2514      0.523      0.480      0.631      -0.774       1.277
ar.L3         -0.2122      0.142     -1.497      0.134      -0.490       0.066
ma.L1         -1.3449      0.624     -2.154      0.031      -2.568      -0.121
ma.L2          0.0798      1.220      0.065      0.948      -2.312       2.472
ma.L3          0.2853      0.639      0.446      0.655      -0.968       1.539
ma.S.L11      -0.7053      0.188     -3.751      0.000      -1.074      -0.337
sigma2      1.387e+10    2.3e-10   6.02e+19      0.000    1.39e+10    1.39e+10
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 0.12
Prob(Q):                              0.93   Prob(JB):                         0.94
Heteroskedasticity (H):               0.59   Skew:                            -0.01
Prob(H) (two-sided):                  0.13   Kurtosis:                         3.17
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 2.4e+36. Standar
d errors may be unstable.
```

In [151…

```
results_sarima323.plot_diagnostics()
plt.show()
```



In [152…
```
sarima323_pred = results_sarima323.get_forecast(steps=66)
```

In [153…
```
sarima323_mean = sarima323_pred.predicted_mean
```

In [154…
```
print(sarima323_mean)
```

```
2016-01-01    383870.019145
2016-02-01    447722.719112
2016-03-01    380639.134129
2016-04-01    337930.439762
2016-05-01    330961.354598
                    ...
2021-02-01    137663.491829
2021-03-01    137610.709843
2021-04-01    155047.822454
2021-05-01    134137.723974
2021-06-01    131662.113511
Freq: MS, Name: predicted_mean, Length: 66, dtype: float64
```

In [155…
```
sarima323_meandf = pd.DataFrame(sarima323_mean, columns = ['predicted_mean'])
sarima323_meandf["predicted_mean"] = sarima323_meandf["predicted_mean"].astype("float")
sarima323_out = pd.merge(sarima323_meandf, rrabs_testdf, left_index = True, right_index =
sarima323_out['errors'] = abs(sarima323_out['predicted_mean'] - sarima323_out['RiverRunOff
sarima323_out['percent_error'] = (sarima323_out['errors']/sarima323_out['RiverRunOffAbsklm
print(sarima323_out.head())
mape_testsarima323 = sarima323_out['percent_error'].mean()
```

```
            predicted_mean  RiverRunOffAbsklm        errors  percent_error
2016-01-01   383870.019145             480257  96386.980855      20.069875
2016-02-01   447722.719112             491587  43864.280888       8.922994
2016-03-01   380639.134129             462244  81604.865871      17.654067
2016-04-01   337930.439762             405251  67320.560238      16.612065
2016-05-01   330961.354598             366784  35822.645402       9.766687
```

In [156…
```
mae_testsarima323 = mean_absolute_error(rrabs_test, sarima323_mean)
print('Mean_Absolute_Error_323: %f' % mae_testsarima323)
mse_testsarima323 = mean_squared_error(rrabs_test, sarima323_mean)
print('Mean_Squared_Error_323: %f' % mse_testsarima323)
rmse_testsarima323 = mse_testsarima323**(1/2)
print('Root_Mean_Squared_Error_323: %f' % rmse_testsarima323)
rmspe_testsarima323 = (rmse_testsarima323 / sarima323_out['RiverRunOffAbsklm'].mean())*10(
```

```
print('Root_Mean_Squared_Percentage_Error_323: %f' % rmspe_testsarima323)
print('Mean_Absolute_Percentage_Error_sarima323: %f' % mape_testsarima323)
```

```
Mean_Absolute_Error_323: 95145.081030
Mean_Squared_Error_323: 13043539553.692945
Root_Mean_Squared_Error_323: 114208.316482
Root_Mean_Squared_Percentage_Error_323: 34.455683
Mean_Absolute_Percentage_Error_sarima323: 30.326299
```

In [157…

```
model_sarima124 = SARIMAX(rrabs_train, order = (1,2,4), seasonal_order = (0,1,1,11))
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.p
y:471: ValueWarning: No frequency information was provided, so inferred frequency MS will
be used.
  self._init_dates(dates, freq)
```

In [158…

```
results_sarima124 = model_sarima124.fit()
```

```
C:\Users\Rejoice van der Walt\Anaconda3\lib\site-packages\statsmodels\tsa\statespace\sarim
ax.py:978: UserWarning: Non-invertible starting MA parameters found. Using zeros as starti
ng parameters.
  warn('Non-invertible starting MA parameters found.'
```

In [159…

```
residuals_sarima124 = results_sarima124.resid
```

In [160…

```
mae_sarima124 = np.mean(np.abs(residuals_sarima124))
```

In [161…

```
print(mae_sarima124)
```

```
95058.58414116365
```

In [162…

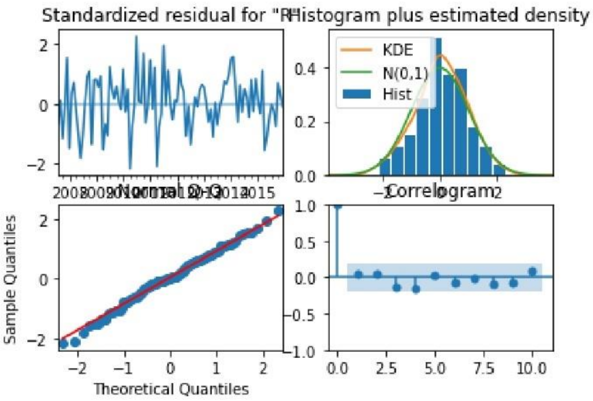
```
print(results_sarima124.summary())
```

```
                                        SARIMAX Results
==========================================================================================
==
Dep. Variable:                   RiverRunOffAbsklm   No. Observations:                  1
14
Model:              SARIMAX(1, 2, 4)x(0, 1, [1], 11)   Log Likelihood              -1313.4
94
Date:                             Sun, 05 Feb 2023   AIC                          2640.9
89
Time:                                     10:27:14   BIC                          2659.2
95
Sample:                                 07-01-2006   HQIC                         2648.3
99
                                      - 12-01-2015

Covariance Type:                               opg

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8919      0.172     -5.180      0.000      -1.229      -0.554
```

```
ma.L1          -0.3855      0.210     -1.832      0.067     -0.798      0.027
ma.L2          -0.9820      0.288     -3.410      0.001     -1.546     -0.418
ma.L3           0.1795      0.155      1.159      0.247     -0.124      0.483
ma.L4           0.2062      0.132      1.558      0.119     -0.053      0.466
ma.S.L11       -0.5662      0.132     -4.286      0.000     -0.825     -0.307
sigma2      1.394e+10   5.09e-12   2.74e+21      0.000   1.39e+10   1.39e+10
===================================================================================
Ljung-Box (L1) (Q):                   0.26   Jarque-Bera (JB):                 0.36
Prob(Q):                              0.61   Prob(JB):                         0.84
Heteroskedasticity (H):               0.58   Skew:                            -0.14
Prob(H) (two-sided):                  0.11   Kurtosis:                         2.89
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.15e+38. Standa
rd errors may be unstable.
```

In [163…

```
results_sarima124.plot_diagnostics()
plt.show()
```



In [164…

```
sarima124_pred = results_sarima124.get_forecast(steps=66)
```

In [165…

```
sarima124_mean = sarima124_pred.predicted_mean
```

In [166…

```
print(sarima124_mean)
```

```
2016-01-01    376988.239117
2016-02-01    457907.249391
2016-03-01    357013.572990
2016-04-01    295409.005368
2016-05-01    288002.175890
                  ...
2021-02-01    -46994.172660
2021-03-01    -48662.731419
2021-04-01    -17593.183551
2021-05-01    -28214.332975
2021-06-01    -35659.326341
Freq: MS, Name: predicted_mean, Length: 66, dtype: float64
```

In [167…

```
sarima124_meandf = pd.DataFrame(sarima124_mean, columns = ['predicted_mean'])
sarima124_meandf["predicted_mean"] = sarima124_meandf["predicted_mean"].astype("float")
```

```
sarima124_out = pd.merge(sarima124_meandf, rrabs_testdf, left_index = True, right_index =
sarima124_out['errors'] = abs(sarima124_out['predicted_mean'] - sarima124_out['RiverRunOff
sarima124_out['percent_error'] = (sarima124_out['errors']/sarima124_out['RiverRunOffAbsklm
print(sarima124_out.head())
mape_testsarima124 = sarima124_out['percent_error'].mean()
```

```
            predicted_mean  RiverRunOffAbsklm         errors  percent_error
2016-01-01   376988.239117             480257  103268.760883      21.502812
2016-02-01   457907.249391             491587   33679.750609       6.851229
2016-03-01   357013.572990             462244  105230.427010      22.765126
2016-04-01   295409.005368             405251  109841.994632      27.104682
2016-05-01   288002.175890             366784   78781.824110      21.479079
```

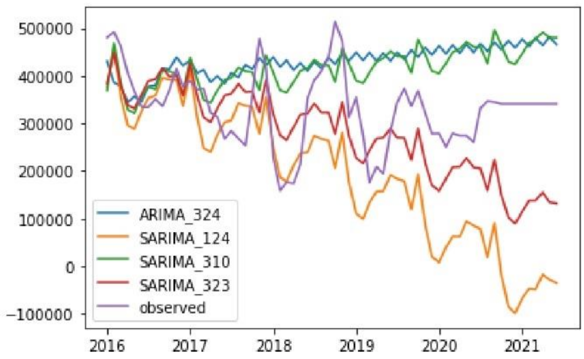
In [168…

```
mae_testsarima124 = mean_absolute_error(rrabs_test, sarima124_mean)
print('Mean_Absolute_Error_124: %f' % mae_testsarima124)
mse_testsarima124 = mean_squared_error(rrabs_test, sarima124_mean)
print('Mean_Squared_Error_124: %f' % mse_testsarima124)
rmse_testsarima124 = mse_testsarima124**(1/2)
print('Root_Mean_Squared_Error_124: %f' % rmse_testsarima124)
rmspe_testsarima124 = (rmse_testsarima124 / sarima124_out['RiverRunOffAbsklm'].mean())*10
print('Root_Mean_Squared_Percentage_Error_124: %f' % rmspe_testsarima124)
print('Mean_Absolute_Percentage_Error_sarima124: %f' % mape_testsarima124)
```

```
Mean_Absolute_Error_124: 154980.322536
Mean_Squared_Error_124: 39674884481.309288
Root_Mean_Squared_Error_124: 199185.552893
Root_Mean_Squared_Percentage_Error_124: 60.092597
Mean_Absolute_Percentage_Error_sarima124: 46.709007
```

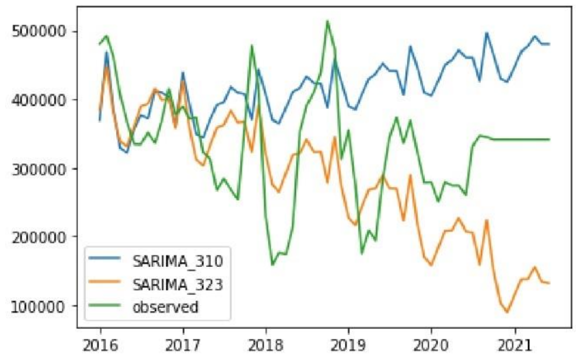
In [169…

```
plt.plot(arima_mean324, label='ARIMA_324')
plt.plot(sarima124_mean, label='SARIMA_124')
plt.plot(sarima310_mean, label='SARIMA_310')
plt.plot(sarima323_mean, label='SARIMA_323')
plt.plot(rrabs_test, label='observed')
plt.legend()
plt.show()
```



In [170…

```
plt.plot(sarima310_mean, label='SARIMA_310')
plt.plot(sarima323_mean, label='SARIMA_323')
plt.plot(rrabs_test, label='observed')
plt.legend()
plt.show()
```



Above is a comparison of the ARIMA(3, 2, 4) model and the SARIMA (3, 2, 3) model against the observed time series. The SARIMA seems to follow the observed time series better, and seems to model the observed time series better but because we are extrapolating the data too far into the future (66months ahead) the SARIMA enters into the negative river runoff abstraction which is senseless.

Now it's time to put the chosen models into practice to make future forecasts. Lets predict!

To make future forecasts we use the get_forecast() method of the result object. We choose the number of steps after the end of the test data to forecast up to. Forecasting too far into the future results in meaningless results. These are out of sample dynamical forecasts (Predictions). It is more difficult to make precise long-term forecasts because the shock terms add up. The further into the future the predictions go, the more uncertain they become. Therefore predictions should be as close to the data points that were used to train the model as possible. Perhaps 6 to 12 months ahead.

The result is the average of all the possible predictions that could be made by the model. The Upper and lower limits are an indicator of the uncertainty associated with our predicted mean and the wider the more uncertain our prediction is.

In [171…
```
forecast_object124 = results_sarima124.get_forecast(steps=36)
```

In [172…
```
mean124 = forecast_object124.predicted_mean
```

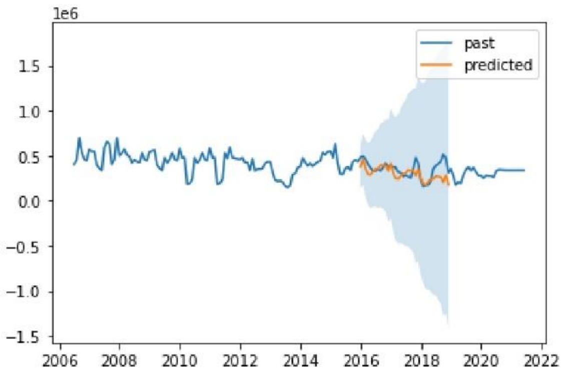
In [173…
```
conf_int124 = forecast_object124.conf_int()
```

In [174…
```
dates = mean124.index
```

In [175…
```
plt.figure()
```

Out[175…
```
<Figure size 432x288 with 0 Axes>

<Figure size 432x288 with 0 Axes>
```

In [176…
```
plt.plot(rrabs.index, rrabs, label='past')
plt.plot(dates, mean124, label='predicted')
plt.fill_between(dates, conf_int124.iloc[:,0], conf_int124.iloc[:,1], alpha=0.2)
plt.legend()
plt.show()
```



In [177…
```
print(mean124.iloc[-1])
```
```
176670.16894333257
```

In [178…
```
print(conf_int124.iloc[-1])
```
```
lower RiverRunOffAbsklm   -1.414134e+06
upper RiverRunOffAbsklm    1.767474e+06
Name: 2018-12-01 00:00:00, dtype: float64
```

In [179…
```
forecast_object310 = results_sarima310.get_forecast(steps=36)
```

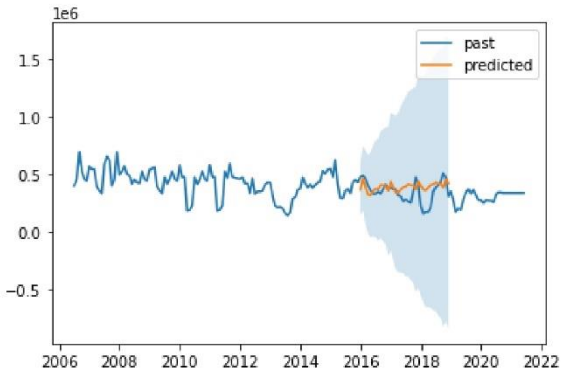
In [180…
```
mean310 = forecast_object310.predicted_mean
```

In [181…
```
conf_int310 = forecast_object310.conf_int()
```

In [182…
```
dates = mean310.index
```

In [183…
```
plt.figure()
```
Out[183…
```
<Figure size 432x288 with 0 Axes>
```
```
<Figure size 432x288 with 0 Axes>
```

In [184…
```
plt.plot(rrabs.index, rrabs, label='past')
plt.plot(dates, mean310, label='predicted')
plt.fill_between(dates, conf_int310.iloc[:,0], conf_int310.iloc[:,1], alpha=0.2)
plt.legend()
plt.show()
```



In [185…

```
print(mean310.iloc[-1])
```

```
425388.41797308467
```

In [186…

```
print(conf_int310.iloc[-1])
```

```
lower RiverRunOffAbsklm   -8.413815e+05
upper RiverRunOffAbsklm    1.692158e+06
Name: 2018-12-01 00:00:00, dtype: float64
```

Saving model objects

Once you have fit a model in this way, you may want to save it and load it latter. You can do this using the joblib package.

In [187…

```
import joblib
```

In [188…

```
import pickle
with open('model_323.pkl', 'wb') as file:
    pickle.dump(model_sarima323, file)
```

In [189…

```
import pickle
with open('model_324.pkl', 'wb') as file:
    pickle.dump(model_324, file)
```

In [190…

```
import pickle
with open('model_sarima124.pkl', 'wb') as file:
    pickle.dump(model_sarima124, file)
```

In [191…

```
import pickle
with open('model_310.pkl', 'wb') as file:
    pickle.dump(model_sarima310, file)
```

The above model objects have been saved in the workbook and can be called to do new preictions in the future.

Latter on when we want to make new predictions, we can load these model again. When new data becomes available, these models can either be updated or be rerun.

In [ ]:

```
```

E2. Supervised machine learning models

Stellenbosch River Runoff Abstraction in kl/m Forecasting using Supervised Machine Learning with Python

Problem Statement:

Design predictive models with the use of machine learning algorithms to forecast River Runoff Abstraction in kl/m using weather information associated with historical quantities of River Runoff Abstraction in kl/m.

Importing libraries:

The first step in any Data Analysis step is importing the necessary libraries.

In [1]:
```
import pandas as pd
```

In [2]:
```
import numpy as np
```

In [3]:
```
import matplotlib.pyplot as plt
```

In [4]:
```
import seaborn as sns
```

In [5]:
```
import tensorflow as tf
```

Load Data Set:

Dataset can be loaded using a method read_csv().

In [6]:
```
StellWater=pd.read_csv('StellWaterClimate2.csv')
```

The shape property is used to find the dimensions of the dataset.

In [7]:
```
print(StellWater.shape)
```

```
(180, 6)
```

Number of columns: 6

Number of rows: 180

Number of Independent Columns: 5

Number of Dependent Column: 1 RoRabs

In [8]:
```
print(StellWater.head())
```

```
        Date    RoRabs     mtmin      mtmax      mtave  spre
0   7/1/2006  404000.0  8.658065  16.922581  12.790323  71.4
1   8/1/2006  455000.0  7.932258  17.738710  12.835484  56.2
```

```
2   9/1/2006  697000.0  10.323333  20.873333  15.598333  20.0
3  10/1/2006  529664.0  11.274194  22.380645  16.827419  37.2
4  11/1/2006  458241.0  13.906667  24.553333  19.230000  37.7
```

In [9]:
```
StellWater.drop('mtave', axis = 1, inplace = True)
StellWater.head()
```

Out[9]:

	Date	RoRabs	mtmin	mtmax	spre
0	7/1/2006	404000.0	8.658065	16.922581	71.4
1	8/1/2006	455000.0	7.932258	17.738710	56.2
2	9/1/2006	697000.0	10.323333	20.873333	20.0
3	10/1/2006	529664.0	11.274194	22.380645	37.2
4	11/1/2006	458241.0	13.906667	24.553333	37.7

Data Preprocessing:

Real-world data is often messy, incomplete, unstructured, inconsistent, redundant, sprinkled with wacky values some of which are not even there. So, without deploying any Data Preprocessing techniques, it is almost impossible to gain insights from raw data.

Data preprocessing is a process of converting raw data to a suitable format to extract insights. It is the first and foremost step in the Data Science life cycle. Data Preprocessing makes sure that data is clean, organized and ready-to-feed to the Machine Learning model.

A concise summary of a Dataset:

In [10]:
```
print(StellWater.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Date    180 non-null    object
 1   RoRabs  180 non-null    float64
 2   mtmin   175 non-null    float64
 3   mtmax   175 non-null    float64
 4   spre    179 non-null    float64
dtypes: float64(4), object(1)
memory usage: 7.2+ KB
None
```

• Dataset has two data types: float64, object

• Except for the Date, Run-of-River abstraction columns, every column has missing values.

Descriptive Statistics:

Let's generate descriptive statistics for the dataset using the function describe() in pandas.

Descriptive Statistics: are used to summarize and describe the features of the data in a meaningful way to extract insights. It uses two types of statistics to describe or summarize data:

• Measures of central tendency and

• Measures of spread (variation)

In [11]:
```
print(StellWater.describe(exclude=[object]))
```

```
              RoRabs        mtmin        mtmax        spre
count     180.000000   175.000000   175.000000   179.000000
mean   393396.853653    12.225765    22.686007    37.415642
std    114303.595717     3.456344     3.557681    35.767688
min    146477.000000     6.090323    16.406452     0.500000
25%    333976.250000     9.171828    19.200269     8.800000
50%    396150.000000    12.236667    23.083871    26.000000
75%    469445.750000    15.587097    26.203226    58.000000
max    697000.000000    18.635484    29.725806   182.400000
```

In [12]:
```
print(StellWater.describe(include=[object]))
```

```
             Date
count         180
unique        180
top      7/1/2006
freq            1
```

In [13]:
```
categorical_features = [column_name for column_name in StellWater.columns if StellWater[co
print("Number of Categorical Features: {}".format(len(categorical_features)))
print("Categorical Features: ",categorical_features)
```

```
Number of Categorical Features: 1
Categorical Features:  ['Date']
```

In [14]:
```
numerical_features = [column_name for column_name in StellWater.columns if StellWater[colu
print("Number of Numerical Features: {}".format(len(numerical_features)))
print("Numerical Features: ",numerical_features)
```

```
Number of Numerical Features: 4
Numerical Features:  ['RoRabs', 'mtmin', 'mtmax', 'spre']
```

Cardinality check for categorical features:

• The accuracy, performance of a classifier not only depends on the model that we use, but also depends on how we preprocess data, and what kind of data we're feeding to the classifier to learn.

• Many Machine learning algorithms like Linear Regression, Logistic Regression, k-nearest neighbors, etc. can handle only numerical data, so encoding categorical data to numeric becomes a necessary step. But before jumping into encoding, check the cardinality of each categorical feature.

• Cardinality: The number of unique values in each categorical feature is known as cardinality.

• A feature with a high number of distinct/ unique values is a high cardinality feature. A categorical feature with hundreds of zip codes is the best example of a high cardinality feature.

• This high cardinality feature poses many serious problems, like it will increase the number of dimensions of data when that feature is encoded. This is not good for the model.

• There are many ways to handle high cardinality, one would be feature engineering and the other is simply dropping that feature if it doesn't add any value to the model.

In [15]:
```
#the cardinality for Categorical features:

for each_feature in categorical_features:
 unique_values = len(StellWater[each_feature].unique())
 print("Cardinality(no. of unique values) of {} are: {}".format(each_feature, unique_value
```

```
Cardinality(no. of unique values) of Date are: 180
```

Date column has high cardinality which poses several problems to the model in terms of efficiency and also dimensions of data increase when encoded to numerical data.

Feature Engineering of Date column to decrease high cardinality:

In [16]:
```
# Feature Engineering of Date column to decrease high cardinality:

StellWater['Date'] = pd.to_datetime(StellWater['Date'])
StellWater['year'] = StellWater['Date'].dt.year
StellWater['month'] = StellWater['Date'].dt.month
StellWater['day'] = StellWater['Date'].dt.day
```

In [17]:
```
print(StellWater.head())
```

```
         Date    RoRabs      mtmin      mtmax  spre  year  month  day
0  2006-07-01  404000.0   8.658065  16.922581  71.4  2006      7    1
1  2006-08-01  455000.0   7.932258  17.738710  56.2  2006      8    1
2  2006-09-01  697000.0  10.323333  20.873333  20.0  2006      9    1
3  2006-10-01  529664.0  11.274194  22.380645  37.2  2006     10    1
4  2006-11-01  458241.0  13.906667  24.553333  37.7  2006     11    1
```

In [18]:
```
StellWater.drop('Date', axis = 1, inplace = True)
StellWater.head()
```

Out[18]:

	RoRabs	mtmin	mtmax	spre	year	month	day
0	404000.0	8.658065	16.922581	71.4	2006	7	1
1	455000.0	7.932258	17.738710	56.2	2006	8	1
2	697000.0	10.323333	20.873333	20.0	2006	9	1
3	529664.0	11.274194	22.380645	37.2	2006	10	1
4	458241.0	13.906667	24.553333	37.7	2006	11	1

In [19]:
```
StellWater.drop('day', axis = 1, inplace = True)
StellWater.head()
```

Out[19]:

	RoRabs	mtmin	mtmax	spre	year	month
0	404000.0	8.658065	16.922581	71.4	2006	7

	RoRabs	mtmin	mtmax	spre	year	month
1	455000.0	7.932258	17.738710	56.2	2006	8
2	697000.0	10.323333	20.873333	20.0	2006	9
3	529664.0	11.274194	22.380645	37.2	2006	10
4	458241.0	13.906667	24.553333	37.7	2006	11

In [20]:

```
print(StellWater.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   RoRabs   180 non-null    float64
 1   mtmin    175 non-null    float64
 2   mtmax    175 non-null    float64
 3   spre     179 non-null    float64
 4   year     180 non-null    int64
 5   month    180 non-null    int64
dtypes: float64(4), int64(2)
memory usage: 8.6 KB
None
```

Handling Missing Values:

Machine learning algorithms can't handle missing values and cause problems. So they need to be addressed in the first place. There are many techniques to identify and impute missing values.

If a dataset contains missing values and loaded using pandas, then missing values get replaced with NaN(Not a Number) values. These NaN values can be identified using methods like isna() or isnull() and they can be imputed using fillna(). This process is known as Missing Data Imputation.

In [21]:

```
# Handling Missing values in Numerical features:

numerical_features = [column_name for column_name in StellWater.columns if StellWater[colu
StellWater[numerical_features].isnull().sum()
```

Out[21]:

```
RoRabs     0
mtmin      5
mtmax      5
spre       1
year       0
month      0
dtype: int64
```

fancyimpute Package

'fancyimpute' is a package containing several advanced imputation techniques that use machine learning algorithms to impute missing values.

In the simplest approaches, we used imputation techniques like mean, median and mode imputations or interpolation. In these techniques, only the respective column is utilized for computing and imputing missing values.

In contrast, the advanced imputation techniques use other columns as well to predict the missing values and impute them. Think of it as fitting a machine learning model to predict the missing values in a column using the remaining columns.

There are two very important techniques, namely, KNN or K Nearest Neighbor imputation and MICE or Multiple Imputation by Chained Equations imputation. We will apply the KNN imputation.

K-Nearest Neighbour Imputation

The KNN imputation technique uses the K-Nearest Neighbor algorithm for predicting the missing values. The KNN algorithm finds the most similar data points using all the non-missing features for a data point and calculates the average of these similar points to fill the missing feature. Here, K specifies the number of similar or nearest points to consider.

In [22]:
```
# Import KNN from fancyimpute

from fancyimpute import KNN
```

In [23]:
```
# Copy StellWater to StellWater_knn_imputed

StellWater_knn_imputed = StellWater.copy(deep=True)
```

In [24]:
```
# Initialize KNN

knn_imputer = KNN()
```

In [25]:
```
# Impute using fit_tranform on diabetes_knn_imputed

StellWater_knn_imputed.iloc[:, :] = knn_imputer.fit_transform(StellWater_knn_imputed)
```

```
Imputing row 1/180 with 0 missing, elapsed time: 0.115
Imputing row 101/180 with 0 missing, elapsed time: 0.128
```

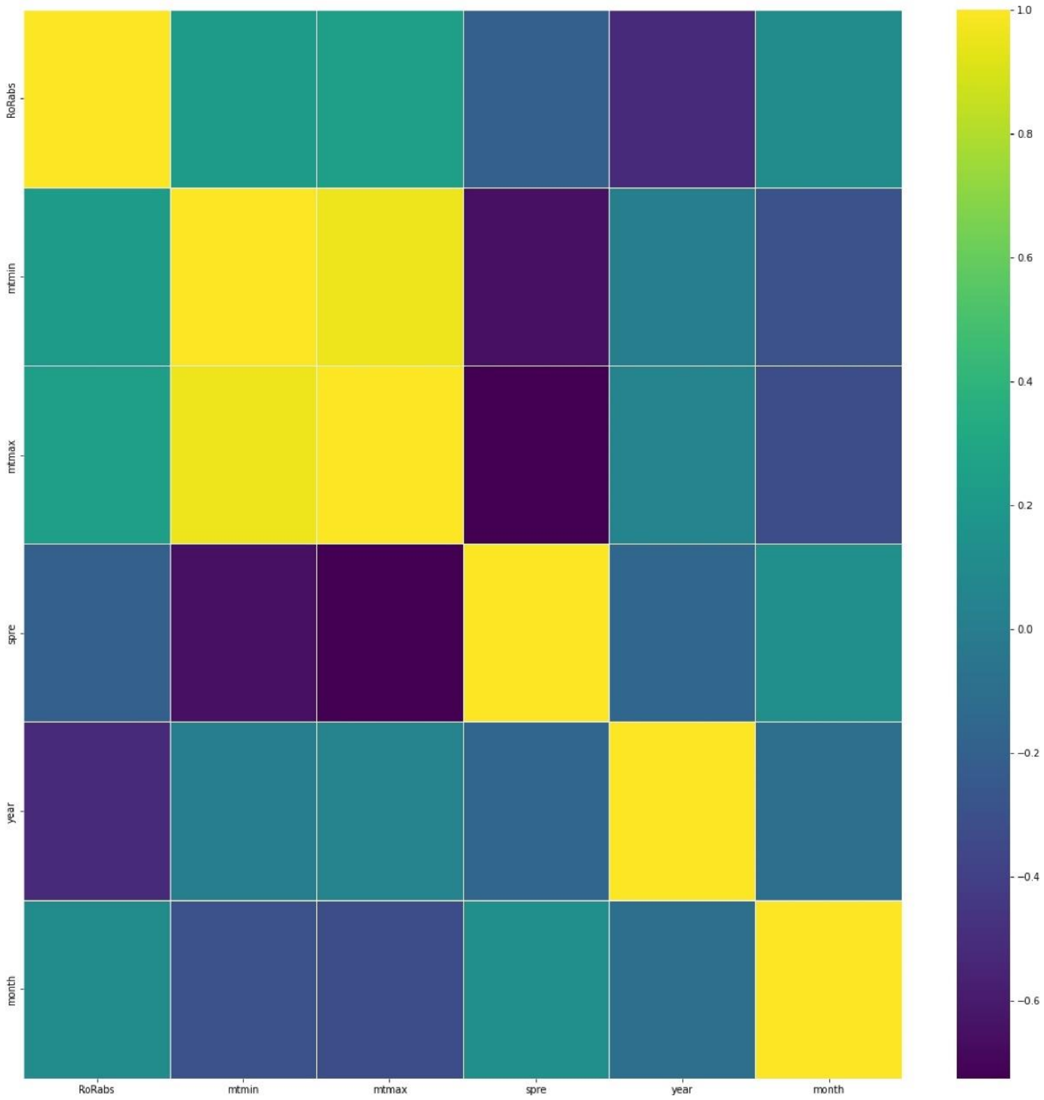
In [26]:
```
StellWater_knn_imputed.head()
```

Out[26]:

	RoRabs	mtmin	mtmax	spre	year	month
0	404000.0	8.658065	16.922581	71.4	2006.0	7.0
1	455000.0	7.932258	17.738710	56.2	2006.0	8.0
2	697000.0	10.323333	20.873333	20.0	2006.0	9.0
3	529664.0	11.274194	22.380645	37.2	2006.0	10.0
4	458241.0	13.906667	24.553333	37.7	2006.0	11.0

In [27]:
```
plt.figure(figsize=(20,20))
sns.heatmap(StellWater_knn_imputed.corr(), linewidths=0.5, annot=False, fmt=".2f", cmap =
```

Out[27]:
```
<AxesSubplot:>
```



In [28]:

```
numerical_features = [column_name for column_name in StellWater_knn_imputed.columns if Ste
StellWater_knn_imputed[numerical_features].isnull().sum()
```

Out[28]:

```
RoRabs     0
mtmin      0
mtmax      0
spre       0
year       0
month      0
dtype: int64
```

In [29]:

```
StellWater_knn_imputed.head()
```

Out[29]:

	RoRabs	mtmin	mtmax	spre	year	month
0	404000.0	8.658065	16.922581	71.4	2006.0	7.0

	RoRabs	mtmin	mtmax	spre	year	month
1	455000.0	7.932258	17.738710	56.2	2006.0	8.0
2	697000.0	10.323333	20.873333	20.0	2006.0	9.0
3	529664.0	11.274194	22.380645	37.2	2006.0	10.0
4	458241.0	13.906667	24.553333	37.7	2006.0	11.0

Exploratory Data Analysis

Exploratory Data Analysis(EDA) is a technique used to:

• analyze,

• visualize,

• investigate,

• interpret,

• discover and

• summarize data.

It helps the researcher to:

• extract trends,

• patterns, and

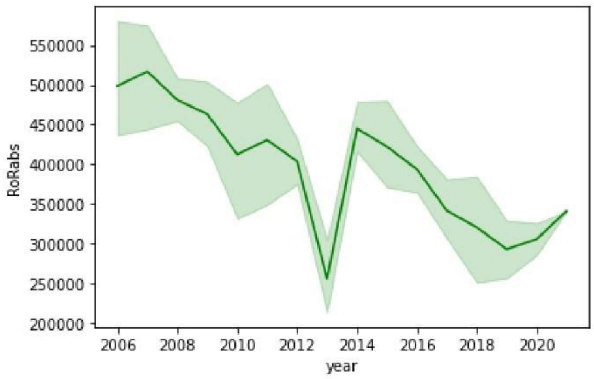
• relationships in data.

In [30]:

```
#Bi-variate Analysis: year vs RoRabs:

sns.lineplot(data=StellWater_knn_imputed,x='year',y='RoRabs',color='green')
```

Out[30]:

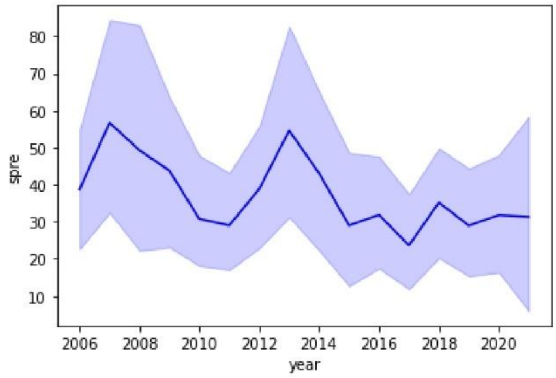
```
<AxesSubplot:xlabel='year', ylabel='RoRabs'>
```



The general trend depicted by the above plot is one of declining Run-off-River abstraction over the years!

In [31]:
```
#Bi-variate Analysis: year vs spre:

sns.lineplot(data=StellWater_knn_imputed,x='year',y='spre',color='blue')
```

Out[31]: <AxesSubplot:xlabel='year', ylabel='spre'>

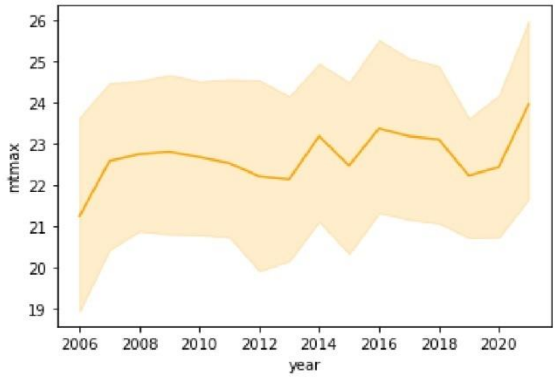


The general trend of the sum of precipitation has been a declining one over the years just like the Run-off-River abstraction, however of concern in comparing these two plots above is the fact that the Run-off-abstraction was lowest in 2013 but the sum of precititation seems to peak about that same period.

This contrary to expectation as one would expect that more water is collected in the years when precipitation was high. What happened in 2013?¶

In [32]:
```
#Bi-variate Analysis: year vs mtmax:

sns.lineplot(data=StellWater_knn_imputed,x='year',y='mtmax',color='orange')
```

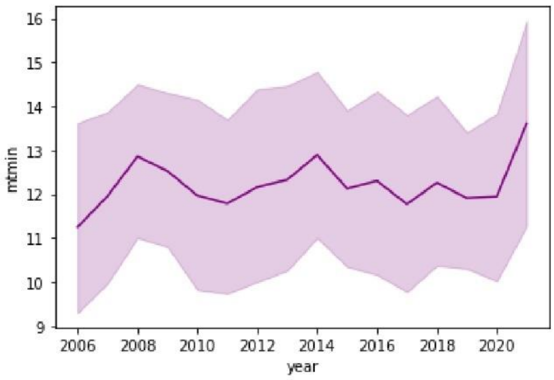
Out[32]: <AxesSubplot:xlabel='year', ylabel='mtmax'>



The general trend is one of rising monthly maximum temperatures over the years perhaps pointing towards climate change concerns dominating international discourse currently!

In [33]:
```
#Bi-variate Analysis: year vs mtmin:

sns.lineplot(data=StellWater_knn_imputed,x='year',y='mtmin',color='purple')
```

Out[33]: <AxesSubplot:xlabel='year', ylabel='mtmin'>



The general trend points to a suttle decline in the monthly minimum temperatures over the years.

The general trend for the average temperature over the years appears relatively flat from 2008 to 2018 with an indication of rising from 2018 on wards.

In [34]:
```
StellWater_knn_imputed['yearmonth'] = StellWater_knn_imputed['year'].astype(str) + StellWa
```

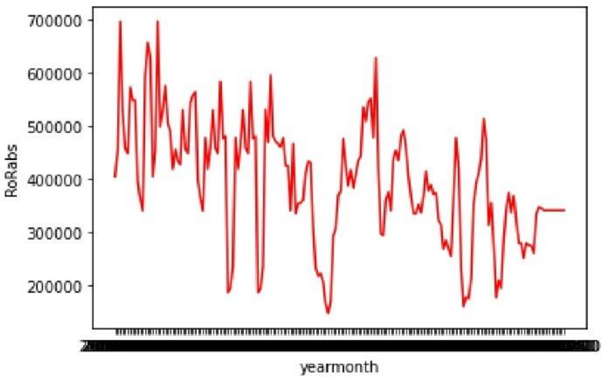
In [35]:
```
StellWater_knn_imputed.head()
```

Out[35]:

	RoRabs	mtmin	mtmax	spre	year	month	yearmonth
0	404000.0	8.658065	16.922581	71.4	2006.0	7.0	2006.07.0
1	455000.0	7.932258	17.738710	56.2	2006.0	8.0	2006.08.0
2	697000.0	10.323333	20.873333	20.0	2006.0	9.0	2006.09.0
3	529664.0	11.274194	22.380645	37.2	2006.0	10.0	2006.010.0
4	458241.0	13.906667	24.553333	37.7	2006.0	11.0	2006.011.0

In [36]:
```
sns.lineplot(data=StellWater_knn_imputed,x='yearmonth',y='RoRabs',color='red')
```

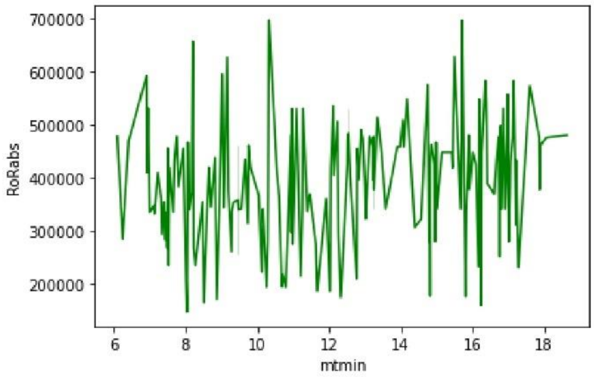
Out[36]: <AxesSubplot:xlabel='yearmonth', ylabel='RoRabs'>



In [37]:
```
#Bi-variate Analysis: mtmin vs Run-of-River abstraction...2:
```

```
sns.lineplot(data=StellWater_knn_imputed,x='mtmin',y='RoRabs',color='green')
```

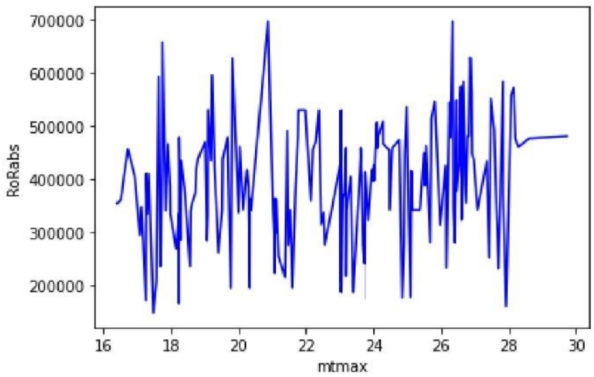
Out[37]: <AxesSubplot:xlabel='mtmin', ylabel='RoRabs'>



In [38]:

```
#Bi-variate Analysis: mtmax vs RoRabs:

sns.lineplot(data=StellWater_knn_imputed,x='mtmax',y='RoRabs',color='blue')
```

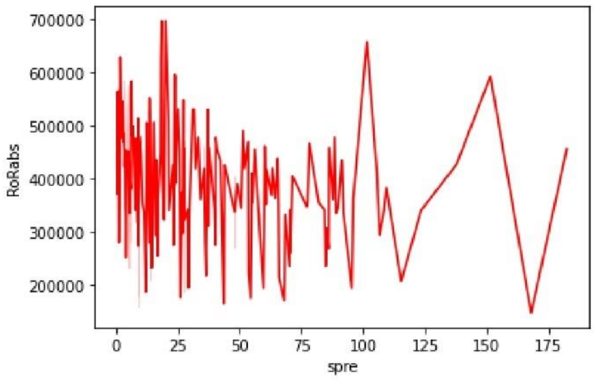
Out[38]: <AxesSubplot:xlabel='mtmax', ylabel='RoRabs'>



In [39]:

```
#Bi-variate Analysis: spre vs RoRabs:

sns.lineplot(data=StellWater_knn_imputed,x='spre',y='RoRabs',color='red')
```

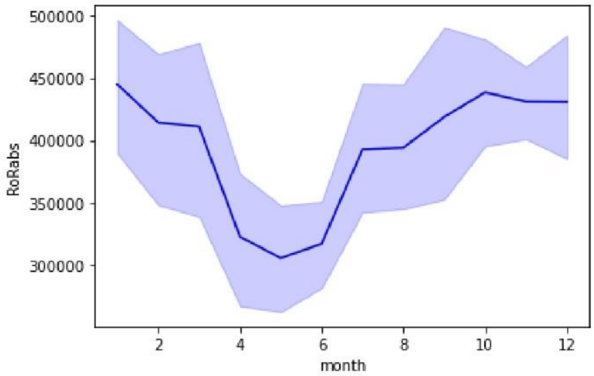
Out[39]: <AxesSubplot:xlabel='spre', ylabel='RoRabs'>



In [40]:

```
#Bi-variate Analysis: month vs RoRabs:
sns.lineplot(data=StellWater_knn_imputed,x='month',y='RoRabs',color='blue')
```

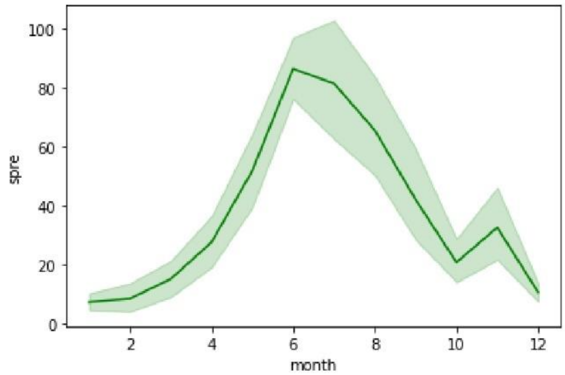
Out[40]: <AxesSubplot:xlabel='month', ylabel='RoRabs'>



In [41]:

```
#Bi-variate Analysis: month vs spre:
sns.lineplot(data=StellWater_knn_imputed,x='month',y='spre',color='green')
```

Out[41]: <AxesSubplot:xlabel='month', ylabel='spre'>



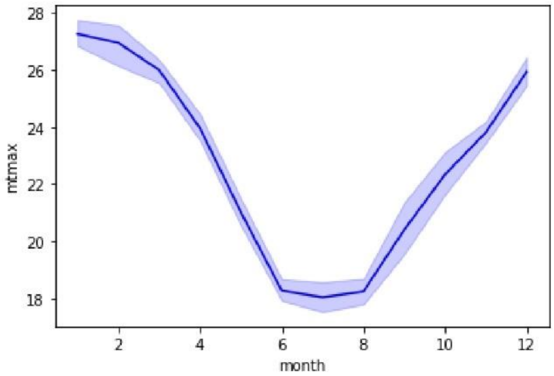
The plot of the Run-off-River abstraction is quite unexpected and concerning more if interpreted together with the month versus precipitation plot. The indication is that the Run-off-River abstraction is lowest during the months when the precipitation is the highest. This is contrary to the expectation.

One would expect the Run-off-River abstraction to be hughest during the rain season and lowest during the dry season. Even it the abstraction methods were inefficient, surely more water would be collected during the rain season than during the dry seaon umless if water is diverted and hence not accounted for during the periods of high precipitation.

In [42]:

```
#Bi-variate Analysis: month vs mtmax:

sns.lineplot(data=StellWater_knn_imputed,x='month',y='mtmax',color='blue')
```

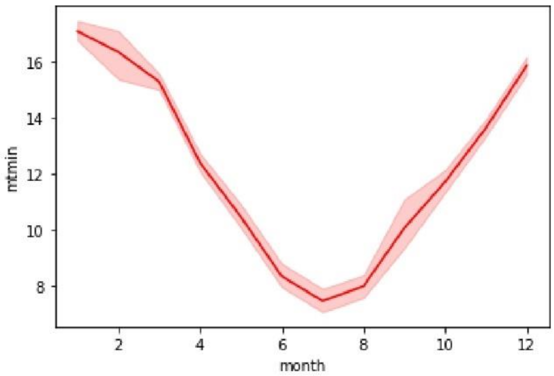
Out[42]: <AxesSubplot:xlabel='month', ylabel='mtmax'>



In [43]:

```
#Bi-variate Analysis: month vs mtmin:

sns.lineplot(data=StellWater_knn_imputed,x='month',y='mtmin',color='red')
```

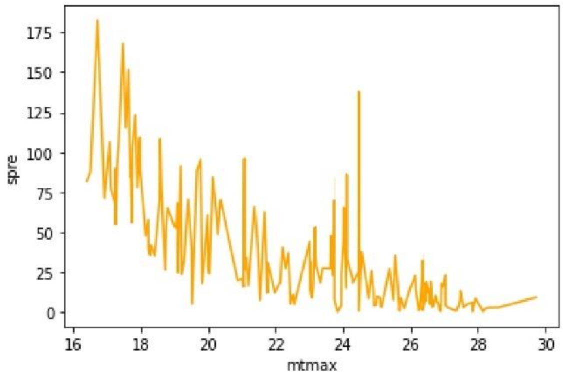
Out[43]: <AxesSubplot:xlabel='month', ylabel='mtmin'>



In [44]:

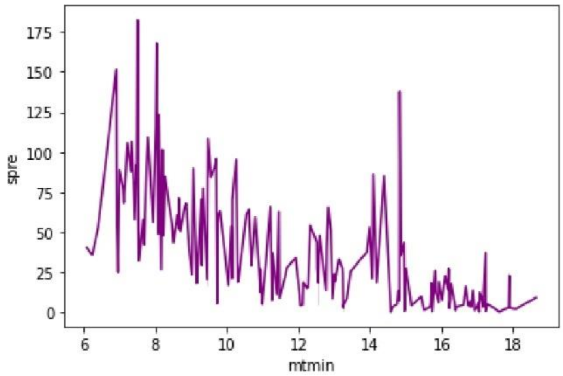
```
#Bi-variate Analysis: mtmax vs spre:

sns.lineplot(data=StellWater_knn_imputed,x='mtmax',y='spre',color='orange')
```

Out[44]: <AxesSubplot:xlabel='mtmax', ylabel='spre'>



In [45]:
```
#Bi-variate Analysis: mtmin vs spre:

sns.lineplot(data=StellWater_knn_imputed,x='mtmin',y='spre',color='purple')
```

Out[45]: <AxesSubplot:xlabel='mtmin', ylabel='spre'>



In [46]:
```
print(StellWater_knn_imputed.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   RoRabs     180 non-null    float64
 1   mtmin      180 non-null    float64
 2   mtmax      180 non-null    float64
 3   spre       180 non-null    float64
 4   year       180 non-null    float64
 5   month      180 non-null    float64
 6   yearmonth  180 non-null    object
dtypes: float64(6), object(1)
memory usage: 10.0+ KB
None
```

In [47]:
```
rrabs = StellWater_knn_imputed.copy(deep=True)
```

In [48]:
```
print(rrabs.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   RoRabs     180 non-null    float64
 1   mtmin      180 non-null    float64
 2   mtmax      180 non-null    float64
 3   spre       180 non-null    float64
 4   year       180 non-null    float64
 5   month      180 non-null    float64
 6   yearmonth  180 non-null    object 
dtypes: float64(6), object(1)
memory usage: 10.0+ KB
None
```

In [49]:

```
rrabs.drop('yearmonth', axis = 1, inplace = True)
rrabs.head()
```

Out[49]:

	RoRabs	mtmin	mtmax	spre	year	month
0	404000.0	8.658065	16.922581	71.4	2006.0	7.0
1	455000.0	7.932258	17.738710	56.2	2006.0	8.0
2	697000.0	10.323333	20.873333	20.0	2006.0	9.0
3	529664.0	11.274194	22.380645	37.2	2006.0	10.0
4	458241.0	13.906667	24.553333	37.7	2006.0	11.0

In [50]:

```
rrabs['year'] = rrabs['year'].astype('int')
print(rrabs.head())
```

```
     RoRabs      mtmin      mtmax  spre  year  month
0  404000.0   8.658065  16.922581  71.4  2006    7.0
1  455000.0   7.932258  17.738710  56.2  2006    8.0
2  697000.0  10.323333  20.873333  20.0  2006    9.0
3  529664.0  11.274194  22.380645  37.2  2006   10.0
4  458241.0  13.906667  24.553333  37.7  2006   11.0
```

In [51]:

```
rrabs["month"] = rrabs["month"].astype("int")
print(rrabs.dtypes)
```

```
RoRabs    float64
mtmin     float64
mtmax     float64
spre      float64
year        int32
month       int32
dtype: object
```

In [52]:

```
print(rrabs.head())
```

```
     RoRabs      mtmin      mtmax  spre  year  month
0  404000.0   8.658065  16.922581  71.4  2006      7
1  455000.0   7.932258  17.738710  56.2  2006      8
2  697000.0  10.323333  20.873333  20.0  2006      9
3  529664.0  11.274194  22.380645  37.2  2006     10
4  458241.0  13.906667  24.553333  37.7  2006     11
```

Machine Learning:

Machine learning is the process whereby computers learn to make decisions from data without being explicitly programmed. Supervised learning is a type of machine learning where the values to be predicted are already known, and a model is built with the aim of accurately predicting values of previously unseen data.

Supervised learning uses features to predict the value of a target variable, such as predicting a basketball player's position based on their points per game. In our case here, the quantity to be predicted (target) is the Run-off-River abstraction and the features we will be using to do that are the monthly minimum temperature (mtmin), monthly maximum temperature (mtmax), monthhly average temperature (mtave) and the month extracted from the Date.

Regression is used to predict continuous values. We will be using decision tree ensemble training to do the regression. In regression, the target variable is continuous. In other words, the output of your model is a real value.

Decision trees are supervised learning models used for problems involving

• classification and

• regression.

Tree models present a high flexibility that comes at a price:

• on one hand, trees are able to capture complex non-linear relationships;

• on the other hand, they are prone to memorizing the noise present in a dataset.

By aggregating the predictions of trees that are trained differently, ensemble methods take advantage of the flexibility of trees while reducing their tendency to memorize noise. Ensemble methods are used across a variety of fields and have a proven track record of winning many machine learning competitions.

Ensemble learning can be summarized as follows:

• -As a first step, different models are trained on the same dataset.

• -Each model makes its own predictions.

• -A meta-model then aggregates the predictions of individual models and outputs a final prediction.

• -The final prediction is more robust and less prone to errors than each individual model.

• -The best results are obtained when the models are skilful but in different ways meaning that if some models make predictions that are way off, the other models should compensate these errors.

In such case, the meta-model's predictions are more robust.

Boosting refers to an ensemble method in which several models are trained sequentially with each model learning from the errors of its predecessors.

In [53]:

```
from sklearn.ensemble import AdaBoostRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error as MSE
from sklearn.model_selection import train_test_split
```

In [54]:
```
# Separate the target variable (RoRabs) and the features (predictors):

X = rrabs[["year", "mtmin", "mtmax", "spre", "month"]].values
y = rrabs["RoRabs"].values
print(X.shape, y.shape)
```

```
(180, 5) (180,)
```

In order to obtain an unbiased estimate of a model's performance, you must evaluate it on an unseen test set.

To do so, first split the data into 80% train and 20% test using train_test_split().

Set the parameter stratify to y in order for the train and test sets to have the same proportion of class labels as the unsplit dataset.

In [55]:
```
# Set seed for reproducibility:

SEED = 1

# Split data into 80% train and 20% test:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=SEED)

print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

```
(144, 5) (36, 5) (144,) (36,)
```

In [56]:
```
X_testdf = pd.DataFrame(X_test, columns = ['year', 'mtmin', 'mtmax', 'spre', 'month'])
X_testdf["month"] = X_testdf["month"].astype("int")
X_testdf["year"] = X_testdf["year"].astype("int")
print(X_testdf.head())
```

```
   year      mtmin      mtmax   spre  month
0  2016  14.854839  25.551613   35.6      3
1  2015  13.456667  24.470000   25.6     11
2  2014  12.116129  24.980645    4.8     10
3  2019  14.790323  24.900000   13.5     12
4  2009   9.473333  18.580000  108.4      6
```

In [57]:
```
X_traindf = pd.DataFrame(X_train, columns = ['year', 'mtmin', 'mtmax','spre', 'month'])
X_traindf["month"] = X_traindf["month"].astype("int")
X_traindf["year"] = X_traindf["year"].astype("int")
print(X_traindf.head())
```

```
   year      mtmin      mtmax  spre  month
0  2019  12.770000  23.026667  13.9      4
1  2010   7.520000  18.560000  70.2      6
2  2013   8.503226  18.229032  43.6      7
3  2016  12.590000  23.646667  48.0      4
4  2021  10.151613  20.367742  70.4      5
```

In [58]:
```
y_traindf = pd.DataFrame(y_train, columns = ['RoRabs'])
y_traindf["RoRabs"] = y_traindf["RoRabs"].astype("float")
print(y_traindf.head())
```

```
     RoRabs
0  208665.0
1  234205.0
2  163722.0
```

```
3   405251.0
4   340779.0
```

In [59]:
```
y_testdf = pd.DataFrame(y_test, columns = ['RoRabs'])
y_testdf["RoRabs"] = y_testdf["RoRabs"].astype("float")
print(y_testdf.head())
```

```
     RoRabs
0  462244.0
1  453707.0
2  535226.0
3  278396.0
4  339402.0
```

Adaboost Model:

AdaBoost stands for Adaptive Boosting.

In AdaBoost,

• each predictor pays more attention to the instances wrongly predicted by its predecessor by constantly changing the weights of training instances. Furthermore,

• each predictor is assigned a coefficient alpha that weighs its contribution in the ensemble's final prediction. Alpha depends on the predictor's training error.

An important parameter used in training is the learning rate, eta.

• Eta is a number between 0 and 1;

• it is used to shrink the coefficient alpha of a trained predictor.

It's important to note that there's a trade-off between eta and the number of estimators.

A smaller value of eta should be compensated by a greater number of estimators.

In [60]:
```
# ADABOOST MODEL

# Instantiate dt
dt = DecisionTreeRegressor(max_depth=2, random_state=SEED)

# Instantiate ada
ada = AdaBoostRegressor(base_estimator=dt, n_estimators=180, random_state=SEED)

# Fit ada to the training set
ada.fit(X_train, y_train)
```

Out[60]:
```
AdaBoostRegressor(base_estimator=DecisionTreeRegressor(max_depth=2,
                                                       random_state=1),
                  n_estimators=180, random_state=1)
```

In [61]:
```
from sklearn.metrics import mean_absolute_error
```

In [62]:
```
from sklearn.metrics import mean_squared_error
```

In [63]:
```
# Predict train set labels
```

```
y_trainpred_ada = ada.predict(X_train)

# Convert y_pred_ada from Numpy array to Data Frame
adapredstrain_col = pd.DataFrame()
adapredstrain_col['adaboosttrainpredictions'] = y_trainpred_ada.tolist()
adapredstrain_col['recordedtrainRoRabs'] = y_train
adapredstrain_col['errors'] = abs(adapredstrain_col['recordedtrainRoRabs'] - adapredstrai
adapredstrain_col['percent_error'] = (adapredstrain_col['errors']/adapredstrain_col['reco
adaboostpredstrain_out = pd.merge(X_traindf, adapredstrain_col, left_index = True, right_
print(adaboostpredstrain_out.head())

# Compute test set MSE
mse_train_ada = MSE(y_train, y_trainpred_ada)

# Compute test set RMSE
rmse_train_ada = mse_train_ada**(1/2)

# Print rmse_test_ada
print('Train set RMSE of ada: {:.3f}'.format(rmse_train_ada))
```

```
   year      mtmin      mtmax  spre  month  adaboosttrainpredictions  \
0  2019  12.770000  23.026667  13.9      4             294193.002333
1  2010   7.520000  18.560000  70.2      6             313425.934682
2  2013   8.503226  18.229032  43.6      7             313425.934682
3  2016  12.590000  23.646667  48.0      4             312300.402700
4  2021  10.151613  20.367742  70.4      5             293513.264227

   recordedtrainRoRabs         errors  percent_error
0             208665.0   85528.002333      40.988188
1             234205.0   79220.934682      33.825467
2             163722.0  149703.934682      91.437885
3             405251.0   92950.597300      22.936550
4             340779.0   47265.735773      13.869909
Train set RMSE of ada: 81784.082
```

In [64]:

```
mae_train_ada = mean_absolute_error(y_train, y_trainpred_ada)
print('Mean_Absolute_Error_trainada: %f' % mae_train_ada)
print('Mean_Squared_Error_trainada: %f' % mse_train_ada)
print('Root_Mean_Squared_Error_trainada: %f' % rmse_train_ada)
rmspe_train_ada = (rmse_train_ada / adapredstrain_col['recordedtrainRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_trainada: %f' % rmspe_train_ada)
mape_train_ada = (mae_train_ada / adapredstrain_col['recordedtrainRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_trainada: %f' % mape_train_ada)
```

```
Mean_Absolute_Error_trainada: 68738.529315
Mean_Squared_Error_trainada: 6688636014.493271
Root_Mean_Squared_Error_trainada: 81784.081669
Root_Mean_Squared_Percentage_Error_trainada: 21.199225
Mean_Absolute_Percentage_Error_trainada: 17.817691
```

In [65]:

```
# Predict test set labels
y_pred_ada = ada.predict(X_test)

# Convert y_pred_ada from Numpy array to Data Frame
adapreds_col = pd.DataFrame()
adapreds_col['adaboostpredictions'] = y_pred_ada.tolist()
adapreds_col['recordedRoRabs'] = y_test
adapreds_col['errors'] = abs(adapreds_col['recordedRoRabs'] - adapreds_col['adaboostpredic
adapreds_col['percent_error'] = (adapreds_col['errors']/adapreds_col['recordedRoRabs'])*1(
adaboostpreds_out = pd.merge(X_testdf, adapreds_col, left_index = True, right_index = True
print(adaboostpreds_out.head())

# Compute test set MSE
mse_test_ada = MSE(y_test, y_pred_ada)
```

```
# Compute test set RMSE
rmse_test_ada = mse_test_ada**(1/2)

# Print rmse_test_ada
print('Test set RMSE of ada: {:.3f}'.format(rmse_test_ada))
```

```
   year     mtmin      mtmax   spre  month  adaboostpredictions  \
0  2016  14.854839  25.551613   35.6      3        313425.934682
1  2015  13.456667  24.470000   25.6     11        409802.987952
2  2014  12.116129  24.980645    4.8     10        409802.987952
3  2019  14.790323  24.900000   13.5     12        399510.917717
4  2009   9.473333  18.580000  108.4      6        437186.933333

   recordedRoRabs         errors  percent_error
0        462244.0  148818.065318      32.194699
1        453707.0   43904.012048       9.676732
2        535226.0  125423.012048      23.433655
3        278396.0  121114.917717      43.504547
4        339402.0   97784.933333      28.810948
Test set RMSE of ada: 86568.683
```

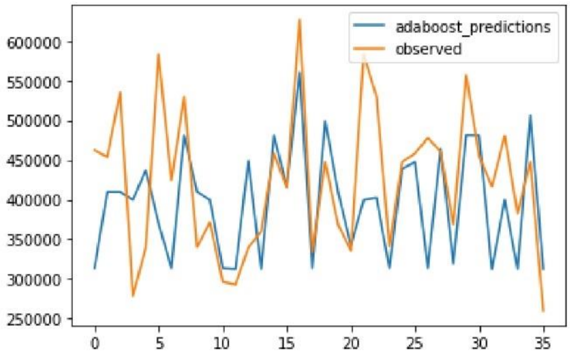
In [66]:

```
mae_test_ada = mean_absolute_error(y_test, y_pred_ada)
print('Mean_Absolute_Error_testada: %f' % mae_test_ada)
print('Mean_Squared_Error_testada: %f' % mse_test_ada)
print('Root_Mean_Squared_Error_testada: %f' % rmse_test_ada)
rmspe_test_ada = (rmse_test_ada / adapreds_col['recordedRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_testada: %f' % rmspe_test_ada)
mape_test_ada = (mae_test_ada / adapreds_col['recordedRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_testada: %f' % mape_test_ada)
```

```
Mean_Absolute_Error_testada: 68190.706520
Mean_Squared_Error_testada: 7494136839.101922
Root_Mean_Squared_Error_testada: 86568.682785
Root_Mean_Squared_Percentage_Error_testada: 20.425238
Mean_Absolute_Percentage_Error_testada: 16.089091
```

In [67]:

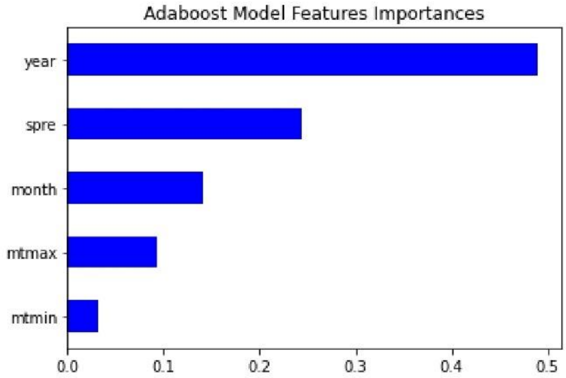
```
plt.plot(adapreds_col['adaboostpredictions'], label='adaboost_predictions')
plt.plot(adapreds_col['recordedRoRabs'], label='observed')
plt.legend()
plt.show()
```



In [68]:

```
# Feature importances according to the Adaboost model

# Create a pd.Series of features importances
ada_importances = pd.Series(data=ada.feature_importances_,
                        index= X_traindf.columns)
```

```
# Sort importances
ada_importances_sorted = ada_importances.sort_values()

# Draw a horizontal barplot of importances_sorted
ada_importances_sorted.plot(kind='barh', color='blue')
plt.title('Adaboost Model Features Importances')
plt.show()
```



Gradient Boosting Model (GBM)

Gradient Boosting is a popular boosting algorithm that has a proven track record of winning many machine learning competitions.

In gradient boosting, each predictor in the ensemble corrects its predecessor's error. In contrast to AdaBoost, the weights of the training instances are not tweaked.

Instead, each predictor is trained using the residual errors of its predecessor as labels.

The ensemble consists of N trees.

Tree1 is trained using the features matrix X and the dataset labels y.

The predictions labeled y_1hat are used to determine the training set residual errors r_1hat.

Tree2 is then trained using the features matrix X and the residual errors of Tree1 as labels.

The predicted residuals r_1hat are then used to determine the residuals of residuals which are labeled r_2hat.

This process is repeated until all of the N trees forming the ensemble are trained.

Shrinkage

An important parameter used in training gradient boosted trees is shrinkage.

In this context, shrinkage refers to the fact that the prediction of each tree in the ensemble is shrinked after it is multiplied by a learning rate eta which is a number between 0 and 1.

Similarly to AdaBoost, there's a trade-off between eta and the number of estimators.

• Decreasing the learning rate needs to be compensated by increasing the number of estimators in order for the ensemble to reach a certain performance.

In [69]:
```
# GRADIENT BOOSTING MODEL

from sklearn.ensemble import GradientBoostingRegressor

# Instantiate gbm
gbm = GradientBoostingRegressor(max_depth=2,
            n_estimators=200,
            random_state=SEED)

# Fit gb to the training set
gbm.fit(X_train, y_train)
```

Out[69]:
```
GradientBoostingRegressor(max_depth=2, n_estimators=200, random_state=1)
```

In [70]:
```
# Predict train set labels
y_trainpred_gbm = gbm.predict(X_train)

# Convert y_pred_ada from Numpy array to Data Frame
gbmpredstrain_col = pd.DataFrame()
gbmpredstrain_col['gbmboosttrainpredictions'] = y_trainpred_gbm.tolist()
gbmpredstrain_col['recordedtrainRoRabs'] = y_train
gbmpredstrain_col['errors'] = abs(gbmpredstrain_col['recordedtrainRoRabs'] - gbmpredstrair
gbmpredstrain_col['percent_error'] = (gbmpredstrain_col['errors']/gbmpredstrain_col['recor
gbmboostpredstrain_out = pd.merge(X_traindf, gbmpredstrain_col, left_index = True, right_i
print(gbmboostpredstrain_out.head())

# Compute test set MSE
mse_train_gbm = MSE(y_train, y_trainpred_gbm)

# Compute test set RMSE
rmse_train_gbm = mse_train_gbm**(1/2)

# Print rmse_test_ada
print('Train set RMSE of gbm: {:.3f}'.format(rmse_train_gbm))
```

```
   year     mtmin      mtmax  spre  month  gbmboosttrainpredictions  \
0  2019  12.770000  23.026667  13.9      4             230635.234488
1  2010   7.520000  18.560000  70.2      6             277033.858310
2  2013   8.503226  18.229032  43.6      7             215445.725361
3  2016  12.590000  23.646667  48.0      4             356020.280845
4  2021  10.151613  20.367742  70.4      5             301247.744270

   recordedtrainRoRabs        errors  percent_error
0             208665.0  21970.234488      10.528950
1             234205.0  42828.858310      18.286910
2             163722.0  51723.725361      31.592410
3             405251.0  49230.719155      12.148204
4             340779.0  39531.255730      11.600262
Train set RMSE of gbm: 32079.627
```

In [71]:
```
mae_train_gbm = mean_absolute_error(y_train, y_trainpred_gbm)
print('Mean_Absolute_Error_traingbm: %f' % mae_train_gbm)
print('Mean_Squared_Error_traingbm: %f' % mse_train_gbm)
print('Root_Mean_Squared_Error_traingbm: %f' % rmse_train_gbm)
rmspe_train_gbm = (rmse_train_gbm / gbmpredstrain_col['recordedtrainRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_traingbm: %f' % rmspe_train_gbm)
mape_train_gbm = (mae_train_gbm / gbmpredstrain_col['recordedtrainRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_traingbm: %f' % mape_train_gbm)
```

```
Mean_Absolute_Error_traingbm: 25136.592121
Mean_Squared_Error_traingbm: 1029102456.210484
Root_Mean_Squared_Error_traingbm: 32079.626809
Root_Mean_Squared_Percentage_Error_traingbm: 8.315349
Mean_Absolute_Percentage_Error_traingbm: 6.515648
```

In [72]:

```
# Predict test set labels
y_pred_gbm = gbm.predict(X_test)

# Convert y_pred_gbm from Numpy array to Data Frame
gbmpreds_col = pd.DataFrame()
gbmpreds_col['gbmpredictions'] = y_pred_gbm.tolist()
gbmpreds_col['recordedRoRabs'] = y_test
gbmpreds_col['errors'] = abs(gbmpreds_col['recordedRoRabs'] - gbmpreds_col['gbmpredictions
gbmpreds_col['percent_error'] = (gbmpreds_col['errors']/gbmpreds_col['recordedRoRabs'])*1(
gbmpreds_out = pd.merge(X_testdf, gbmpreds_col, left_index = True, right_index = True)
print(gbmpreds_out.head())

# Compute test set MSE
mse_test_gbm = MSE(y_test, y_pred_gbm)

# Compute test set RMSE
rmse_test_gbm = mse_test_gbm**(1/2)

# Print rmse_test_ada
print('Test set RMSE of gbm: {:.3f}'.format(rmse_test_gbm))
```

```
   year      mtmin      mtmax   spre  month  gbmpredictions  recordedRoRabs  \
0  2016  14.854839  25.551613   35.6      3   369880.395416        462244.0
1  2015  13.456667  24.470000   25.6     11   475276.472876        453707.0
2  2014  12.116129  24.980645    4.8     10   433396.975716        535226.0
3  2019  14.790323  24.900000   13.5     12   368249.640873        278396.0
4  2009   9.473333  18.580000  108.4      6   390437.894226        339402.0

          errors  percent_error
0   92363.604584      19.981569
1   21569.472876       4.754053
2  101829.024284      19.025426
3   89853.640873      32.275478
4   51035.894226      15.037005
Test set RMSE of gbm: 99579.861
```

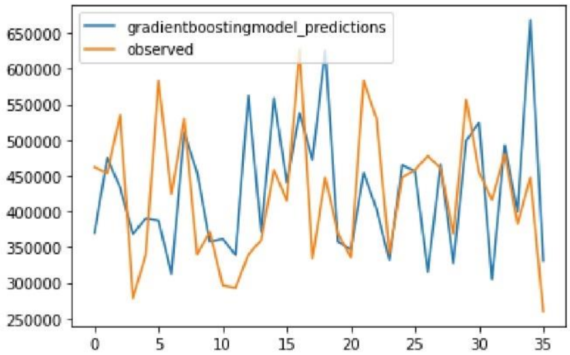
In [73]:

```
mae_test_gbm = mean_absolute_error(y_test, y_pred_gbm)
print('Mean_Absolute_Error_testgbm: %f' % mae_test_gbm)
print('Mean_Squared_Error_testgbm: %f' % mse_test_gbm)
print('Root_Mean_Squared_Error_testgbm: %f' % rmse_test_gbm)
rmspe_test_gbm = (rmse_test_gbm / gbmpreds_col['recordedRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_testgbm: %f' % rmspe_test_gbm)
mape_test_gbm = (mae_test_gbm / gbmpreds_col['recordedRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_testgbm: %f' % mape_test_gbm)
```

```
Mean_Absolute_Error_testgbm: 76818.888615
Mean_Squared_Error_testgbm: 9916148702.120998
Root_Mean_Squared_Error_testgbm: 99579.860926
Root_Mean_Squared_Percentage_Error_testgbm: 23.495129
Mean_Absolute_Percentage_Error_testgbm: 18.124846
```

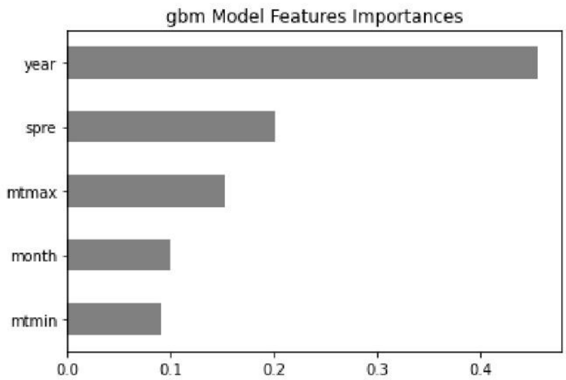
In [74]:

```
plt.plot(gbmpreds_col['gbmpredictions'], label='gradientboostingmodel_predictions')
plt.plot(gbmpreds_col['recordedRoRabs'], label='observed')
plt.legend()
plt.show()
```



In [75]:

```
# Feature importances according to the gbm model

# Create a pd.Series of features importances
gbm_importances = pd.Series(data=gbm.feature_importances_,
                        index= X_traindf.columns)

# Sort importances
gbm_importances_sorted = gbm_importances.sort_values()

# Draw a horizontal barplot of importances_sorted
gbm_importances_sorted.plot(kind='barh', color='gray')
plt.title('gbm Model Features Importances')
plt.show()
```



Stochastic Gradient Boosting (SGB)

Gradient boosting involves an exhaustive search procedure.

Each tree in the ensemble is trained to find the best split-points and the best features.

This procedure may lead to CARTs that use the same split-points and possibly the same features.

To mitigate these effects, you can use an algorithm known as stochastic gradient boosting.

• In stochastic gradient boosting, each CART is trained on a random subset of the training data.

• This subset is sampled without replacement.

• Furthermore, at the level of each node, features are sampled without replacement when choosing the best split-points. As a result, this creates further diversity in the ensemble and the net effect is adding more variance to the ensemble of trees.

Stochastic Gradient Boosting: Training

• First, instead of providing all the training instances to a tree, only a fraction of these instances are provided through sampling without replacement.

• The sampled data is then used for training a tree.

• However, not all features are considered when a split is made. Instead, only a certain randomly sampled fraction of these features are used for this purpose.

• Once a tree is trained, predictions are made, and the residual errors can be computed. These residual errors are multiplied by the learning rate eta and are fed to the next tree in the ensemble.

• This procedure is repeated sequentially until all the trees in the ensemble are trained.

• The prediction procedure for a new instance in stochastic gradient boosting is similar to that of gradient boosting.

In [76]:
```
# Stochastic Gradient Boosting Model

# Instantiate sgbr
sgbr = GradientBoostingRegressor(max_depth=2,
            subsample=0.9,
            max_features=0.02,
            n_estimators=200,
            random_state=SEED)

# Fit sgbr to the training set
sgbr.fit(X_train, y_train)
```

Out[76]:
```
GradientBoostingRegressor(max_depth=2, max_features=0.02, n_estimators=200,
                          random_state=1, subsample=0.9)
```

In [77]:
```
# Predict train set labels
y_trainpred_sgbr = sgbr.predict(X_train)

# Convert y_pred_ada from Numpy array to Data Frame
sgbrpredstrain_col = pd.DataFrame()
sgbrpredstrain_col['sgbrboosttrainpredictions'] = y_trainpred_sgbr.tolist()
sgbrpredstrain_col['recordedtrainRoRabs'] = y_train
sgbrpredstrain_col['errors'] = abs(sgbrpredstrain_col['recordedtrainRoRabs'] - sgbrpredst
sgbrpredstrain_col['percent_error'] = (sgbrpredstrain_col['errors']/sgbrpredstrain_col['re
sgbrboostpredstrain_out = pd.merge(X_traindf, sgbrpredstrain_col, left_index = True, right
print(sgbrboostpredstrain_out.head())

# Compute test set MSE
mse_train_sgbr = MSE(y_train, y_trainpred_sgbr)

# Compute test set RMSE
rmse_train_sgbr = mse_train_sgbr**(1/2)

# Print rmse_test_ada
print('Train set RMSE of sgbr: {:.3f}'.format(rmse_train_sgbr))
```

```
   year      mtmin      mtmax  spre  month  sgbrboosttrainpredictions  \
0  2019  12.770000  23.026667  13.9      4              240507.912109
```

```
1  2010   7.520000  18.560000  70.2      6             281957.564367
2  2013   8.503226  18.229032  43.6      7             248877.764457
3  2016  12.590000  23.646667  48.0      4             335375.236492
4  2021  10.151613  20.367742  70.4      5             307535.619401

   recordedtrainRoRabs         errors  percent_error
0             208665.0  31842.912109      15.260303
1             234205.0  47752.564367      20.389216
2             163722.0  85155.764457      52.012414
3             405251.0  69875.763508      17.242589
4             340779.0  33243.380599       9.755114
Train set RMSE of sgbr: 41409.760
```

In [78]:

```
mae_train_sgbr = mean_absolute_error(y_train, y_trainpred_sgbr)
print('Mean_Absolute_Error_trainsgbr: %f' % mae_train_sgbr)
print('Mean_Squared_Error_trainsgbr: %f' % mse_train_sgbr)
print('Root_Mean_Squared_Error_trainsgbr: %f' % rmse_train_sgbr)
rmspe_train_sgbr = (rmse_train_sgbr / sgbrpredstrain_col['recordedtrainRoRabs'].mean())*1(
print('Root_Mean_Squared_Percentage_Error_trainsgbr: %f' % rmspe_train_sgbr)
mape_train_sgbr = (mae_train_sgbr / sgbrpredstrain_col['recordedtrainRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_trainsgbr: %f' % mape_train_sgbr)
```

```
Mean_Absolute_Error_trainsgbr: 33150.917895
Mean_Squared_Error_trainsgbr: 1714768202.170695
Root_Mean_Squared_Error_trainsgbr: 41409.759745
Root_Mean_Squared_Percentage_Error_trainsgbr: 10.733810
Mean_Absolute_Percentage_Error_trainsgbr: 8.593038
```

In [79]:

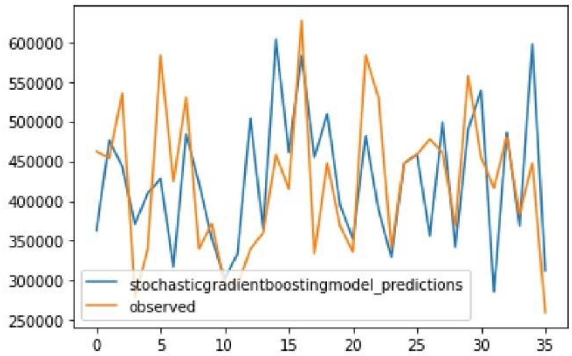
```
# Predict test set labels
y_pred_sgbr = sgbr.predict(X_test)

# Convert y_pred_sgbr from Numpy array to Data Frame
sgbrpreds_col = pd.DataFrame()
sgbrpreds_col['sgbrpredictions'] = y_pred_sgbr.tolist()
sgbrpreds_col['recordedRoRabs'] = y_test
sgbrpreds_col['errors'] = abs(sgbrpreds_col['recordedRoRabs'] - sgbrpreds_col['sgbrpredict
sgbrpreds_col['percent_error'] = (sgbrpreds_col['errors']/ sgbrpreds_col['recordedRoRabs']
sgbrpreds_out = pd.merge(X_testdf, sgbrpreds_col, left_index = True, right_index = True)
print(sgbrpreds_out.head())

# Compute test set MSE
mse_test_sgbr = MSE(y_test, y_pred_sgbr)

# Compute test set RMSE
rmse_test_sgbr = mse_test_sgbr**(1/2)

# Print rmse_test_ada
print('Test set RMSE of sgbr: {:.3f}'.format(rmse_test_sgbr))
```

```
   year      mtmin      mtmax   spre  month  sgbrpredictions  recordedRoRabs  \
0  2016  14.854839  25.551613   35.6      3    362983.441609        462244.0
1  2015  13.456667  24.470000   25.6     11    476276.971003        453707.0
2  2014  12.116129  24.980645    4.8     10    443618.024584        535226.0
3  2019  14.790323  24.900000   13.5     12    370636.709947        278396.0
4  2009   9.473333  18.580000  108.4      6    409653.131877        339402.0

         errors  percent_error
0  99260.558391      21.473628
1  22569.971003       4.974570
2  91607.975416      17.115756
3  92240.709947      33.132915
4  70251.131877      20.698503
Test set RMSE of sgbr: 83598.637
```

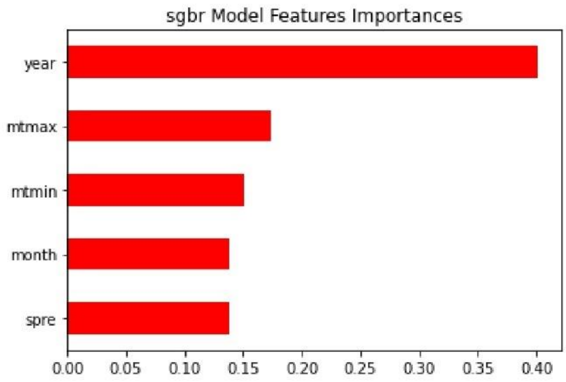
In [80]:
```
mae_test_sgbr = mean_absolute_error(y_test, y_pred_sgbr)
print('Mean_Absolute_Error_testsgbr: %f' % mae_test_sgbr)
print('Mean_Squared_Error_testsgbr: %f' % mse_test_sgbr)
print('Root_Mean_Squared_Error_testsgbr: %f' % rmse_test_sgbr)
rmspe_test_sgbr = (rmse_test_sgbr / sgbrpreds_col['recordedRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_testsgbr: %f' % rmspe_test_sgbr)
mape_test_sgbr = (mae_test_sgbr / sgbrpreds_col['recordedRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_testsgbr: %f' % mape_test_sgbr)
```

```
Mean_Absolute_Error_testsgbr: 66805.183689
Mean_Squared_Error_testsgbr: 6988732125.629733
Root_Mean_Squared_Error_testsgbr: 83598.637104
Root_Mean_Squared_Percentage_Error_testsgbr: 19.724478
Mean_Absolute_Percentage_Error_testsgbr: 15.762187
```

In [81]:
```
plt.plot(sgbrpreds_col['sgbrpredictions'], label='stochasticgradientboostingmodel_predict:
plt.plot(sgbrpreds_col['recordedRoRabs'], label='observed')
plt.legend()
plt.show()
```



In [82]:
```
# Feature importances according to the sgbr model

# Create a pd.Series of features importances
sgbr_importances = pd.Series(data=sgbr.feature_importances_,
                        index= X_traindf.columns)

# Sort importances
sgbr_importances_sorted = sgbr_importances.sort_values()

# Draw a horizontal barplot of importances_sorted
sgbr_importances_sorted.plot(kind='barh', color='red')
plt.title('sgbr Model Features Importances')
plt.show()
```



Model Tuning: Hyperparameter Tuning

The hyperparameters of a machine learning model are parameters that are not learned from data. They should be set prior to fitting the model to the training set.

In [83]:

```
# Instantiate a DecisonTreeRegressor:

dt2 = DecisionTreeRegressor(random_state = SEED)
print(dt2.get_params())
```

```
{'ccp_alpha': 0.0, 'criterion': 'mse', 'max_depth': None, 'max_features': None, 'max_leaf_
nodes': None, 'min_impurity_decrease': 0.0, 'min_impurity_split': None, 'min_samples_lea
f': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'random_state': 1, 'splitt
er': 'best'}
```

In [84]:

```
# Import GridSearchCV
from sklearn.model_selection import GridSearchCV

# Import roc_auc_score from sklearn.metrics
from sklearn.metrics import roc_auc_score

# Define params_dt
params_dt = {'max_depth': [2, 3, 4, 5, 6],
             'min_samples_leaf': [0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18],
             'max_features': [0.02, 0.04, 0.06, 0.08]}

# Instantiate grid_dt
grid_dt = GridSearchCV(estimator=dt2,
                       param_grid=params_dt,
                       scoring='neg_mean_squared_error',
                       cv=10,
                       n_jobs=-1)

# Fit 'grid_dt' to the training data
grid_dt.fit(X_train, y_train)
```

Out[84]:

```
GridSearchCV(cv=10, estimator=DecisionTreeRegressor(random_state=1), n_jobs=-1,
             param_grid={'max_depth': [2, 3, 4, 5, 6],
                         'max_features': [0.02, 0.04, 0.06, 0.08],
                         'min_samples_leaf': [0.04, 0.06, 0.08, 0.1, 0.12, 0.14,
                                              0.16, 0.18]},
             scoring='neg_mean_squared_error')
```

In [85]:

```
# Extract best hyperparameters from 'grid_dt'
best_hyperparams = grid_dt.best_params_
print('Best hyperparameters:\n', best_hyperparams)

# Extract best CV score from 'grid_dt'
best_CV_score = grid_dt.best_score_
print('Best CV accuracy'.format(best_CV_score))

# Extract best model from 'grid_dt'
best_model = grid_dt.best_estimator_

# Evaluate test set accuracy
test_acc = best_model.score(X_test, y_test)
print('Test set accuracy of best model: {:.3}'.format(test_acc))

# Predict test set labels
y_pred_best_model = best_model.predict(X_test)

# Compute rmse_test
rmse_test_best_model = MSE(y_test, y_pred_best_model)**(1/2)

# Print rmse_test
print('Test RMSE of best model: {:.3f}'.format(rmse_test_best_model))
```

```
Best hyperparameters:
 {'max_depth': 5, 'max_features': 0.02, 'min_samples_leaf': 0.04}
Best CV accuracy
Test set accuracy of best model: -0.119
Test RMSE of best model: 96534.523
```

In [86]:

```
# Predict train set labels
y_trainpred_gridsearchbest = best_model.predict(X_train)

# Convert y_pred_ada from Numpy array to Data Frame
gridsearchbestpredstrain_col = pd.DataFrame()
gridsearchbestpredstrain_col['gridsearchbestboosttrainpredictions'] = y_trainpred_gridsear
gridsearchbestpredstrain_col['recordedtrainRoRabs'] = y_train
gridsearchbestpredstrain_col['errors'] = abs(gridsearchbestpredstrain_col['recordedtrainRo
gridsearchbestpredstrain_col['percent_error'] = (gridsearchbestpredstrain_col['errors'] /
gridsearchbestboostpredstrain_out = pd.merge(X_traindf, gridsearchbestpredstrain_col, left
print(gridsearchbestboostpredstrain_out.head())

# Compute test set MSE
mse_train_gridsearchbest = MSE(y_train, y_trainpred_gridsearchbest)

# Compute test set RMSE
rmse_train_gridsearchbest = mse_train_gridsearchbest**(1/2)

# Print rmse_test_ada
print('Train set RMSE of gridsearchbest: {:.3f}'.format(rmse_train_gridsearchbest))
```

```
   year      mtmin      mtmax  spre  month  \
0  2019  12.770000  23.026667  13.9      4
1  2010   7.520000  18.560000  70.2      6
2  2013   8.503226  18.229032  43.6      7
3  2016  12.590000  23.646667  48.0      4
4  2021  10.151613  20.367742  70.4      5

   gridsearchbestboosttrainpredictions  recordedtrainRoRabs         errors  \
0                        240485.007875             208665.0   31820.007875
1                        317641.207735             234205.0   83436.207735
2                        416826.055556             163722.0  253104.055556
3                        240485.007875             405251.0  164765.992125
4                        317641.207735             340779.0   23137.792265
```

```
   percent_error
0      15.249327
1      35.625289
2     154.593797
3      40.657763
4       6.789677
Train set RMSE of gridsearchbest: 91103.178
```

In [87]:

```
mae_train_gridsearchbest = mean_absolute_error(y_train, y_trainpred_gridsearchbest)
print('Mean_Absolute_Error_traingridsearchbest: %f' % mae_train_gridsearchbest)
print('Mean_Squared_Error_traingridsearchbest: %f' % mse_train_gridsearchbest)
print('Root_Mean_Squared_Error_traingridsearchbest: %f' % rmse_train_gridsearchbest)
rmspe_train_gridsearchbest = (rmse_train_gridsearchbest / gridsearchbestpredstrain_col['re
print('Root_Mean_Squared_Percentage_Error_traingridsearchbest: %f' % rmspe_train_gridsearc
mape_train_gridsearchbest = (mae_train_gridsearchbest / gridsearchbestpredstrain_col['reco
print('Mean_Absolute_Percentage_Error_traingridsearchbest: %f' % mape_train_gridsearchbest
```

```
Mean_Absolute_Error_traingridsearchbest: 71647.355810
Mean_Squared_Error_traingridsearchbest: 8299789103.832258
Root_Mean_Squared_Error_traingridsearchbest: 91103.178341
Root_Mean_Squared_Percentage_Error_traingridsearchbest: 23.614825
Mean_Absolute_Percentage_Error_traingridsearchbest: 18.571687
```

In [88]:

```
# Predict test set labels
y_pred_gridsearchbest = best_model.predict(X_test)

# Convert y_pred_gridsearchbest from Numpy array to Data Frame
gridsearchbmpreds_col = pd.DataFrame()
gridsearchbmpreds_col['gridsearchbmpredictions'] = y_pred_gridsearchbest.tolist()
gridsearchbmpreds_col['recordedRoRabs'] = y_test
gridsearchbmpreds_col['errors'] = abs(gridsearchbmpreds_col['recordedRoRabs'] - gridsearch
gridsearchbmpreds_col['percent_error'] = (gridsearchbmpreds_col['errors']/ gridsearchbmpre
gridsearchbmpreds_out = pd.merge(X_testdf, gridsearchbmpreds_col, left_index = True, right
print(gridsearchbmpreds_out.head())

# Compute test set MSE
mse_test_gridsearchbest = MSE(y_test, y_pred_gridsearchbest)

# Compute test set RMSE
rmse_test_gridsearchbest = mse_test_gridsearchbest**(1/2)

# Print rmse_test_ada
print('Test set RMSE of gridsearchbestmodel: {:.3f}'.format(rmse_test_gridsearchbest))
```

```
   year      mtmin      mtmax   spre  month  gridsearchbmpredictions  \
0  2016  14.854839  25.551613   35.6      3            349000.823909
1  2015  13.456667  24.470000   25.6     11            439853.111111
2  2014  12.116129  24.980645    4.8     10            383400.666667
3  2019  14.790323  24.900000   13.5     12            439853.111111
4  2009   9.473333  18.580000  108.4      6            267360.500000

   recordedRoRabs         errors  percent_error
0        462244.0  113243.176091      24.498571
1        453707.0   13853.888889       3.053488
2        535226.0  151825.333333      28.366584
3        278396.0  161457.111111      57.995485
4        339402.0   72041.500000      21.226009
Test set RMSE of gridsearchbestmodel: 96534.523
```

In [89]:

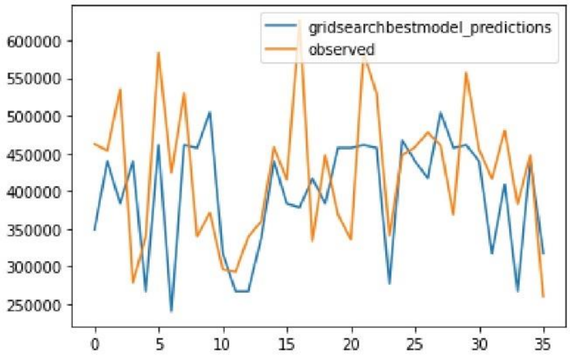
```
mae_test_gridsearchbest = mean_absolute_error(y_test, y_pred_gridsearchbest)
print('Mean_Absolute_Error_testgridsearchbest: %f' % mae_test_gridsearchbest)
print('Mean_Squared_Error_testgridsearchbest: %f' % mse_test_gridsearchbest)
print('Root_Mean_Squared_Error_testgridsearchbest: %f' % rmse_test_gridsearchbest)
rmspe_test_gridsearchbest = (rmse_test_gridsearchbest / gridsearchbmpreds_col['recordedRoR
```

```
print('Root_Mean_Squared_Percentage_Error_testgridsearchbest: %f' % rmspe_test_gridsearchk
mape_test_gridsearchbest = (mae_test_gridsearchbest / gridsearchbmpreds_col['recordedRoRak
print('Mean_Absolute_Percentage_Error_testgridsearchbest: %f' % mape_test_gridsearchbest)
```

```
Mean_Absolute_Error_testgridsearchbest: 80217.796197
Mean_Squared_Error_testgridsearchbest: 9318914094.032612
Root_Mean_Squared_Error_testgridsearchbest: 96534.522809
Root_Mean_Squared_Percentage_Error_testgridsearchbest: 22.776604
Mean_Absolute_Percentage_Error_testgridsearchbest: 18.926793
```

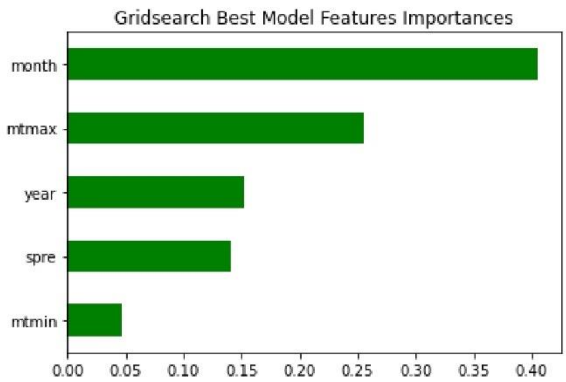
In [90]:

```
plt.plot(gridsearchbmpreds_col['gridsearchbmpredictions'], label='gridsearchbestmodel_prec
plt.plot(gridsearchbmpreds_col['recordedRoRabs'], label='observed')
plt.legend()
plt.show()
```



In [91]:

```
# Feature importances according to the grid search best random forests model

# Create a pd.Series of features importances
gridsearch_bestmodel_importances = pd.Series(data=best_model.feature_importances_,
                           index= X_traindf.columns)

# Sort importances
gridsearch_bestmodel_importances_sorted = gridsearch_bestmodel_importances.sort_values()

# Draw a horizontal barplot of importances_sorted
gridsearch_bestmodel_importances_sorted.plot(kind='barh', color='green')
plt.title('Gridsearch Best Model Features Importances')
plt.show()
```



In [92]:
```
# Tuning a Random Forest Regressor
from sklearn.ensemble import RandomForestRegressor

# Set seed for reproducibility
SEED = 1

# Instantiate a random forests regressor 'rf'
rf = RandomForestRegressor(random_state = SEED)

rf.get_params()
```

Out[92]:
```
{'bootstrap': True,
 'ccp_alpha': 0.0,
 'criterion': 'mse',
 'max_depth': None,
 'max_features': 'auto',
 'max_leaf_nodes': None,
 'max_samples': None,
 'min_impurity_decrease': 0.0,
 'min_impurity_split': None,
 'min_samples_leaf': 1,
 'min_samples_split': 2,
 'min_weight_fraction_leaf': 0.0,
 'n_estimators': 100,
 'n_jobs': None,
 'oob_score': False,
 'random_state': 1,
 'verbose': 0,
 'warm_start': False}
```

In [93]:
```
# Define the dictionary 'params_rf'
params_rf = {'n_estimators': [100, 350, 500],
             'max_features': ['log2', 'auto', 'sqrt'],
             'min_samples_leaf': [2, 10, 30]
                         }
```

In [94]:
```
# Instantiate grid_rf
grid_rf = GridSearchCV(estimator=rf,
                       param_grid=params_rf,
                       scoring='neg_mean_squared_error',
                       cv=10,
                       verbose=1,
                       n_jobs=-1)
```

In [95]:
```
# Fit 'grid_rf' to the training set
grid_rf.fit(X_train, y_train)
```

```
Fitting 10 folds for each of 27 candidates, totalling 270 fits
```

Out[95]:
```
GridSearchCV(cv=10, estimator=RandomForestRegressor(random_state=1), n_jobs=-1,
             param_grid={'max_features': ['log2', 'auto', 'sqrt'],
                         'min_samples_leaf': [2, 10, 30],
                         'n_estimators': [100, 350, 500]},
             scoring='neg_mean_squared_error', verbose=1)
```

In [96]:
```
# Extract best hyperparameters from 'grid_rf'
best_rf_hyperparams = grid_rf.best_params_
print('Best hyperparameters:\n', best_rf_hyperparams)
```

```
Best hyperparameters:
 {'max_features': 'auto', 'min_samples_leaf': 2, 'n_estimators': 500}
```

In [97]:
```
# Extract the best estimator
best_rf_model = grid_rf.best_estimator_
```

In [98]:
```
# Predict train set labels
y_trainpred_best_rf = best_rf_model.predict(X_train)

# Convert y_pred_ada from Numpy array to Data Frame
gridsearchrfpredstrain_col = pd.DataFrame()
gridsearchrfpredstrain_col['gridsearchrfboosttrainpredictions'] = y_trainpred_best_rf.tol
gridsearchrfpredstrain_col['recordedtrainRoRabs'] = y_train
gridsearchrfpredstrain_col['errors'] = abs(gridsearchrfpredstrain_col['recordedtrainRoRabs
gridsearchrfpredstrain_col['percent_error'] = (gridsearchrfpredstrain_col['errors'] / gric
gridsearchrfboostpredstrain_out = pd.merge(X_traindf, gridsearchrfpredstrain_col, left_inc
print(gridsearchrfboostpredstrain_out.head())

# Compute test set MSE
mse_train_gridsearchrf = MSE(y_train, y_trainpred_best_rf)

# Compute test set RMSE
rmse_train_gridsearchrf = mse_train_gridsearchrf**(1/2)

# Print rmse_test_ada
print('Train set RMSE of gridsearchrf: {:.3f}'.format(rmse_train_gridsearchrf))
```

```
   year      mtmin      mtmax  spre  month  gridsearchrfboosttrainpredictions  \
0  2019  12.770000  23.026667  13.9      4                      243543.441914
1  2010   7.520000  18.560000  70.2      6                      295116.722687
2  2013   8.503226  18.229032  43.6      7                      230292.706782
3  2016  12.590000  23.646667  48.0      4                      353519.529489
4  2021  10.151613  20.367742  70.4      5                      312790.223979

   recordedtrainRoRabs        errors  percent_error
0             208665.0  34878.441914      16.715042
1             234205.0  60911.722687      26.007866
2             163722.0  66570.706782      40.660819
3             405251.0  51731.470511      12.765291
4             340779.0  27988.776021       8.213175
Train set RMSE of gridsearchrf: 41055.473
```

In [99]:
```
mae_train_gridsearchrf = mean_absolute_error(y_train, y_trainpred_best_rf)
print('Mean_Absolute_Error_traingridsearchrf: %f' % mae_train_gridsearchrf)
print('Mean_Squared_Error_traingridsearchrf: %f' % mse_train_gridsearchrf)
print('Root_Mean_Squared_Error_traingridsearchrf: %f' % rmse_train_gridsearchrf)
rmspe_train_gridsearchrf = (rmse_train_gridsearchrf / gridsearchrfpredstrain_col['recordec
print('Root_Mean_Squared_Percentage_Error_traingridsearchrf: %f' % rmspe_train_gridsearch
mape_train_gridsearchrf = (mae_train_gridsearchrf / gridsearchrfpredstrain_col['recordedt
print('Mean_Absolute_Percentage_Error_traingridsearchrf: %f' % mape_train_gridsearchrf)
```

```
Mean_Absolute_Error_traingridsearchrf: 30997.405483
Mean_Squared_Error_traingridsearchrf: 1685551881.785399
Root_Mean_Squared_Error_traingridsearchrf: 41055.473226
Root_Mean_Squared_Percentage_Error_traingridsearchrf: 10.641976
Mean_Absolute_Percentage_Error_traingridsearchrf: 8.034827
```

In [100…
```
# Predict test set labels
y_pred_best_rf = best_rf_model.predict(X_test)

# Evaluate test set accuracy
test_acc = best_rf_model.score(X_test, y_test)
print('Test set accuracy of best rf model: {:.3}'.format(test_acc))

# Compute mse_test
mse_test_best_rf = MSE(y_test, y_pred_best_rf)
```

```
# Compute rmse_test
rmse_test_best_rf = MSE(y_test, y_pred_best_rf)**(1/2)

# Print rmse_test
print('Test RMSE of best rf model: {:.3f}'.format(rmse_test_best_rf))
```

```
Test set accuracy of best rf model: 0.42
Test RMSE of best rf model: 69485.978
```

In [101…

```
# Convert y_pred_gridsearchbest from Numpy array to Data Frame
gridsearchrfpreds_col = pd.DataFrame()
gridsearchrfpreds_col['gridsearchrfpredictions'] = y_pred_best_rf.tolist()
gridsearchrfpreds_col['recordedRoRabs'] = y_test
gridsearchrfpreds_col['errors'] = abs(gridsearchrfpreds_col['recordedRoRabs'] - gridsearch
gridsearchrfpreds_col['percent_error'] = (gridsearchrfpreds_col['errors']/ gridsearchrfpre
gridsearchrfpreds_out = pd.merge(X_testdf, gridsearchrfpreds_col, left_index = True, right
print(gridsearchrfpreds_out.head())

# Compute rmse_test
rmse_test_best_rf = MSE(y_test, y_pred_best_rf)**(1/2)

# Print rmse_test
print('Test RMSE of best rf model: {:.3f}'.format(rmse_test_best_rf))
```

```
   year     mtmin      mtmax   spre  month  gridsearchrfpredictions  \
0  2016  14.854839  25.551613   35.6      3            375820.239796
1  2015  13.456667  24.470000   25.6     11            453389.021709
2  2014  12.116129  24.980645    4.8     10            408218.071687
3  2019  14.790323  24.900000   13.5     12            354421.477954
4  2009   9.473333  18.580000  108.4      6            433489.146077

   recordedRoRabs         errors  percent_error
0        462244.0   86423.760204      18.696567
1        453707.0     317.978291       0.070085
2        535226.0  127007.928313      23.729776
3        278396.0   76025.477954      27.308395
4        339402.0   94087.146077      27.721447
Test RMSE of best rf model: 69485.978
```

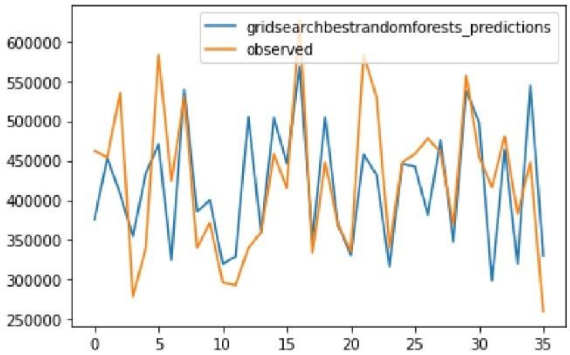
In [102…

```
mae_test_gridsearchrf = mean_absolute_error(y_test, y_pred_best_rf)
print('Mean_Absolute_Error_testgridsearchrf: %f' % mae_test_gridsearchrf)

print('Mean_Squared_Error_testgridsearchrf: %f' % mse_test_best_rf)
print('Root_Mean_Squared_Error_testgridsearchrf: %f' % rmse_test_best_rf)
rmspe_test_gridsearchrf = (rmse_test_best_rf / gridsearchrfpreds_col['recordedRoRabs'].mea
print('Root_Mean_Squared_Percentage_Error_testgridsearchrf: %f' % rmspe_test_gridsearchrf)
mape_test_gridsearchrf = (mae_test_gridsearchrf / gridsearchrfpreds_col['recordedRoRabs'].
print('Mean_Absolute_Percentage_Error_testgridsearchrf: %f' % mape_test_gridsearchrf)
```

```
Mean_Absolute_Error_testgridsearchrf: 54022.276758
Mean_Squared_Error_testgridsearchrf: 4828301083.407607
Root_Mean_Squared_Error_testgridsearchrf: 69485.977603
Root_Mean_Squared_Percentage_Error_testgridsearchrf: 16.394700
Mean_Absolute_Percentage_Error_testgridsearchrf: 12.746155
```

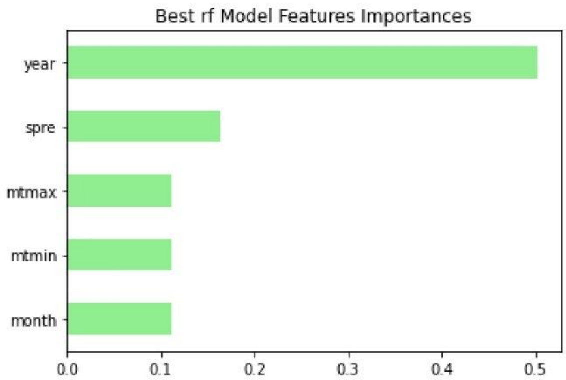
In [103…

```
plt.plot(gridsearchrfpreds_col['gridsearchrfpredictions'], label='gridsearchbestrandomfore
plt.plot(gridsearchrfpreds_col['recordedRoRabs'], label='observed')
plt.legend()
plt.show()
```



In [104…

```
# Feature importances according to the grid search best random forests model

# Create a pd.Series of features importances
rf_importances = pd.Series(data=best_rf_model.feature_importances_,
                        index= X_traindf.columns)

# Sort importances
rf_importances_sorted = rf_importances.sort_values()

# Draw a horizontal barplot of importances_sorted
rf_importances_sorted.plot(kind='barh', color='lightgreen')
plt.title('Best rf Model Features Importances')
plt.show()
```



NEURAL NETWORKS

In a nutshell, a neural network is a machine learning algorithm that is fetched with training data through its input layer to then predict a value at its output layer.

Each connection from one neuron to another has an associated weight, w.

Each neuron, except the input layer which just holds the input value, also has an extra weight and we call this the bias weight, b.

During feed-forward our input gets

• transformed by weight multiplications and additions at each layer,

• the output of each neuron can also get transformed by the application of what we call an activation function.

Learning in neural networks consists of tuning the weights or parameters to give the desired output. One way of achieving this is by using the famous gradient descent algorithm and applying weight updates incrementally via a process known as back-propagation.

In [121...
```
from tensorflow.keras.models import Sequential
```

In [122...
```
from tensorflow.keras.layers import Input, Dense
```

In [123...
```
train_X = X_traindf[['year', 'mtmin', 'mtmax', 'spre', 'month']]
```

In [124...
```
train_y = y_traindf['RoRabs']
```

In [125...
```
test_X = X_testdf[['year', 'mtmin', 'mtmax','spre', 'month']]
```

In [126...
```
test_y = y_testdf['RoRabs']
```

In [127...
```
# Install a pip package in the current Jupyter kernel
import sys
!{sys.executable} -m pip install h5py --upgrade pip
```

```
'C:\Users\Rejoice' is not recognized as an internal or external command,
operable program or batch file.
```

In [128...
```
model_seq = Sequential()
```

In [129...
```
model_seq.add(Dense(100, input_shape=(5,), activation='relu'))
```

In [130...
```
model_seq.add(Dense(100,activation='relu'))
model_seq.add(Dense(100,activation='relu'))
```

In [131...
```
model_seq.add(Dense(1,))
```

In [132...
```
model_seq.compile(optimizer = 'adam', loss = 'mae')
```

In [133...
```
from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint
```

In [134...
```
# Define early_stopping_monitor
early_stopping_monitor = EarlyStopping(patience=200)
```

In [135...
```
mc = ModelCheckpoint('best_model.h5', monitor='val_loss', mode='min', save_best_only=True)
```

In [136...

```
history = model_seq.fit(train_X, train_y,
                        epochs = 5000,
                        validation_split=0.10,
                        callbacks=[early_stopping_monitor, mc],
                       verbose=False)
```

In [137…

```
from keras.models import load_model
```

In [138…

```
NN_2 = load_model('best_model.h5')
```

In [139…

```
train_mae_nn_2 = NN_2.evaluate(train_X, train_y, verbose=False)
test_mae_nn_2 = NN_2.evaluate(test_X, test_y, verbose=False)
print('Train: %.3f, Test: %.3f' % (train_mae_nn_2, test_mae_nn_2))
```

```
Train: 95829.797, Test: 82297.594
```

In [140…

```
print("Final loss value:",model_seq.evaluate(test_X, test_y))
```

```
2/2 [==============================] - 0s 86ms/step - loss: 82774.2031
Final loss value: 82774.203125
```

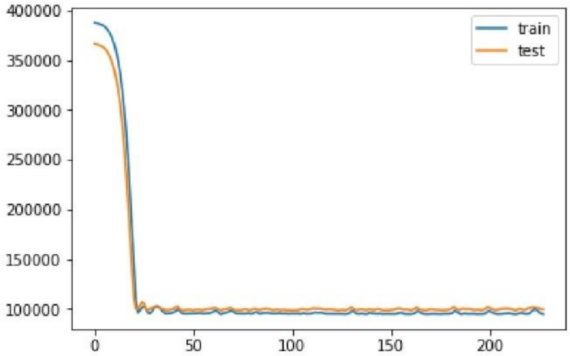
In [141…

```
train_mae = model_seq.evaluate(train_X, train_y, verbose=False)
test_mae = model_seq.evaluate(test_X, test_y, verbose=False)
print('Train: %.3f, Test: %.3f' % (train_mae, test_mae))
```

```
Train: 95347.195, Test: 82774.203
```

In [142…

```
# plot training history
plt.plot(history.history['loss'], label='train')
plt.plot(history.history['val_loss'], label='test')
plt.legend()
plt.show()
```



In [143…

```
y_trainpred_model_seq = model_seq.predict(train_X)
nntrainpreds_col = pd.DataFrame(y_trainpred_model_seq, columns = ['nntrainpredictions'])
print(nntrainpreds_col.head())
```

```
5/5 [==============================] - 1s 20ms/step
   nntrainpredictions
0        377202.06250
1        371042.96875
2        373237.62500
```

```
3        374986.31250
4        373782.56250
```

In [144…

```
nntrainpreds_col['recordedRoRabs'] = y_traindf['RoRabs']
nntrainpreds_col['errors'] = abs(nntrainpreds_col['recordedRoRabs'] - nntrainpreds_col['nn
nntrainpreds_col['percent_error'] = (nntrainpreds_col['errors'] / nntrainpreds_col['record
nntrainpreds_out = pd.merge(train_X, nntrainpreds_col, left_index = True, right_index = Tr
print(nntrainpreds_out.head())

# Compute mse_test
mse_train_nn = MSE(train_y, y_trainpred_model_seq)

# Compute rmse_test
rmse_train_nn = MSE(train_y, y_trainpred_model_seq)**(1/2)

# Print rmse_test
print('Train RMSE of best nn model: {:.3f}'.format(rmse_train_nn))
```

```
   year      mtmin      mtmax  spre  month  nntrainpredictions  \
0  2019  12.770000  23.026667  13.9      4        377202.06250
1  2010   7.520000  18.560000  70.2      6        371042.96875
2  2013   8.503226  18.229032  43.6      7        373237.62500
3  2016  12.590000  23.646667  48.0      4        374986.31250
4  2021  10.151613  20.367742  70.4      5        373782.56250

   recordedRoRabs         errors  percent_error
0        208665.0  168537.06250      80.769205
1        234205.0  136837.96875      58.426579
2        163722.0  209515.62500     127.970355
3        405251.0   30264.68750       7.468134
4        340779.0   33003.56250       9.684741
Train RMSE of best nn model: 118069.133
```

In [145…

```
mae_train_nn = mean_absolute_error(train_y, y_trainpred_model_seq)
print('Mean_Absolute_Error_trainnn: %f' % mae_train_nn)
print('Mean_Squared_Error_trainnn: %f' % mse_train_nn)
print('Root_Mean_Squared_Error_trainnn: %f' % rmse_train_nn)
rmspe_train_nn = (rmse_train_nn / nntrainpreds_col['recordedRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_trainnn: %f' % rmspe_train_nn)
mape_train_nn = (mae_train_nn / nntrainpreds_col['recordedRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_trainnn: %f' % mape_train_nn)
```

```
Mean_Absolute_Error_trainnn: 95347.190530
Mean_Squared_Error_trainnn: 13940320273.726135
Root_Mean_Squared_Error_trainnn: 118069.133450
Root_Mean_Squared_Percentage_Error_trainnn: 30.604661
Mean_Absolute_Percentage_Error_trainnn: 24.714914
```

In [146…

```
y_testpred_model_seq = model_seq.predict(test_X)
nntestpreds_col = pd.DataFrame(y_testpred_model_seq, columns = ['nntestpredictions'])
print(nntestpreds_col.head())
```

```
2/2 [==============================] - 1s 10ms/step
   nntestpredictions
0       376298.28125
1       377298.96875
2       377860.96875
3       379150.00000
4       369303.00000
```

In [147…

```
nntestpreds_col['recordedRoRabs'] = y_testdf['RoRabs']
nntestpreds_col['errors'] = abs(nntestpreds_col['recordedRoRabs'] - nntestpreds_col['nntes
nntestpreds_col['percent_error'] = (nntestpreds_col['errors'] / nntestpreds_col['recordedR
```

```
nntestpreds_out = pd.merge(test_X, nntestpreds_col, left_index = True, right_index = True)
print(nntestpreds_out.head())

# Compute rmse_test
mse_test_nn = MSE(test_y, y_testpred_model_seq)

# Compute rmse_test
rmse_test_nn = MSE(test_y, y_testpred_model_seq)**(1/2)

# Print rmse_test
print('Test RMSE of best nn model: {:.3f}'.format(rmse_test_nn))
```

```
   year      mtmin      mtmax   spre  month  nntestpredictions  \
0  2016  14.854839  25.551613   35.6      3       376298.28125
1  2015  13.456667  24.470000   25.6     11       377298.96875
2  2014  12.116129  24.980645    4.8     10       377860.96875
3  2019  14.790323  24.900000   13.5     12       379150.00000
4  2009   9.473333  18.580000  108.4      6       369303.00000

   recordedRoRabs         errors  percent_error
0        462244.0   85945.71875      18.593150
1        453707.0   76408.03125      16.840831
2        535226.0  157365.03125      29.401604
3        278396.0  100754.00000      36.190894
4        339402.0   29901.00000       8.809907
Test RMSE of best nn model: 102637.276
```

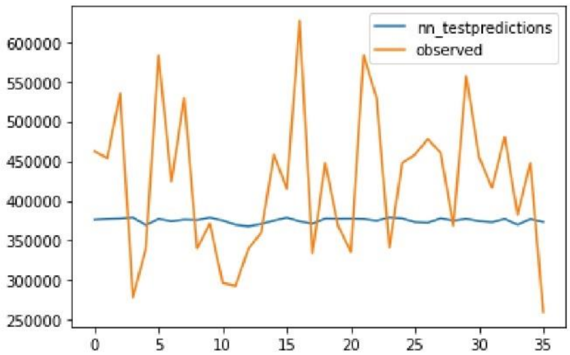
In [148…

```
mae_test_nn = mean_absolute_error(test_y, y_testpred_model_seq)
print('Mean_Absolute_Error_testnn: %f' % mae_test_nn)
print('Mean_Squared_Error_testnn: %f' % mse_test_nn)
print('Root_Mean_Squared_Error_testnn: %f' % rmse_test_nn)
rmspe_test_nn = (rmse_test_nn / nntestpreds_col['recordedRoRabs'].mean())*100
print('Root_Mean_Squared_Percentage_Error_testnn: %f' % rmspe_test_nn)
mape_test_nn = (mae_test_nn / nntestpreds_col['recordedRoRabs'].mean())*100
print('Mean_Absolute_Percentage_Error_testnn: %f' % mape_test_nn)
```

```
Mean_Absolute_Error_testnn: 82774.205729
Mean_Squared_Error_testnn: 10534410386.037407
Root_Mean_Squared_Error_testnn: 102637.275812
Root_Mean_Squared_Percentage_Error_testnn: 24.216503
Mean_Absolute_Percentage_Error_testnn: 19.529959
```

In [149…

```
plt.plot(nntestpreds_col['nntestpredictions'], label='nn_testpredictions')
plt.plot(nntestpreds_col['recordedRoRabs'], label='observed')
plt.legend()
plt.show()
```



THE BEST MODELS:

On the basis of the RMSE the best model is the best_rf_model obtained by tuning the hyperparameters of the random forests model with cross validation.

This is followed closely by the sgbr model ada.

With the Adaboost coming at third place

We will save these two models and recommend them for production! We will load them latter for scoring new data to get the predictions for R0Rabs.

In [158…
```
import joblib
```

In [159…
```
import pickle
with open('best_rf_model', 'wb') as file:
    pickle.dump(best_rf_model, file)
```

In [160…
```
import pickle
with open('sgbr', 'wb') as file:
    pickle.dump(sgbr, file)
```

In [161…
```
import pickle
with open('ada', 'wb') as file:
    pickle.dump(ada, file)
```

In [162…
```
import pickle
with open('model_seq', 'wb') as file:
    pickle.dump(model_seq, file)
```

```
Keras weights file (<HDF5 file "variables.h5" (mode r+)>) saving:
...layers\dense
......vars
.........0
.........1
...layers\dense_1
......vars
.........0
.........1
...layers\dense_2
......vars
.........0
.........1
...layers\dense_3
......vars
.........0
.........1
...metrics\mean
......vars
.........0
.........1
...optimizer
......vars
.........0
.........1
.........10
```

```
.........11
.........12
.........13
.........14
.........15
.........16
.........2
.........3
.........4
.........5
.........6
.........7
.........8
.........9
...vars
Keras model archive saving:
File Name                                             Modified             Size
config.json                                    2023-02-10 19:51:27         2223
metadata.json                                  2023-02-10 19:51:27           64
variables.h5                                   2023-02-10 19:51:27       275360
```

FUTURE PREDICTIONS OF RoRabs

To make predictions using machine learning models, you need to feed the model with exactly the features you used during training in the same format. In our case we will need to give the model the mtmin, mtmax, mtave, spre and month for a new data set in order to get the prediction for RoRabs for the prevailing conditions.

This is the difference between time series model predictions that can be extrapolated into the future on the basis of the prevailing trend.

PYCARET

In [61]:

```
rrabs.head()
```

Out[61]:

	RoRabs	mtmin	mtmax	mtave	spre	year	month
0	404000.0	8.658065	16.922581	12.790323	71.4	2006	7
1	455000.0	7.932258	17.738710	12.835484	56.2	2006	8
2	697000.0	10.323333	20.873333	15.598333	20.0	2006	9
3	529664.0	11.274194	22.380645	16.827419	37.2	2006	10
4	458241.0	13.906667	24.553333	19.230000	37.7	2006	11

In [62]:

```
# create a sequence of numbers
rrabs['Series'] = np.arange(1,len(rrabs)+1)
```

In [63]:

```
# drop unnecessary columns and re-arrange
rrabs_pc = rrabs[['Series', 'year', 'month', 'RoRabs']]
```

In [64]:

```
rrabs_pc.head()
```

Out[64]:

	Series	year	month	RoRabs

	Series	year	month	RoRabs
0	1	2006	7	404000.0
1	2	2006	8	455000.0
2	3	2006	9	697000.0
3	4	2006	10	529664.0
4	5	2006	11	458241.0

Sample rows after extracting features

Something to note here is that the train-test-split for time-series data is special. Because you cannot change the order of the table, you have to ensure that you don't sample randomly as you want your test data to contain points that are in the future from the points in the train data (time always moves forward).

In [67]:
```
# split data into train-test set
train_pc = rrabs_pc[rrabs_pc['year'] < 2016]
test_pc = rrabs_pc[rrabs_pc['year'] >= 2016]

# check shape
train_pc.shape, test_pc.shape
```

Out[67]:
```
((114, 4), (66, 4))
```

We will use PyCaret; an open-source, low-code machine learning library in Python that automates machine learning workflows. To use PyCaret, you have to install it using pip.

In [68]:
```
# install pycaret
pip install pycaret
```

```
  Cell In [68], line 2
    pip install pycaret
        ^
SyntaxError: invalid syntax
```

In [ ]:

Appendix F:

Publications

1. Malisa, R., Schwella, E. & Theletsane, K.I. 2018. Urban waste water governance in South Africa: A case study of Stellenbosch. *International Journal of Environmental and Ecological Engineering*, 12(10):617-631.

2. Malisa-Van der Walt, R. & Taigbenu, A. 2022. Policy, laws, and guidelines of wastewater reuse for agricultural purposes in developing countries. In M. Nasr & A.M. Negm (Eds.). *Cost-Efficient Wastewater Treatment Technologies*. Cham: Springer International Publishing. pp. 1-24.

3. Malisa-Van der Walt, R., Babi, T.F. & Taigbenu, A. Short- and intermediate time horizon urban water demand forecasting for Stellenbosch Municipality. [To be submitted for publication.]