THE APPLICATION OF MACHINE LEARNING FOR GROUNDWATER LEVEL PREDICTION IN THE STEENKOPPIES COMPARTMENT OF THE GAUTENG AND NORTH WEST DOLOMITE AQUIFER, SOUTH AFRICA

Kirsty Gibson

Submitted in fulfilment of the requirements for the degree

Magister Scientiae in Geohydrology

in the

Faculty of Natural and Agricultural Sciences

(Institute for Groundwater Studies)

at the

University of the Free State

Supervisor: Professor Kai Witthüser

November 2020

DECLARATION

I, Kirsty Gibson, hereby declare that the dissertation hereby submitted by me to the Institute for Groundwater Studies in the Faculty of Natural and Agricultural Sciences at the University of the Free State, in fulfilment of the degree of Magister Scientiae, is my own independent work. It has not previously been submitted by me to any other institution of higher education. In addition, I declare that all sources cited have been acknowledged by means of a list of references.

I furthermore cede copyright of the dissertation and its contents in favour of the University of the Free State.

Kirsty Gibson

November 2020

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following people for their guidance during this thesis:

- Kai Witthüser for the opportunity to be a part of the Big Data Analytics and Transboundary Water Collaboration and for his vast knowledge in mathematics and groundwater modelling.
- Helen Seyler for the opportunity to be a part of the Big Data Analytics and Transboundary Water Collaboration and for all her sensible advice, patience and support throughout the project. Thank you for also showing me that a run in the mountains helps clear the head.
- All funders and members of the Big Data Analytics and Transboundary Water Collaboration
- Dr. Stefan Broda and his colleague Maximilian Nölscher, of the Federal Institute for Geosciences and Natural Resources (BGR) in Berlin, for their technical help with the code, when the results were too good to be true.
- Prof Ritesh Ajoodha, a lecturer from the University of the Witwatersrand who was a part of the the Big Data Analytics and Transboundary Water Collaboration and provided insight on machine learning modelling.

Most importantly thank you to my family and Craig Rennie for their incredible support, especially at the times when I was under stress. Without them, completing my masters thesis would not have been possible.

ABSTRACT

Groundwater in the Steenkoppies compartment of the Gauteng and North West dolomite aquifer is extensively used for agriculture practices that can potentially lead to groundwater storage depletion, threatening groundwater sustainability in the compartment. Groundwater models are needed to describe the complex, interdependent relationships occurring in a groundwater system. Groundwater levels represent the response of an aquifer to changes in storage, recharge, discharge and hydrological stresses. Groundwater levels in an aquifer are, therefore, useful to identify limits and unacceptable impacts on an aquifer and to use this information to implement sustainable groundwater management decisions. Conventionally, numerical techniques are used for groundwater modelling. The use of machine learning techniques for groundwater modelling is relatively new in South Africa. Unlike numerical models, machine learning models are data-driven and learn the behaviour of the aquifer system from measured values without needing an understanding of the internal structure and physical processes of an aquifer. In this study, Neural Network Autoregression (NNAR) was applied to obtain groundwater level predictions in the Steenkoppies compartment of the Gauteng and North West Dolomite Aquifer in South Africa. Multiple variables (rainfall, temperature, groundwater usage and spring discharge from the Maloney's Eye spring) were chosen as input parameters to facilitate groundwater level predictions. The importance of each of these inputs to aid the prediction of groundwater levels was assessed using the mutual information index. The NNAR model was also used to predict groundwater levels under scenarios of change (change in recharge and abstraction). The coefficient of determination, mean squared error, root mean squared error and mean absolute error was used to evaluate the predictions made by the NNAR. The results showed that the NNAR could be used to make groundwater level predictions in 18 boreholes across the Steenkoppies aquifer and that the model can be used to make predictions for scenarios of change. Overall, the NNAR performed well in predicting and simulating groundwater levels in the Steenkoppies aquifer. The transferability of the NNAR to model groundwater levels in different aquifer systems or groundwater levels at different temporal resolutions should be tested to confirm the robustness of the NNAR to predict groundwater levels.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1:
INTRODUCTION

1.1 BACKGROUND

Groundwater is essential, not only for domestic use but also for industrial and agricultural use and supporting natural ecosystems (Beetlestone, 2009). When a supply borehole is drilled, the aim is that the borehole will be able to provide enough water for an extended time and its use will be sustainable for future generations (Konikow and Bredehoeft, 2019). Modelling an aquifer's response to recharge and discharge conditions are critical for water supply and agricultural development (Saatsaz *et al.*, 2011). Through groundwater modelling, groundwater resources may be managed and used sustainably.

Groundwater systems are dynamic and vary due to short and long term changes in climate, abstraction and land use (Kenda *et al.*, 2018). Since groundwater is below the surface, it is often challenging to quantify the impacts of hydrological stresses, such as abstraction and climate change, on an aquifer (Alley and Leake, 2004). An aquifer will respond to hydrological stresses in one of two ways. Either the groundwater system will reach a new equilibrium state, where recharge and discharge balance, allowing the aquifer to exist indefinitely or the stresses on the aquifer are so large that a new equilibrium cannot be reached resulting in a finite span on the aquifer's life (Bredehoeft and Durbin, 2009). Groundwater models are used as a tool to identify an ideal balance between groundwater abstraction and adverse effects on the aquifer to preserve aquifer sustainability (Coppola *et al.*, 2003).

In hydrogeology, physical-based numerical techniques are most commonly used for groundwater modelling, especially since the development of powerful computers and user-friendly software (Kenda *et al.*, 2018). These numerical models require spatial data on geological and hydrological properties of the aquifer (Kenda *et al.*, 2018). This information is often challenging to define accurately. In the last decade, machine learning has been used as a modern, alternative technique for groundwater modelling (Chang *et al.*, 2016). Unlike numerical models, machine learning models are data-driven and learn the behaviour of the aquifer system from measured values without needing an understanding of the internal structure and physical processes of an aquifer (Coppola *et al.*, 2003; Huang and Tian, 2015).

1.2 RESEARCH MOTIVATION

Groundwater is a vital resource of freshwater globally (Walter, 2010). Rapid population growth, industrialisation, urbanisation and climate change have raised the demand for water (McGill *et al.*, 2019). Surface water resources do not meet this increased demand, making groundwater an essential source of water globally and in southern Africa (McGill *et al.*, 2019). In the South African Development Community (SADC), groundwater resources are heavily relied on with an estimated 60% of the SADC's population utilising this resource for their basic water needs (Groundwater Consultants, 2001). Climate change, population growth and increased demand for irrigation to meet food production requirements suggest that the need for groundwater is going to increase in the future (Altchenko and Villholth, 2013). It is crucial to ensure that there will be a long-term groundwater supply through the implementation of sustainable management protocols to meet these future demands (Saatsaz *et al.*, 2011). In addition, groundwater storage depletion in aquifers is already a significant issue threatening the sustainability of groundwater resources (Konikow and Bredehoeft, 2019).

Groundwater models are used to establish the relationship between hydrological stresses and their impact on aquifer recharge, storage and discharge (Lafare *et al.*, 2016; Kenda *et al.*, 2018). An observed groundwater level time series represents the response of an aquifer to hydrological stresses (Lafare *et al*., 2016). Therefore, there is great value in developing a model that can simulate and predict groundwater levels so that one can understand the impact of hydrological stresses, specifically abstraction, on an aquifer. The model can then be used as a tool to identify limits and unacceptable impacts on an aquifer and to use this information to implement sustainable groundwater management policies and protocols (Sahoo *et al.*, 2017).

Hydrogeologist can use both numerical and machine learning techniques to model groundwater levels. However, the conventionally used numerical models are laborious and have practical limitations as they require extensive quantification of physical properties and aquifer conditions (Zhao *et al.*, 2011). Machine learning models do not need such quantification as they are data-driven and recognise the underlying processes and patterns hidden in historical data sets (Hsu and Gupta, 1995). Machine learning models are most useful to model groundwater systems that are poorly defined, and one does not recognise or fully understand all the existing relationships in a system (French *et al.*, 1992). Machine learning models can then be used to predict future scenarios without having additional expert operator input of the aquifer dynamics (Kenda *et al.*, 2018). Many government departments and organisations, such as the Department of Water and Sanitation (DWS) in South Africa, allow free access to historical data sets that drive the machine

learning. There is great incentive to explore machine learning as an alternative for groundwater level modelling. Machine learning offers hydrogeologists a variety of techniques to model groundwater levels.

The machine learning technique used in this study is the Neural Network Autoregression (NNAR). The NNAR is a type of Artificial Neural Network (ANN). In the literature, ANNs are the most popular machine learning technique used to successfully model groundwater levels (Maier *et al.*, 2010). ANNs are a valuable tool for modelling complex non-linear issues which are difficult to describe with conventional methods, specifically in situations where the outcome is more important than the understanding of the involved process (Wunsch *et al*., 2018). ANNs are also a powerful tool for predicting groundwater levels in highly irregular hydrogeological systems (Lee *et al*., 2019). The popularity of ANNs has resulted in the availability of a large number of open access resources (such as TensorFlow in Python or the Forecast package in R) to assist in learning and building ANN models.

NNARs, are a type of ANNs that are specifically good at modelling time series data. NNARs strength is in their ability to look at previous lagged values of the target time series as well as the current value to make predictions (Wunsch *et al.*, 2018). This allows them to determine time dependant patterns more accurately compared to other machine learning techniques (Da Silva *et al.*, 2017). The NNAR was, therefore, an appropriate machine learning technique to model groundwater levels in the Steenkoppies aquifer.

1.3 AIMS AND OBJECTIVES

This study aimed to test the applicability of machine learning, specifically the Neural Network Autoregression (NNAR), as an alternative technique for modelling groundwater levels.

The objectives of the study were to:

a) Use the NNAR to establish a model capable of simulating and predicting groundwater level change in the Steenkoppies compartment of the Gauteng and North West dolomite aquifer.
b) Assess the success of the performance made by the model statistically, using several error indices (coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE)) to compare the observed and the predicted groundwater levels and graphically assess the model's predictive capability by plotting the observed groundwater levels against those simulated by the model.
c) Use the machine learning model developed to predict groundwater levels for future scenarios not seen in previous data sets.

1.4 CASE STUDY SITE SELECTION

The Gauteng and North West dolomite aquifers are arguably one of the most important aquifers in South Africa (Cobbing *et al.*, 2016). Communities and the surrounding ecosystem heavily utilise this aquifer. Not only does it maintain springs, wetlands and rivers in and around the area, but it also supplies water to several towns (e.g. Groot Marico, Mahikeng, and Ventersdorp) and supports lucrative irrigated commercial agriculture (Cobbing *et al.*, 2016). Volcanic dykes have intruded into the Gauteng and North West dolomite aquifer and divided it into several smaller aquifer compartments (Johnson *et al.*, 2006). One of these compartments is the Steenkoppies Groundwater Management Unit (GMU) which is the case study for this research (Figure 3-1). Throughout this study, the Steenkoppies Groundwater Management Unit (GMU) is referred to as the Steenkoppies compartment.

The groundwater in the Steenkoppies compartment provides water primarily for productive local agriculture, including the largest producer of carrots for export in South Africa (Cobbing, 2018). The farms that are reliant on groundwater employ over 4000 people and have significant links to other sectors of South Africa's economy (Cobbing, 2018). This makes the groundwater in the Steenkoppies aquifer a very valued resource. The Steenkoppies compartment is at risk of groundwater over-use. The historical data sets show a definite decline in groundwater levels and spring discharge as abstraction increased, and the aquifer transitions to a new dynamic

equilibrium. By developing a model that can simulate and predict groundwater levels in the Steenkoppies compartment, one can understand the impact of hydrological stresses, specifically abstraction, to make appropriate sustainable groundwater management decisions in the Steenkoppies aquifer.

Machine learning is data-driven. Therefore, good quality data sets are needed to develop an accurate model. Relative to other aquifers in the SADC region, an extensive historical groundwater level data set is available for the Gauteng and North West dolomite aquifer. Historical data sets for the input variables were also available (for this study, the input variables used were rainfall, temperature and spring discharge). Compared to other compartments in the Gauteng and North West dolomite aquifer, the Steenkoppies compartment has the greatest density of boreholes recording groundwater level readings for an extensive time with the least data gaps. Historical data sets for the input variables are also available from nearby meteorological stations

The available historical groundwater and input variables data makes the Steenkoppies compartment in the Gauteng and North West Dolomite aquifer a worthwhile case study for testing the use of machine learning as an alternative technique for modelling groundwater levels.

CHAPTER 2:
LITERATURE REVIEW

2.1 INTRODUCTION

The machine learning model used in this study (Neural Network Autoregression (NNNAR)) makes use of the relationship between the target variable (groundwater levels) and other input variables to make predictions. The selection of suitable input variables is reliant on one's understanding of the influence and interactions these variables have with groundwater levels. Section 2.2 explains the variables that influence change in groundwater storage and ultimately groundwater levels over time. Section 2.3 describes the definition and concept of groundwater sustainability to understand how hydrogeologists can use groundwater models as a tool to promote sustainable management. Machine learning is a new field in hydrogeology, especially in South Africa. Section 2.4 explains what machine learning is, as well as the different types of machine learning models available to model time series. Section 2.4 also describes the model used in this study and an overview of the investigations where researchers have used machine learning models to model groundwater levels.

2.2 SOURCE OF WATER TO AN AQUIFER

2.2.1 Under Natural Conditions

Prior to groundwater abstraction, it is assumed that there are no long term storage changes in the groundwater system and that recharge balances discharge (Konikow and Bredehoeft, 2019). The recharge going into the groundwater system is equal to the discharge leaving the system. A groundwater system may experience fluctuations in storage on a daily to decadal time scale based on dryer or wetter periods. However, it is assumed that these fluctuations eventually balance out and the groundwater system remains in a state of dynamic equilibrium (Konikow and Bredehoeft, 2019).

The water budget (or balance), first described by Lee (1908) and Meinzer (1931) describes the state of an aquifer in its natural balance over a long period, before groundwater abstraction (Bredehoeft and Durbin, 2009). The concept of the water budget is consistent with the Law of Conservation of Mass and expressed as a simplified mass balance (Equation 2-1) (Healy *et al.*, 2007):

$$\boldsymbol{P + Q_{in} = ET + \Delta S + Q_{out}}$$

Equation 2-1

Where:

P = recharge

Q_{in} = water flow into the groundwater system

ET = evapotranspiration

ΔS = change in water storage

Q_{out} = discharge out of the groundwater system.

The natural balance of an aquifer can be upset by climate change. Climate change can lead to a change in the amount of recharge to an unconfined aquifer due to reduced, sporadic or less intense rainfall, decreasing soil infiltration and deeper percolation of water to the aquifer (Wu *et al.,* 2020). Increased evaporation due to rising temperatures also limits the amount of water available to recharge aquifers (Döll, 2009). The effect of climate change may alter the balance between recharge and discharge, imposing hydrological stress on the groundwater system (Bredehoeft and Durbin, 2009).

2.2.2 Under Groundwater Abstraction Conditions

A pumped borehole is supplied or balanced by water from one or a combination of three potential sources; i) an increase of recharge to the borehole caused by pumping, ii) a decrease of discharge from the aquifer also caused by pumping or iii) a reduction in groundwater storage in the aquifer (Theis, 1940, cited in Konikow and Bredehoeft, 2019).

When groundwater is pumped from a borehole, water is initially removed from storage and a cone of depression is formed. A new discharge is introduced into the system and this creates a gradient in the local hydraulic head adjacent to the borehole, resulting in a flow of water to the borehole. After a time, water to the borehole is no longer supplied by groundwater storage alone, but by an increase in recharge to the aquifer and a decrease in discharge from the aquifer caused by pumping (Theis, 1940, cited in Konikow and Bredehoeft, 2019). "Capture" or the "capture

principal" defines the increase in recharge, and the decrease in discharge caused by pumping (Seyler *et al.,* 2016). “Capture” is the water that naturally would not have entered the groundwater or the decrease in discharge bellow the natural amount due to pumping (Konikow and Bredehoeft, 2019).

If there is sufficient capture, the groundwater system will reach a new equilibrium, and groundwater storage will no longer be the source of water to the borehole as capture balances the pumping, causing no change in the groundwater storage (van der Gun and Lipponen, 2010). This means that the water mass balance is conserved, and the new discharge from pumping is balanced by the increase in recharge and decrease in natural discharge (Figure 2-1). The time it takes to reach the new dynamic equilibrium is called response time (Alley and Leake, 2004). Response time is dependent on the size, boundary conditions and hydraulic properties of the aquifer and can vary from days to centuries (Konikow and Leake, 2014).

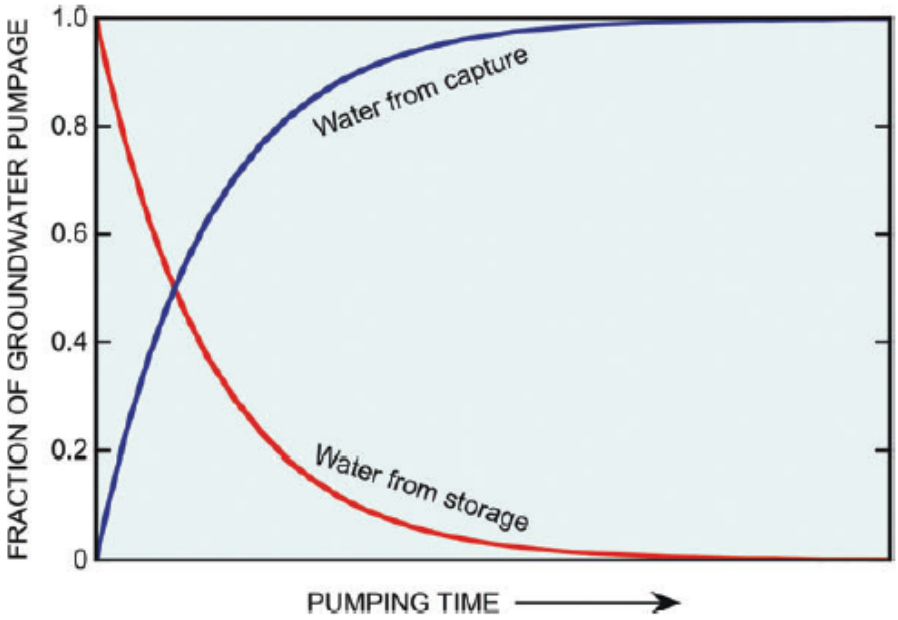


Figure 2-1: Source of water to a borehole over time (Source: Konikow and Leak, 2014).

Theis (1940) noted that in many circumstances, surface water features such as rivers form aquifer boundaries. If such features do not exist or if rivers have limited water supply, they may pose a limit on capture (Konikow and Leake, 2014). This is because a feature such as a river can supply capture to satisfy the increase in recharge caused by pumping.

If the stress on a groundwater system due to pumping is too great, abstraction may not be balanced by capture. As a result, water levels will continue to decline, and the source of water to a borehole will continue to be from storage (Bredehoeft and Durbin, 2009; Konikow and Leake,

2014). If capture is not large enough to balance pumping, and an equilibrium is not reached, groundwater storage may become exhausted (van der Gun, 2017). This may result in unsustainable aquifer usage; however, the definition of groundwater sustainability is case-specific (further explained in section 2.3). The impact humans have on groundwater should be balanced against the social and economic benefits (Seyler *et al*., 2016).

Since pumping can increase recharge rates and decreased discharge rates, one cannot use the basic water budget approaches to determine sustainable pumping (Devlin and Sophocleous, 2005). The Water Budget Myth is mistakenly thinking that to maintain groundwater sustainability, pumping rates should not be greater than natural /virgin recharge rates (Devlin and Sophocleous, 2005). This is incorrect as the water budget does not recognise that pumping alters the recharge and discharge rate of an aquifer (Devlin and Sophocleous, 2005; Seyler *et al., 2016)*. Many authors (Theis, 1940; Brown, 1963 and Bredehoeft, 2002) have provided theoretical proof that the water budget approach does not consider the dynamic nature of an aquifer when pumped and only considers long term steady-state aquifer behaviour.

Bredehoeft *et al.* (1982), discuss that natural recharge rates are not required to determine sustainable pumping rates. However, this is not to say that recharge is irrelevant to groundwater sustainability (Bredehoeft *et al.,* 1982). Sustainability is a broader concept which depends on the changes that occur to the environment and socio-economy, to which recharge is bound to affect (Kendy, 2003; Devlin and Sophocleous, 2005). For example, recharge may affect groundwater quality and ultimately affect the associated ecology (Devlin and Sophocleous, 2005).

In the case of irrigation for agriculture, Alley *et al*., (1999) and others point out that not all the water pumped from the groundwater is used up and lost from the aquifer. In many shallow unconfined aquifer systems, the water used to irrigate crops filtrates back into the aquifer and the net pumping rate is not equal to the total pumped water (Devlin and Sophocleous, 2005). This discrepancy is essential for determining limits on groundwater abstraction for irrigation. It is not pumping per se that depletes the groundwater storage, but the amount of pumped groundwater that evaporates (Kendy, 2003).

2.3 GROUNDWATER SUSTAINABILITY

2.3.1 Definitions of Groundwater Sustainability

The use of groundwater in Africa has resulted in significant economic growth; however, the exploitation of this resource has led to reduced quantity and quality of the groundwater resources which threaten groundwater sustainability (Walter, 2010). As Davies *et al.* (2013) states, groundwater resources are often poorly managed and poorly understood, due to an "out of sight, out of mind" association with groundwater (Davies *et al.*, 2013). Groundwater is not strictly renewable. In most cases, abstracted water will return to the hydrological cycle; however, the replenishment period may take several generations (Hiscock *et al.,* 2002). Groundwater policies from governments are often inconsistent with groundwater flow and timescales, making it very difficult to manage and enforce groundwater sustainability (Gleeson *et al.*, 2012).

The definition of groundwater sustainability has changed significantly, from the older 20th-century concept of safe yield, which only encompasses the physical system of an aquifer, to the more recent idea of groundwater sustainability, which recognises the interactions between people, economics and the environment (Alley and Leake, 2004; Devlin and Sophocleous, 2005; Gleeson *et al.*, 2012).

Safe yield was first defined by Lee (1915) as the quantity of water that can be abstracted permanently without depleting the storage reserve (Alley and Leake, 2004). Safe yield is equivalent to "maintainable aquifer yield" defined by Seyler *et al*., (2016) and will be used throughout this study. Alley *et al.*, (1999) defined groundwater sustainability as the "development and use of groundwater in a manner that can be maintained for an indefinite time without causing unacceptable environmental, economic, or social consequences". Declaring an aquifer yield as sustainable assumes that the impacts associated with the abstraction (such as reduced natural discharge) are socially, economically, and environmentally acceptable (Seyler *et al*., 2016).

2.3.2 Sustainable Groundwater Use

Safe yield depends on how much of the groundwater pumped can be captured from increased recharge and decreased discharge (Alley and Leake, 2004). The safe yield does not necessarily reflect sustainable groundwater use (Seyler *et al*., 2016). To estimate the safe yield, one should determine the following conditions/ changes to the aquifer: the yield of each source of water to the borehole (recharge and discharge as capture and aquifer storage), the extent of storage depletion (long term reduction in groundwater levels) and the response time (Seyler *et al*., 2016). Based on these conditions, one can also monitor the socioeconomic and environmental changes to determine if this yield is acceptable and reflects groundwater sustainability (Seyler *et al*., 2016). If the changes are acceptable, then the pumping at the safe yield can be considered sustainable for groundwater use (Alley and Leake, 2004). The safe yield should be updated through adaptive management practices as many variables can influence the safe yield and the sustainability considerations over time. This is the basis for the capture principal to sustainable groundwater use.

Hiscock *et al.* (2019), Gleeson (2012) and others recognise that groundwater should not be used in a way that inflicts negative consequences to either the economy, society, or the environment. Alternatively, Foster (200) and Price (2002) reason that aquifer storage can be over-exploited if there is a beneficial outcome (Hiscock *et al.*, 2002). For example, overexploitation of the High Plains Aquifer in the United States caused severe water-level declines in some aquifer compartments, however, the over-exploitation of this aquifer has transformed this area into a world-leading agricultural hub (Hiscock *et al.*, 2002). All groundwater abstractions have an impact, and one needs to determine what level of impact is acceptable (Seyler *et al*., 2016).

2.4 MACHINE LEARNING

Machine learning is an evolved subfield of artificial intelligence (Marinósdóttir, 2019). Mohr *et al.* (2018), defines machine learning as algorithms that learn and improve from experience in the form of data collected to make accurate predictions. Machine learning is a broad and growing field and has become a powerful tool for predictive analysis (Ye, 2015). Figure 2-2 summarises the various categories and types of machine learning models applicable to this study, which are explained below.

Machine learning algorithms are grouped into two categories; supervised and unsupervised algorithms (Da Silva *et al.*, 2017). In unsupervised learning, the algorithm infers patterns from data that is not known or labelled (Da Silva *et al.*, 2017). The goal is to learn the underlying structure of the data. The model does not have a reference for what the output data should look like, making it impossible to conduct model training. In order to extract meaningful information, the model explores the structure of the data (Da Silva *et al.*, 2017).

Supervised learning algorithms are used when input and output variables are clearly labelled. The goal is to learn patterns and correlations between variables from previous experience (training data) and use that to make predictions on the unseen or unknown data (test data). There are two subcategories of data that can be modelled in supervised learning; classification in which a model aims to predict categorical or class labels, and regression in which the models attempt to predict a continuous output (Agatonovic-Kustrin and Beresford, 2000). In this study, regression data is modelled using supervised learning.

There is a wide range of machine learning techniques for predictive analysis. One of the more popular techniques is the artificial neural network (ANN) (Maier and Dandy, 2000).

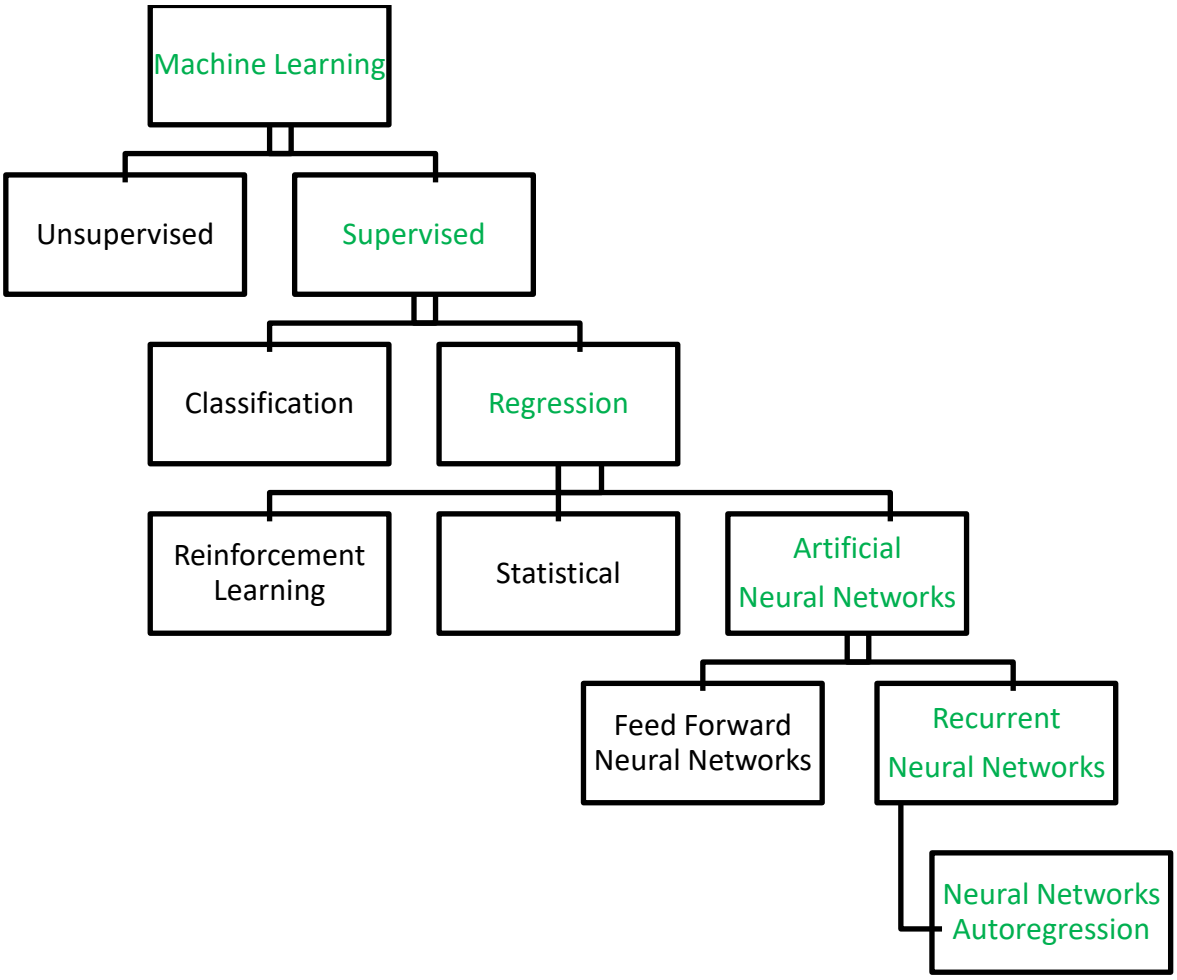


Figure 2-2: Flow chart summarising the classification of some of the machine learning techniques available. The green text highlights those applicable to this study.

2.4.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are a type of machine learning algorithm designed to simulate the way the human brain analyses and processes information (Agatonovic-Kustrin and Beresford, 2000). The human brain can memorise, and recognise patterns and still generalise in a wide variety of circumstances. The ability of the brain to do this is the driving factor behind the development of artificial neural networks which attempt to mimic biological neural systems (Engelbrecht, 2007). ANNs were first conceptualised in 1943 by McCulloch and Pitts; nowadays, the use of ANNs is widely used in hydrogeology, mainly for research (Maier *et al.*, 2010).

The basic architecture of an ANN can be divided into three main layers namely; the input layer, where data is received from the external environment; hidden layers (or "black box"), where the majority of the internal processing of the ANN occurs such as extracting the patterns associated with the data; and the output layer, which produces the network outputs from the

processing performed in the hidden layers (Figure 2-3) (Da Silva *et al.*, 2017). Each layer of an ANN comprises of nodes that are linked by weighted connections (Da Silva *et al.*, 2017). A node receives a weighted input to which a bias is added or subtracted to gauge the input to a useful range to improve the ANN's performance (Shahain *et al.*, 2008). The result after the bias is added or subtracted is passed through an activation function to the neurons in the next layer (Shahain *et al.*, 2008). The activation function allows the ANN to make complex connections between the network's inputs and outputs (Da Silva *et al.*, 2017). By changing the activation function and the number of hidden layers one alters the degree to which the data is non-linear (Shahin *et al.*, 2008). Depending on how the connections between the neurons are organised, an ANN is classified as a feedforward neural network (FFNN) or a recurrent neural network (RNN).

ANN models take on the following form (Maier *et al.*, 2010):

$$Y = f(X, W) + \varepsilon$$

Equation 2-2

Where:

Y = vector of model outputs
X = vector of model inputs
W = vector of model parameters (connection weights)
f (*) = functional relationship between model outputs, inputs and parameters
ε = vector of model errors

If the connections between the neurons are in one direction, from inputs to outputs, the ANN is an FFNN (Figure 2-3A) (Agatonovic-Kustrin and Beresford, 2000). FFNN architecture only uses the current timestep fed into the model. It does not retain a record of the previous output values as there is no connection from the output to the input neurons (Agatonovic-Kustrin and Beresford, 2000).

If the connection between the neurons is in both a forward and backwards direction, the ANN is classified as an RNN (Figure 2-3B) (Da Silva *et al.*, 2017). RNNs are different from FFNNs as the output of one layer directs back into the input of the same or previous layer and stored as a hidden state in the model (Agatonovic-Kustrin and Beresford, 2000). RNNs have great success in determining time dependant patterns and can be employed on time-variant systems (Brezak *et al.*, 2012). The chronological information stored in the RNNs hidden state extents several time stamps and cascades forward to influencing the predictions (Taver *et al.*, 2015). Therefore, an

event downstream in time is a function of one or more events before the current timestamp. This added memory allows RNNs to determine time dependant patterns more accurately compared to the FFNNs (Brezak *et al.* 2012). RNN have the potential to be successfully applied to model hydrogeological processes and produce reasonable solutions, even when there is limited information about the physical properties of the aquifer. One drawback to modelling using RNNs is that the prediction error increases with time. There is an error on the predicted output, which is fed back into the model, causing an accumulation of errors as predictions are made further into the future. This is known as "the problem of vanishing gradients" and as a result, less accurate predictions are generated the further into the future one predicts (Scardapane and Wang, 2017). There are many different types of RNNs available for groundwater modelling.

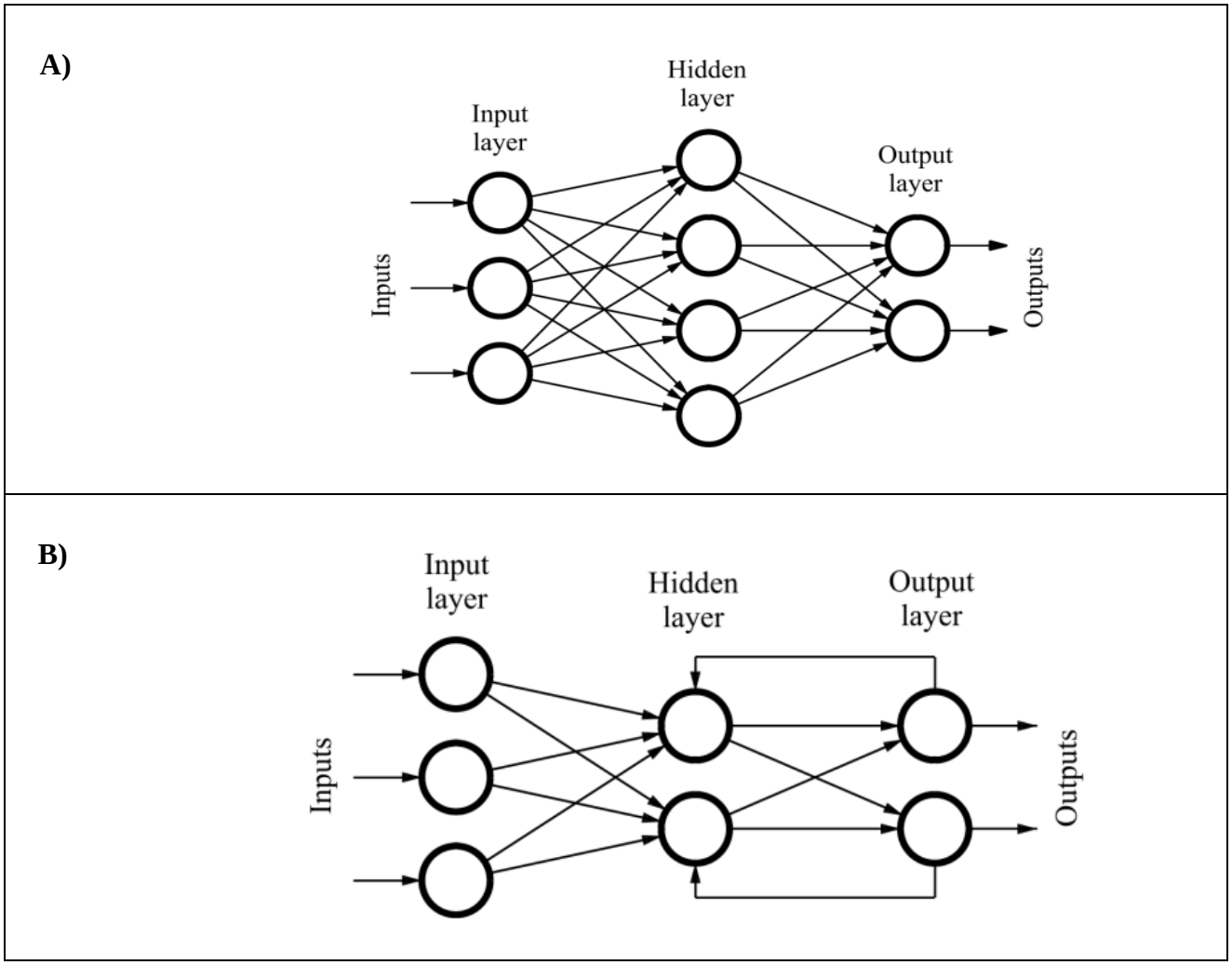


Figure 2-3: Comparison between the architecture of two types of Artificial Neural Networks: A) Feed Forward Neural Network (FFNN) where the connections between the nodes are in a forwards direction, B) Recurrent Neural Network (RNN) where the connections between the nodes are in a forward and backwards direction, and the output is directed back into the input of the same or previous layer and stored in a hidden state (Source: Quiza and Davim, 2011)

2.4.1.1 Neural Network Autoregression (NNAR)

Neural Network Autoregression (NNAR) is one variant of the RNN. Although RNNs have limitations in making long-term predictions (the problem of vanishing gradients) (Scardapane and Wang, 2017), NNARs usually provide better results than conventional RNNs for predicting groundwater levels as these models keep information about the data two or three times longer compared to standard RNNs (Wunsch *et al.*, 2018). Like an RNN, the NNAR model's next output is not only dependant on the present inputs but also previous output signals of the target variable (also known as lagged inputs) (Izady *et al.*, 2013). However, unlike an RNN, the NNAR's previous timestamps are not stored in a hidden state but given as another input to the model (Figure 2-4) (Izady *et al.*, 2013).

The internal processing of the NNAR model is a neural network with a linear combination function and a non-linear activation function (Equation 2-3 and Equation 2-4) (Yoon *et al.*, 2011; Khalek *et al.*, 2016). The inputs to the model are put through the linear function, and the result is then passed through the non-linear sigmoid activation function (Khalek *et al.*, 2016)

The linear combination function at node j in the hidden layer is defined as below (Yoon *et al.*, 2011):

$$y_j = f\left(\sum_{i=1}^{N} w_{ji} x_i + b_j\right)$$

Equation 2-3

Where:

x_i = i^{th} input from the node in the previous layer

y_j = input value at the present node, b_j is the bias

w_{ji} = weight connecting x_i and y_j,

f = sigmoid activation function

N = number of nodes in the previous layer

The sigmoid activation function for target variable y, is formulated as (Khalek *et al.*, 2016):

$$f(y) = \frac{1}{1 + e^{-y}}$$

Equation 2-4

The values for the weights and bias are "learned" from the data. To begin with, the values for the weights are chosen randomly and then updated using the training data so that the overall predictive error is minimised (Zhang and Hu, 1998). Consequently, the NNAR (or any ANN) will produce slightly sifferent predictions each time the model is run (Scardapane and Wang, 2017). To obtain a robust evaluation of the models ability to make predictions, one should run the NNAR several times, and the final results should be an average of the several model runs.

The general formula for an NNAR model fitted to the data is given in Equation 2-5 (Ruiz *et al.*, 2016):

$$y(t) = h(\mathrm{x(t-1), y(t-1), y(t-2), \ldots, y(t-p))} + \mathrm{e(t)}$$

Equation 2-5

Where:

h = neural network used to fit the data

y = target time series to predict

x = input variable used to predict the target

p = lagged inputs of the target time series

$e(t)$ = the error

The model makes predictions recursively. To predict one step, the model simply uses the available historical inputs. For predicting two steps, the model simply uses the one-step prediction as an input, along with the historical data (Hyndman and Athanasopoulos, 2014). This process proceeds until all the required predictions are calculated.

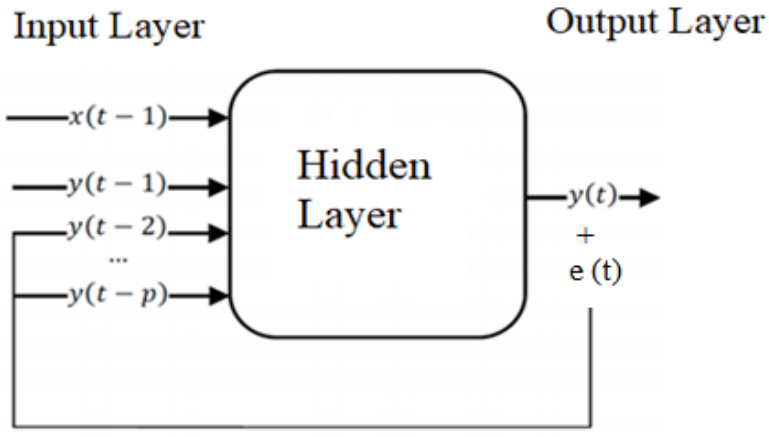


Figure 2-4: Basic architecture of the Neural Network Autoregression (NNAR) where the connections between the nodes are in a forward and backwards direction, and the outputs are fed as a new input to the model (Source: Ruiz *et al.*, 2016)

2.4.2 Modelling Timeseries Using Artificial Neural Networks

The basic concept to model a time series using ANNs is to predict the target variable, y, assuming it has a relationship with input variables, x (Wickham and Grolemund, 2016). In this study, we forecast y, the groundwater levels, using the x variables rainfall spring discharge, temperature and groundwater usage.

The first decision to make when developing an ANN model is to decide how to utilise the time series at hand to create the model and assess its performance. This is known as the model pipeline. It is common practice to split the time series of both the target and the input variables into two, the training, and the test subsets (Kuhun and Johnson, 2019). Timeseries order must be maintained when split to ensure that the model picks up the patterns in the data (Adhikari and Agrawal, 2013).

The training data set is used to develop the model and allow for the model to recognise and "learn" the patterns in the target variable (y) and the relationships between the target variable and input variables (x) (Wickham and Grolemund, 2016). During training, the model makes use both the time series from the input and target variables to optimise parameters such as the number of hidden layers, the weighted connection between layers and the number of nodes in each layer, so that the model can predict the target variable with the lowest error. This is called hyperparameter tuning. Although one can set the model parameters manually, hyperparameter tuning is typically carried out by the ANN adaptively based on the data fed to the model (Wickham and Grolemund, 2016).

There are many methods in which one can feed the training data to the ANN algorithm that may improve the hyperparameter tuning ability of the ANN. One of these methods is cross-validation. Cross-validation is an effective method as it increases the generalisability of the model (Kuhun and Johnson, 2019). Cross-validation was used in this study and explained in Chapter 4. Once the model training is complete, and the hyperparameters of the model are optimised, the model simulates the target variable for the training data. This allows one to see if the data is overfitting or underfitting. Overfitting is when the model learns the noise in the training data so accurately (fits all peaks at troughs exactly) that it negatively affects the performance of the model when predicting new data (Wickham and Grolemund, 2016). Underfitting is when the model is unable to simulate the training data or predict new data (Wickham and Grolemund, 2016). If simulated target output for the training data shows a good but not exact fit to the data, then the model is ready to make predictions.

The developed ANN model does not "see" the target variable it aims to predict but make predictions by observing the consistencies and patterns in the test data set of the input variables and provide predictions based on what the model has learnt during training (Adhikari and Agrawal, 2013). The test data should remain unused until the final model is developed. The target variable test data is an unbiased assessment of the model's performance (Adelabu *et al*., 2015). The test data of the target variable is used to assess the model's predictive capability by statistically and graphically comparing the predictions made by the model with the actual measured values of the target variable (Pani *et al*., 2019). If the model successfully predicts the target variable with a suitable error level, then one can deploy the model and use it to make additional predictions. If the model does not successfully predict the target variable and the error is too large, then the model pipeline would need to adjust until the model can successfully predict the target variable.

2.4.3 Application of Artificial Neural Networks in Hydrogeology

In the last decade, neural networks have been used and applied for time series prediction tasks due to their ability to model complex non-linear functions (Gao and Er, 2005). The practical implementation of machine learning techniques for groundwater modelling has not yet taken off but explored in research (Kenda *et al.*, 2018). Some of the machine learning techniques successfully used to model groundwater levels include, FFNN by Aziz and Wong (1992); support vector machines applied by Yoon *et al.*, (2011); and autoregressive (AR) models, used by authors such as Shirmohammadi *et al.*, (2013) and Wunsch *et al*., (2018).

In a review, Maier and Dandy (2000) assessed 43 papers that made use of neural networks for the prediction and forecasting of water resources. Most authors using ANNs for hydrogeological applications made use of FFNN, however recently there has been more investigations on the use of RNN, and the NNAR, for hydrogeological modelling (Shirmohammadi *et al.*, 2013; Wunsch *et al.*, 2018).

Sreekanth et al. (2009), used FFNN to develop a model that could predict monthly groundwater level fluctuations in the Maheshwaram watershed in India. In this study, the author used monthly groundwater levels between 2000-2006 from 22 wells in the study area. The ANNs predicted groundwater with a high degree of accuracy ($R^2 = 0.93$). This study concluded that satisfactory groundwater level predictions could be made using limited groundwater records.

A neural network was applied in India by Nayak *et al.* (2006), which used monthly rainfall, irrigation and canal release to predict groundwater levels. The model produces groundwater level predictions for up to four months. However, after two months, the error in the model starts to accumulate (the problem of vanishing gradients) and the performance of the model deteriorates.

Shamsuddin *et al.* (2017), and other authors have used FFNN to forecast groundwater levels on a daily temporal resolution. Shamsuddin *et al*. (2017) showed that FFNNs are capable of modelling daily groundwater levels using rainfall, temperature, streamflow, and river water level data as inputs along with values of groundwater level.

Lee (2019), predicted hourly groundwater levels using FFNN to find the impact of natural factors and anthropogenic factors, such as artificial recharge and pumping, on groundwater level. Lee (2019), trained an FFNN using the back-propagation algorithm. Lee (2019) successfully used the model to predict groundwater levels with acceptable errors. However, slightly higher errors were encountered in the wells where there was a more significant anthropogenic influence on the groundwater.

Daliakopoulos and Coulibaly (2005), made use of different ANNs to predict groundwater level fluctuations 18 months ahead. The three ANNs used in this study were FFNN, RNN and the Radial basis function network (RBF). All the models performed sufficiently. However, the model that performed the best was the FFNN trained with the Gradient descent with momentum and adaptive learning rate back-propagation algorithm (GDX). GDX is an activation function that attempts to reduce the effect of the vanishing gradient problem.

Chang *et al.* (2016), is one of the few studies where groundwater level change at an aquifer scale was modelled at a monthly temporal resolution. This study used a combination of a Self Organised Map (SOM) and Nonlinear Autoregressive with Exogenous Inputs (NARX) (similar model to the NNAR) for predicting basin-scale groundwater level in the Zhuoshui River basin in Taiwan. Monthly data sets from 203 groundwater stations, 32 rainfall stations and six flow stations between 2000 and 2013 were used to for modelling. The results indicate that the NARX, can predict reliable groundwater level predictions (R^2 >0.9 for training and test cases) at a basin scale. The finding by Chang *et al.* (2016), shows the applicability of an NARX for large scale environmental systems.

Guzman *et al.* (2017) used an autoregressive neural network, simmilar to the NNAR, to predict daily groundwater levels up to three months ahead. The predictions become less accurate over time, and the best performance is shown for predictions 15 days ahead with a minimal error of less than 0.001 m between measured and predicted values. Comparable to Guzman *et al*., (2017),

Wunsch *et al.*, (2016) also made use of an autoregressive neural network model, the NARX, to model groundwater level at a weekly temporal resolution. Groundwater levels from boreholes both influenced and uninfluenced by nearby pumping from three different aquifer types, (karst, fractured and porous) were modelled using rainfall and temperature time series as inputs. By only using two climatic variables (rainfall and temperature) as model inputs, the study shows how the NARX can be used in all areas with simple, readily available input variables. The model performed best in predicting groundwater levels from the karst aquifers. The wells that were influenced by pumping performed worse than those uninfluenced by pumping. Overall, the NARX showed promising results on groundwater predictions with only two input variables.

Table 2-1: Summary of studies where researchers used ANNs to model hydrogeological systems (Feed Forward Naural Network (FFNN), Recurrent Neural Network (RNN), Radial Basis Function (RBF), Nonlinear Autoregressive Network With Exogenous Inputs (NARX)).

Author	Model algorithm	Target variable and Input variables	Temporal resolution	Total sample size	Prediction length (months)	Accuracy
Sreekanth *et al.* (2009)	FFNN	Target: Groundwater levels Input: Rainfall, evapotranspiration, temperature, humidity, river discharge	Monthly	72	24	R^2 = 0.93 RMSE= 4.50 m
Nayak *et al.* (2006)	FFNN	Target: Groundwater levels Input: Rainfall, canal releases	Monthly	96	4	RMSE=0.32-0.58 m
Shamsuddin *et al.* (2017)	FFNN	Target: Groundwater levels Input: Rainfall, temperature, river stage, stream flow rate	Daily	406	4	R^2 = 0.55-0.75 MSE=0.03-0.1 m RMSE= 0.01m
Lee *et al.* (2019)	FFNN	Target: Groundwater levels Input: Rainfall, groundwater pumping, artificial recharge	Hourly	8712	3.4	R^2= >0.9 NSE= >0.75 RMSE=3-6 cm

Daliakopoulos and Coulibaly (2005)	FFNN RNN RBF	Target: Groundwater levels Input: Rainfall, temperature	Monthly	180	18	**FFNN:** R^2 = 0.59-0.99 RMSE= 2.11-9.84 m **RNN:** R^2 = 0.61-0.91, RMSE= 3.31-9.32 m RBF: R^2 = 0.74 RMSE= 5.23 m
Chang *et al*. (2016)	NARX	Target: Groundwater levels Input: Rainfall, river flow	Monthly	165	33	R^2 = 0.68-0.97 MSE=0.25-0.96 m RMSE= 0.34-1.18 m
Wunsch *et al.* (2018)	NARX	Target: Groundwater levels Input: Rainfall, temperature	Weekly	1624 - 3640 (Borehole dependant)	54	R^2 = 0.18-0.91 RMSE=0.10-1.45 m RMSEr = 0.03-0.51 % NSE=0.26-0.94 m
Guzman *et al*. (2017)	NARX	Target: Groundwater levels Input: Rainfall, previous groundwater levels	Daily	2920	1 - 3	R^2 = 0.83-0.92 MSE=0.001-1.002 m NSE=0.84-0.96 m

CHAPTER 3:
SITE DESCRIPTION OF THE STEENKOPPIES COMPARTMENT

3.1 LOCATION AND SETTING

The Gauteng and North West dolomite aquifer underlay the Gauteng and North West Province (DWA, 2009). This aquifer comprises of 300 million-year-old Malmani dolomite formations of the Chuniespoort Group (Wiegmans *et al.*, 2013). The Malmani dolomites have undergone karstification to become one of South Africa's most high yielding aquifers (DWA, 2009). Dykes intruded into the dolomites, forming boundaries impermeable to groundwater flow. These dykes have created separate hydrogeological aquifer compartments (Wiegmans *et al.*, 2013). The focus of this study is the Steenkoppies compartment. The Steenkoppies compartment is situated west of Tarlton, South Africa (26°02'S to 26°13' S, 27°29' E to 27°39' E) and covers an area of approximately 312 km^2 (Figure 3-1) (Holland *et al.*, 2009).

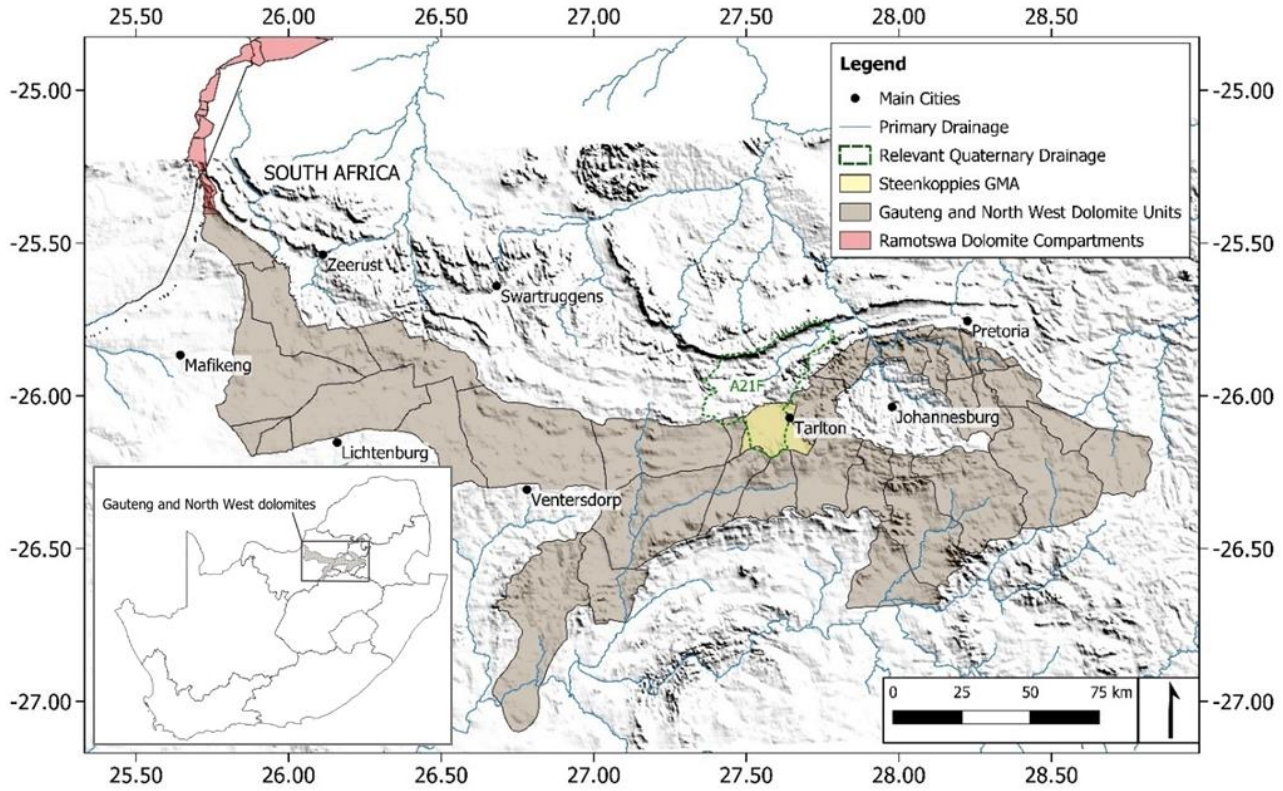


Figure 3-1: Map showing the location of the Gauteng and North West dolomite aquifer (highlighted in grey) and the Steenkoppies compartment (highlighted in yellow). The inset shows the position of the Gauteng and North West dolomite aquifer in South Africa.

3.2 TOPOGRAPHY, DRAINAGE AND LAND COVER

The topography in the Steenkoppies compartment is characterised by undulating plains (Vahrmeijer *et al.*, 2013). The altitude of the Steenkoppies compartment varies from 1700 m above mean sea level (mamsl) in the south-east to 1495 mamsl in the north and west of the compartment. The altitude dips down to an even lower altitude to 1200 mamsl at the flood plains of the Magalies River in the far north of the compartment (Figure 3-2).

Besides the non-perennial Brandvlei River, there are no surface water drainage features or wetlands in the Steenkoppies compartment (Figure 3-2). This indicates that most of the rainfall drains directly into the aquifer (Vahrmeijer *et al.*, 2013).

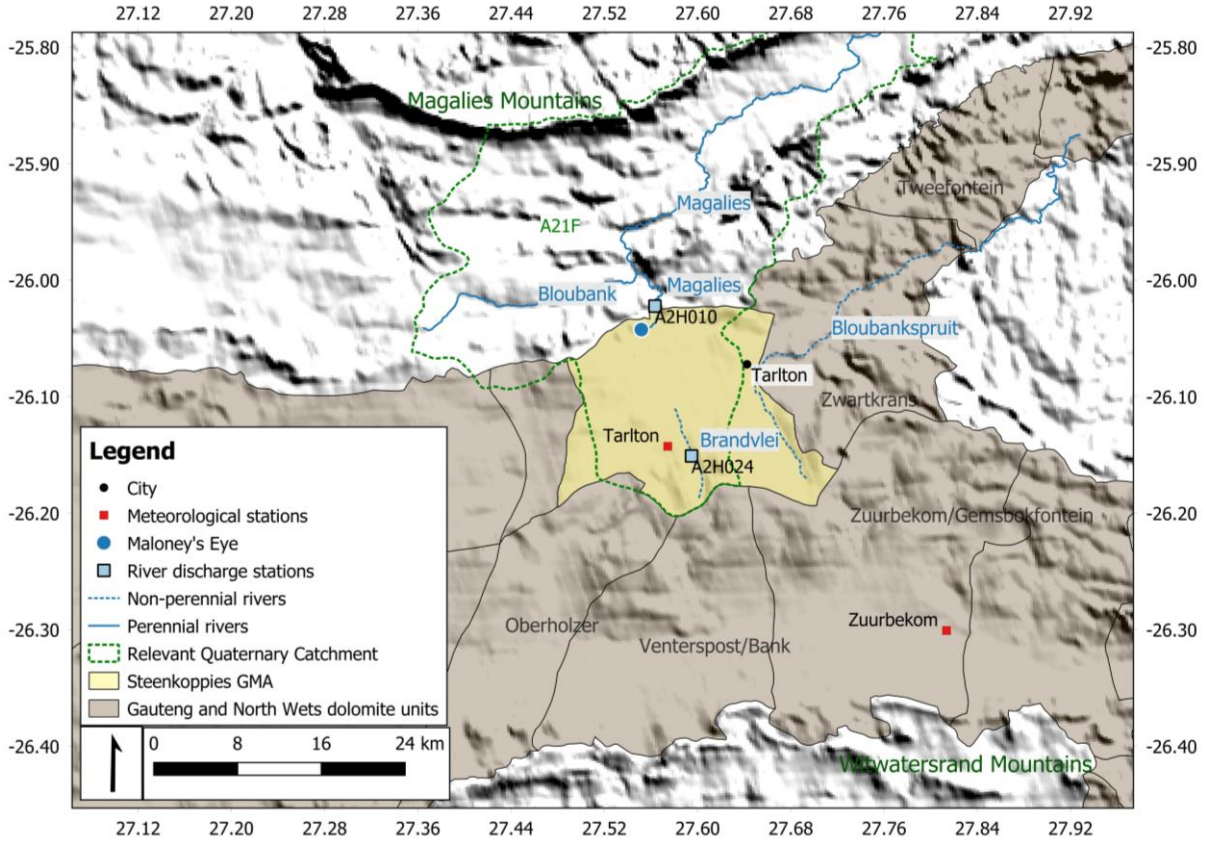


Figure 3-2: Location of the Steenkoppies compartment boundary, relevant hydrology, meteorological stations and cities.

Figure 3-3 shows land cover over the Steenkoppies compartment, according to the South African National Land-cover data set of 2018. From the south-west to the north-east of the Steenkoppies compartment, both pivot irrigated, and rain-fed agriculture dominates. The south-east of the Steenkoppies compartment is occupied by residential settlements, both formal and informal equally represented. Minor planted forest covers the south of the compartment. In the north of the Steenkoppies compartment, "commercial" dominates. The South African National Land-cover data set of 2018 defines "commercial" as the non-residential areas used for business and commerce. Overall, agricultural practices cover most of the in the Steenkoppies compartment.

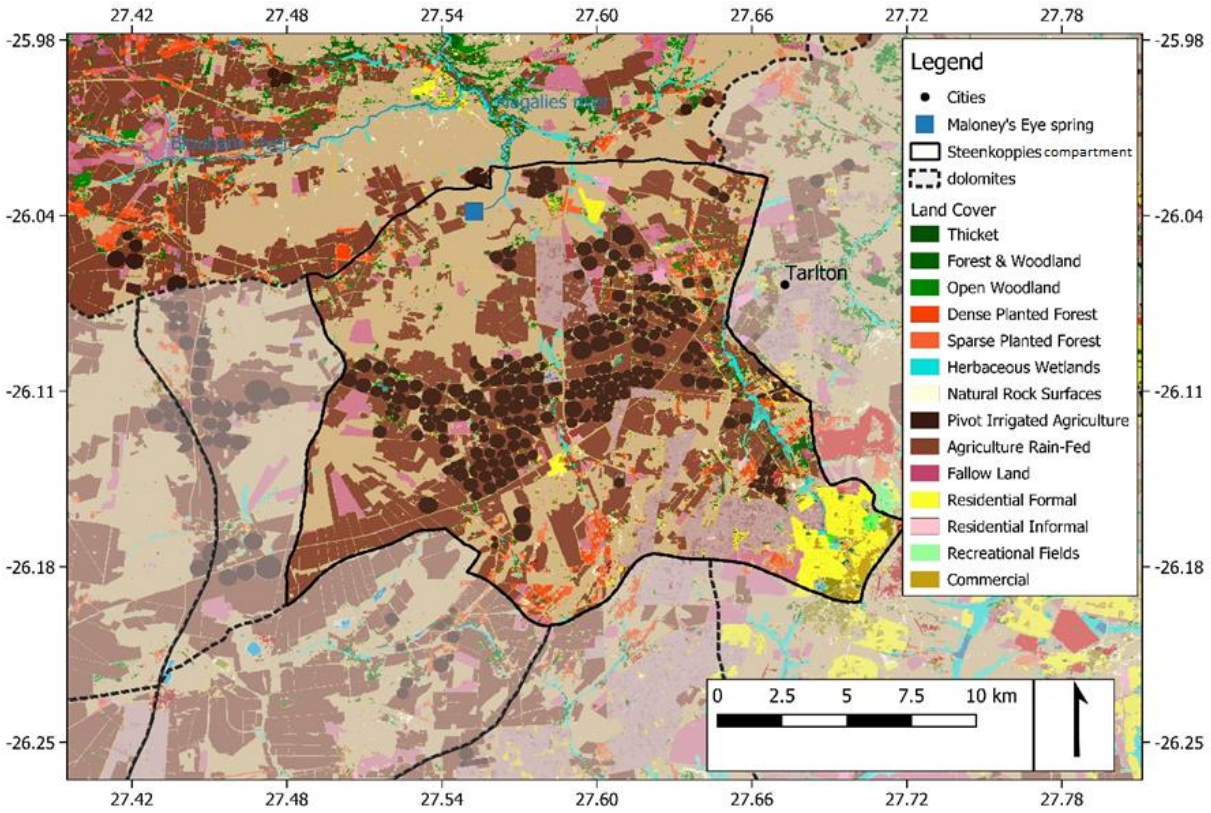


Figure 3-3: Landcover over the Steenkoppies compartment (shapefile from the South African National Land-cover data set of the South African Spatial Data Infrastructure 2018)

3.3 GEOLOGY

The stratigraphy of the Steenkoppies compartment is outlined in Table 3-1. The Steenkoppies compartment comprises of the north to north-west dipping rocks of the Chuniespoort Group and Malmani Subgroup. The Malmani Subgroup is subdivided into chert-rich and chert-poor formations of which the chert rich formations form the higher-yielding aquifers (Department of Water and Sanitation, 2009). The conglomerates and shales of the Rooihoogte and Timeball Hill Formations overlay the Malmani Dolomites and form the Northern boundary of the Steenkoppies compartment. The quartzites and shales of the Black Reef Formation lie beneath the Malmani dolomites and form the southern border of the Steenkoppies compartment. Mafic dykes intruded into the Malmani dolomites, the Tarlton East and Tarlton West Dyke, and create the eastern boundary of the compartment while the Eigendom Dyke forms the western border (Vahrmeijer *et al.* 2013). The Wolwerkrans dyke intruded across the centre of the compartment with an east-west trend, however, is not thought to be a substantial barrier to groundwater flow. (Holland *et al.*, 2009).

Table 3-1: Stratigraphy of the Gauteng and North West dolomites (Source: Vahrmeijer *et al.* 2013)

Super Group	Group	Formation	Thickness	Lithology
Transvaal	Pretoria	Rayton	120	Shale, quartzite
		Magaliesburg	300	Quartzite
		Silverton	600	Shale
		Daspoort	80-95	Quartzite
		Strubenkop	105-120	Slate
		Hekpoort	340-550	Andesite
		Timbal Hill	270-660	Shale, Diamictite, Klapperkop Quartzite and ferruginous Quartzite
		Rooigate	10-150	Quartzite, Shale, Bevets Conglomerate Member and Breccia
	Chuniespoort	Frisco	30-158	Chert-free dolomite with some primary limestone oolitic bands Chert increases to the top
		Fccles	490	Chert-rich dark dolomite with stromatolitic and oolitic bands. Chet increases to the top
		Lyttelton	220-290	Chert-free dark dolomite with large stromatolites and sometimes with wad
		Monte Christo	740	Alternate layers of chert-rich and chert-poor light coloured dolomite with stromatolites and oolites
		Oaktree	190-330	Chert-poor dark dolomite with interbedded layers of carbonaceous shale at the base
		Black reef	11-30	Shale and Quartzite. Arkosic Grit
Wiwatersrand	Central Rand	-	2880	Aranaceous, rudaceous rocks
	West Rand		5150	Quartzite, reddish and ferruginous magnetic shales
	Dominion		?	Quartzite, conglomerate, shale interbedded lava
Basement Complex				

3.4 CLIMATE

South Africa is a semi-arid country, with highly seasonal and unevenly distributed rainfall (Vahrmeijer *et al.*, 2013). The climate in the Steenkoppies compartment area is characterised by cool, dry winters and warm, wet summers with 80% of the rainfall occurring as thunderstorms (Seyler *et al.*, 2016). This type of climate is typical of the South African highveld.

3.4.1 Rainfall

Figure 3-4 presents annual rainfall for the Steenkoppies compartment according to data from meteorological stations maintained by the Agricultural Research Council (ARC) and the South African Weather Service (SAWS). A two-year moving average of the annual rainfall (grey dotted line) and the mean annual precipitation (MAP, red dotted line) is also presented (Figure 3-4). The annual rainfall recorded in the Steenkoppies compartment from 1909 to 2019 was the lowest in 1935 (323,2 mm) and the highest in 1929 (1081,3 mm) with a MAP of 668,57 mm.

According to Vahrmeijer *et al.* (2013), the Steenkoppies region experienced two periods of meteorological drought from 1990-1994 and again from 2002-2007. These two drought events are apparent in the historical rainfall time series, with the moving average of the rainfall time series well below the MAP during these periods (Figure 3-4). Table 3-2 compares the MAP to four individual years of annual rainfall. The annual rainfall in 1994 was 87% of the MAP while in 2004 the annual rainfall was only 73% which confirms the drought events described by Vahrmeijer *et al.* (2013). Table 3-1 further illustrates that there was another potential drought between 2015 - 2017 as the annual rainfall recorded in 2015 and 2017 was 71% and 63% of the MAP, respectively.

There appears to be a cyclical pattern in the rainfall recorded in the Steenkoppies compartment. Consecutive years of higher rainfall, above or near to the MAP, are followed by years of lower rainfall, below the MAP. The green and blue arrows demonstrate an example of high rainfall years and low rainfall years, respectively (Figure 3-4). Between 1966-1981 the Steenkoppies compartment received consecutive years of higher rainfall. During this time only three years recorded rainfall slightly below the MAP and twelve years of rainfall is above the MAP. From 1982-1986 the compartment received successive years of lower rainfall as all the years recorded annual rainfall below the MAP. The cycle was repeated with high rainfall between 1987-1989 and lower rainfall between 1990-1994. The moving average helps to visualise the cyclical pattern in the time series as it represents a smoothed out rainfall curve. When the moving average is

above the MAP one can distinguish years of higher rainfall and when it drops below the MAP the years of lower rainfall can be seen. The cyclical pattern does not repeat at fixed intervals throughout the entire time series. From the moving average curve and the annual rainfall in Figure 3-4,

Table 3-3 was formed outlining the cycles of higher and lower rainfall years throughout the time series. Figure 3-5 shows the monthly rainfall in the Steenkoppies aquifer (grey bars) with the trend of the monthly rainfall (blue line). In the last ten years, there has been a decline in the number of monthly rainfall peaks (Figure 3-5). Only nine months recorded rainfall above 100 mm in the last five years (2014 – 2019), compared to 16 months of rainfall above 100 mm the five years before (2009-2014). This is worth mentioning as groundwater in the Steenkoppies aquifer is recharged (signified as an increase in groundwater levels) from rainfall when more than 100 mm rain falls in a month (further explained in 3.7.1).

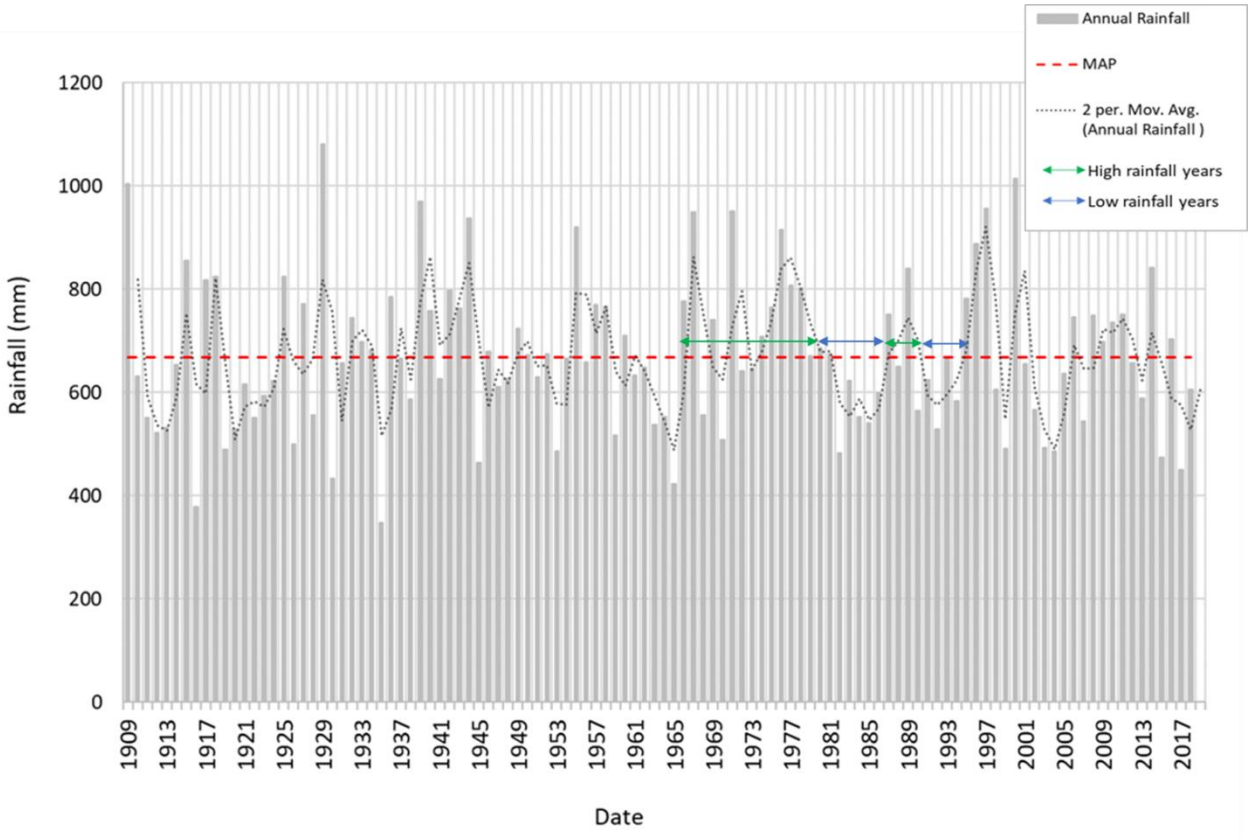


Figure 3-4: Annual rainfall in the Steenkoppies compartment (grey bars). A two-year moving average (grey dotted line), and the Mean Annual Precipitation (MAP) (the red dotted line) are also represented—An example of high and low rainfall years are represented as the green and blue arrows, respectively.

Table 3-2: Annual rainfall compared to the MAP in the Steenkoppies aquifer

Year	Annual rainfall (mm)	% of MAP
MAP (1908-2019)	668	100
1994	583	87
2004	486	73
2015	474	71
2017	450	63

Table 3-3: Cycles of higher and lower rainfall in the Steenkoppies compartment

Date	Part of rainfall cycle	Years
1910-1913	low rainfall	3
1914-1918	high rainfall	4
1919-1924	low rainfall	5
1925-1929	high rainfall	4
1930-1935	low rainfall	5
1936-1944	high rainfall	8
1945-1953	low rainfall	8
1954-1958	high rainfall	4
1959-1965	low rainfall	6
1966-1981	high rainfall	15
1982-1986	low rainfall	4
1987-1989	high rainfall	2
1990-1994	low rainfall	4
1995-1997	high rainfall	2
1998-1999	low rainfall	1
2000-2001	high rainfall	1
2002-2004	low rainfall	2
2005-2012	high rainfall	7
2013-2018	low rainfall	5

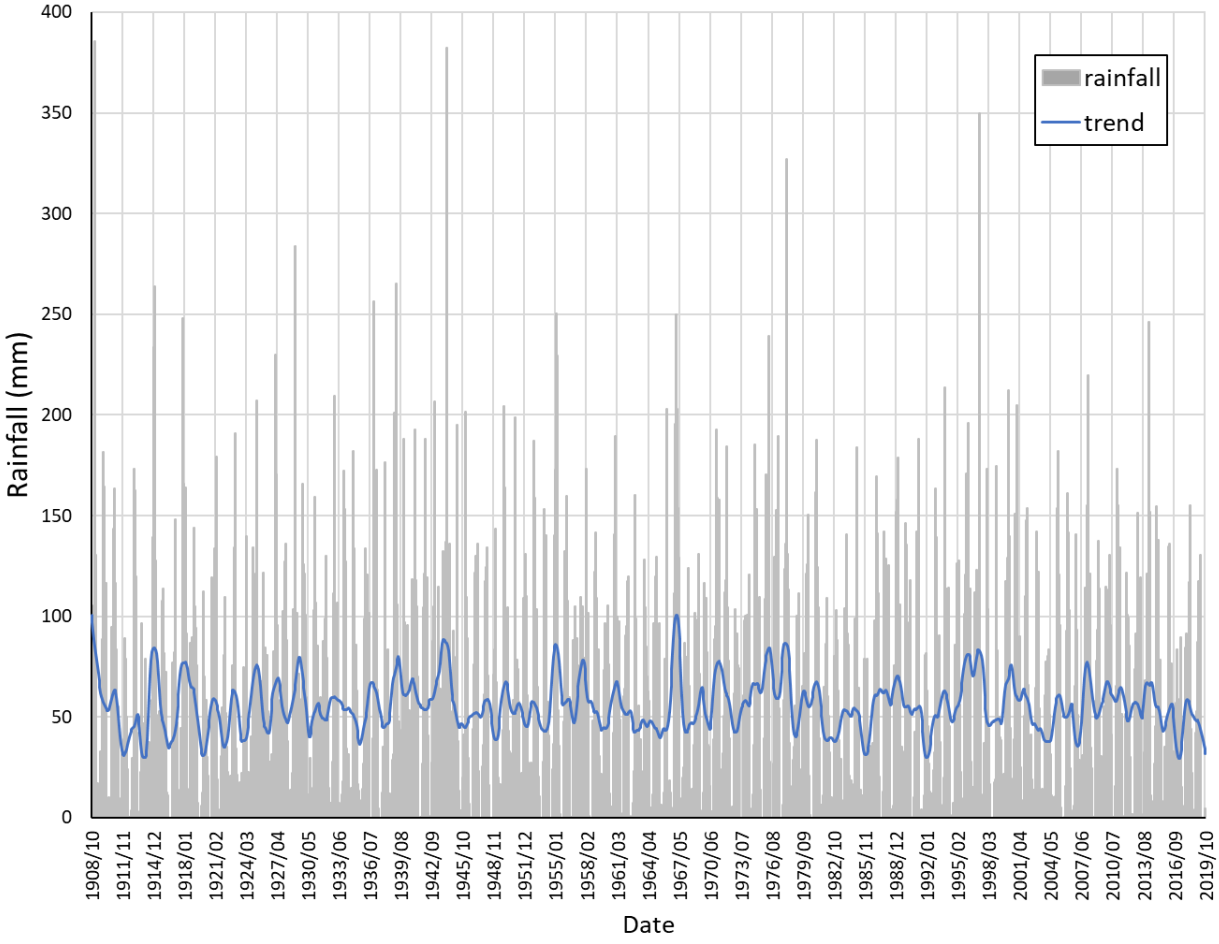


Figure 3-5: Monthly rainfall for the Steenkoppies aquifer (grey bars) and the trend of the monthly rainfall time series (blue line).

3.4.2 Temperature

The temperature in the Steenkoppies compartment has followed a consistent seasonal trend (). In the spring and summer months (September - March) the mean monthly temperature fluctuated between 15°C - 24°C and in the autumn to winter months (April - August) fluctuated between 4°C - 19°C. This is according to data collected from a meteorological station in the Steeenkoppies compartment maintained by the ARC (Figure **3-6**). The twelve-month moving average in Figure **3-6** shows that the monthly average temperature increased from 1950 to 2019. Higher temperatures will cause higher evapotranspiration rates and lower soil water content (Döll, 2009). This may result in less water recharging to the aquifer and thus reduction in groundwater storage.

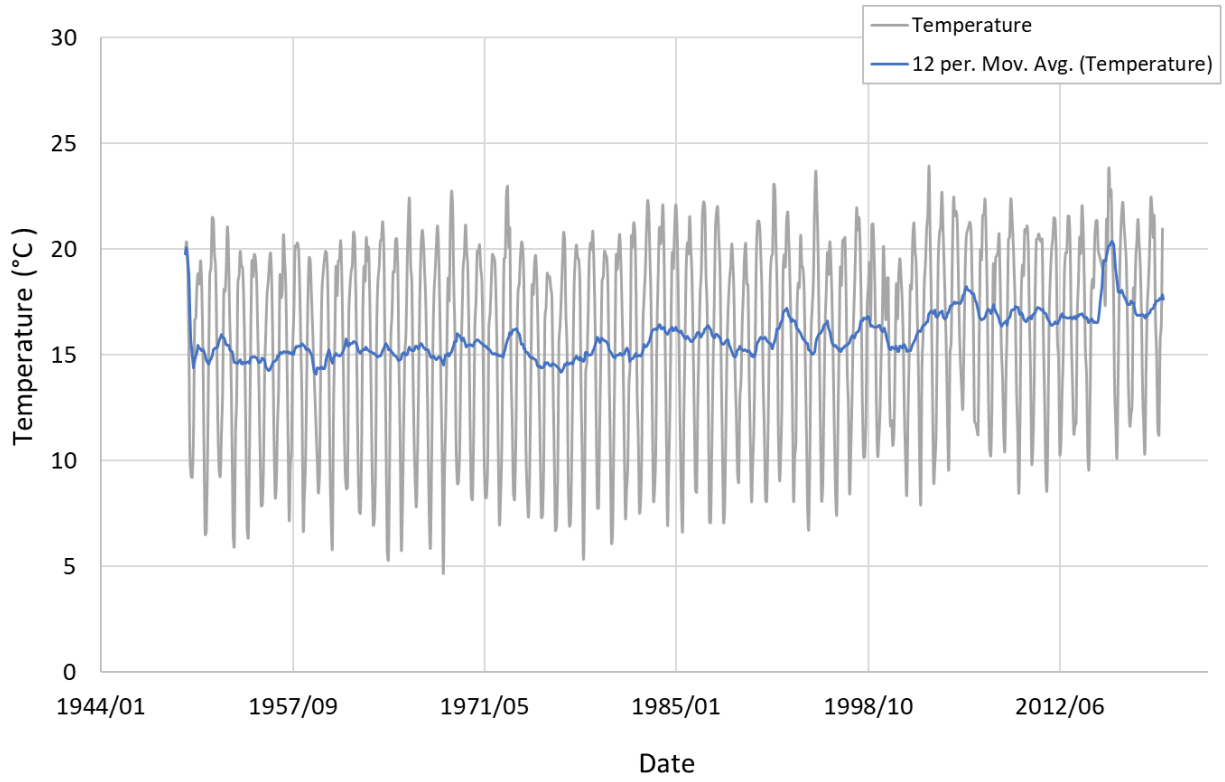


Figure 3-6: Mean monthly temperature in the Steenkoppies compartment (grey line) and the twelve month moving average (blue line).

3.5 HYDROLOGY AND SPRING FLOW

The Steenkoppies compartment lies within the quaternary drainage region A21F. The perennial spring, Maloney's Eye is situated 750 m south of the northern boundary of the Steenkoppies compartment at 1495 masl (Figure 3-2). The Maloney's Eye is located at the intersection of an east-west striking fault zone in the Malmani Subgroup that is overlain by geology of lower permeability, forcing the groundwater to the surface. The Maloney's Eye serves as the only natural significant outlet for the groundwater in the Steenkoppies compartment (Holland *et al.*, 2009). Maloney's Eye flows out into the Magalies River that feeds the Hartbeespoort Dam. Between 1908-2017 the mean monthly discharge from Maloney's Eye spring varies from a low of 0.051 cubic centimetres per second (m^3/s) recorded in April 2007 to a high of 1.035 m^3/s recorded in March 1979 (Figure 3-7).

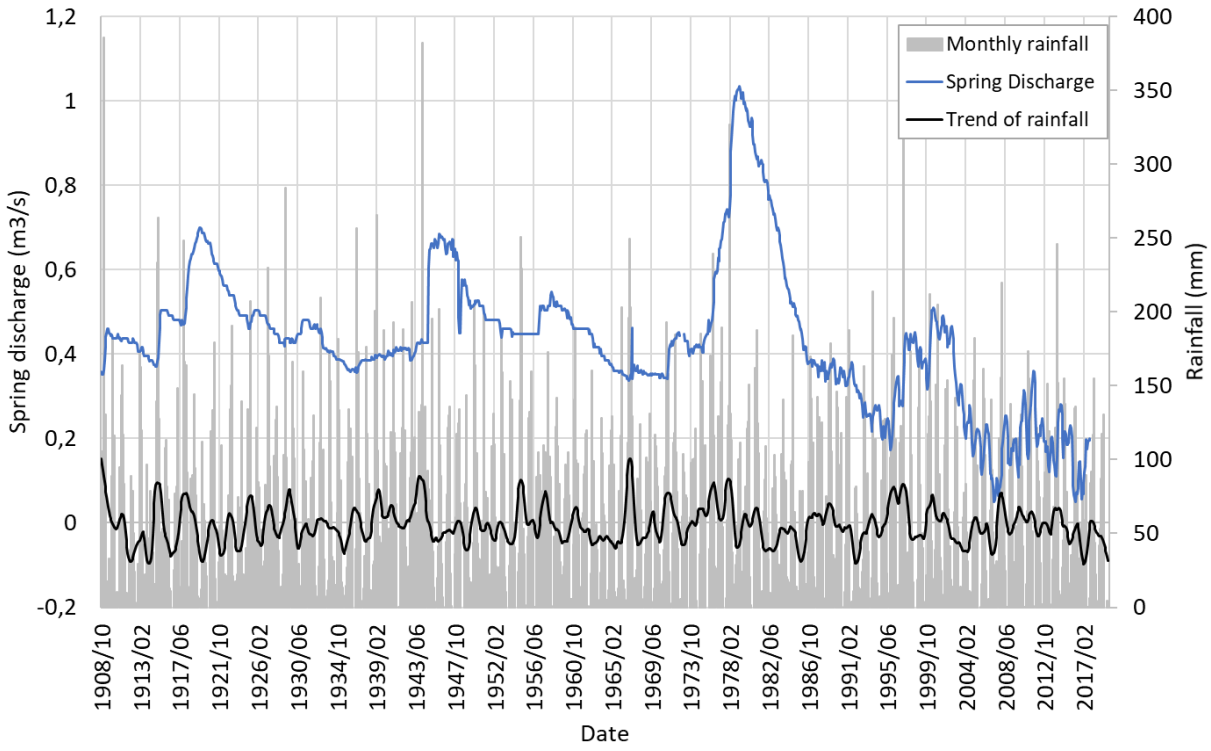


Figure 3-7: Mean monthly discharge recorded at the Maloney's Eye Spring (left axis) and monthly rainfall with the trend (right axis) over time.

There are six significant peaks in the spring discharge time series in 1919, 1945, 1979, 1997, 2001 and 2011 (Figure 3-7). The peaks are a delayed, response to rainfall (Figure 3-7 and Figure 3-8). The spring discharge peaks coincide with the end of high rainfall cycles (Figure 3-8 and

Table 3-3). The longer the period of high rainfall, the larger the peak in the spring discharge. For example, the highest average spring discharge peak was recorded in 1979 (1.01 m3/s) towards the end of 15 years of high rainfall between 1966-1981 (Figure 3-8). In 1945, the average spring discharge peaked at a lower rate (0.67 m3/s) after eight years of high rainfall recorded from 1936-1944 (Figure 3-8).

The spring discharge decreases substantially from 1998 (Figure 3-8). Two peaks after 1987 (seen in 2001 and 2011) are of a much lower magnitude compared to the peaks recorded pre 1987 (Figure 3-8). The last peak recorded in the average spring discharge in 2011 was very low (0.28 m^3/s) although it follows seven years of high rainfall between 2005 and 2012. It is uncertain whether this peak was related to rainfall or associated with another external factor acting on the spring discharge, such as groundwater abstraction causing a decrease in discharge. Vahrmeijer *et al.* (2013), used the cumulative rainfall departure (CRD) method to evaluate the relationship between rainfall and spring discharge and confirmed that there is a good connection between the two. However, in 1987 the actual measured discharge was lower than the simulated discharge from the CRD, indicating that other external factors not accounted for (such as abstraction) influenced the discharge from 1987. This explains the lower discharge rates recorded after 1987.

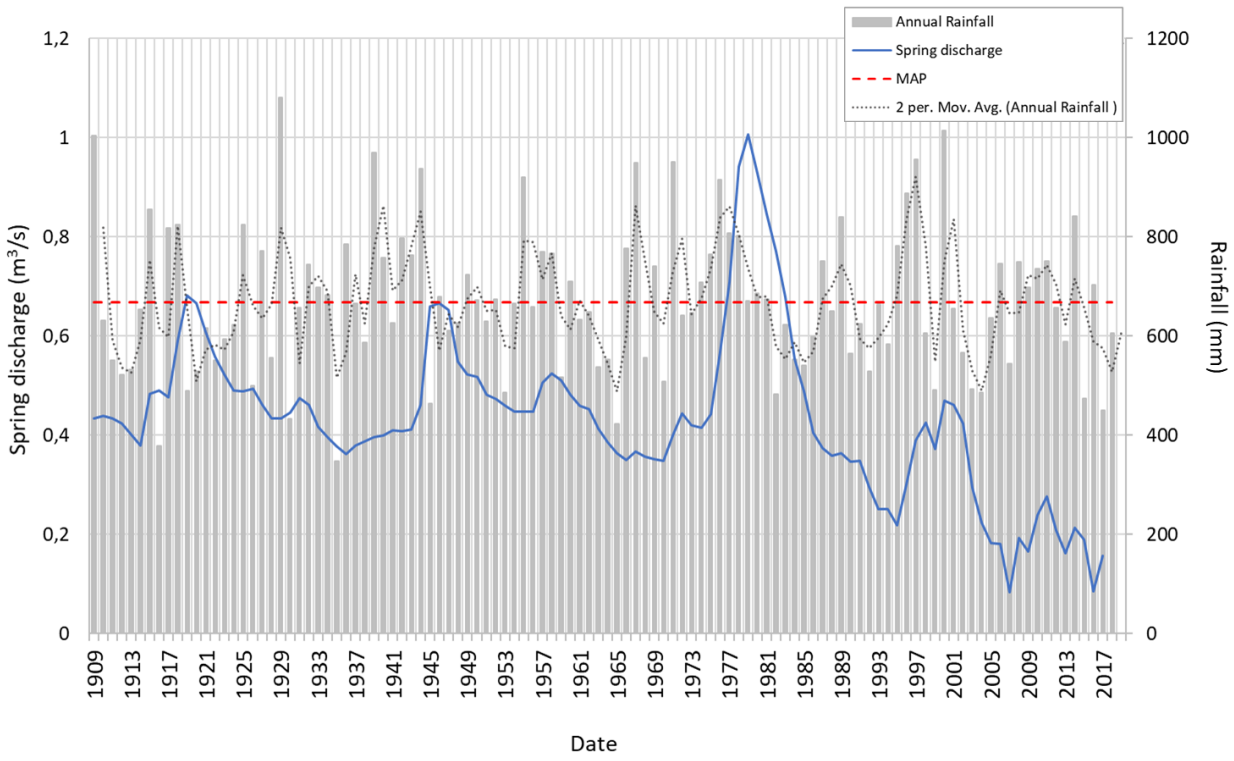


Figure 3-8: Mean annual spring discharge from the Maloney's Eye, represented as the blue line (left axis). Mean annual rainfall represented as the grey bars with a two-year moving average shown as the grey dotted line, and the Mean Annual Precipitation (MAP) represented as the red dotted line (right axis).

3.6 GROUNDWATER LEVELS AND FLOW DIRECTION

For this study, historical groundwater level data was requested from the National Groundwater Archive (NGA) and HYDSTRA databases which are maintained by the Department of Water and Sanitation (DWS). The groundwater level is measured in meters above ground level (magl) and as such, represented as a negative value. Majority of the boreholes in Steenkoppies compartment are situated in the north-west of the compartment. Fewer boreholes are located in the north-east and south-east of the compartment and even less boreholes in the south-west of the compartment (Figure 3-9).

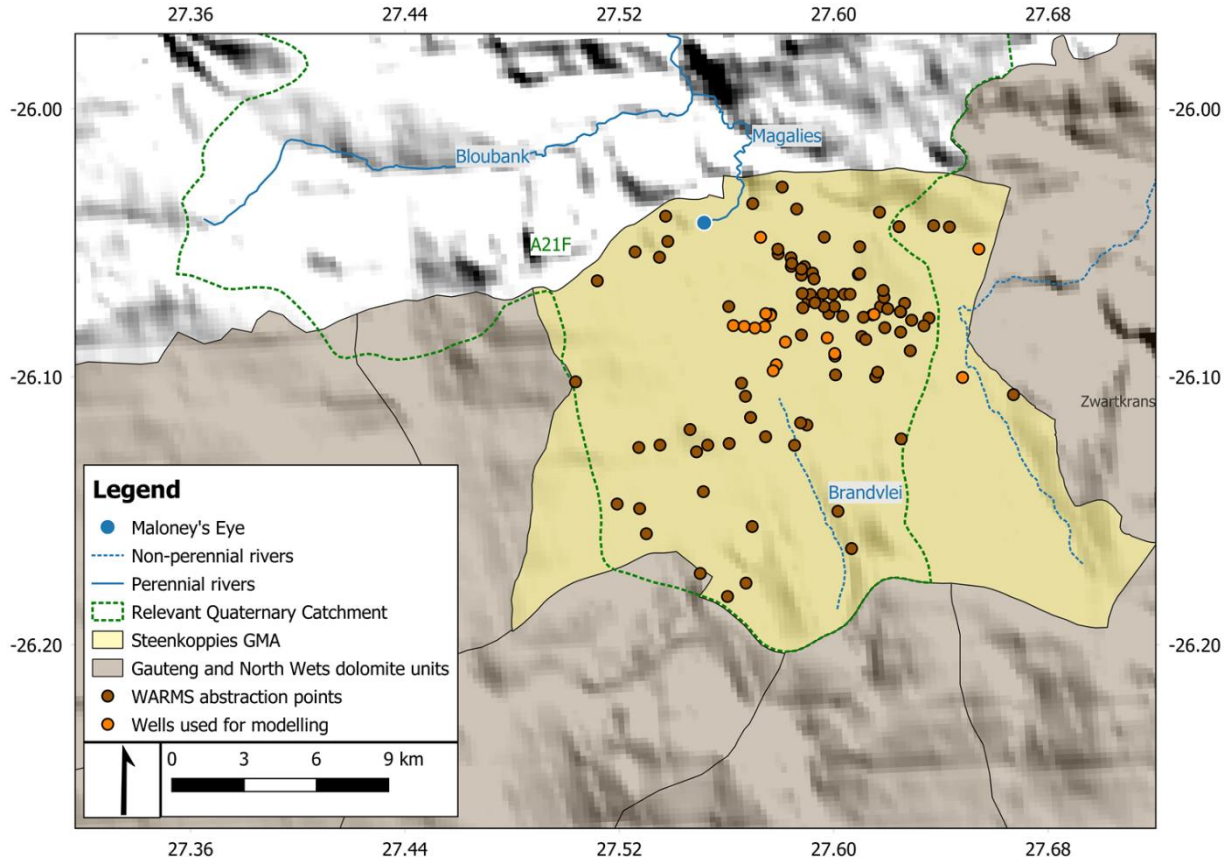


Figure 3-9: Position of the boreholes in the Steenkoppies aquifer (Source of data: WARMS)

Monthly groundwater level time series recorded from boreholes in the Steenkoppies compartment are shown in Figure 3-10. The groundwater levels in the Steenkoppies compartment range from -82,12 meters above ground level (magl) to -28.76 magl. Although each groundwater level time series increases and decreases at different magnitudes, a similar trend and timing of the groundwater level increases and decreases are recorded throughout the compartment (Figure 3-10).

The groundwater levels show an increase from approximately 1984 to 1991 after which the groundwater levels show a decrease until 1994. The groundwater levels once again increase from 1994 to 2001, with a slight decline seen in between this time at around 1999. The groundwater levels decrease another time until 2007, increase until 2011 and then show a decrease up until the end of the time series (Figure 3-10). These increases and decreases in the groundwater level time series coincide with the rainfall cycles outlined in Figure 3-4 and Table 3-3. For visual clarity, only six groundwater level time series are shown with the monthly rainfall and the monthly rainfall trend (Figure 3-11). The peaks and the troughs in the groundwater levels are a slightly delayed and muted version of the monthly rainfall (Figure 3-11).

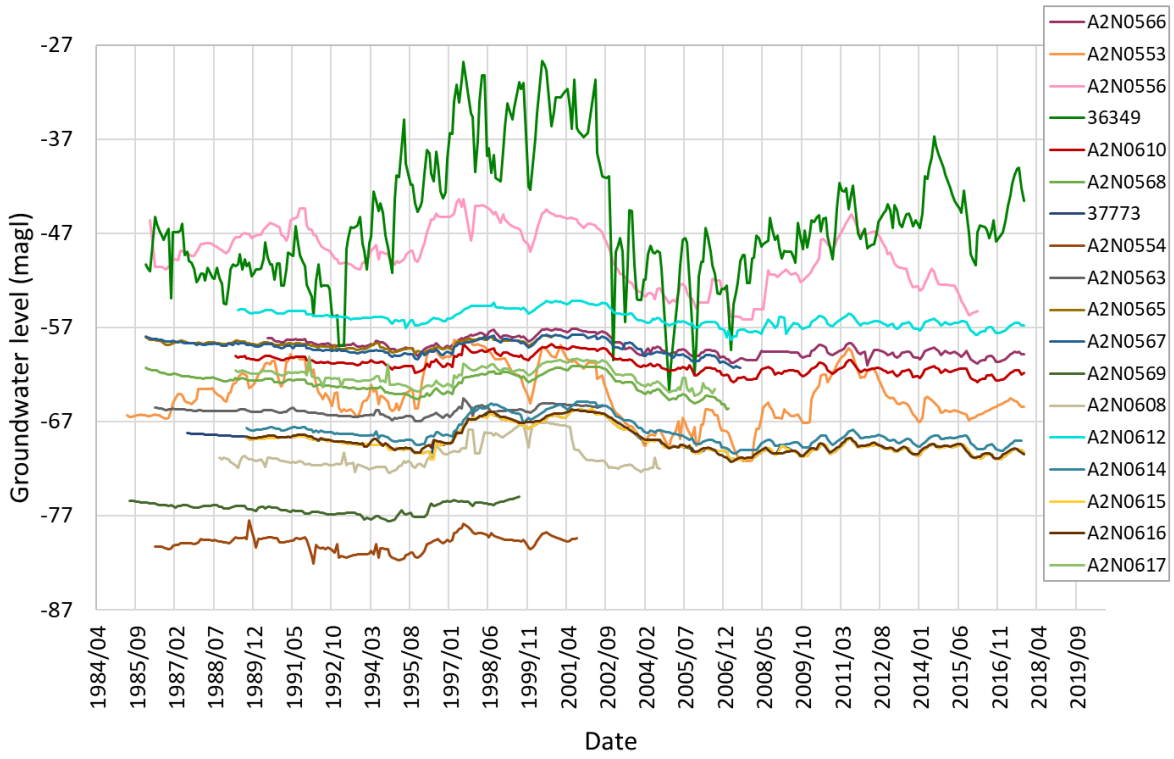


Figure 3-10: Monthly groundwater levels recorded from boreholes in the Steenkoppies compartment.

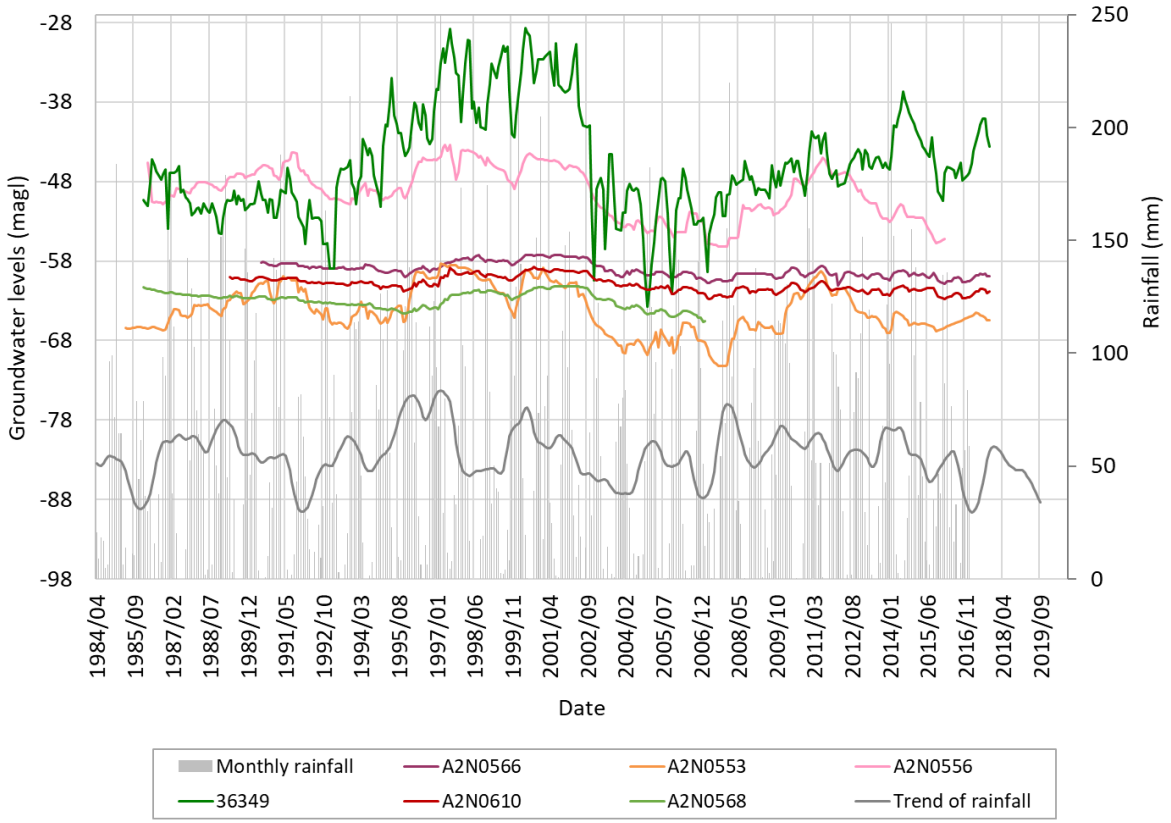


Figure 3-11: Monthly groundwater level (left axis) and monthly rainfall with the trend (right axis) over time.

Figure 3-12 shows the six groundwater level time series with the monthly mean spring discharge from the Maloney's Eye. In a conceptual hydrogeological model Holland *et al.* (2009) described the flow of groundwater in the Steenkoppies compartment as towards the north to discharge at the Maloney's Eye. The Maloney's Eye spring is the only natural outlet for the groundwater in the Steenkoppies compartment. Since the hydraulic gradient between the surrounding aquifer and the Maloney's Eye is what drives spring discharge, groundwater level will directly translate to changes in the spring discharge rate. This is also confirmed in the historical time series (Figure 3-12). The pattern of the spring discharge time series is very similar to the groundwater level time series (Figure 3-12). The spring discharge records significant peaks in 1997, 2001 and 2011, which coincides with the increases in the groundwater level time series. The spring discharge represents a response to the fluctuations in the groundwater levels throughout the compartment.

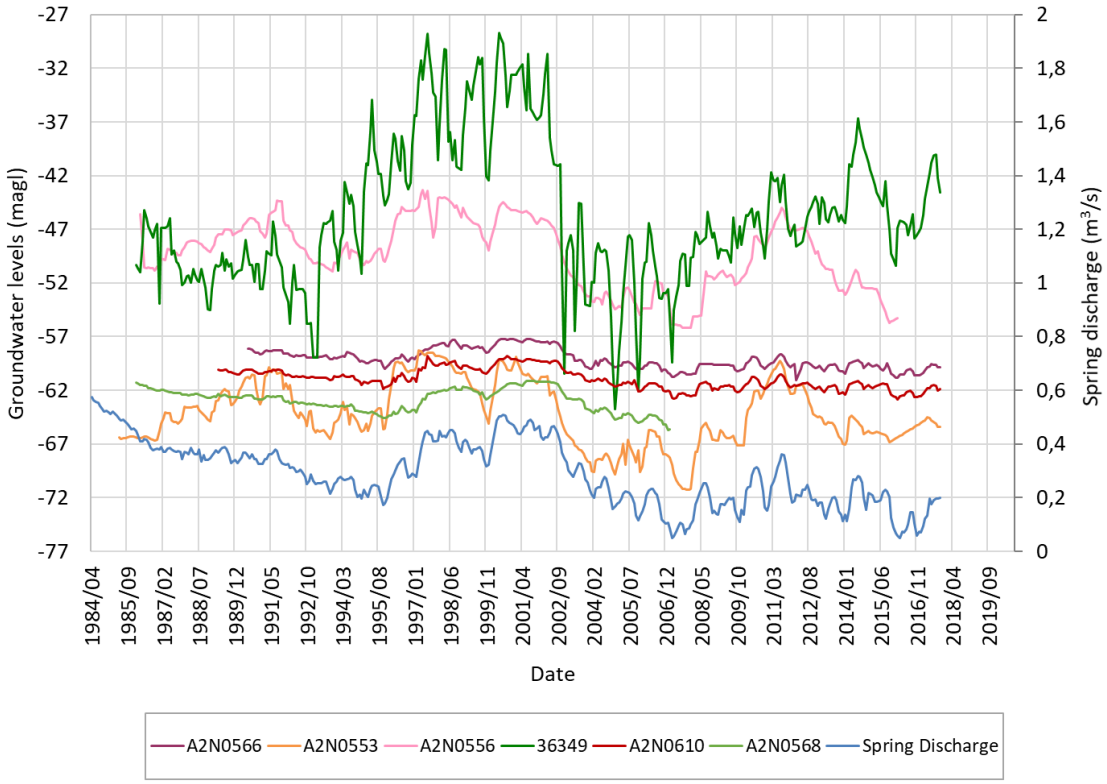


Figure 3-12: Monthly groundwater level (left axis) and monthly spring discharge (right axis) over time.

Table 3-4 shows the summary of the groundwater level data sets used for modelling in the Steenkoppies compartment. Table 3-4 describes a summary of the rainfall and spring discharge data sets. The majority (13 out of the 18 boreholes) record declining groundwater levels. The spring discharge and rainfall time series also record declining groundwater levels. Since groundwater levels are a muted response to rainfall, and spring discharge is a muted response to groundwater levels, the decrease in the rainfall could be the cause of the decline in the groundwater levels and spring discharge. However, declining rainfall may not be the only variable contributing to declining groundwater levels and spring discharge in the Steenkoppies aquifer. As described in section 3.5, the study using the CRD method by Vahrmeijer *et al.* (2013) supports the idea that there may be external factors such as groundwater abstraction influencing the groundwater levels in the Steenkoppies aquifer.

Table 3-4: Summary of the groundwater level time series in the Steenkoppies (Ground water levels (GWL), Meters above ground level (magl)).

Borehole ID	Latitude	Longitude	Start date (yyyy/mm)	Start GWL (magl)	End date	End GWL (magl)	Years Active	Trend	Trend Gradient	Annual increase (m)
A2N0566	-26.07665	27.57651	1990/06	-58.16	2017/10	-59.88	27	decline	-0.0002	-0.06
A2N0553	-25.90942	25.89306	1985/05	-66.37	2017/10	-65.4	32	incline	-0.0002	0.03
A2N0610	-26.08706	27.58199	1989/04	-60.05	2017/10	-61.85	28	decline	-0.0002	-0.06
A2N0610	-26.05231	27.65417	1986/03	-45.60	2016/02	-55.26	29	decline	-0.0005	-0.33
36349	-26.10025	27.64812	1986/01	-50.3	2017/10	-43.59	31	incline	0.0002	0.22
A2N0617	-26.09555	27.578611	1989/04	-61.61	2006/08	-63.55	17	decline	<-0.0005	-0.11
A2N0612	-26.08129	27.57433	1989/09	-55.14	2017/10	-56.81	28	decline	-0.0002	-0.06
A2N0616	-26.08089	27.56255	1989/09	-68.56	2017/10	-70.5	17	decline	-0.0002	-0.11
A2N0615	-26.07671	27.61508	1989/09	-68.65	2017/10	-70.32	28	decline	-0.0002	-0.06
A2N0614	-26.0818	27.57062	1989/09	-67.73	2017/09	-69.08	27	decline	-0.0002	-0.05
A2N0608	-26.08549	27.5976	1989/09	-70.81	2005/08	-72	15	decline	0.0003	-0.08
A2N0568	-26.09776	27.57742	1986/01	-61.29	2007/02	-65.61	21	decline	-0.0002	-0.21
A2N0567	-26.07711	27.57643	1986/01	-57.99	2007/07	-61.33	21	decline	-0.0001	-0.16
37773	-26.08122	27.56661	1987/07	-68.26	2017/10	-70.32	30	decline	-0.0002	-0.07
A2N0554	-26.05302	27.62917	1986/05	-80.34	2001/08	-79.39	15	incline	< 0.0003	0.06
A2N0563	-26.04796	27.57273	1986/05	-65.51	2002/07	-65.32	16	incline	< 0.0003	0.01
A2N0565	-26.07643	27.57467	1986/01	-58.02	2000/09	-58.02	14	decline	< -0.0005	-
A2N0569	-26.09134	27.60041	1986/06	-75.44	2000/07	-75.01	14	incline	< 0.0003	0.03
Rainfall	-	-	1908/10	105,6 mm	2019/10	0,25 mm	111	decline	< -0.0005	-3,59 mm
Spring discharge	-26.02236	27.56336	1908/10	0,36 m3/s	2017/10	0,2 m3/s	99	decline	<-0.0005	-0,002 m3/s

Table 3-5: A summary of the monthly rainfall and spring discharge (Ground water levels (GWL)).

Borehole ID	Latitude	Longitude	Start date (yyyy/mm)	Start GWL	End date	End GWL	Years Active	Trend	Trend Gradient	Annual increase
Rainfall	-	-	1908/10	105,6 mm	2019/10	0,25 mm	111	decline	< -0.0005	-3,59 mm
Spring discharge	-26.02236	27.56336	1908/10	0,36 m3/s	2017/10	0,2 m3/s	99	decline	<-0.0005	-0,002 m3/s

3.7 SOURCES AND SINKS

3.7.1 Direct Recharge

The primary source of recharge to the groundwater in the Steenkoppies aquifer is from rainfall (Cobbing *et al.*, 2016). Recharge for the Steenkoppies compartment ranges from 9% to 21% of the mean annual precipitation (Holland *et al*., 2009). The graphs shown above demonstrate that groundwater levels are influenced and recharged by rainfall (e.g. Figure 3-11). Another source of recharge in the Steenkoppies compartment is the effluent discharge from the Randfontein Sewage Works, and return flow from irrigation (Cobbing *et al.*, 2016).

Vahrmeijer *et al.* (2013) describe how there is an exponential relationship between rainfall and recharge to the aquifer in the Steenkoppies compartment. Moderate recharge occurs during average rainfall events; however, during increased rainfall events, the recharge to the aquifer increases disproportionately to the increase in rainfall. This makes sense as sufficient soil moisture would already be acquired with sizeable monthly rain, increasing the hydraulic conductivity and maximising infiltration of rainfall to the aquifer (Vahrmeijer *et al.*, 2013). By analysing the data closely, it was found that monthly rainfall greater than 100 mm result in significant recharge to the aquifer (increase in groundwater levels).

3.7.2 Abstraction

The Water use Authorisation and Registration Management System (WARMS) data set recorded a groundwater usage total of 78,28 million cubic meters per annum (million m^3/a) from registered users in the Steenkoppies compartment (Table 3-6). The dominant water use sector is agricultural for both irrigation and livestock irrigation using a total of 39.36 and 37.03 million m^3/a, respectively (Table 3-6). This makes up 76% of the total registered groundwater use in the Steenkoppies aquifer. It should be utlined that Table 3-6 for the volume of registered groundwater use for livestock watering is not realistic and could be a mistake from the WARMS records.

Table 3-6: Registered groundwater use on the Steenkoppies compartment (WARMS, 2019)

Water use sector	Number registrations	Groundwater use (million m3/a)	Percentage of total usage
Agriculture (irrigation)	180	30,81	39,36
Agriculture (watering livestock)	12	28,99	37,03
Mining	1	0,06	0,08
Industry (urban)	3	11,07	14,14
Industry (non-urban)	3	7,35	9,39
total	199	78,28	100,00

Groundwater use data was gathered from various literature, reports and surveys by Seyler et al. (2016) (Table 3-7 and Figure 3-13). The data set from literature compiled by Seyler et al. (2016) reported slightly higher groundwater usage in the Steenkoppies compartment compared to the WARMS data. A sharp increase is recorded in the WARMS data set from 9.75 to 19.05 million cubic meters per annum (million m^3/a) in 1999, and the literature data set shows an increase in groundwater abstraction from 13.45 to 25.55 million m^3/a in 1996 (Figure 3-13). The groundwater usage data from literature also indicates a rise in groundwater usage in 1986 from 3.95 to 13.45 million m^3/a.

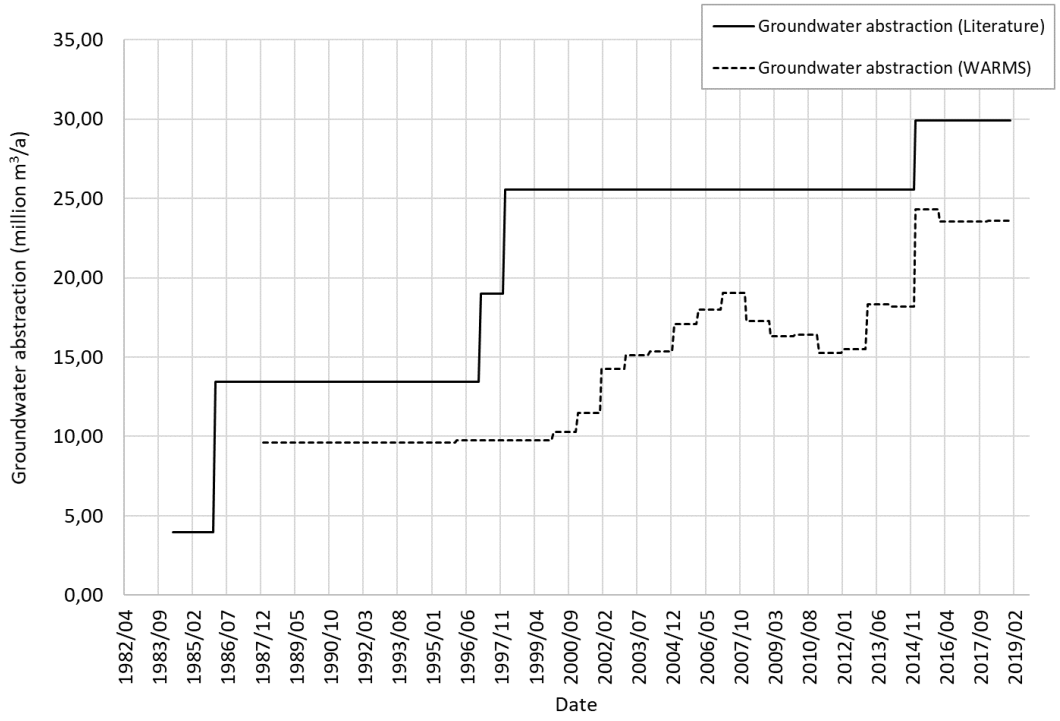


Figure 3-13: Groundwater usage in the Steenkoppies compartment. The dotted line represents registered groundwater (WARMS, 2019), the solid line represents groundwater usage from literature (Seyler et al., 2019).

Table 3-7: Groundwater usage in the Steenkoppies aquifer estimated from literature (Seyler et al., 2019)

Reference	Year	Abstraction (million m3/a)
Hobbs (1980)	1980	3,95
Bredenkamp et al (1986)	1986	13,45
Barnard (1997)	1997	19
Schoeman and Partners (2016)	1998	25,55
Schoeman and Partners (2016)	2015	29,92

The incline in groundwater usage between 1986 and 1996 coincides with a decline in spring discharge (Figure 3-14). As mentioned, Vahrmeijer *et al.* (2013) found that in 1987 the actual measured spring discharge was lower than the simulated discharge (using the CRD method). This is just one year after the first peak in groundwater usage in 1986, indicating that there are other external factors such as abstraction influencing the discharge. The decrease in spring discharge can be attributed to a delayed impact of groundwater usage (Figure 3-14).

In section 3.5 it was established that spring discharge in the Maloney's Eye spring is a muted response to groundwater levels in the Steenkoppies compartment. Therefore, if an impact of abstraction on spring discharge is recognised, then groundwater usage also has an impact on groundwater levels. A few years after the two peaks in groundwater usage, the groundwater levels begin to decline in both 1989 and 2000 confirming that the groundwater usage has a delayed influence on the groundwater levels in the Steenkoppies compartment (Figure 3-15).

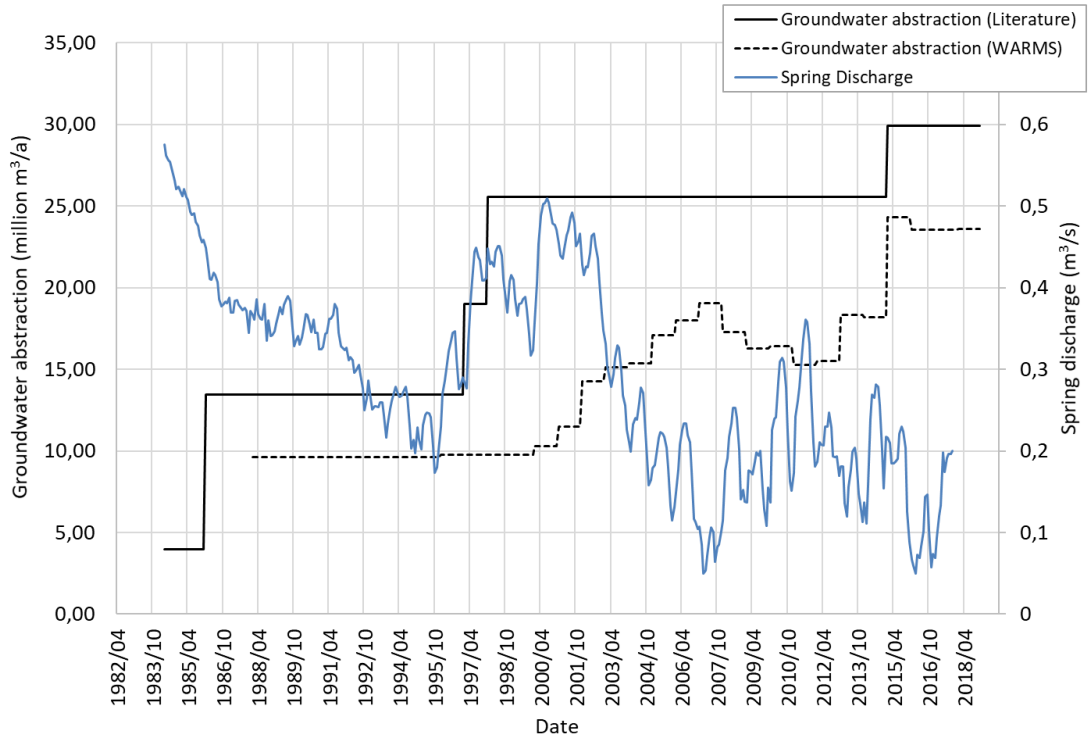


Figure 3-14: Monthly spring discharge (left axis) and groundwater usage in the Steenkoppies compartment (right axis).

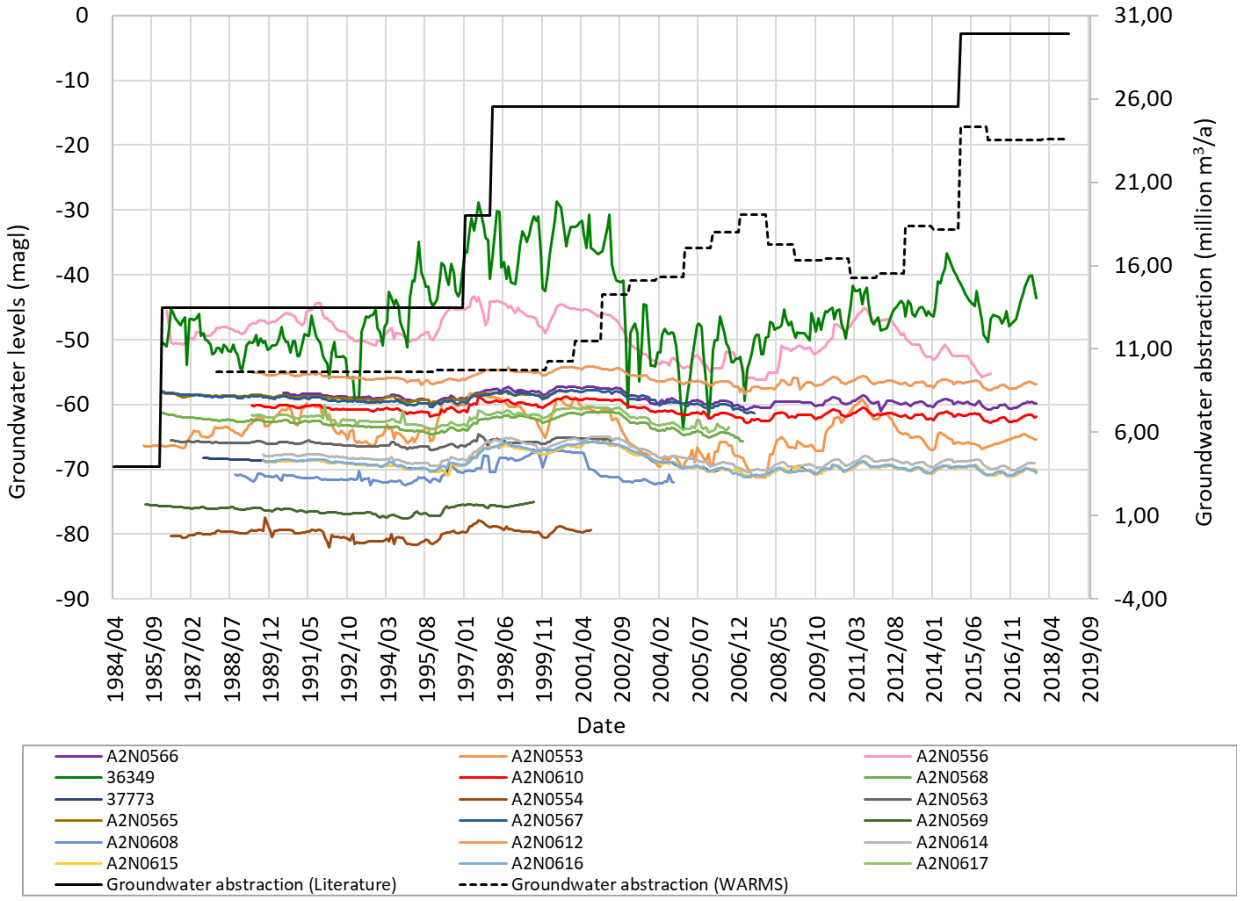


Figure 3-15: Monthly groundwater levels (left axis) and groundwater usage in the Steenkoppies compartment (right axis).

CHAPTER 4:
METHODOLOGY

4.1 MODEL INPUT VARIABLES

This study makes use of the relationship between the target variable, and the input variables to assist predictions. Suitable input parameters were chosen based on the data available and on how related and how much influence the input variables have on the target variable, groundwater levels.

The hydrogeological knowledge described in Section 2.2 and the knowledge of the Steenkoppies compartment outlined in Chapter 3 describe how groundwater levels in the Steenkoppies compartment are influenced by groundwater recharge, evapotranspiration, natural and artificial discharge. Rainfall, temperature, spring discharge and groundwater usage are used as substitutes for these parameters.

Other variables may influence groundwater levels (e.g. vertical interaction from neighbouring aquifer compartments or evapotranspiration). However, other parameters are difficult to measure and not monitored regularly. Historical time series for the chosen input variables at are readily available from government departments in South Africa at reasonable quality.

4.2 DATA ACQUISITION AND PREPROCESSING

Time series for the target and input variables in the study area were gathered from literature, organisations and government departments in South Africa. Data was requested for the maximum time frame available. This was to ensure the machine learning model could train on adequate data to capture the long-term relationships that exist between the groundwater levels and the input variables.

In machine learning, data pre-processing is a crucial step in which raw data is transformed or encoded to produce a preffered data set for modelling (García *et al.*, 2016). Measured data often has some errors or inconsistencies. Analysing data that has not been carefully screened for such issues can provide misleading results (Kotsiantis and Kanellopoulos, 2006). Some of the errors in a data set could have been due to human error, mistakes in the data collection process or imprecisions with the measuring equipment. It is imperative that each data set is filtered and the errors in the data are rectified so that the best possible data for the region is used as inputs to the

machine learning model. The IBM programmer George Fuechsel coined the term "garbage in garbage out", to which he references machine learning algorithms that simply process data they are fed (García *et al.*, 2016). Therefore, pre-processing is a vital first step to ensure the success of a machine learning model. A summary of the final data acquired for modelling is displayed at the end of the section 4.2 in Table 4-2.

4.2.1 Groundwater Levels

In the Steenkoppies aquifer, the NGA database contained historical groundwater levels from 166 boreholes, and the HYDSTRA database had historical groundwater level data sets from 36 boreholes. There was a large extent of duplication between the two databases. Both the databases included monthly groundwater level measurements over the period 1940-2019 with very infrequent weekly and daily measurements also recorded. Many boreholes had incomplete data sets with numerous missing values and some contained less than 10 years worth of data.

4.2.1.1 Data Deduplication

The NGA is the original database, containing groundwater level data from boreholes owned and monitored by both the DWS and consultancies. The NGA database became too large and could no longer handle the quantity of data. Thus, the HYDSTRA database was created, in the hopes that all the NGA data would be transferred to the new HYDSTRA database that could handle the large, growing quantity of data.

However, when the HYDSTRA database was created, the borehole IDs were changed and only included data from boreholes owned and monitored by the DWS and not those monitored by consultancies. Not all the information from the NGA was transferred to the HYDSTRA database. For this reason, groundwater levels from both the NGA and HYDSTRA databases were acquired, and hence a significant degree of duplication exists between the two data sets.

The coordinates of each of the boreholes were examined to identify duplicate groundwater level data sets. The groundwater level data for the Steenkoppies aquifer from the NGA and HYDSTRA databases were merged into one data frame, and analyses were performed on the combined data set.

In the merged data set, there were 202 unique borehole ID's; however, there were only 141 unique coordinates. This indicates that between the NGA and HYDSTRA databases, 61 boreholes had the same coordinates and are thus duplicates. The duplicate boreholes holding the

most comprehensive and complete data set were kept while the shorter data set, containing more missing values, were removed from the data set.

4.2.1.2 Boreholes Selected for Modelling

The final boreholes selected for modelling were based on three criteria; (i) the geology into which the boreholes are drilled, (ii) the time frame over which groundwater levels were recorded and (iii) the percentage missing data in the data set.

This study aims to model groundwater levels in the dolomites of the Steenkoppies aquifer compartment. It is, therefore, imperative to ensure that the boreholes chosen for modelling are measuring the groundwater levels in the Steenkoppies dolomite aquifer. The 1:1 000 000 scale geological map of the Republic of South Africa (published by the Council for Geoscience (CGS)), and shapefile showing the dolomitic compartments of the Gauteng and North West dolomite aquifer (published by Holland *et al.*, 2009), was uploaded to QGIS Geographic Information System (QGIS.org, 2020). The boreholes that were not situated in both the Steenkoppies compartment and Malmani dolomite lithology were removed from the data set.

The study used only the boreholes recording more than ten years of monthly groundwater level measurements so that the models were able to capture the long-term trends in the data, and for the machine learning models to train on sufficient data.

The last criteria were the percentage of missing data in the groundwater level data sets. Boreholes, with no more than 50% missing data and gaps no longer than five consecutive months, were selected for modelling. This ensured that the missing data could be interpolated realistically.

A summary of the pre-processing actions applied to the groundwater level data set and the number of boreholes remaining after each pre-processing action are presented in Table 4-1. After pre-processing, groundwater level time series from 18 boreholes in the Steenkoppies compartment were of suitable quality for modelling.

Table 4-1: Summary of pre-processing actions on GWL data set

GWL pre-processing procedure	Action	Number of boreholes after pre-processing action
1	Merged NGA & HYDSTRA	202
2	Duplicates removed	141
3	Dolomite lithology confirmed	85
4	Boreholes with > 10 years data	34
5	Boreholes < 50% missing months of data	24
6	Data gaps < 5 consecutive months	18

4.2.2 Spring Discharge

Historical spring discharge data were obtained from gauge-stations maintained by the DWS. In the Steenkoppies aquifer, gauge-station A2H010 measured the discharge of the Maloney's Eye spring (Figure 3-2). The gauge station recorded monthly average river flow in cubic centimetres per second (m^3/s). Gauge station A2H010 is no longer active today; however, it holds an extensive historical data set recording monthly discharge values from August 1908 to December 2017.

4.2.3 Rainfall

Historical monthly rainfall data was acquired from a study by Holland *et al.* (2009). Extensive historical rainfall data are not recorded from meteorological stations in the Steenkoppies. Holland *et al.* (2009) therefore, compiled a representative monthly rainfall time series from 1908 to 2009 from four meteorological stations near to the Steenkoppies compartment, maintained by the South African Weather Service (SAWS) and the Agricultural Research Council (ARC). All four stations showed a similar mean annual precipitation (MAP) where the deviation between the stations was less than 10%. Holland *et al.* (2009) compiled the time series by calculating the weighted average (squared inverse distance weighting method) of all the monthly precipitation. This time series was augmented with data from the Tarlton station on the Deodar farm situated in the Steenkoppies aquifer (Figure 3-2) to create a final rainfall data set from October 1908 to October 2019.

A paper by Xu *et al.* (2020) looked at the decomposed signals of rainfall as inputs for forecasting rainfall. Xu *et al.* (2020) found that the decomposed signals of rainfall produced better predictions compared to actual measured rainfall. A similar observation is seen in this study.

Preliminary model results showed a lack of influence of rainfall on groundwater levels. Therefore, Seasonal-Trend decomposition using LOESS (STL) time series decomposition (Equation 4-1), was used to extract the trend of the rainfall time series (Figure 4-1). The trend of the rainfall time series was used as an input to improve model performance.

STL is a filtering procedure that uses LOESS (locally estimated scatterplot smoothing) to decompose a time series into smoothed estimates of three components namely, trend, seasonality and residual (Cleveland *et al.*, 1990). The original time series data set, is broken down into the addition of these three components, according to Equation 4-1 (Cleveland *et al.*, 1990):

$$\boldsymbol{Y_t = T_t + S_t + R_t}$$

Equation 4-1

Where Y_t represents the original data set and T_t, S_t, R_t refers to the trend, seasonality and residuals components respectively for time t.

The monthly rainfall time series was broken down into the components according to Equation 4-1, and the trend of the rainfall time series (T_t) was extracted and used as the model input for rainfall. The trend of the rainfall time series is represented as the blue line in (Figure 4-1).

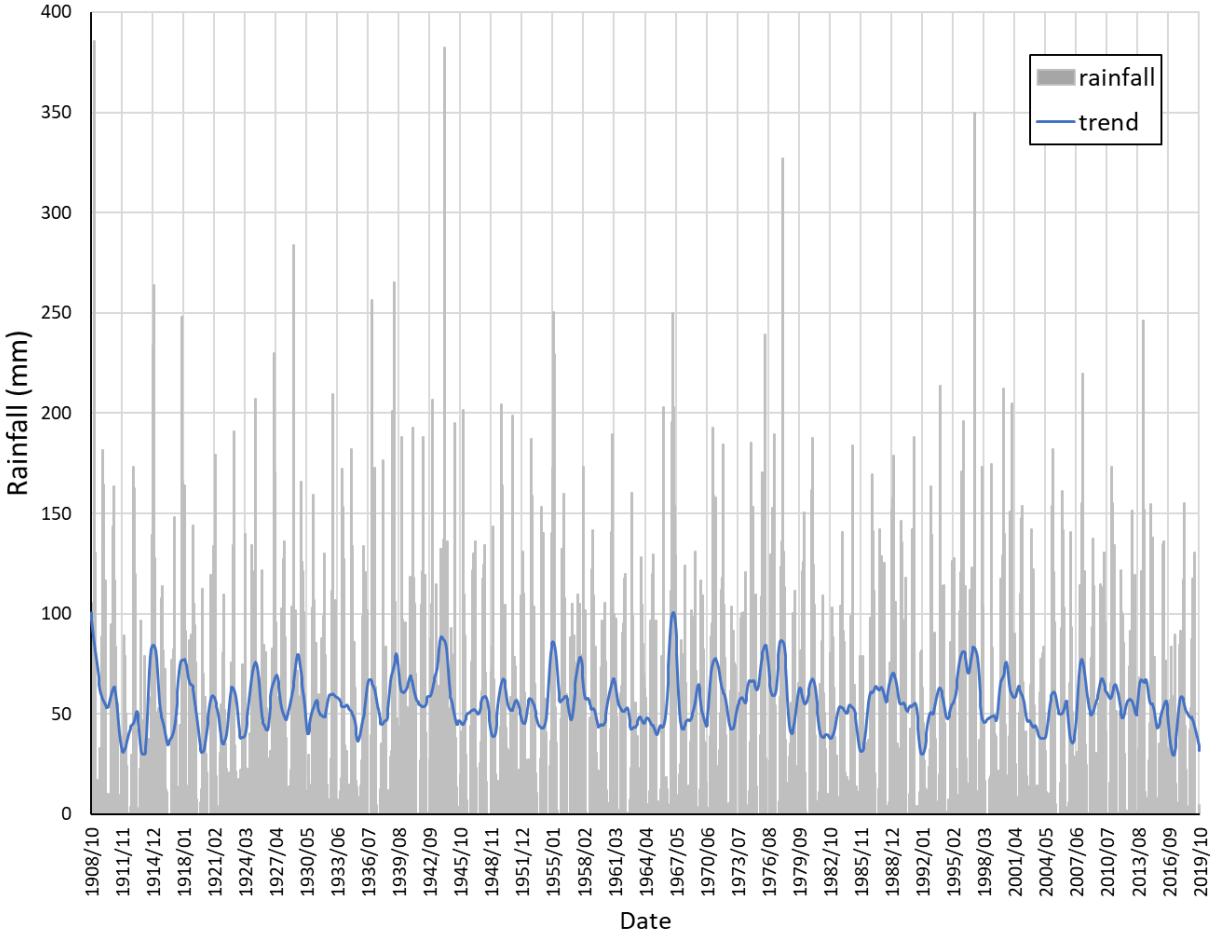


Figure 4-1: Monthly rainfall trend (blue line) extracted from monthly rainfall (grey bars) using STL time series decomposition.

4.2.4 Temperature

Historical temperature data set was acquired from the SAWS and ARC. The two meteorological stations recording temperature closest to the boreholes in the Steenkoppies aquifer was the Tarlton station and the Zuurbekom station (+/- 30 km southeast of the Steenkoppies aquifer) (Figure 3-2). The meteorological stations record daily temperature in ˚C. The Tarlton station recorded daily temperature from November 2003 to October 2019. The Zuurbekom station recorded daily temperature from January 1950 to August 2019.

Data from the Tarlton station was representative of the actual daily temperature of the study area as it is situated in the Steenkoppies aquifer. However, to increase the length of the Tarlton data set, the temperature data from the Zuurbekom was used to back extend the time series to January 1950.

Although there is very little topography that could cause a considerable change in temperature over distance in the vicinity of the study area, a linear regression on the overlapping data between the Tarlton and Zuurbekom data sets (November 2003 to August 2019) was calculated to assess the correlation between temperatures recorded at these two stations. Between the two data sets, there is a high linear correlation ($R^2 = 0.96$). Both stations also showed a similar mean annual temperature where the standard deviation between the stations was less than 10%. For these reasons, the Zuurbekom data set was used as a realistic representation of the temperature in the Steenkoppies aquifer.

4.2.5 Groundwater Usage

Two data sets for groundwater usage in the Steenkoppies compartment was acquired, the groundwater usage data set from the Water Use Authorisation and Registration Management System (WARMS) database maintained by the DWS and groundwater usage from literature collected by Seyler *et al.* (2016) (Figure 3-13). Determining actual groundwater usage from registered users is challenging.

There are many inaccuracies in the WARMS database such as duplicate entries, incomplete entries and entries that are not updated or inaccurate coordinates etc. The registrations in the WARMS data set may be an overestimate to secure supply, or may an underestimate to secure a licence. Additionally, there may be numerous unlicensed groundwater users as not all users may be registered. The groundwater usage data from the literature may not be completely accurate as it is estimations gathered from the available literature.

Although there are potential inaccuracies in both groundwater usage datasets, the WAMRS data set was thought to be less accurate based on the unrealistic value recorded for the registered groundwater used for irrigation livestock. The groundwater usage gathered from literature by Seyler *et al*. (2016), was also longer than the WARMS data set. The shorter data set would limit the amount of data used for model training. Therefore the groundwater usage gathered from literature by Seyler *et al*. (2016), was chosen as a model input.

4.2.6 Data Organisation

After gathering the historical time series for the target and input variables, data sets were created for each borehole. Data were aggregated to a monthly temporal resolution. The rainfall values were summed monthly, and for temperature, discharge and groundwater usage, the arithmetic mean was applied. The period for the target groundwater levels and input variables were selected to show the maximum time frame possible but also to show only minor data gaps.

4.2.7 Interpolation

The gaps in the data set were relatively small with a maximum of 6 missing values (which equates to 6 months) and were interpolated linearly. The two most common methods to interpolate groundwater level data sets are the linear and cubic spline interpolation (use by authors such as Wusch *et al.*, (2018) and Daliakopoulos *et al.*, (2005) respectively). A study by North and Livingstone (2013) compare the linear and cubic spline methods for interpolation of lake water column profiles. North and Livingstone (2013) created artificial "pseudo-gaps" of various sizes in the data which were filled using the two methods. The performance of the two methods was assessed by statistic evaluation of the error between the measured values and the interpolated values. The result of the study suggests that the linear interpolation method interpolates the gaps most accurately.

Since every data set is different, a similar experiment to North and Livingston (2013) was conducted on the data used in this study, to ensure that the linear interpolation was a suitable interpolation method. "Pseudo-gaps" of various sizes (1 to 6-month gaps) were created. Both linear and cubic spline interpolation were tested to interpolate the pseudo-gaps. The experiment confirmed that the linear interpolation method performed best on average between the measured values and the interpolated values (Linear interpolation $R^2 = 0.98$ compared to cubic spline method $R^2 = 0.95$).

4.2.8 Summary of Data for Modelling

Table 4-2: Summary of the final data sets acquired for modelling

Variable	Source	Station / Gauge number	Units	The time for which the data was available	Missing data (%)	Temporal resolution
Groundwater level	NGA, HYDSTRA	-	magl	varying time periods between 1940 -2019	< 50%	Monthly
Spring discharge	DWS flow gauges	A2H010	m^3/s	Oct 1908 - Oct 2017	<1%	Monthly
Groundwater usage	Seyler et al. (2016)	Licenced abstraction / indications from literature	million m^3/a	1980 -2019	Extrapolated from literature values	Monthly
Rainfall	Holland *et al.* (2009), SAWS, ARC	Compilation of 5 stations	mm	Oct 1908 - Oct 2019	<1%	Monthly
Temperature	SAWS, ARC	Tarlton and Zuurbekom station	˚C	Jan 1950 - Oct 2019	3%	Monthly

4.3 MUTUAL INFORMATION (MI)

The input variables (x) are not equally informative in predicting the target variable (y). This is due to some variables being noisier or having less of a relationship and correlation with the target data set. The mutual information (MI) was used to quantify how informative each input variable (rainfall, temperature, spring discharge and groundwater usage) was in predicting the target variable (groundwater levels). MI measures how much information (measured in units called bits) can be obtained from one variable given another (Ross, 2014). MI is also called "information gain" and linked to the concept of entropy that quantifies the amount of information there is in a random variable (Kraskov *et al.*, 2004). MI measures how much, on average, the knowledge of the x variable reduce the uncertainty of the y variable (MacKay, 2005). MI between the input and target variables, x and y, was calculated using Scikit-learn (Pedregosa *et al.,* 2011) in Python (Python Core Team, 2015) according to Equation 4-2 (MacKay, 2005):

$$I\ (x; y) = H(x) - H\ (x \mid y)$$

Equation 4-2

Where

x = the input variable

y = the target variable (groundwater level)

$I(x; y)$ = the mutual information for x and y

$H(x)$ = the entropy of x

$H\ (x \mid y)$ = the conditional entropy for x given y

MI is a measure of mutual dependence between two random variables, therefore, I (x; y) = I (y; x), and I (x; y) ≥ 0.

If x and y are independent, then information about y cannot be obtained from x, and the MI equals zero. If x and y are deterministic functions of one another, then all the information about y can be obtained from x, and the MI equals one (Ross, 2014).

MI is a generalised version of correlation and handles nonlinear dependencies and continuous random values. The MI can indicate how useful the input parameter will be in predicting groundwater levels. The higher the MI between the input and target variable (groundwater levels), the more information is shared between the two variables. The input variable will have a high correlation with the groundwater levels and will be more informative in predicting the groundwater levels as the patterns

the model learns from the input variable data set will be useful for predicting groundwater levels (Lee *et al*., 2016).

If a low MI is recorded, then there is a low correlation, and little information is shared between the two variables. The input variable will be less informative for predicting groundwater levels. A low MI may also indicate that the input variable data set may be too noisy and the information from the input variable with respect to groundwater levels cannot be recorded. The MI between the groundwater levels and the target variable, helped to understand the groundwater level predictions made by the model.

It should be noted that MI does not model the influence of the input variables on the target variable. The influence between the variables in the study should be simulated using probabilistic graphical modelling. An example of such a model is shown in a paper by Ajoodha and Rosman (2020), where a Bayesian network was used to learn the conditional probability distribution between all the features and the target variable.

4.4 MODEL PIPELINE AND SET UP

The NNAR model (as described in section 2.4.1.1) was chosen to simulate and predict groundwater levels. The NNAR model is fit using the nnetar function in the Forecast Package in R (Hyndman and Athanasopoulos, 2014; R Core Team, 2017). Other options were the NARX package in python (Xie and Wang, 2018), however this did not allow to add any x varibables. The NARX package in Matlab could havalso been used (MATLAB, 2010) however, it is not open access as one would need to purchase Matlab. The nnetar function was chosen as it is open access, simple to implement and allows to use x variables as inputs. Figure 4-2 summarises the applied methodological approach taken to model and predict groundwater levels using the NNAR.

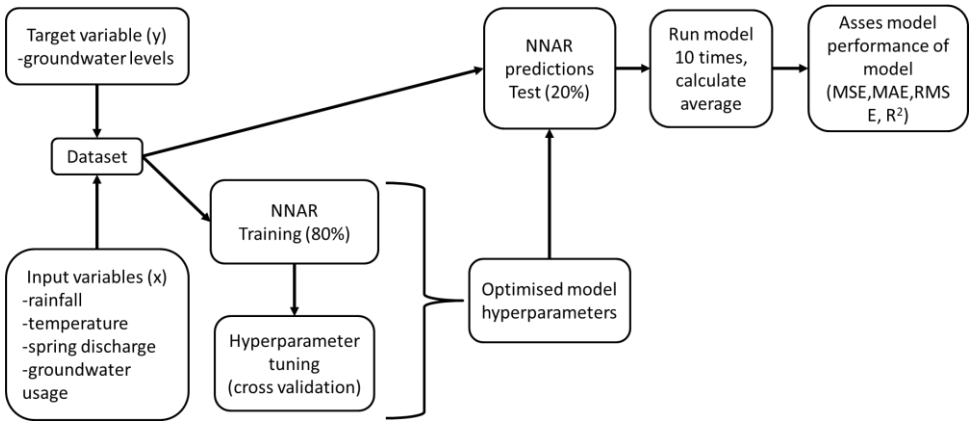


Figure 4-2: The applied methodological approach taken to model and predict groundwater levels using an NNAR (Neural Network Autoregression) in the present study

A similar approach to the "v-fold cross-validation" described in the book by Kuhun and Johnson (2019) is used to train, select optimal hyperparameters and test the model. This approach is implemented to prevent overfitting and improve the generalisability of the model (Kuhun and Johnson, 2019). This approach conducts an initial split of the time-series into two, 80% for training and 20% for the testing (e.g. Figure 4-3). Once the time series were split, the data were transformed though normalisation according to Equation 4-3. Normalising the data speeds the learning of the model, converts all the variables to the same scale and to a normal distribution (Pedregosa et al., 2011). The 20% used for model testing remains unseen until the final model is developed. The training data set is further split into five different resampling slices (Figure 4-4). Each slice excludes parts of the time series and is split with a 90%:10% ratio into analysis and assessment subsets. The model was trained on the analysis set, and assessment set was predicted using the model. The performance of the model was summarised using the mean squared error, and the statistics were saved. This proceeds on each slice so that in the end, there were five estimates of performance for each model, and each was calculated on a different assessment set. The cross-validation estimate of performance was calculated by averaging the five-individual metrics. This process was repeated with multiple models of different hyperparameter combinations. It is important to note that each model run during the cross-validation process were mutually exclusive and contained different instances. The hyperparameter combination producing predictions with the lowest error is saved and applied to make predictions on the initial split.

Normalisation applied to the time series after split (Pedregosa *et al.*, 2011):

$$Z = \frac{x - \mu}{\sigma}$$

Equation 4-3

Where:

Z= the standard score after normalisation

x = observed value

μ = mean of the sample

σ = standard deviation of the sample

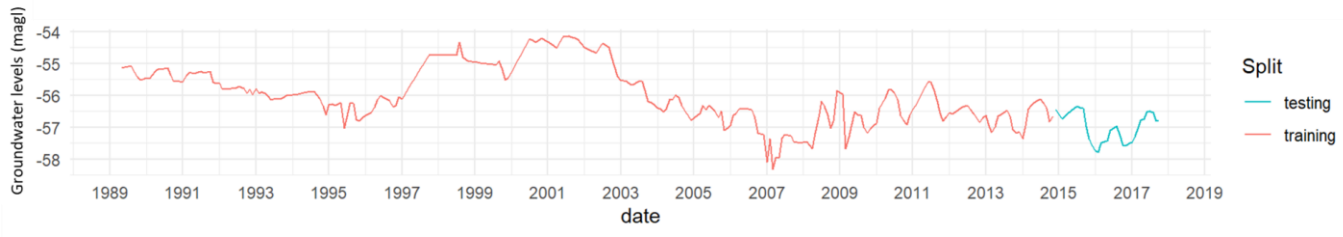


Figure 4-3: Example of the initial split on data which was used for the final prediction.

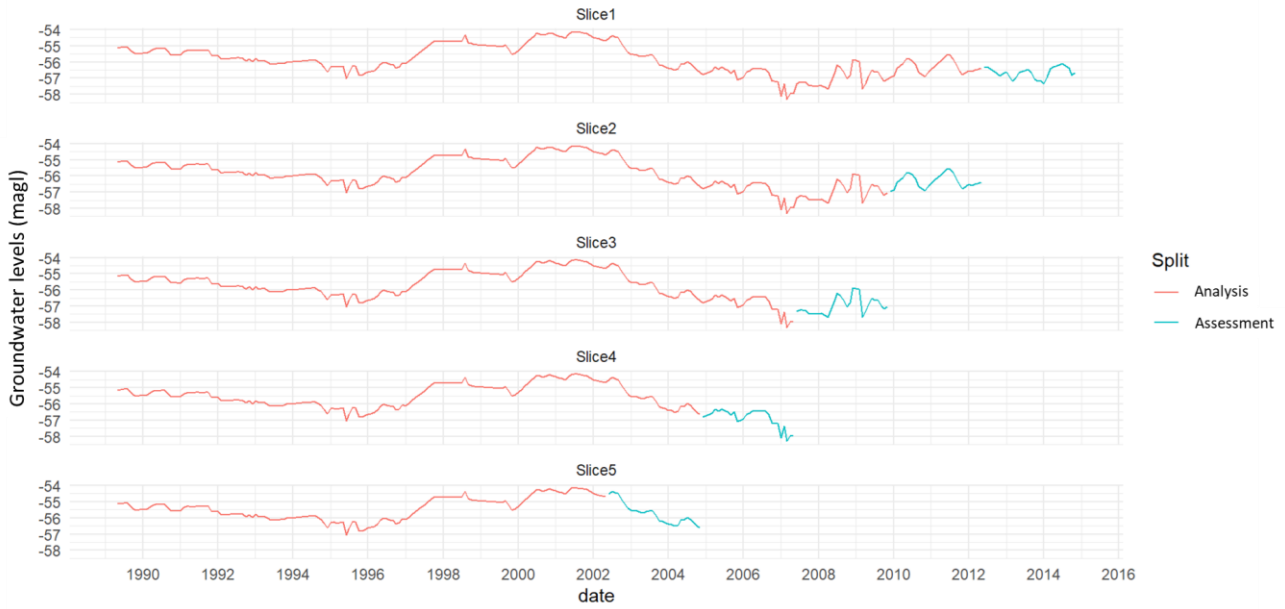


Figure 4-4: Example of cross-validation resampling strategy which was used for model training and hyperparameter optimisation.

There is an element of randomness in the predictions (since weights are initially chosen at random) (Scardapane and Wang, 2017). Therefore, in this study, once the optimal hyperparameters were selected (through the cross-validation process), the model was trained and tested on the initial split ten times. The results from the ten model runs were averaged, and the standard deviation at each timestamp between the ten model runs was calculated according to Equation 4-4 (Lee et al., 2015)

$$Standard\ deviation = \sqrt{\frac{\sum_{i=1}^{n}(\overline{x} - x_i)^2}{n-1}}$$

Equation 4-4

Where:

n = sample size (equal to ten for the ten model runs)

$\overline{x}$ = the mean of the samples (values from the 10 model runs)

x_i = the observation value

The performance of the model is then assessed using the average groundwater levels simulated from the ten runs. This ensures that the results displayed in this study are repeatable and a good representation of the model's ability to predict groundwater levels in the study area. Ideally, the more model runs, the more realistic representation of the model's ability to predict groundwater levels would be, however, there is a trade off between time and computer capacity. Ten model runs were chose as optimal as it captures the majority of the randomness in the model's output without taking too much time and computer capacity. The upper and lower limit of the standard deviation at each timestamp across the ten model runs were graphed to show the variability in the model to predict groundwater level during ten independent runs of the model.

4.4.1 Model Evaluation Criteria

The model performance was evaluated both graphically and statistically. The observed groundwater levels were plotted against those simulated by the model (both train and test). The upper and lower limit of the standard deviation at each time stamp across the ten model runs for each groundwater level prediction were also plotted. This was to show variability of the model to predict groundwater levels. Several error indices were used to compare the observed and the simulated groundwater levels statistically. The error indices were only calculated for the predictions made by the model (simulated test), as the predictive capability of the model was the criteria being evaluated. However, both the simulated train and test were shown graphically to visualise how the model fits the training data and how well the model predicted groundwater levels.

The error indices used in this study were the coefficient of determination (R^2), mean squared error model (MSE), root mean squared error (RMSE), and the mean absolute error (MAE), which are expressed below. It applies that y signifies the groundwater levels, $\hat{y}$ is the predicted groundwater levels, O_i and $\bar{O}$ represent the observed value and the mean observed value respectively, P_i and $\bar{P}$ represent the simulated values and mean simulated value respectively and *n* is the number of samples available. Each error indices have pros and cons. For this reason, four error indices are used in this study. The limitations of one measure can be addressed by an alternative method.

R^2 is the squared value of the Bravais Pearson correlation coefficient (Krause et al., 2005). R^2 ranges from 0 to 1. Since the values for R^2 range from 0 to 1 it more interpretable. Values greater than 0.5,

are typically considered as satisfactory. However, it should be noted that the R^2 value can be deceiving as it measures correlation and not accuracy. R^2 is formulated as Equation **4-5** (Moriasi et al., 2007).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})\,(P_I - \bar{P})}{\sqrt{\sum_{i-1}^{n}(O_i - \bar{O})^2}\,\sqrt{\sum_{i-1}^{n}(P_i - \bar{P})^2}} \right)^2$$

Equation 4-5

Mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE) indicate the error in unites or squared units of the simulated model output against the observed groundwater levels in meters (m). It assesses the model accuracy. The larger the value, the worse the fit. A value of zero indicates a perfect fit (Moriasi *et al.* 2007). The MSE, RMSE, MAE are formulated as (Moriasi *et al.* 2007):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2$$

Equation 4-6

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}$$

Equation 4-7

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|$$

Equation 4-8

4.5 APPLICATION OF MODEL PIPELINE FOR SCENARIO TESTING

The NNAR model was also used to predict groundwater levels under possible scenarios of change. Scenarios were created to mimic possible future changes in rainfall and groundwater abstraction (Table 4-3).

To simulate groundwater predictions for scenarios 1-3 (Table 4-3), the models were trained using the original training data set (80% of the time series). The test data set for the input parameters were then altered accordingly so that the model would receive input parameters that mimic a specific scenario. The model then uses the "scenario input parameters" to make groundwater level predictions based on the particular scenario.

For example, when groundwater level predictions were generated for scenario 1a (Table 4-3), the test data set (last 20% of the time series) of the rainfall time series was altered from the actual measured values to the values that mimic the scenario being tested. In scenario 1a months with > 100 mm of rainfall during the test data, then 100 mm were subtracted from that month. The model will then use the input variable altered to mimic the scenario with the other input variables altered, as well as what the model has learnt during training to make groundwater level predictions.

The scenarios tested are summarised in Table 4-3 showing the shortened ID, name and description assigned to each scenario tested. Scenario 1a-c aim to simulate scenarios where there is a change in rainfall peaks and an overall decrease in the rainfall. A scenario of increased groundwater abstraction is simulated by Scenario 2a.

Scenario 3a and b tests the ability of the model to make long-term predictions. Scenario 3a aimed to test if the NNAR model is capable of making a 15-year prediction. The training to test split ratio was altered so that the model only trains on 50% of the available time series, and the remaining 50% is used to make groundwater level predictions. It should be noted that only the train to test split ratio was changed and the input parameters were not altered in any way (i.e. the actual measured values for the input parameters were used).

Scenario 3b aimed to test the model's ability to make a 30-year prediction. The complete data set (100%) of the actual measured values was used for model training. Groundwater levels were then predicted 30 years into the future using long-term averages for the input variables. The long-term averages were calculated by generating an average for each month of the year from the entire time series of each input variable. The average for each month of the year was then repeated until a 30-year time series was created for each input variable.

Table 4-3: Description of the scenarios applied to the developed models for the Steenkoppies aquifer

Scenario ID	Name	Description
1a	Decrease rainfall peaks	If monthly rainfall > 100mm then minus 100mm. No alteration to spring discharge, temperature and groundwater usage input variables.
1b	Decrease rainfall	Half the monthly rainfall. No alteration to spring discharge, temperature and groundwater usage input variables.
1c	Increase rainfall peaks	If monthly rainfall > 100mm then add 200mm. No alteration to spring discharge, temperature and groundwater usage input variables.
2	Increase groundwater abstraction	Double the monthly abstraction. No alteration to spring discharge, temperature and rainfall input variables.
3a	Long term prediction with reduced training	Change in training/test split to 50%/50% to model a +/- 15-year prediction. No alteration to input variables.
3b	Long term prediction 30 years	100% data for training, predict groundwater levels for 30 years using long term averages of input variables.

CHAPTER 5:
RESULTS AND DISCUSSION

5.1 MUTUAL INFORMATION (MI)

The measures of mutual information (MI) for the five input variables with respect to groundwater levels in the Steenkoppies compartment, with the mean ($\bar{x}$) and standard deviation (sd), are presented in Table 5-1. MI was calculated using Scikit-learn (Pedregosa *et al.*, 2011) in Python (Python Core Team, 2015) according to Equation 4-2. In the Steenkoppies compartment, spring discharge recorded the highest MI with respect to groundwater levels ($\bar{x}$ =0.78, sd = 0.19) signifying that among all the input variables, spring discharge is the most informative input variable to predicting groundwater levels. A strong relationship was expected, given that the flow of groundwater in the Steenkoppies compartment is towards the Maloney's Eye spring as the hydraulic gradient between the surrounding aquifer and the Maloney's Eye is what drives spring discharge (Holland *et al.*, 2009). Hence, changes in groundwater level directly translate to changes in the spring discharge rate.

Recharge from rainfall causes and increase in groundwater levels in the Steenkoppies compartment (Holland *et al*., 2009). Groundwater levels are a muted version of the monthly rainfall, and the peaks and troughs seen in the groundwater levels correspond to the cyclical pattern seen in the rainfall (see Figure 3-11). It was, therefore, expected that rainfall would be one of the variables with a higher MI with respect to groundwater levels. However, a low MI between the rainfall and groundwater levels was recorded ($\bar{x}$ = 0.04, sd = 0.04).

This unusual result prompted the use of the rainfall trend (the decomposed version of the rainfall time series) as an input variable. Using decomposed signals of a time series, such as the trend, is common practice in machine learning (e.g. Xu *et al.*, 2020). From the MI values, it is clear that the rainfall trend showed a more informative output ($\bar{x}$ = 0.42, sd = 0.09). This indicates that the information shared between the rainfall data set and groundwater levels were not detected due to the noise in the rainfall time series. As described in the site description (section 3.7.1) larger rainfall events (> 100 mm per month) recharge the aquifer and correlate with groundwater level increases. Smaller rainfall events (< 100 mm per month) were not informative with regards to groundwater levels. The smaller rainfall events could, therefore, have been detected as noise and distorted the ability of the MI to pick up the relationship that exists between groundwater levels and rainfall. This confirms that the trend of the rainfall time series should be used as a model input over the actual rainfall time series as it is a more informative input variable for predicting the groundwater levels.

Groundwater abstraction is the third most informative variable with respect to groundwater levels in the Steenkoppies compartment ($\bar{x}$ = 0.37, sd = 0.07). Whilst it is suspected that sustained groundwater use across the compartment has contributed to an overall decline in groundwater levels (Chapter 3), whether this pattern would be detectable in the models was uncertain before undertaking the research given the stepwise nature of the groundwater use data (Figure 3-13), and the comparatively noisy groundwater level. However, there is a degree of information shared between the groundwater level and groundwater usage dataset.

The temperature has the lowest MI with respect to groundwater levels ($\bar{x}$ = 0.02, sd = 0.02) and therefore, shares the least amount of information with groundwater levels compared to the other four input features. Similar to a study by Wunsch et al. (2018), the temperature time series was used as a substitute for evapotranspiration. This was because the evapotranspiration recorded for the Steenkoppies aquifer held only five years worth of data, enabling this dataset to be used as a model input. Temperature and evapotranspiration are positively correlated. It is expected that an increase in the temperature will drive higher evapotranspiration (Döll, 2009). As per the water balance equation (Equation 2-1), there is a correlation between evapotranspiration and groundwater levels, as increased evapotranspiration contributes to a decrease in groundwater storage (represented as a decrease in groundwater levels) (Healy *et al.*, 2007).

Although an increase in temperature is recorded in the Steenkoppies compartment (section 3.4.2), a low MI recorded show that temperature (substitute for evapotranspiration in this study) was not very informative of the groundwater levels in the Steenkoppies aquifer. This may be attributed to the fact that most of the rainfall in the Steenkoppies compartment occurs as thunderstorm events (Seyler *et al*., 2016). During thunderstorm events, atmospheric temperature drops and air humidity is at 100 %, therefore, evaporation is not possible (Gulliver *et al*., 2010). The rainfall is thus not evaporated, but recharged to the aquifer (depending on the surface could also contribute to runoff). Water pumped from the groundwater that is not used up by the crop, can filtrate back into the aquifer (assuming the aquifer is unconfined like the Steenkoppies) and the net pumping rate is not equal to the total pumped water (Devlin and Sophocleous, 2005). Pumping does not deplete the groundwater storage, but the amount of abstracted groundwater that evaporates, used up by the crop or taken away or other uses (Kendy, 2003). Modern farming practices have also leaned towards irrigating in ways that reduce evapotranspiration of water (e.g. watering during the evening and early morning, drip irrigation) (Qin *et al.,* 2016). These could be reasons why a significant relationship of temperature on groundwater levels could not be picked up. However, evapotranspiration is a very complex process and involves other factors such as vegetation type, irrigation schedules, soil types etc. which was not investigated in this study.

Table 5-1: The mutual information (MI) of each input feature with respect to the target variable (groundwater levels) calculated using Scikit-learn (Pedregosa *et al.*, 2011) in Python (Python Core Team, 2015) according to Equation 4-2.

	Borehole ID	Temperature	Rainfall trend	Rainfall	Spring Discharge	Groundwater Abstraction
1	A2N0566	0.01	0.39	0.02	1	0.4
2	A2N0553	<0.01	0.21	0.06	0.5	0.26
3	A2N0556	0.05	0.29	0.05	0.54	0.2
4	36349	0.06	0.35	0.09	0.46	0.33
5	A2N0610	<0.01	0.49	0.04	1	0.34
6	A2N0568	<0.01	0.5	0.07	0.74	0.36
7	37773	<0.01	0.39	0.02	0.59	0.35
8	A2N0554	<0.01	0.37	<0.01	0.73	0.38
9	A2N0567	0.06	0.41	0.05	1	0.45
10	A2N0569	0.02	0.53	<0.01	0.81	0.46
11	A2N0608	<0.01	0.48	0.01	0.89	0.51
12	A2N0612	0.01	0.44	0.01	0.86	0.37
13	A2N0614	<0.01	0.36	0.04	0.72	0.34
14	A2N0616	<0.01	0.35	0.06	0.65	0.41
15	A2N0617	<0.01	0.53	<0.01	0.96	0.33
16	A2N0563	0.02	0.5	0.09	1	0.38
17	A2N0615	<0.01	0.38	<0.01	0.59	0.35
18	A2N0565	0.05	0.52	0.15	1	0.44
	Mean ($\bar{x}$)	0.02	0.42	0.04	0.78	0.37
	Standard deviation (sd)	0.02	0.09	0.04	0.19	0.07

5.2 MODEL SIMULATION AND PREDICTION

This section shows the results of the NNAR model to simulate (training) and predict (test) groundwater levels from 18 boreholes in the Steenkoppies compartment. The statistical performance of the NNAR model to predict groundwater levels across 18 boreholes in the Steenkoppies compartment was assessed using five error indices which are presented, along with the mean ($\bar{x}$) and standard deviation (sd), in Table 5-2. The performance of the NNAR was also assessed graphically. The graphical results from six boreholes are displayed and discussed in this chapter, and the additional 12 graphs are presented in the Appendix.

The error indices indicate that the performance of the NNAR varied between the boreholes (Table 5-2). The coefficient of determination (R^2) varied considerably across all 18 boreholes ($\bar{x}$ = 0.19 and sd = 1.78). Although R^2 ranged from 0 to 1, five of the 18 models predicted groundwater levels with a negative R^2. The reason for this lies within the computation of the statistic. R^2 compares the fit of

the chosen model against the mean of the observed data as a horizontal straight line (null hypothesis) (Moriasi et al., 2007). If the NNAR fits worse than the horizontal straight line, then a negative R^2 is obtained. The performance of the NNAR model showed relatively low errors across the 18 models for the remaining metrics (MAE, MSE, RMSE). Majority of the models predicted groundwater levels with acceptably low errors, well below 1 m, and only four models recorded errors greater than 1 m (Table 5-2).

In this study, low R^2 values do not necessarily correlate with high error values (MSE, RMSE and MAE). For example, although borehole A2N0553 recorded a higher R^2 value compared to borehole A2N0556. The MAE, MSE and RMSE recorded higher error metrics for borehole A2N0553 compared to A2N0556 (A2N0553: $R^{2\,=}$ 0.20, MSE = 6.50, RMSE = 2.54, MAE = 2.14; A2N0556: $R^{2\,=}$ 0.05, MSE = 4.01, RMSE = 2.00, MAE = 1.67) (Table 5-2). Wunsch et al. (2018) used NARX (Nonlinear Neural Networks with Exogenous inputs) to model groundwater levels. Simmilarly, Wunsch et al. (2018) also found that high R^2 values did not always correlate with the other error metrics used in the study, namely a low RMSE and high Nash Sutcliffe values. This highlights the importance of incorporating multiple error indices for assessing model accuracy.

Overall, according to the statistical performance metrics, it was clear that the NNAR model performed best in predicting groundwater levels from borehole A2N0568 ($R^{2\,=}$ 0.81, MSE = 0.09, RMSE = 0.30, MAE = 0.23) and performed the worst predicting groundwater levels from borehole 36349 ($R^{2\,=}$ -1.19, MSE = 23.04, RMSE = 4.75, MAE = 4.03).

Table 5-2: Summary of the metrics used to assess NNAR performance

	Well ID	R^2	MSE (m)	RMSE (m)	MAE (m)	STDV across ten model tuns
1	A2N0566	0.49	0.05	0.23	0.19	<0.01
2	A2N0553	0.20	6.50	2.54	2.14	<0.01
3	A2N0556	0.05	4.01	2.00	1.67	<0.01
4	36349	-1.19	23.04	4.75	4.03	0.65
5	A2N0610	0.60	0.06	0.24	0.19	0.04
6	A2N0568	0.81	0.09	0.30	0.23	0.04
7	37773	0.41	0.08	0.28	0.24	<0.01
8	A2N0554	-0.16	0.27	0.52	0.42	0.02
9	A2N0563	0.76	0.05	0.22	0.15	<0.01
10	A2N0565	-1.70	0.05	0.22	0.18	0.01
11	A2N0567	0.56	0.09	0.29	0.24	0.04
12	A2N0569	-0.31	0.06	0.25	0.20	0.02
13	A2N0608	-6.93	4.21	1.98	1.77	0.15
14	A2N0612	0.47	0.06	0.25	0.21	<0.01
15	A2N0614	0.14	0.16	0.40	0.37	<0.01
16	A2N0615	0.25	0.13	0.36	0.30	<0.01
17	A2N0616	-0.07	0.16	0.39	0.33	<0.01
18	A2N0617	0.59	0.16	0.40	0.28	0.02
	Average	-0.40	1.91	0.73	0.61	0.07
	Standard deviation	1.78	5.54	1.28	1.03	0.21

Graphically it is clear that the groundwater level predictions made by the NNAR model vary depending on the nature (length, fluctuation, change over time) of the target groundwater level time series (Figure 5-1). Similar to the statistical performance (Table 5-2), the graphs showed the best groundwater level predictions made by the NNAR for borehole A2N0568 and the worst predictions made were for groundwater levels from borehole 36349 (Figure 5-1).

The predictions made by the NNAR underestimated the groundwater level peaks from boreholes A2N556, A2N553, 36349 and A2N0568, indicating that the NNAR had difficulties in predicting extreme events (Figure 5-1). This observation suggests that the transformation applied to the data did not entirely remove the local variations in the function being mapped as the data was negatively skewed (Sudheer *et al.*, 2003). As per the model pipeline (section 4.4), the data was transformed before training and testing. This was done to scale all the variables to the same range and to convert the data to a normal distribution. However, it seems that even after transformation, the target variable still had a slight negative skew (tail towards lower values) and as a result, the NNAR matched that pattern to minimize errors causing an underestimation in the groundwater levels.

The NNAR model gave the best predictions of the observed target groundwater levels when the test data set was representative of the training data set. Artificial Neural Network (ANNs), like the NNAR, use the training data to learn the patterns and relationships in the datasets (Adhikari and Agrawal, 2013). Therefore, if the fluctuations and patterns in the test data are very different to the training data, then the model will struggle to make predictions as it would not have seen groundwater level patterns like those in the test data set (Crowther and Cox, 2005). The more similar the test dataset is to the training dataset, the better the model can predict. This is demonstrated for borehole A2N0568 Figure 5-1. The fluctuations in the observed groundwater levels of the training dataset (1987/09 to 1995/12) are very similar to the fluctuations in the observed groundwater levels in the test data set (2004/03 to 2006/11). The model very accurately predicted groundwater levels in this borehole A2N0568 (R^2 = 0.81, MSE = 0.09, RMSE = 0.30, MAE = 0.23). The model would have learnt the predictive rule from the patterns during training and therefore, was able to make accurate predictions when a similar pattern occurs in the test data. However, in reality, training and test datasets will not always have similar patterns. This highlights the importance of informative input variables to assist predictions.

When modelling using ANNs like the NNAR, each input variable is chosen as it provides information that facilitates the prediction of the target variable (Lee *et al*., 2016). Therefore, the stronger the relationship between the input variable and the target variable, the more useful and informative the input variable is and the better predictions can be made by the model (Lee *et al*., 2016). The NNAR models produced more accurate groundwater level predictions in boreholes, where there was a stronger relationship between the groundwater levels and the input variables.

This is evident in the groundwater level predictions made at borehole A2N0566 and A2N553 (Table 5-2 and Figure 5-1). Referring back to the MI in Table 5-1, it is evident that the MI of each input feature with respect to groundwater levels for borehole A2N0566 was higher than for borehole A2N553. As a result, the model was able to make better groundwater level predictions at borehole A2N0566 compared to borehole A2N553 as the input variables were more informative in prediction groundwater levels from borehole A2N0566.

An extract of the MI scores for borehole A2N0566 and A2N553 from Table 5-1 is shown below:

Borehole ID	Temperature	Rainfall trend	Rainfall	Spring Discharge	Groundwater Abstraction
A2N0566	0.01	0.39	0.02	1	0.4
A2N0553	<0.01	0.21	0.06	0.5	0.26

The NNAR model showed overfitting at borehole 36349 (Figure 5-1). Overfitting is when a model fits closely to the training data resulting in a model that has learnt to reproduce noise and peculiarities in the data (such as large changes over a short time period) and fails to learn the general predictive rule (Wickham and Grolemund, 2016). Overfitting causes the accuracy of the model to deteriorate when predicting anything outside the range of hydrogeological conditions that occurred during the training period (Zanotti *et al.*, 2019). Normally, overfitting is avoided by training using cross-validation, as implemented in this study, however, from the 18 groundwater level time series modelled, the NNAR model still showed overfitting in one model (Figure 5-1). The NNAR model fitted almost exactly to all the peaks and troughs in the groundwater level training data for borehole 36349. The model then made poor groundwater levels predictions (R^2 = -1.19, MSE = 23.04, RMSE = 4.75, MAE = 4.03) (Table 5-2). Zanotti *et al.* (2019) proposed that alternative to cross-validation, the parameters such as the weights and bias, the number of neurons in the hidden layer of the model should be altered to avoid overfitting. Different activation functions or architectures of the neural network could also be changed to reduce overfitting when it is seen to occur (Zanotti *et al.*, 2019).

Another interesting observation is that the groundwater level predictions made by the NNAR were slightly lagged outputs (+/- 3 months) compared to the observed groundwater levels. Studies that made use of autoregressive models, similar to the NNAR to predict groundwater levels clearly show lags in the model results (e.g. Wunsch *et al.*, 2018; Guzman *et al.*, 2017). The lags could have been a result of the recurrent/autoregressive nature of the NNAR. Since the NNAR used lagged inputs to make predictions (Hyndman and Athanasopoulos, 2014), the model observed previous values to make predictions and therefore, outputs a slight lag in the predictions.

The model results additionally demonstrated that the NNAR could be used as a useful tool to fill gaps in the data sets. There was a 6-month data gap in groundwater levels recorded at borehole A2N0566 in 2008, which as per the methods was interpolated linearly (Figure 5-1). Using the input variables, the NNAR model was able to simulate a more realistic groundwater level for that period.

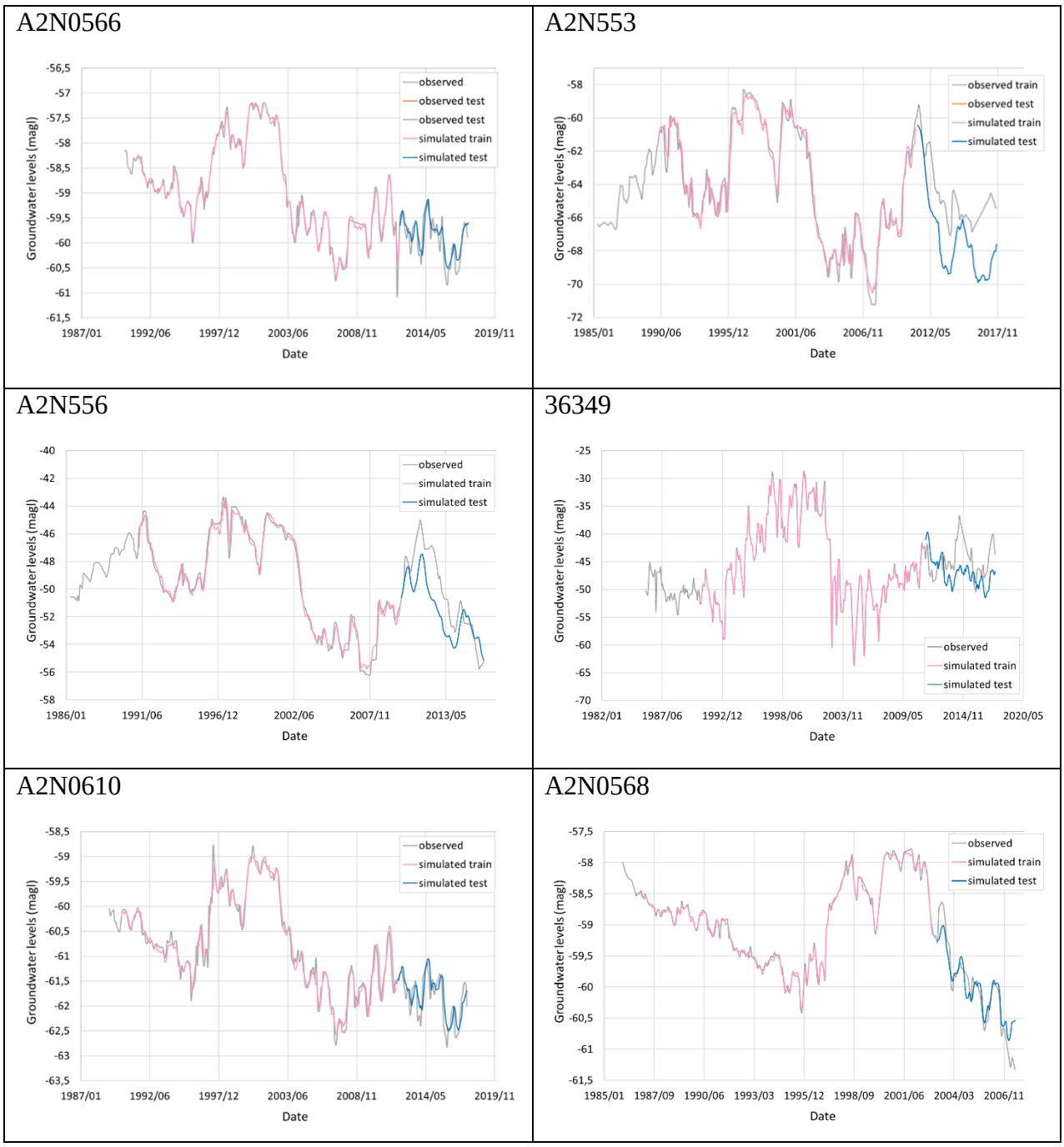


Figure 5-1: Comparison between the observed and simulated groundwater levels for six boreholes in the Steenkoppies compartment, during training and testing phases.

The Standard deviation between the ten independent runs of six models is shown graphically in Figure **5-2**. These results show how ANNs, like the NNAR, produce slightly different predictions each time the model is run. This is because the model weights are chosen at random at the start of model training and then updated using the data so that the overall predictive error is minimised (Zhang and Hu, 1998).

The standard deviation of the predictions across the ten model runs is low for borehole A2N0566, A2N553, A2N556, as the upper and lower limit of the standard deviation matched closely to the average predicted groundwater levels across the ten model runs (Figure **5-2**). The model predictions did not deviate substantially across the ten model runs.

In some cases, the standard deviation across the ten model runs increases with time (Figure **5-2** graph A2N0610 and A2N0568). This is due to the accumulated error in model prediction. Since the model feeds back the predicted groundwater levels at previous time steps as an input into the model, there is an accumulated error as the models predict further into the future (the problem of vanishing gradients) (Scardapane and Wang, 2017).

For borehole 36349, the standard deviation across the ten model runs is large throughout the groundwater level predictions (Figure **5-2**). This may be attributed to the fact that the model overfits to the data during training (Figure 5-1). Therefore, the model is unable to generalise to predict the new test data. This ultimately resulted in significant variability in the predictions made across the ten model runs, as the model had learnt to reproduce noise and peculiarities in the data and fails to generalise to new data.

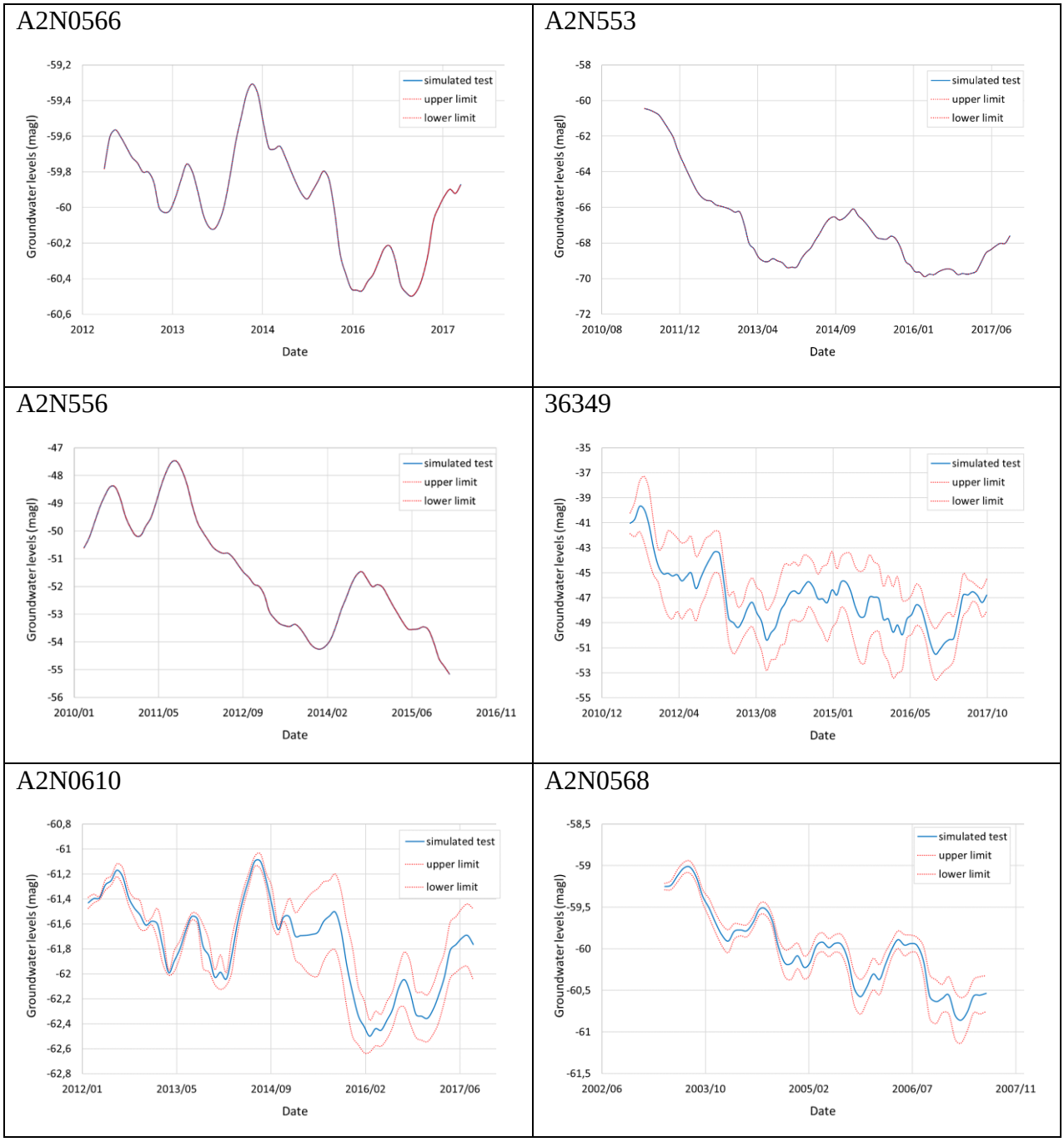


Figure 5-2 Standard deviation error between ten independent runs of each model used to predicted the groundwater levels for six boreholes.

5.3 SCENARIO TESTING

Several scenarios were tested using the models developed for groundwater level predictions at six boreholes. This was done to test how the groundwater levels in the Steenkoppies would respond under specific scenarios of change.

Scenario 1a and b mimicked a reduction in rainfall. Scenario 1a simulated a reduction in rainfall events over 100 mm (subtracted 100 mm from monthly rainfall over 100 mm) and scenario 1b simulated reduction in all rainfall events (monthly rainfall halved). For scenarios, 1a and b, the NNAR model simulated a decrease in groundwater levels at boreholes A2N553, A2N556 and 36349 (Figure 5-3 and Figure 5-4). At these three boreholes, the scenario of decreased rainfall is picked up (learnt) sufficiently by the model. The model recognised the decrease in the rainfall and reacts accordingly by simulating a decrease in groundwater levels.

The models for boreholes A2N0566, A2N0610 and A2N0568 did not show a significant change or drop in the simulated groundwater levels for scenarios 1a and 1b (In both Figure 5-3 and Figure 5-4). The model simulated groundwater levels similar to the observed groundwater levels at these boreholes as the model did not recognise the decrease in rainfall. This may be due to the very high MI recorded between spring discharge and groundwater levels at these boreholes (Table 5-1). The MI between spring discharge and groundwater levels at boreholes A2N0566, A2N0610 and A2N0568 was 1, 1 and 0.74 respectively (Table 5-1). The model may not have been able to simulate the influence of rainfall on groundwater levels at these boreholes because the model placed much weight on the relationship between spring discharge and groundwater levels to make groundwater level predictions. The model, therefore, did not depend on the other input variables to make groundwater level predictions as the information in the spring discharge time series was sufficient to make predictions.

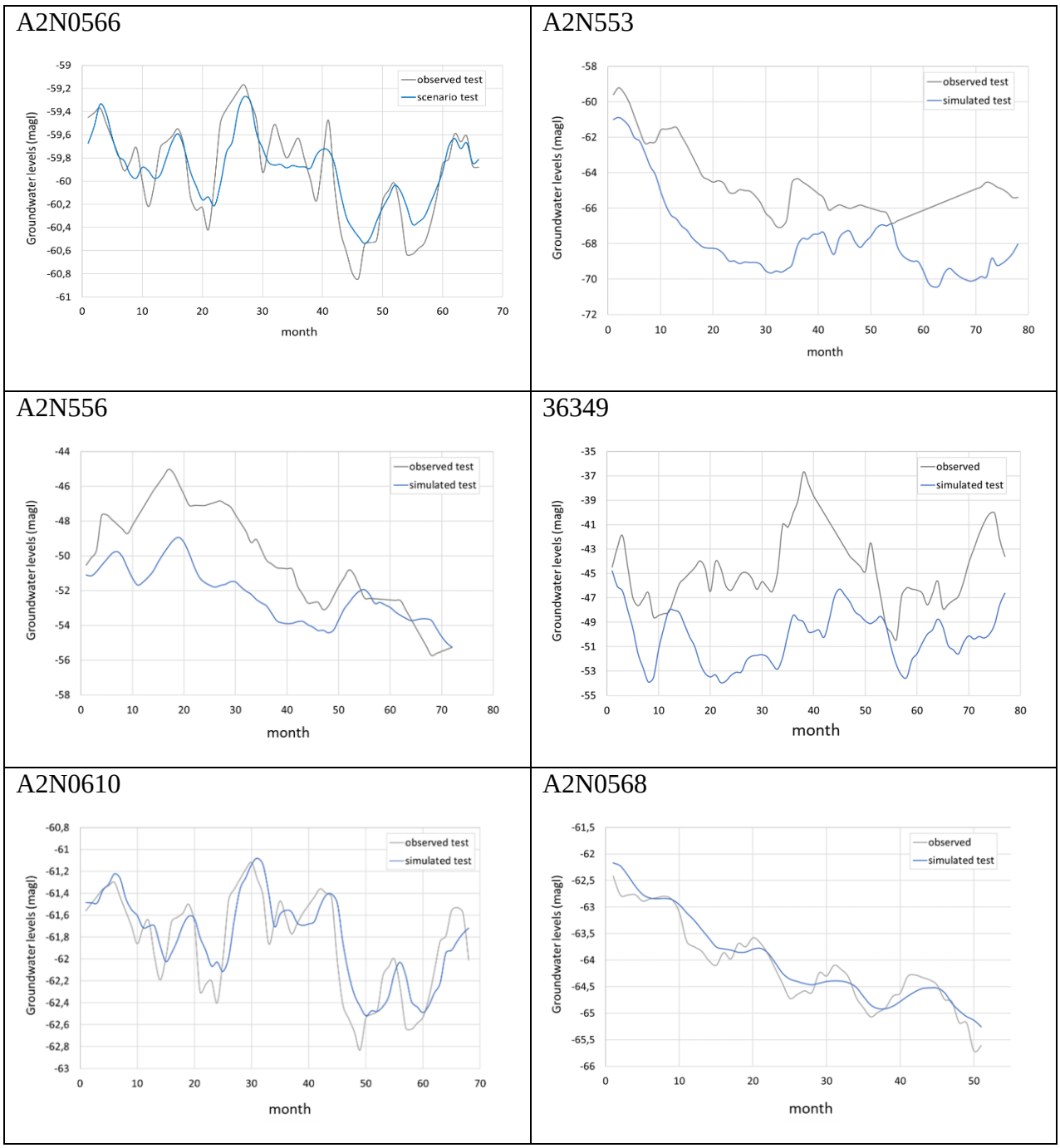


Figure 5-3: Model prediction for scenario 1a, decrease rainfall peaks. The date is in months.

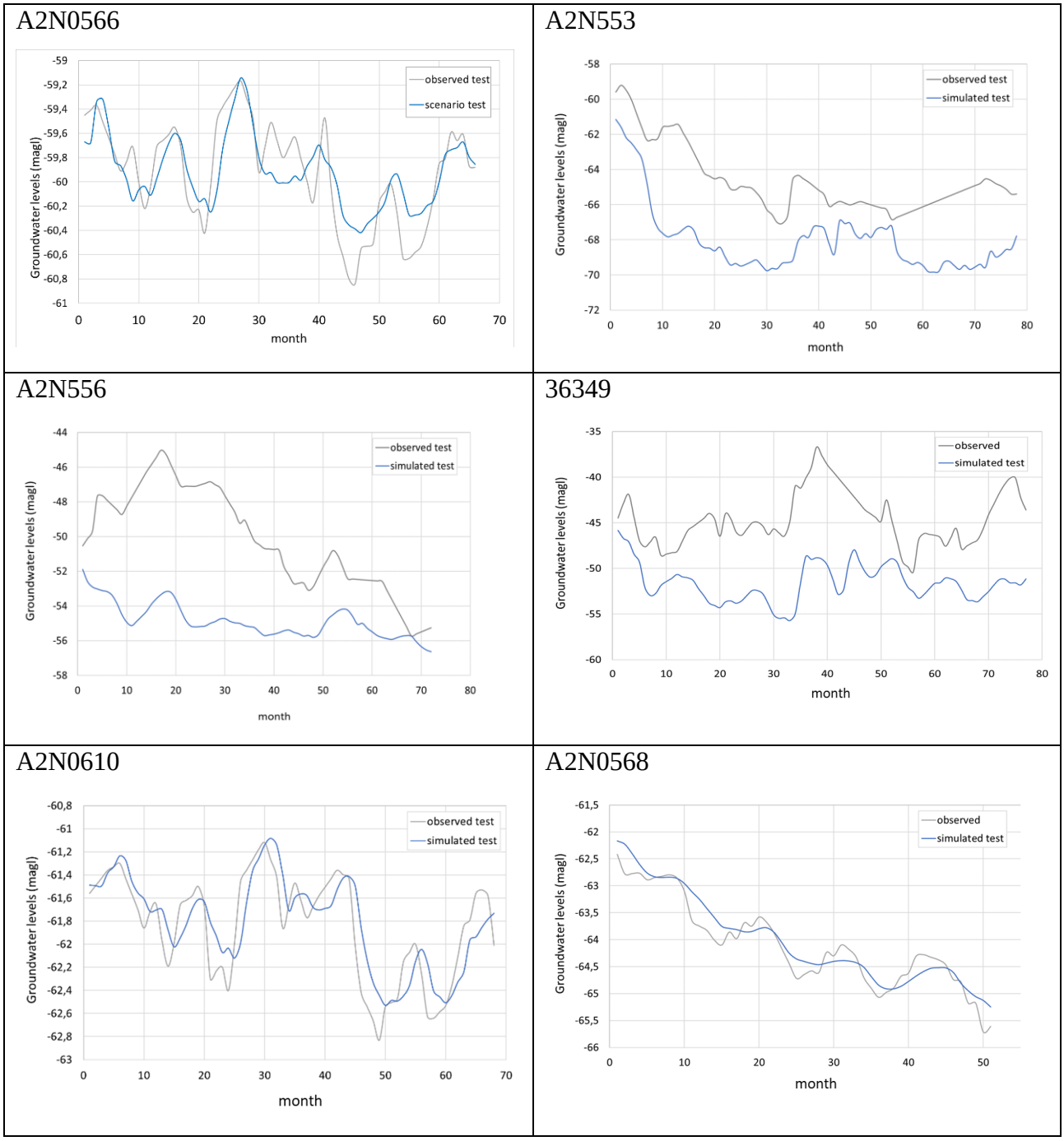


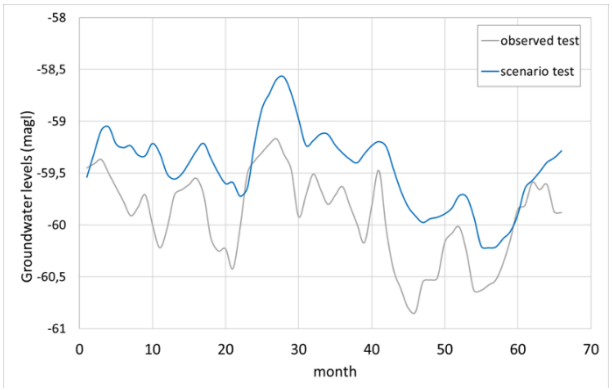
Figure 5-4: Model prediction for scenario 1b, halved monthly rainfall.

Figure 5-5 shows the model results for scenario 1c, where the peaks in rainfall over 100 mm were amplified (200 mm added to monthly rainfall over 100 mm). Similar to that described for scenario 1a and b, the model for boreholes A2N553, A2N556 and 36349 are influenced by the increased rainfall fed into the model, and react accordingly. The model simulated an increase in groundwater levels above the observed values. Boreholes A2N0610 and A2N0568 did not show a significant change or increase in the simulated groundwater levels for scenario 1c. The model simulated groundwater levels similar to the observed groundwater levels at these boreholes as the model does not recognise the increase in rainfall. Once again, this can be attributed to the strong MI between spring discharge and groundwater levels recorded at these boreholes.

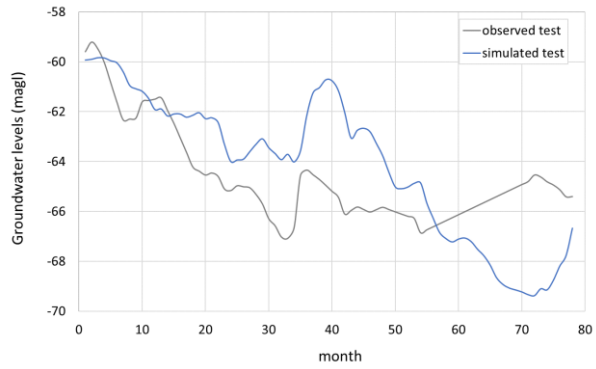
The model simulated an increase in groundwater levels for scenario1c. The model seemed to pick up the influence of the increase in rainfall modelled groundwater levels accordingly but the models do model a change for a decrease in rainfall.

The degree to which the groundwater level decreased in scenario 1a and b and increased in scenario 1c, depends on the degree to which the model picks up the scenario fed to the model. For example, in borehole A2N0566 of Figure 5-5, the simulated groundwater levels rise by less than 1 m, whereas in borehole A2N553, the groundwater levels rise by up to 4 m. The model can pick up the influence of the rainfall at borehole A2N553 better than at borehole A2N0566. This may be due to the differences in the recharge rate at each borehole or possibly because there is larger lag between a rainfall event and recharge to the aquifer. Holland *et al*. (2009) state that the recharge for the Steenkoppies compartment ranges from 9% to 21% of the mean annual precipitation. Fractures in the rock, different soil types, vegetation types and other factors allow for more or less recharge to enter the aquifer in the Steenkoppies compartment. Therefore, if a borehole is in an area where there is a higher recharge, the model's reaction to a change in rainfall scenario will be more substantial compared to a borehole in an area where there is less recharge.

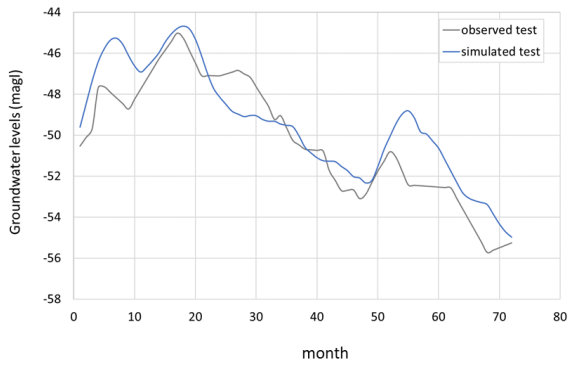
A2N0566



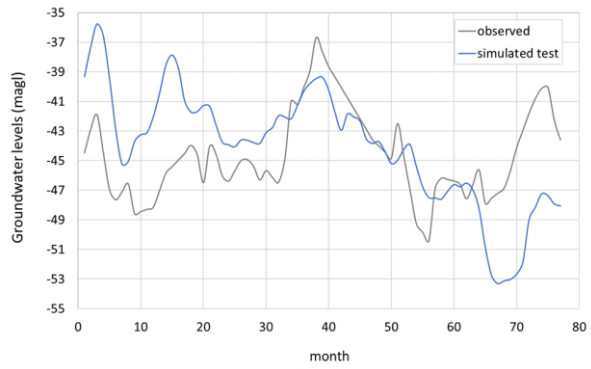
A2N553



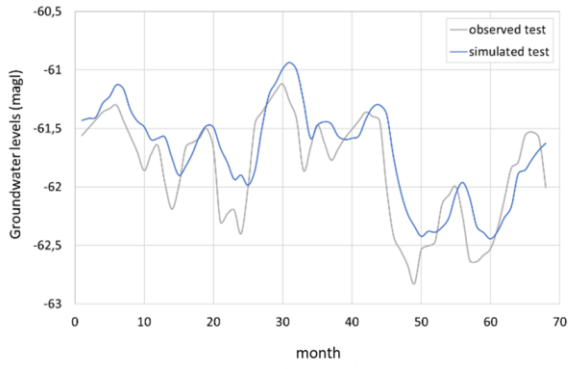
A2N556



36349



A2N0610



A2N0568

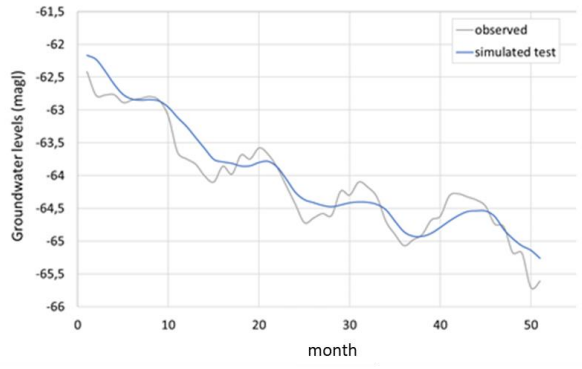


Figure 5-5: Model prediction for scenario 1c increase rainfall peaks

Figure 5-6 shows the effect of scenario 2, increased abstraction, on the groundwater levels (abstraction doubled). It is suspected that groundwater use across the Steenkoppies compartment has contributed to a decline in groundwater levels (Vahrmeijer *et al.*, 2013). If this is true, then the models should simulate a decrease in groundwater levels for scenario 2. It was uncertain whether the pattern between groundwater levels and groundwater usage would be detectable given the data set available for groundwater use (a step-wise increase in abstraction with time).

For boreholes A2N556, 36349, and A2N0610 (Figure 5-6) the model simulates groundwater levels 1-3 m lower than that of the observed groundwater levels. The correlation between increased groundwater abstraction and decreasing groundwater levels was detected. For boreholes A2N566 and A2N553 (Figure 5-6), the model recognises change, but as a linear pattern since the groundwater usage data is this step-wise increase graph. Thus the model simulated a linear groundwater level. The groundwater levels simulated at borehole A2N0568 do not change from the observed groundwater levels. This indicates that the model is unable to pick up the influence in groundwater abstraction for this borehole.

The scenario testing was done to show the potential of the model to predict groundwater levels for the future. Many studies (e.g. Tapoglou *et al.*, 2014; Lee *et al*., 2014) have predicted groundwater levels for scenarios of change using forecasted climate change data. The results from this study show that this would be possible with the NNAR in the Steenkoppies compartment. The NNAR model adequately recognised the relationships between the input variables and groundwater levels. The NNAR was clearly able to recognise there is a change and produces predictions that make sense according to hydrogeological knowledge (outlined in section 2.2) and the knowledge of the Steenkoppies compartment (outlined in Chapter 3).

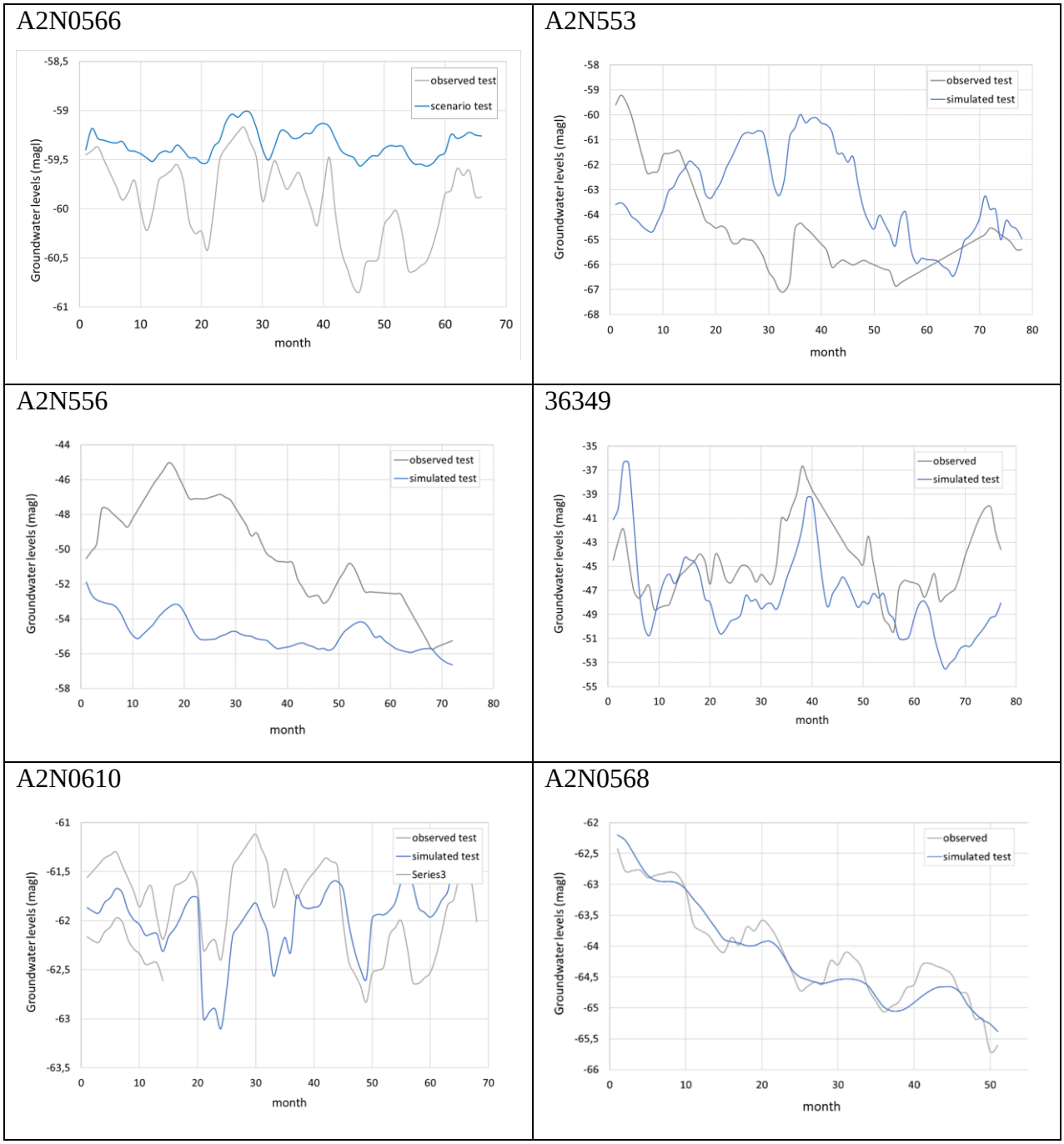


Figure 5-6: Model prediction for scenario 2, increase groundwater abstraction

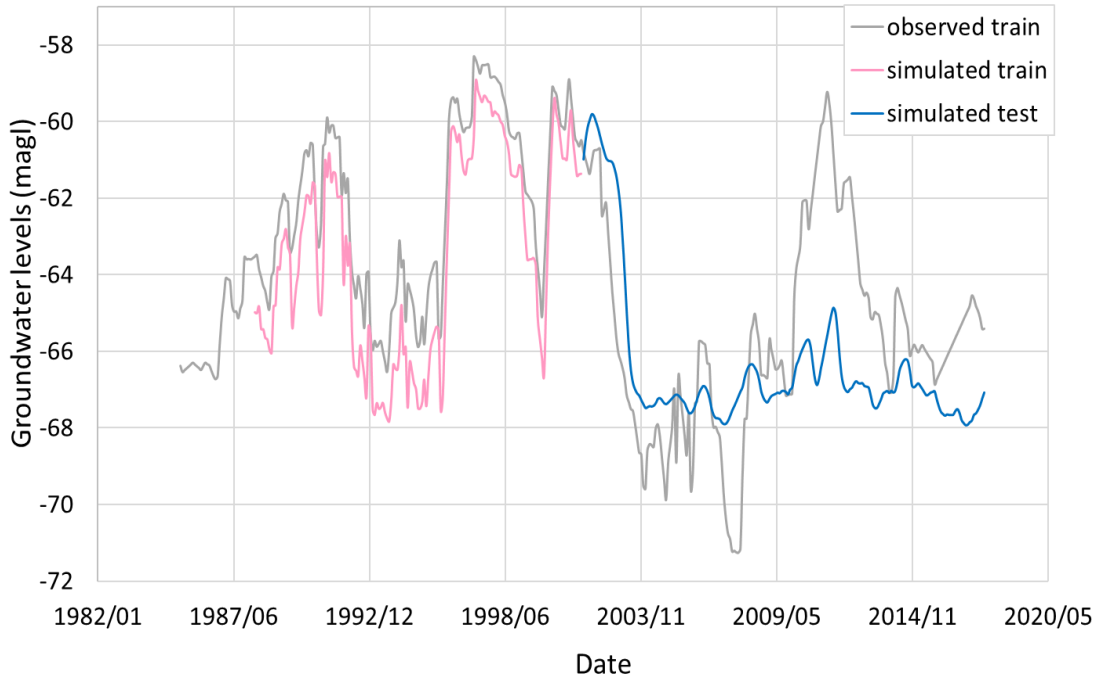
The long-term predictions in Figure 5-7 for scenario 3a demonstrate the impact of reducing the size of the training data set on model accuracy. The model could sufficiently predict groundwater levels for the first two years (2002-2004), however, after that the model accuracy decreased and the simulated output of the model does not match the observed. The small size of the training data set does not allow for the model to sufficiently "learn" the patterns seen in the test data set. ANNs Like the NNAR can optimise the model parameters required to model groundwater levels (hyperparameter tuning) and the predictions made vary based on the data used to train the model (MacKay, 2005; Wickham and Grolemund, 2016). This flexibility power in the model comes at the cost of requiring a lot more training data. Often predictions made by ANNs continue to improve the more data they train on (MacKay, 2005).

The results in scenario 3a (Figure 5-7) also demonstrates the importance of the training to test split. Since the more training data, the better the model is able to predict the target variable (MacKay, 2005); the split should be such that the model has sufficient data to learn from enough training data to make accurate predictions. This is why 80% - 20% training to test split ratio was chosen for this study.

Scenario 3b used the long-term averages of the input variables to predict a long term forecast of groundwater levels (results shown in Figure 5-7). This scenario was only performed on one borehole as it is merely a demonstration to show that it is possible to use the NNAR to make long-term predictions greater than 30 years. It is not realistic as the accuracy of the results cannot be verified. The result also shows that the more data available for training, the further one can predict, for example, a 30-year prediction could be made if approximately 120 years' worth of data were available and an 80% to 20% training to test split ratio was chosen. The importance of a large number of data points to conduct future predictions highlights the critical importance of collecting data proficiently over an extended period.

A2N553

3a



3b

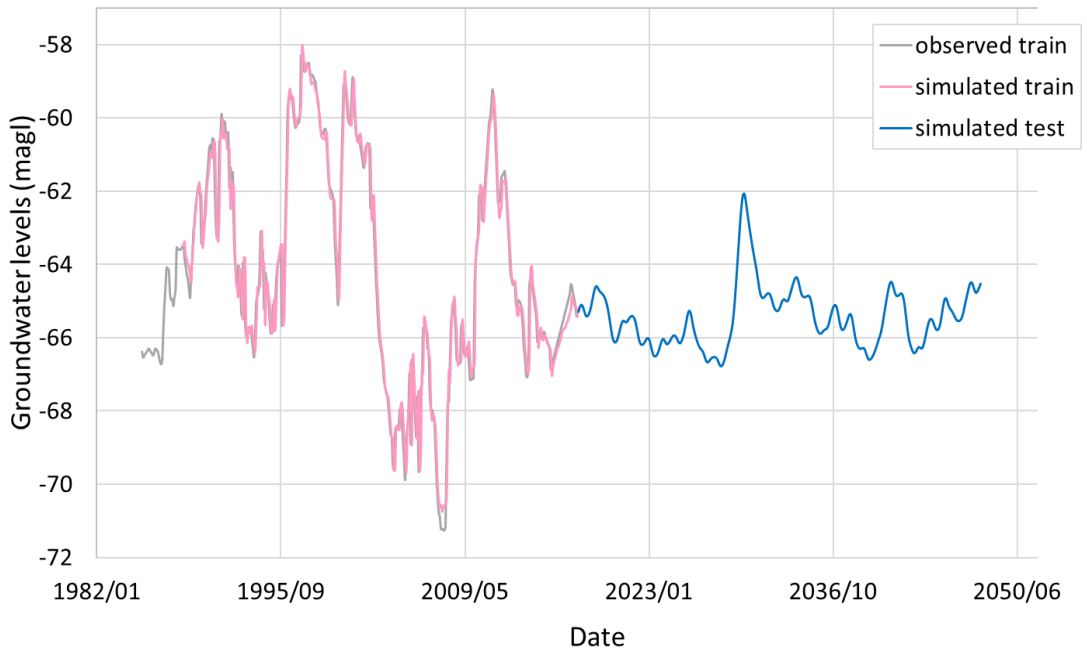


Figure 5-7: Model prediction for scenario 3a and b long term prediction at borehole A2N553

CHAPTER 6:
CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The suitability of the Neural Network Autoregression (NNAR) model to predict groundwater levels in the Steenkoppies compartment has been examined in this study. Additionally, the ability of the NNAR model to make predictions for scenarios of change, not seen in the data, was also examined. The study area is located in the Steenkoppies compartment of the Gauteng and North West dolomite aquifer in South Africa. Eighteen boreholes in the Steenkoppies compartment held sufficient data and were used for modelling. A important assumption of the approach is that groundwater levels in the Steenkoppies compartment are influenced by groundwater recharge, natural and artificial discharge and evapotranspiration. The models developed in the present study used rainfall, spring discharge, groundwater usage and temperature and as substitutes for these parameters due to their known correlations.

Prior to modelling groundwater levels using the NNAR, the mutual information (MI) between groundwater levels and each input variables were calculated. The results showed that spring discharge had the highest MI with respect to groundwater levels, rainfall trend second, groundwater usage third highest and temperature the lowest MI with respect to groundwater levels. These results correlate to the hydrogeological theory (outlined in Section 2.2) and the domain knowledge of the Steenkoppies aquifer (outlined in Chapter 3).

Compared to the results from studies where other forms of ANNs were to predict monthly groundwater levels (e.g. Daliakopoulos and Coulibaly 2005; Sreekanth *et al.,* 2009; Chang *et al*., 2016; Shamsuddin *et al.,* 2017) the results from this study demonstrate that the NNAR was able to learn the long term relationships and dependencies between the historical time series and make reliable groundwater level predictions (Table 5-2). The NNAR model was able to make accurate groundwater levels predictions, on par with the accuracies seen in other studies, at the majority (16 out of 18) of the boreholes in the Steenkoppies aquifer modelled.

The NNAR underestimated the groundwater levels peaks in the data. This observation suggested that the transformation applied to the data did not entirely transform the data to a normal distribution and that there was a slight negative skew to the data when it was modelled. The model then matched that negative skew pattern in the data to minimize errors resulting in an underestimation in the groundwater levels.

The results of the NNAR model also displayed a slight lag in the data, which is also seen in other studies where ANNs similar to the NNAR were used to predict groundwater levels (e.g. Wunsch *et al*., 2018; Guzman *et al*., 2017). This was thought to be a result of the recurrent nature of the NNAR model (i.e. the fact the model looks at lagged outputs to make predictions).

The NNAR model could successfully predict groundwater levels if the patterns in the test dataset were similar to that of the training data. This is because the NNAR makes predictions by learning the patterns and relationships in the training dataset and using this to make predictions. ANNs such as the NNAR learn patterns in the data during the training phase, and based on what it has learnt and the data from the input variables, the model makes predictions on the test data set (Adhikari and Agrawal, 2013). Hence, the patterns in the test data set, which deviate from the patterns in the training data set to decrease the performance as the model, would not have learnt this pattern. The model can only predict the peaks and troughs in the data when they have been previously seen during model training.

The relationship between the input variables and the target variables greatly facilitated the groundwater level predictions. Results in this study found that the better the correlation between the input variables and groundwater levels, the better the model could predict groundwater levels. The NNAR model was also found to be a useful tool to interpolate gaps in the groundwater level datasets in the Steenkoppies aquifer given that the model has sufficient data to train on.

The implication is that the use of the NNAR to predict groundwater levels may be less feasible in an aquifer setting where there is less of a relationship between rainfall and groundwater levels or where this relationship is muted (e.g. confined aquifers or arid areas where recharge is episodic). In this study, spring discharge was advantageous to predict groundwater levels. However, historical spring discharge data is not always available in an aquifer. Therefore, the successful application of the NNAR in different aquifer settings has not been concluded in this study.

The NNAR can be used to make predictions for scenarios of change, as the model was able to predict groundwater levels in response to the data fed to the model for the scenario of change. The NNAR could be used to make groundwater level predictions using climate change data, however, doing this can result in an accumulation of errors as both the forecasted climate change data and the prediction groundwater levels hold errors. This could results in predictions with larger errors.

6.2 RECOMMENDATIONS

The NNAR models in this study have been optimised to model groundwater levels in the Steenkoppies aquifer. The generalisability of the NNAR should be tested to see if they could be optimised to model groundwater levels at different circumstances. For example, similarly structured data (monthly groundwater levels) from a different aquifer system perhaps with a much more muted response to rainfall, such as the Karoo aquifers or data structured very differently such as 3-hourly groundwater levels from a pumped borehole, with abstraction at that same borehole. This would determine whether the NNAR can be used with the same model pipeline developed in this study, whether the model pipeline would need to change or whether an entirely different machine learning technique would need to be used for modelling groundwater levels in the different circumstances.

It is also recommended that further studies should be investigated on using machine learning not only to make predictions at a per borehole scale but aquifer scale. In situations where there are sufficient numbers of boreholes across an aquifer, developing a model per borehole (if each has sufficient historical data), and interpolating the impact from each borehole to derive an aquifer groundwater table is useful.

There is a limitation of scientists proficient in both hydrogeology and computer science in South Africa. Hydrogeologists may find it challenging to develop a machine learning model without having experience in computer coding or knowledge about the machine learning techniques available. During this thesis, I have tought my self computer coding and consider myself a part of the group of scientists developing skills in both hydrogeology and computer science. A “learn by doing” approach was applied to develop the NNAR model used in this study. However, once a model is developed, like the NNAR in this study, it is merely, simple and quick to apply to model groundwater levels. Therefore, the application of machine learning should be more widely investigated and applied.

Data quantity and quality is a limitation to the application of machine learning to model groundwater levels. In South Africa, extensive good quality meteorological and hydrogeological datasets are not available for all aquifer systems. Therefore it is recommended that a concerted effort should be placed on groundwater monitoring in South Africa so that in the future, machine learning techniques can be applied in many aquifer settings in South Africa. Since the NNAR was able to fill data gaps, it is recommended that the groundwater community, with guidance from computer scientists, should look into the use of machine learning to improve groundwater data sets.

REFERENCES

Adelabu, S., Mutanga, O. and Adam, E. (2015). Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto International*, *30*(7), 810-821.

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

Agatonovic-Kustrin, S. and Beresford, R. (2000). Basic concepts of Artificial Neural Network (ANN) modelling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, *22*(5), 717-727.

Ajoodha, R. and Rosman, B. (2020). Discovery of Influence between Processes Represented by Hidden Markov Models. In: *2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, Vancouver, Canada, 9-12 September, pp. 1-7. IEEE.

Alley, W.M. and Leake, S.A. (2004). The journey from safe yield to sustainability. *Groundwater*, *42*(1), 12-16.

Alley, W.M., Reilly, T.E. and Franke, O.L. (1999). *Sustainability of ground-water resources* (Vol. 1186). US Department of the Interior, US Geological Survey.

Altchenko, Y. and Villholth, K.G. (2013). Transboundary aquifer mapping and management in Africa: a harmonised approach. *Hydrogeology Journal*, *21*(7), 1497-1517.

Aziz, A.R.A. and Wong, K.F.V. (1992). A neural-network approach to the determination of aquifer parameters. *Groundwater*, *30*(2), 164-166.

Beetlestone, P. (2009). Challenges to transboundary aquifer management in the SADC region. *SADC Infrastructure and Services Directorate—Water Division: Gaborone, Botswana*.

Bredehoeft, J. and Durbin, T. (2009). Groundwater development—The time to full capture problem. *Groundwater*, *47*(4), 506-514.

Bredehoeft, J.D., S.S. Papadopulos and H.H. Cooper, (1982). Groundwater: The water budget myth. In *Scientific Basis of Water-Resource Management, Studies in Geophysics*, Washington DC: National Academy Press, 51–57.

Bredehoeft, J.D. (2002). The water budget myth revisited: why hydrogeologists model. *Groundwater*, *40*(4), pp.340-345.

Brezak, D., Bacek, T., Majetic, D., Kasac, J. and Novakovic, B. (2012). A comparison of feed-forward and recurrent neural networks in time series forecasting. In: *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, New York, NY, USA, 29-30 March 2012, pp. 1-6. IEEE.

Brown, R.H. (1963). The cone of depression and the area of diversion around a discharging well in an infinite strip aquifer subject to uniform recharge. *US Geological Survey Water-Supply Paper C*, *1545*, C69-C85.

Chang, F.J., Chang, L.C., Huang, C.W. and Kao, I.F. (2016). Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *Journal of Hydrology*, *541*, 965-976.

Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of official statistics*, *6*(1), 3-73.

Cobbing, J. (2018). *The North West dolomite aquifers, South Africa: a stalled opportunity for water security and development* (Vol. 3). International Water Management Institute (IWMI).

Cobbing, J., Eales, K. and Rossouw, T. (2016). *The path to successful water user associations in the north west dolomite aquifers*. WRC Report. No. 2429/1/16. Pretoria, South Africa: Water Research Commission.

Coppola, E., Poulton, M., Charles, E., Dustman, J. and Szidarovszky, F. (2003). Application of artificial neural networks to complex groundwater management problems. *Natural Resources Research*, *12*(4), 303-320.

Crowther, P.S. and Cox, R.J. (2005). A method for optimal division of data sets for use in neural networks. In: *International conference on knowledge-based and intelligent information and engineering systems*, Heidelberg, Berlin, 14 – 16 September pp. 1-7. Springer

Da Silva, I.N., Spatti, D.H., Flauzino, R.A., Liboni, L.H.B. and dos Reis Alves, S.F. (2017). Artificial neural network architectures and training processes. In: *Artificial neural networks* (pp. 21-28). Springer International Publishing, Switzerland.

Daliakopoulos, I.N., Coulibaly, P. and Tsanis, I.K. (2005). Groundwater level forecasting using artificial neural networks. *Journal of hydrology*, *309*(1-4), 229-240.

Davies, J., Robins, N.S., Farr, J., Sorensen, J., Beetlestone, P. and Cobbing, J.E. (2013). Identifying transboundary aquifers in need of international resource management in the Southern African Development Community region. *Hydrogeology Journal*, *21*(2), 321-330.

Department of Water and Sanitation (DWA). (2009). Dolomite Guideline: A short guide to available documents on procedures for developing dolomitic land. DWA, Pretoria.

Devlin, J.F. and Sophocleous, M. (2005). The persistence of the water budget myth and its relationship to sustainability. *Hydrogeology Journal*, *13*(4), 549-554.

Döll, P. (2009). Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. Environmental Research Letters, 4(3), 035006.

Engelbrecht, A.P. (2007). *Computational intelligence: an introduction*. John Wiley & Sons.

Foster, S.S.D. (2000). Assessing and controlling the impacts of agriculture on groundwater—from Barley Barons to Beef Bans. *Quarterly Journal of Engineering Geology and Hydrogeology*, *33*(4), 263-280.

French, M.N., Krajewski, W.F. and Cuykendall, R.R., (1992). Rainfall forecasting in space and time using a neural network. *Journal of hydrology*, *137*(1-4), 1-31.

Gao, Y. and Er, M.J., (2005). NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches. *Fuzzy sets and systems*, *150*(2), 331-350.

García, S., Luengo, J. and Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, *98*, 1-29.

Gleeson, T., Alley, W.M., Allen, D.M., Sophocleous, M.A., Zhou, Y., Taniguchi, M. and VanderSteen, J. (2012). Towards sustainable groundwater use: setting long-term goals, backcasting, and managing adaptively. *Groundwater*, *50*(1), 19-26.

Gulliver, J.S., Erickson, A.J. and Weiss, P., 2010. Stormwater treatment: Assessment and maintenance. *University of Minnesota, St. Anthony Falls Laboratory. Minneapolis, MN*.

Guzman, S.M., Paz, J.O. and Tagert, M.L.M., (2017). The use of NARX neural networks to forecast daily groundwater levels. *Water resources management*, *31*(5), 1591-1603.

Groundwater Consultants. 2001. Development of a Code of Good Practice for Groundwater Development in the SADC Region. Report No.2 (Final). Guidelines for the groundwater development in the SADC region. For Southern African Development Community (SADC) Water Sector Coordination Unit (WSCU). By groundwater consultants November 2001.

Healy, R.W., Winter, T.C., LaBaugh, J.W. and Franke, O.L. (2007). *Water budgets: foundations for effective water-resources and environmental management* (Vol. 1308). Reston, Virginia: US Geological Survey.

Hiscock, K.M., Rivett, M.O. and Davison, R.M. (2002). Sustainable groundwater development. *Geological Society, London, Special Publications*, *193*(1), 1-14.

Holland, M., Wiegmans, F., Cobbing, J. and Witthüser, K.T. (2009). *Geohydrological assessment of the Steenkoppies dolomite compartment.* Project no. 14/14/5/2. Activity 25. Water Geosciences Consulting for Department of Water Affairs. Pretoria

Hsu, K.L., Gupta, H.V. and Sorooshian, S. (1995). Artificial neural network modelling of the rainfall-runoff process. *Water resources research*, *31*(10), 2517-2530.

Huang, M. and Tian, Y. (2015). Prediction of groundwater level for sustainable water management in an arid basin using data-driven models. In: *2015 International Conference on Sustainable Energy and Environmental Engineering*. Shenzhen, China, 20-21 December, Atlantis Press.

Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Izady, A., Davary, K., Alizadeh, A., Nia, A.M., Ziaei, A.N. and Hasheminia, S.M. (2013). Application of NN-ARX model to predict groundwater levels in the Neishaboor Plain, Iran. *Water resources management*, *27*(14), 4773-4794.

Johnson, M.R., Anhaeusser, C.R. and Thomas, R.J. (2006). The Geology of South Africa: Geological Society of South Africa. *Johannesburg and the Council for Geoscience, Pretoria.*

Kenda, K., Čerin, M., Bogataj, M., Senožetnik, M., Klemen, K., Pergar, P., Laspidou, C. and Mladenić, D. (2018). Groundwater modelling with machine learning techniques: Ljubljana polje aquifer. *Multidisciplinary Digital Publishing Institute Proceedings* 2(11), 697.

Kendy, E. (2003). The false promise of sustainable pumping rates. *Ground Water*, *41*(1), 2-4.

Khalek, M.A. and Ali, M.A. (2016). Comparative Study of Wavelet-SARIMA and Wavelet-NNAR Models for Groundwater Level in Rajshahi District. *IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)*, *10*, 1-15.

Konikow, L.F. and Leake, S.A. (2014). Depletion and capture: revisiting "the source of water derived from wells". *Groundwater*, *52*(S1), 100-111.

Konikow. L. and J. Bredehoeft. (2020). Groundwater Resource Development: Effects and Sustainability. In; The Groundwater Project. J. Cherry. *Currently in press.*

Kotsiantis, S.B., Kanellopoulos, D. and Pintelas, P.E. (2006). Data preprocessing for supervised learning. International Journal of Computer Science, 1(2), 111-117.

Kraskov, A., Stögbauer, H. and Grassberger, P. (2004). Estimating mutual information. *Physical review E*, 69(6), 066138.

Krause, P., Boyle, D.P. and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences, European Geosciences Union*, 5, 89-97.

Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer.

Lafare, A.E., Peach, D.W. and Hughes, A.G. (2016). Use of seasonal trend decomposition to understand groundwater behaviour in the Permo-Triassic Sandstone aquifer, Eden Valley, UK. *Hydrogeology Journal*, *24*(1), 141-158.

Lee, C.H., (1908). Water resources of a part of Owens Valley, California. *US Geological Survey Water Supply Paper*, *294*, p.135.

Lee, B., Hamm, S.Y., Jang, S., Cheong, J.Y. and Kim, G.B. (2014). Relationship between groundwater and climate change in South Korea. Geosciences Journal, 18(2), 209-218.

Lee, C.H. (1915). The determination of safe yield of underground reservoirs of the closed basin type. *Transactions, American Society of Civil Engineers* 78, 148-251.

Lee, D.K., In, J. and Lee, S. (2015). Standard deviation and standard error of the mean. *Korean Journal of Anesthesiology*, *68*(3), 220.

Lee, K.Y., Chung, N. and Hwang, S. (2016). Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecological informatics*, 36, 172-180.

Lee, S., Lee, K.K. and Yoon, H. (2019). Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. *Hydrogeology Journal*, *27*(2), 567-579.

MacKay, D.J. (2005). *Information theory, inference and learning algorithms*. Cambridge university press.

Maier, H.R. and Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental modelling & software*, *15*(1), 101-124.

Maier, H.R., Jain, A., Dandy, G.C. and Sudheer, K.P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental modelling & software*, *25*(8), 891-909.

Marinósdóttir, H. (2019). *Applications of different machine learning methods for water level predictions*. Master's thesis, Reykjavík University.

MATLAB, 2010. *version 7.10.0 (R2010a)*, Natick, Massachusetts: The MathWorks Inc.

McGill, B.M., Altchenko, Y., Hamilton, S.K., Kenabatho, P.K., Sylvester, S.R. and Villholth, K.G. (2019). Complex interactions between climate change, sanitation, and groundwater quality: a case study from Ramotswa, Botswana. *Hydrogeology Journal*, *27*(3), 997-1015.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115-133.

Meinzer, O.E. (1931). Outline of methods for estimating ground-water supplies: In contributions to the hydrology of the United States. USGS Water-Supply Paper 638. Reston, Virginia: USGS.

Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900.

Nayak, P.C., Rao, Y.S. and Sudheer, K.P. (2006). Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water resources management*, *20*(1), 77-90.

North, R.P. and Livingstone, D.M. (2013). Comparison of linear and cubic spline methods of interpolating lake water column profiles. *Limnology and Oceanography: Methods*, *11*(4), 213-224.

Pani, L., Karmakar, S., Misra, C. and Dash, S.R. (2019). Multilevel Classification Framework of MRI Data: A Big Data Approach. In *Big Data Analytics for Intelligent Healthcare Management*. Academic Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Price, M. (2002). Who needs sustainability?. *Geological Society, London, Special Publications*, *193*(1), 75-81.

Python Core Team (2015). Python: A dynamic, open source programming language. Python Software Foundation. URL https://www.python.org/.

QGIS.org (2020). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.org

Qin, S., Li, S., Kang, S., Du, T., Tong, L. and Ding, R. (2016). Can the drip irrigation under film mulch reduce crop evapotranspiration and save water under the sufficient irrigation condition?. *Agricultural Water Management*, 177, 128-137.

Quiza, R. and Davim, J.P. (2011). Computational methods and optimization. In *Machining of hard materials* (pp. 177-208). Springer, London.

Ross, B.C. (2014). Mutual information between discrete and continuous data sets. *PloS one*, *9*(2), e87357.

Ruiz, L.G.B., Cuéllar, M.P., Calvo-Flores, M.D. and Jiménez, M.D.C.P. (2016). An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies*, *9*(9), 684.

Saatsaz, M., Chitsazan, M., Eslamian, S. and Sulaiman, W.N.A. (2011). The application of groundwater modelling to simulate the behaviour of groundwater resources in the Ramhormooz Aquifer, Iran. *International Journal of Water*, *6*(1-2), 29-42.

Sahoo, S., Russo, T.A., Elliott, J. and Foster, I. (2017). Machine learning algorithms for modelling groundwater level changes in agricultural regions of the US. *Water Resources Research*, *53*(5), 3878-3895.

Scardapane, S. and Wang, D. (2017). Randomness in neural networks: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2), e1200.

Seyler, H., Witthüser, K. and Holland, M. (2016). *The capture principle approach to sustainable groundwater use incorporating sustainability indicators and decision framework for sustainable groundwater use*. WRC Report. No. 2311/1/17. Pretoria, South Africa: Water Research Commission.

Shahin, M.A., Jaksa, M.B. and Maier, H.R. (2008). State of the art of artificial neural networks in geotechnical engineering. *Electronic Journal of Geotechnical Engineering*, *8*(1), 1-26.

Shamsuddin, M.K.N., Kusin, F.M., Sulaiman, W.N.A., Ramli, M.F., Baharuddin, M.F.T. and Adnan, M.S., (2017). Forecasting of Groundwater Level using Artificial Neural Network by incorporating river recharge and river bank infiltration. In: 2017 *MATEC Web of Conferences*, 18-20 April, Vol. 103, pp. 04007. EDP Sciences.

Shirmohammadi, B., Vafakhah, M., Moosavi, V. and Moghaddamnia, A. (2013). Application of several data-driven techniques for predicting groundwater level. *Water Resources Management*, *27*(2), 419-432.

Sreekanth, P.D., Geethanjali, N., Sreedevi, P.D., Ahmed, S., Kumar, N.R. and Jayanthi, P.K. (2009). Forecasting groundwater level using artificial neural networks. *Current Science*, 933-939.

Sudheer, K.P., Nayak, P.C. and Ramasastri, K.S. (2003). Improving peak flow estimates in artificial neural network river flow models. *Hydrological Processes*, 17(3), 677-686.

Tapoglou, E., Trichakis, I.C., Dokou, Z., Nikolos, I.K. and Karatzas, G.P. (2014). Groundwater-level forecasting under climate change scenarios using an artificial neural network trained with particle swarm optimization. Hydrological Sciences Journal, 59(6), 1225-1239.

Taver, V., Johannet, A., Borrell-Estupina, V. and Pistre, S. (2015). Feed-forward vs recurrent neural network models for non-stationarity modelling using data assimilation and adaptivity. *Hydrological sciences journal*, *60*(7-8), 1242-1265.

Theis, C.V. (1940). The source of water derived from wells. *Civil Engineering*, *10*(5), 277-280.

Vahrmeijer, J.T., Annandale, J.G., Bristow, K.L., Steyn, J.M. and Holland, M. (2013). Drought as a catalyst for change: A case study of the Steenkoppies Dolomitic aquifer. In *Drought in Arid and Semi-Arid Regions* (pp. 251-268). Springer, Dordrecht.

van der Gun, J., Lipponen, A. (2010). Reconciling Groundwater Storage Depletion Due to Pumping with Sustainability. *Sustainability*, 10(2), 3418-3435.

van der Gun, J. (2017). Data, information, knowledge and diagnostics on groundwater. In Advances in groundwater governance (pp. 193-213). CRC Press.

Walter, M. (2010). Managing Transboundary Aquifers: Lessons from the Field. In *Research paper collection, International Conference "Transboundary Aquifers: Challenges and New Directions (ISARM)* (pp. 1-7).

Wickham, H. and Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. " O'Reilly Media, Inc.".

Wiegmans, F., Holland, M. and Janse van Rensburg, H., 2013. *Groundwater Resource Directed Measures for Maloney's Eye Catchment.* WRC Report. No. KV 319/13. Pretoria, South Africa: Water Research Commission.

Wu, W.Y., Lo, M.H., Wada, Y., Famiglietti, J.S., Reager, J.T., Yeh, P.J.F., Ducharne, A. and Yang, Z.L. (2020). Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers. *Nature communications*, *11*(1), 1-9.

Wunsch, A., Liesch, T. and Broda, S. (2018). Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). *Journal of Hydrology*, *567*, 743-758.

Xie, J. and Wang, Q. (2018) Benchmark machine learning approaches with classical time series approaches on the blood glucose level prediction challenge. In *KHD@ IJCAI*.

Xu, L., Chen, N., Zhang, X. and Chen, Z. (2020). A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale. Climate Dynamics, 1-20.

Ye, J., 2015. *Using machine learning for exploratory data analysis and predictive modelling*. Master's thesis, University of Stavanger, Norway.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O. and Lee, K.K. (2011). A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of hydrology*, *396*(1-2), 128-138.

Zanotti, C., Rotiroti, M., Sterlacchini, S., Cappellini, G., Fumagalli, L., Stefania, G.A., Nannucci, M.S., Leoni, B. and Bonomi, T. (2019). Choosing between linear and nonlinear models and avoiding overfitting for short and long term groundwater level forecasting in a linear system. *Journal of Hydrology*, 578,124015.

Zhang, G. and Hu, M.Y. (1998). Neural network forecasting of the British pound/US dollar exchange rate. *Omega*, *26*(4), 495-506.

Zhao, W.G., Wang, H. and Wang, Z.J. (2011). Groundwater level forecasting based on support vector machine. In *Applied Mechanics and Materials* (Vol. 44, pp. 1365-1369). Trans Tech Publications Ltd.

APPENDIX A

Table A-1: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0612 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

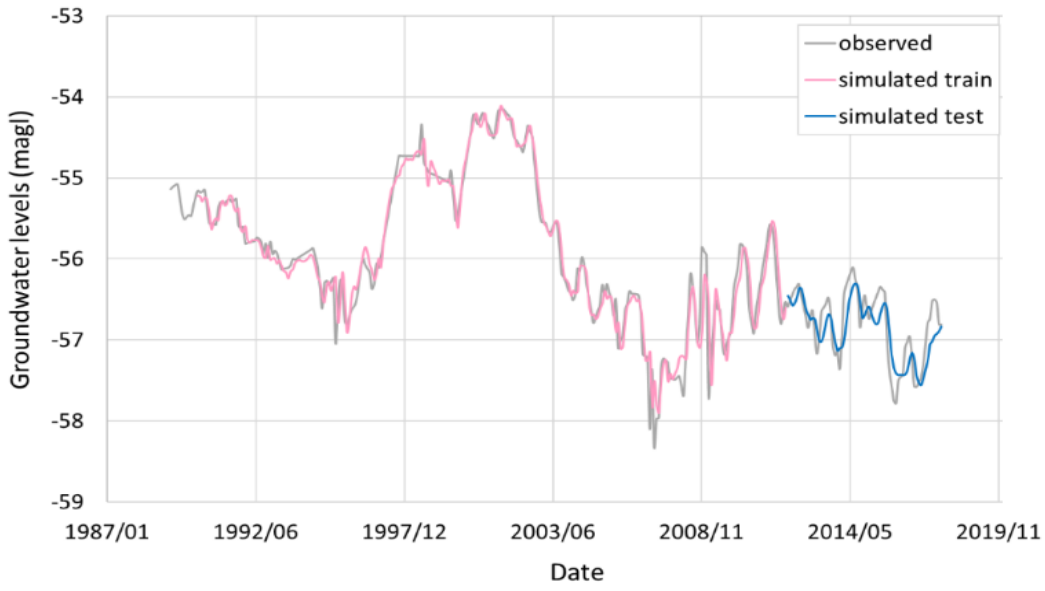
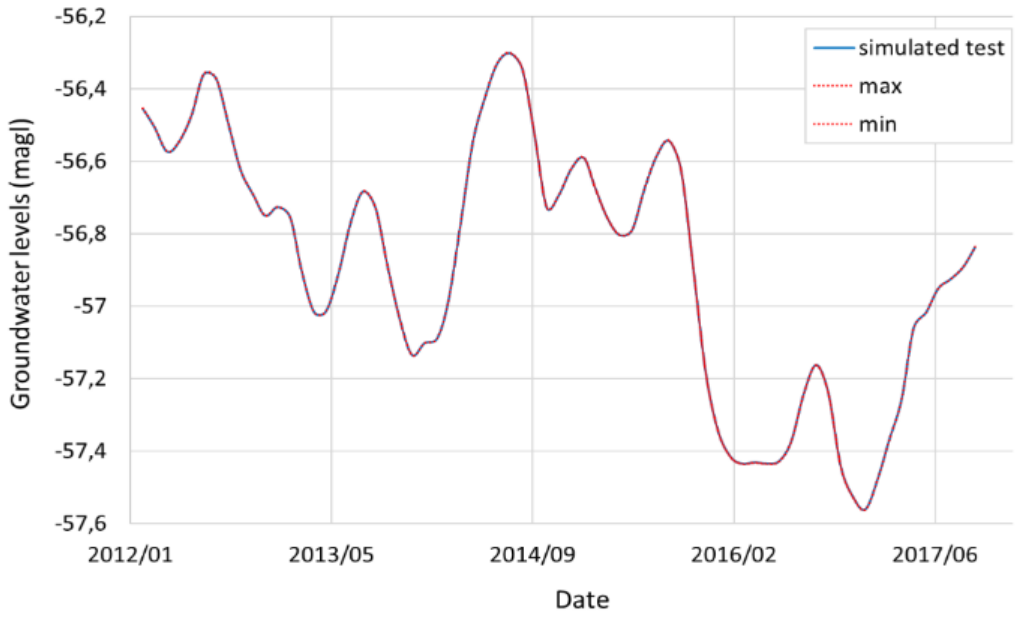
Metrics	A2N0612	
R^2	0.47	
MSE	0.06	
RMSE	0.25	
MAE	0.21	
STDV	<0.01	

Table A-2: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0617 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

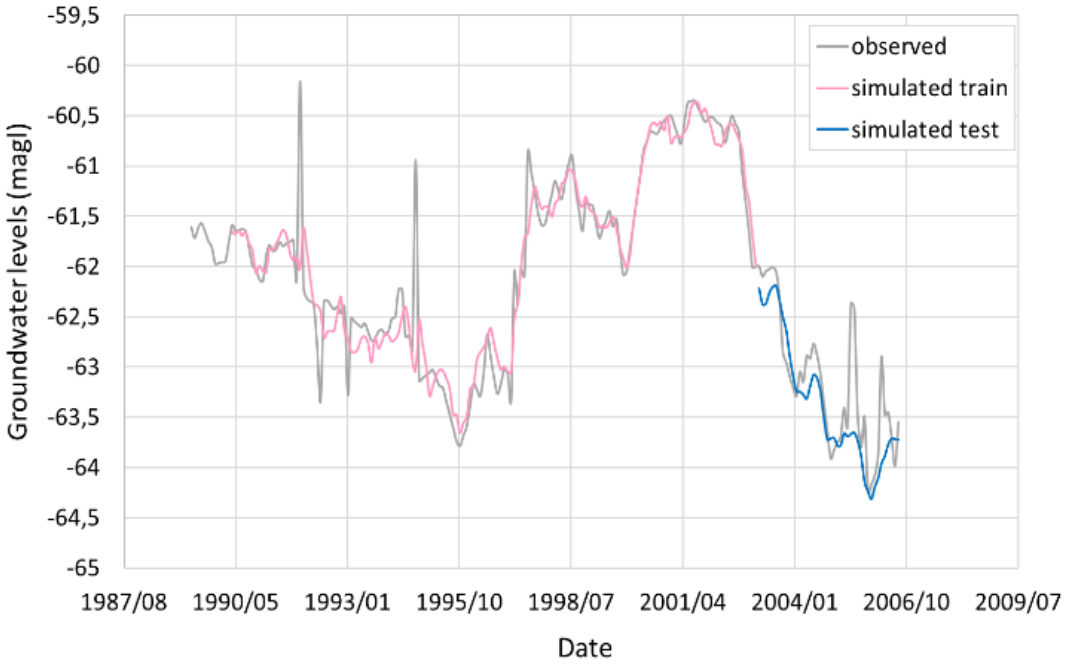
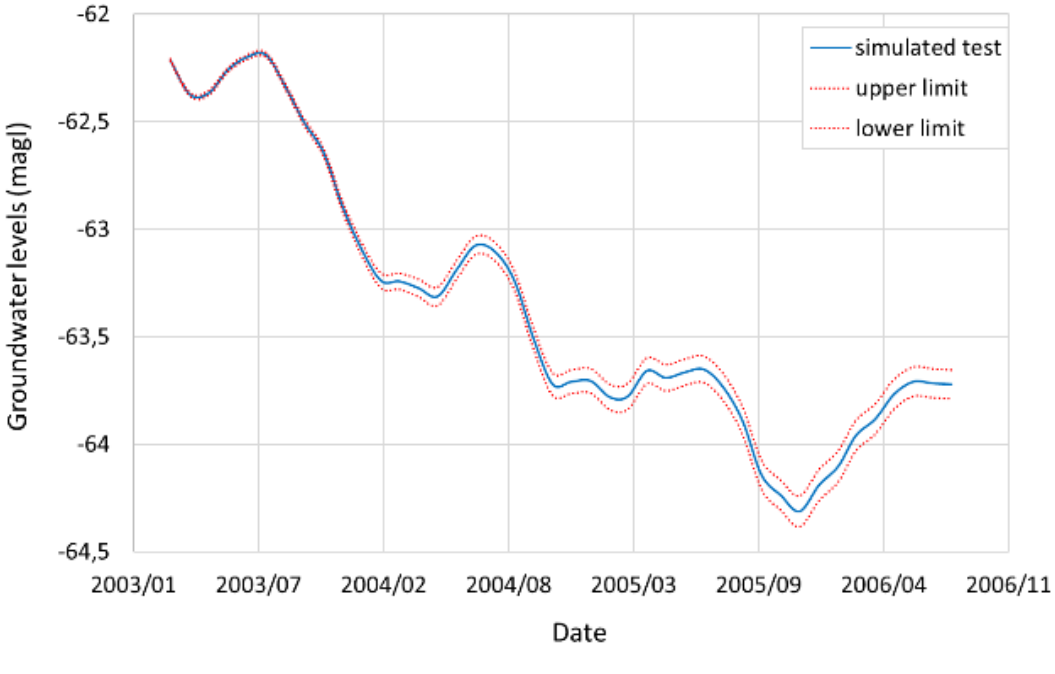
Metrics	A2N0617	
R^2	0.59	
MSE	0.16	
RMSE	0.40	
MAE	0.28	
STDV	0.02	

Table A-3: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0567 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

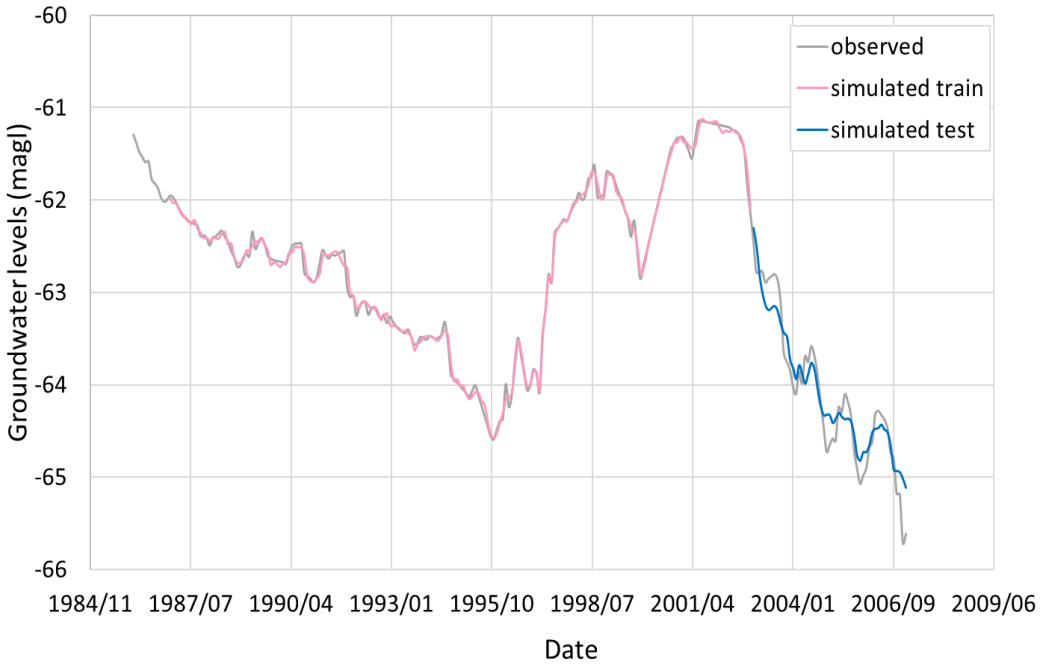
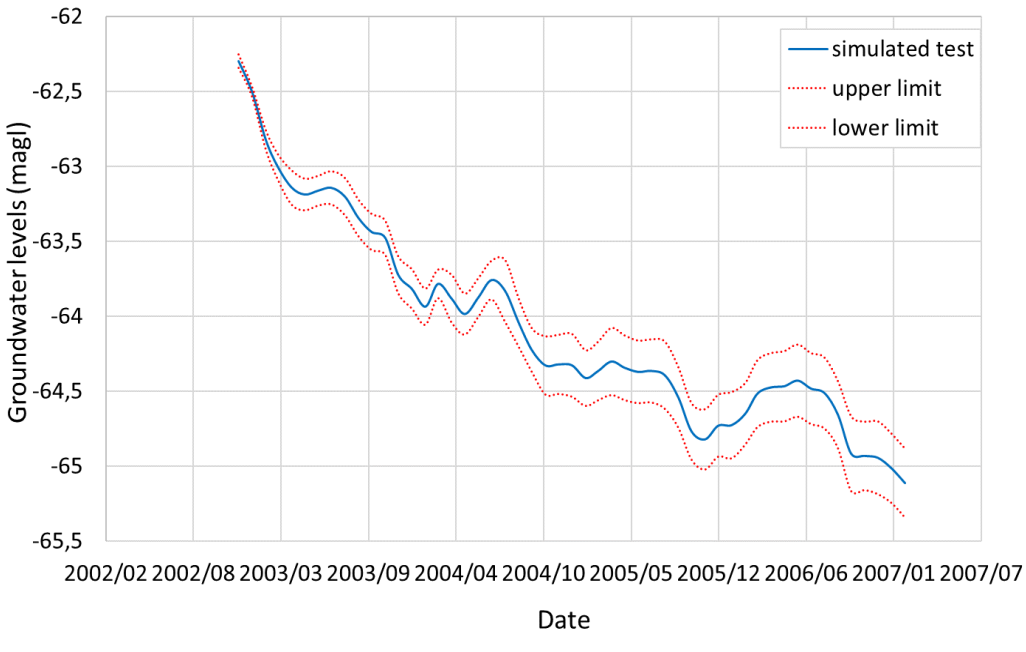
Metrics	A2N0567	
R^2	0.56	
MSE	0.09	
RMSE	0.29	
MAE	0.24	
sd	0.04	

Table A-4: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0616 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

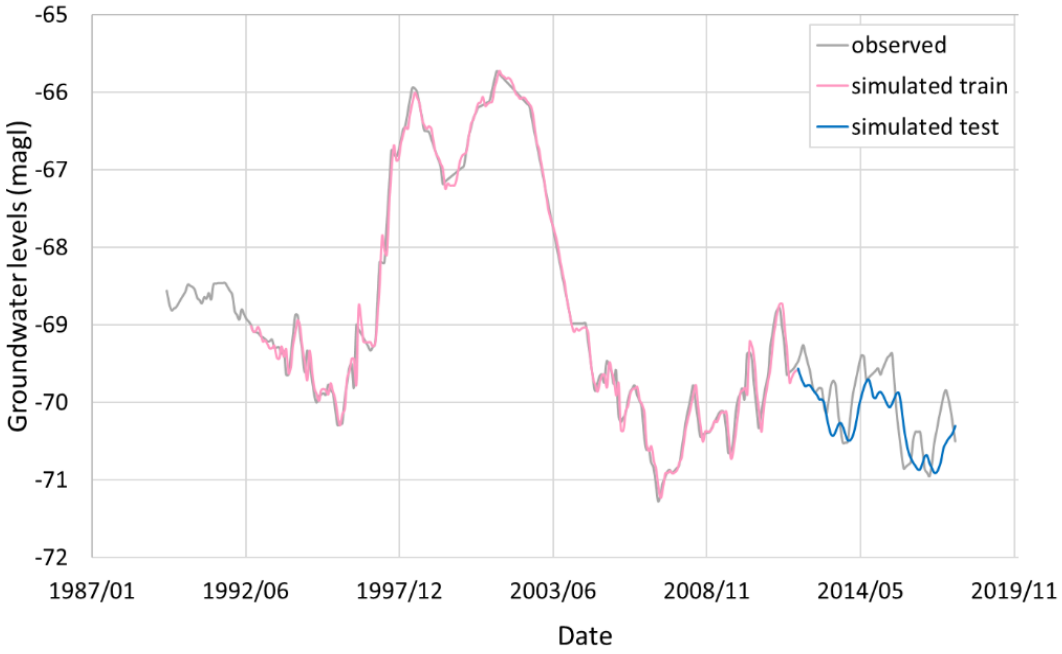
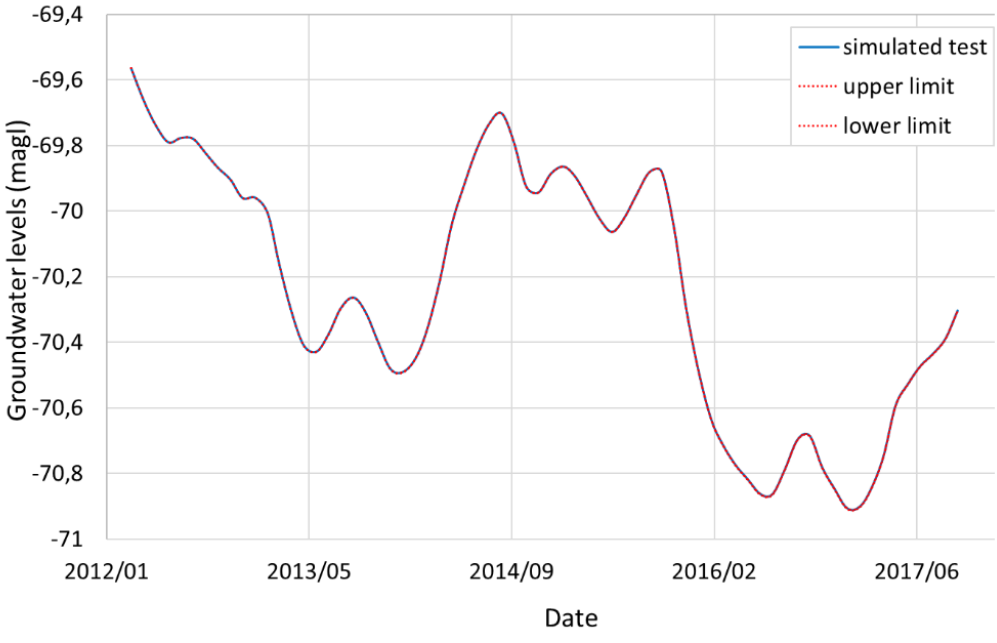
Metrics	A2N0616	
R^2	-0.07	
MSE	0.16	
RMSE	0.39	
MAE	0.33	
sd	<0.01	

Table A-5: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0615 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

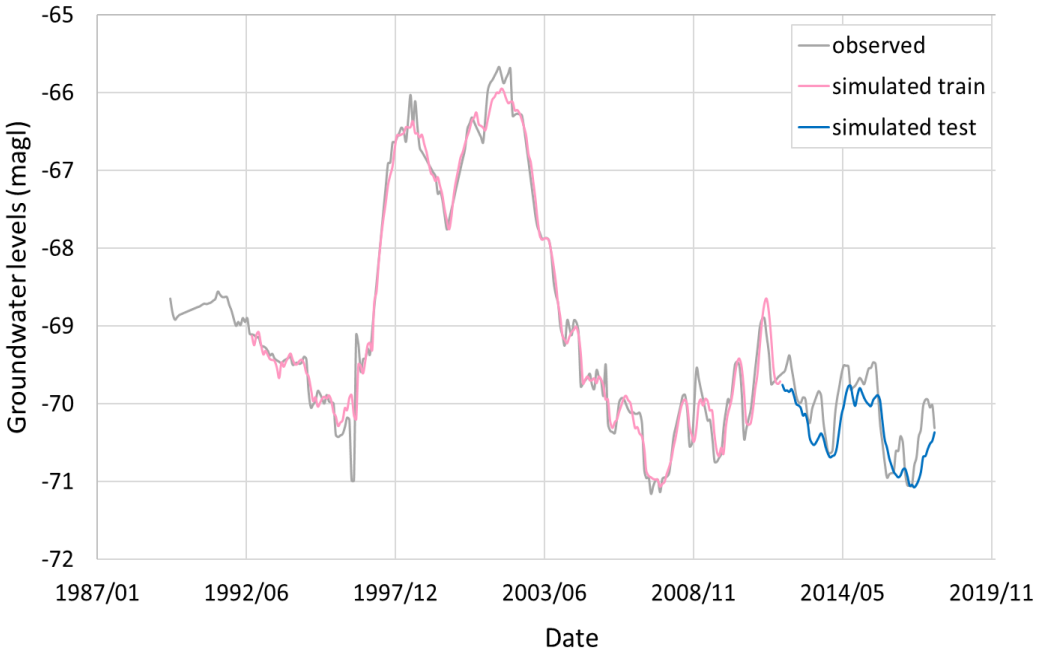
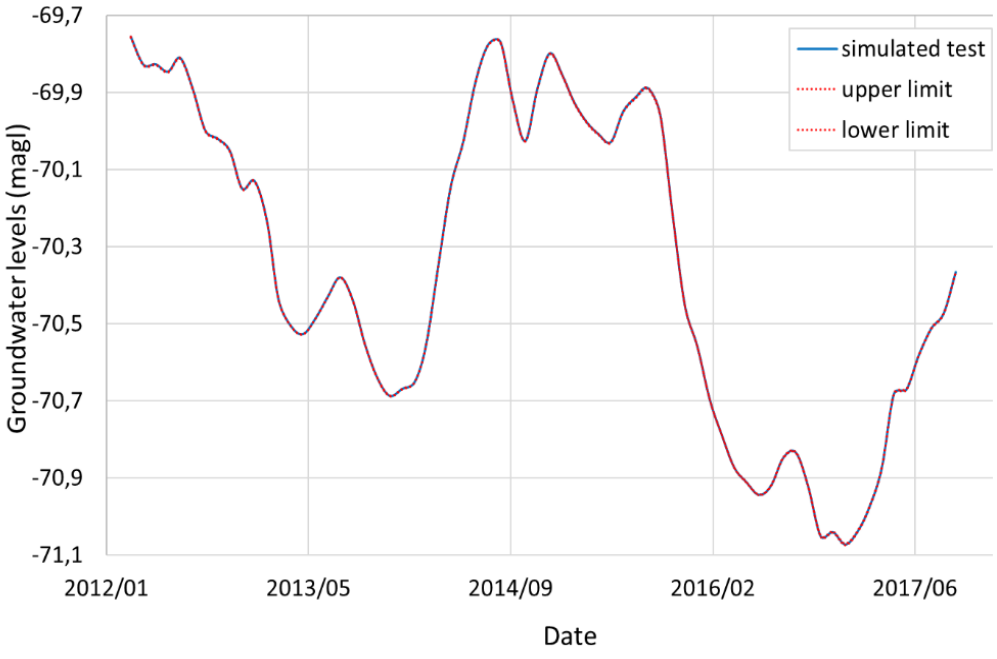
Metrics	A2N0615	
R^2	0.25	
MSE	0.13	
RMSE	0.36	
MAE	0,30	
sd	<0.01	

Table A-6: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0614 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

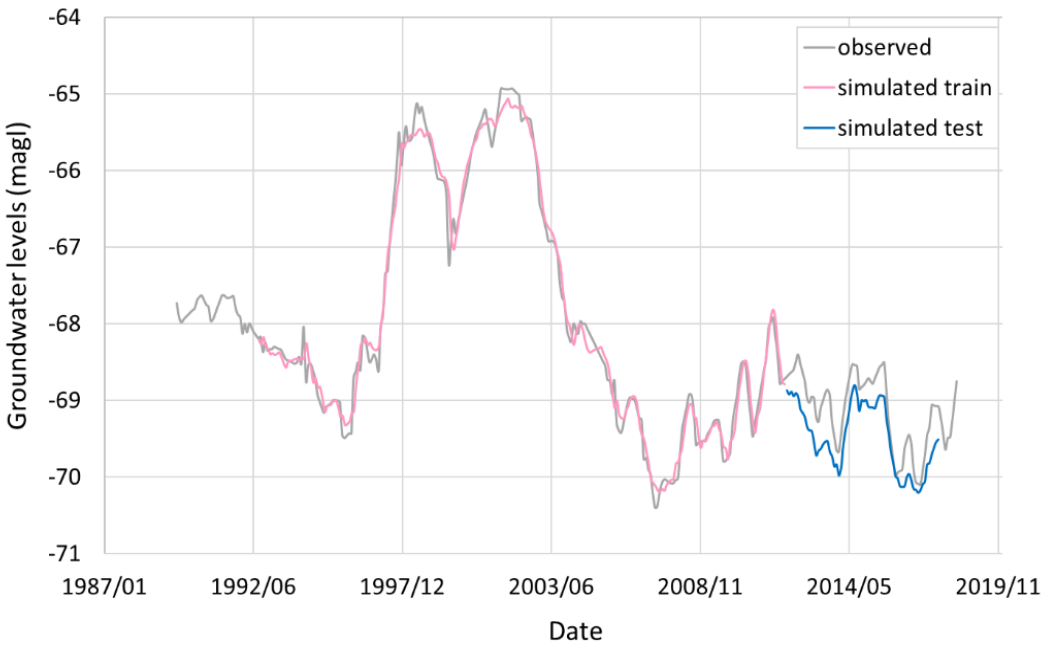
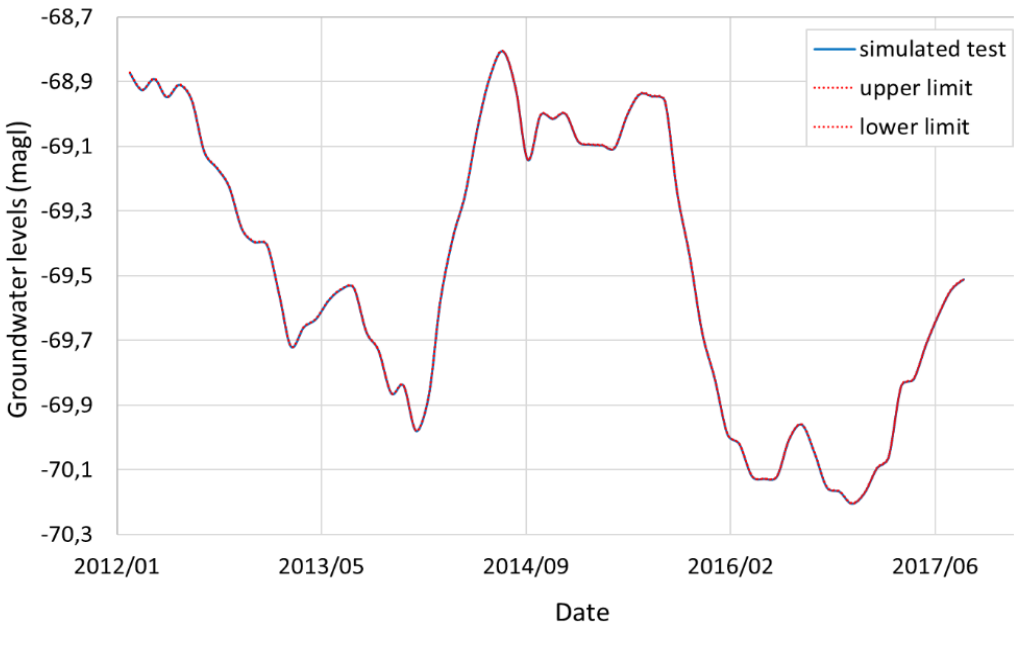
Metrics	A2N0614	
R^2	0.14	
MSE	0.16	
RMSE	0.40	
MAE	0.37	
sd	<0.01	

Table A-7: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0608 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

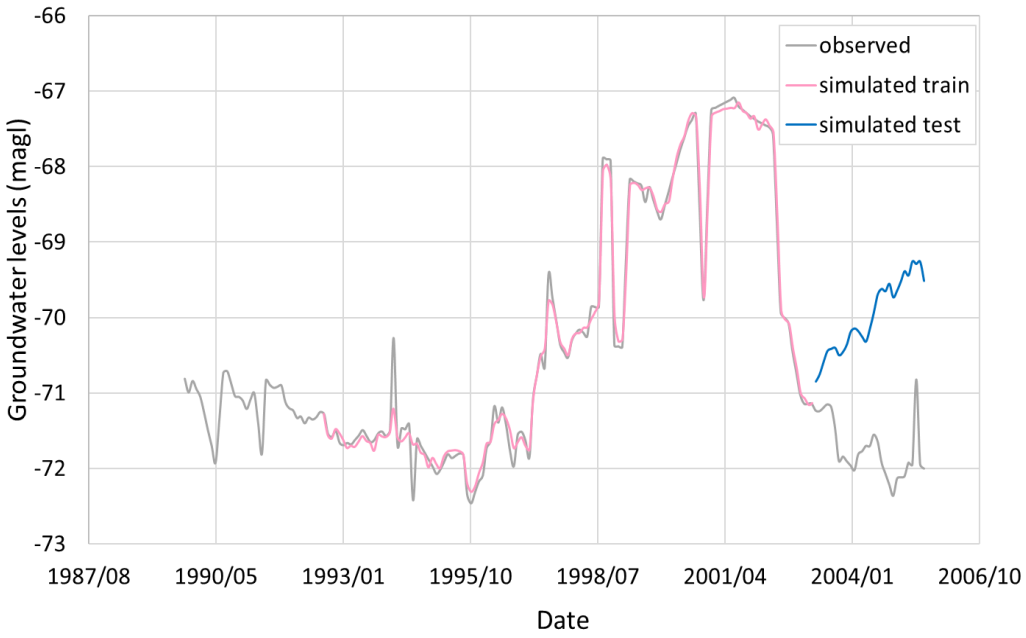
Metrics	A2N0608	
R^2	-6.93	
MSE	4.21	
RMSE	1.98	
MAE	1.77	
sd	0.15	

Table A-8: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole 37773 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

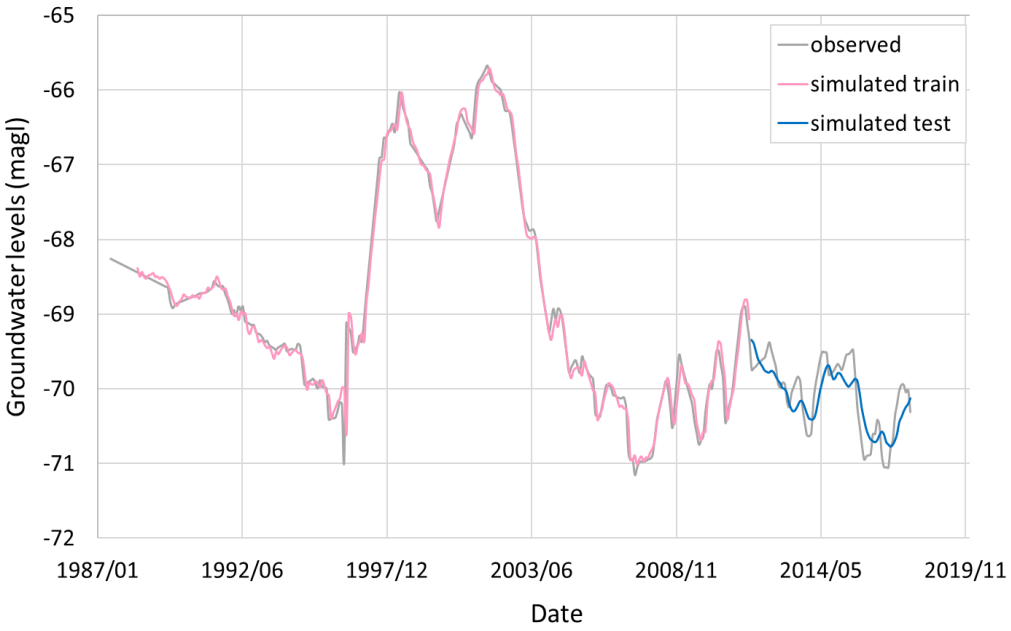
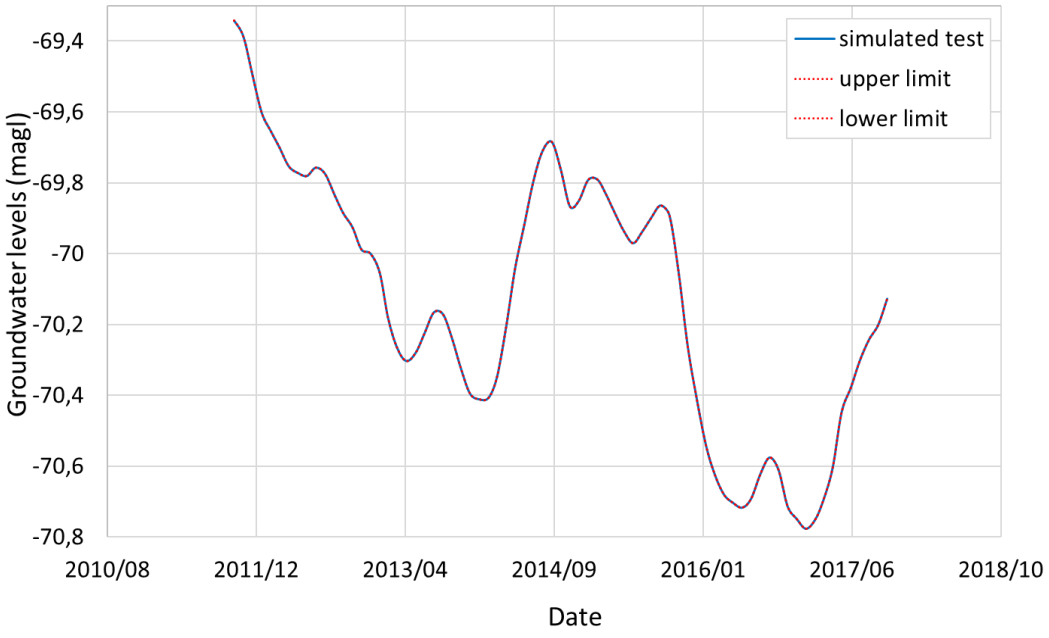
Metrics	37773	
R^2	0.41	
MSE	0.08	
RMSE	0.28	
MAE	0.24	
sd	<0.01	

Table A-9: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0554 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

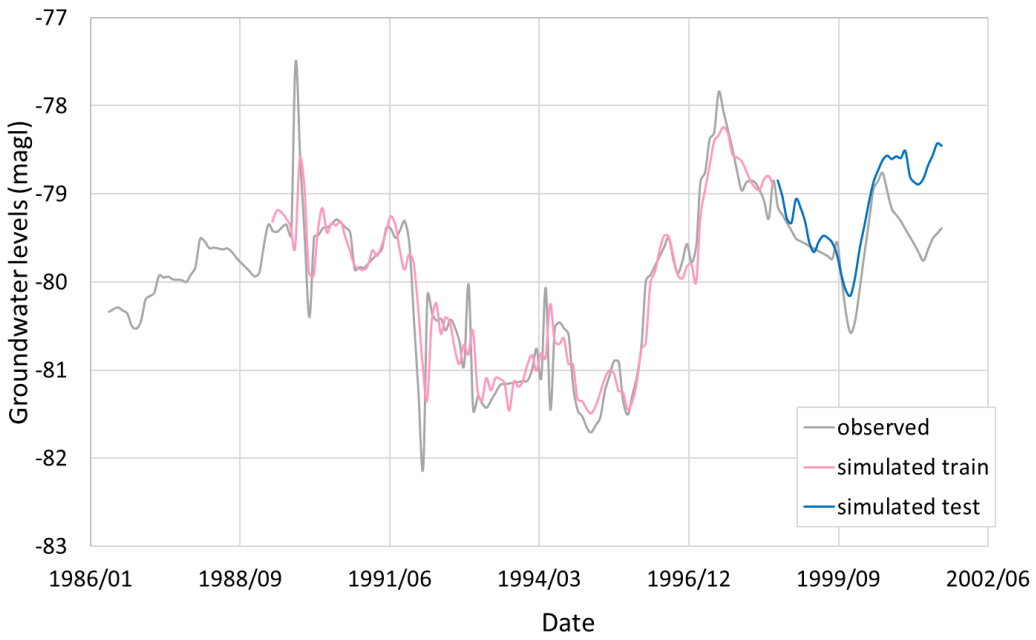
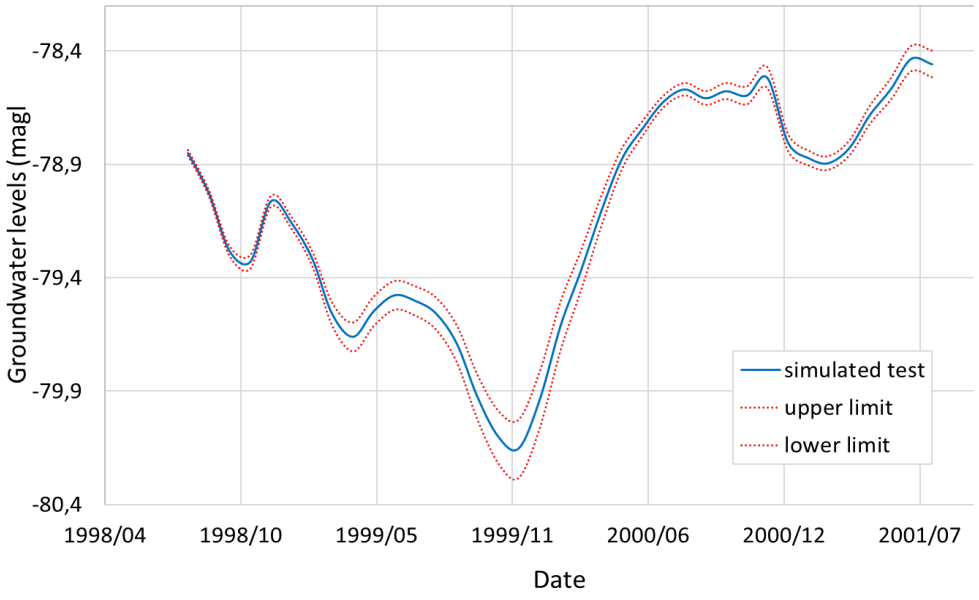
Metrics	A2N0554	
R^2	-0.16	
MSE	0.27	
RMSE	0.52	
MAE	0.42	
sd	0.02	

Table A-10: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0563 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

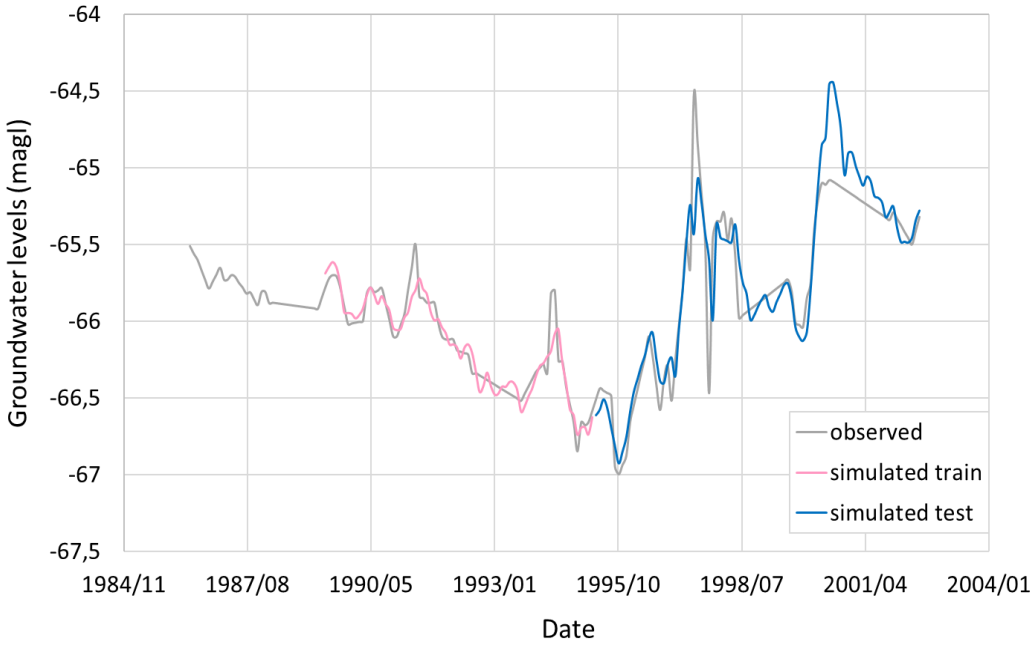
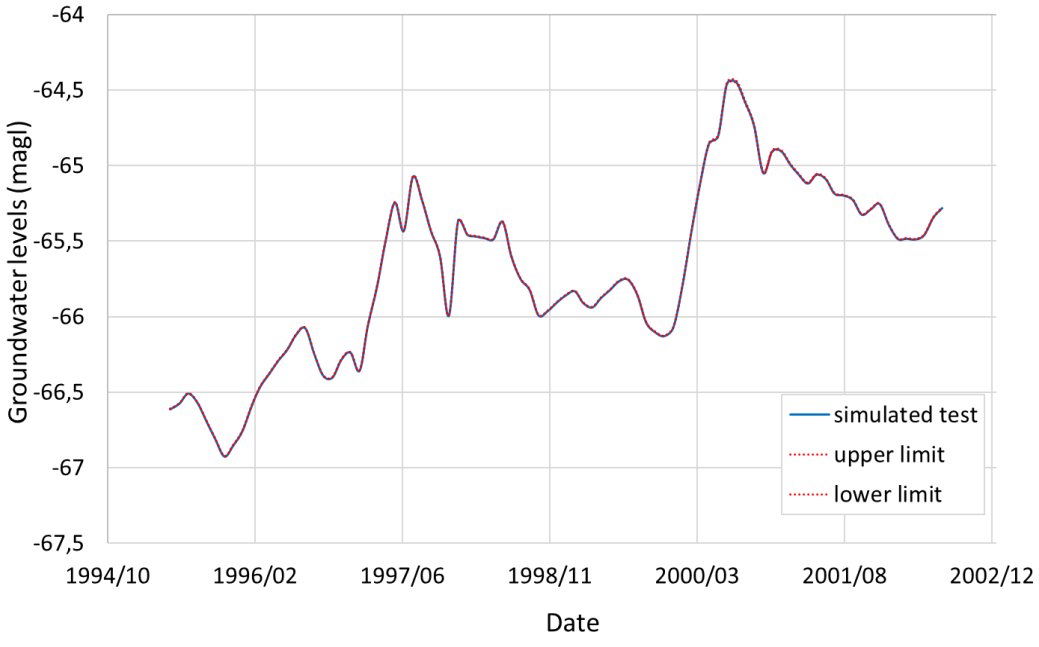
Metrics	A2N0563	
R^2	0.76	
MSE	0.05	
RMSE	0.22	
MAE	0.15	
sd	<0.01	

Table A-11: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0565 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

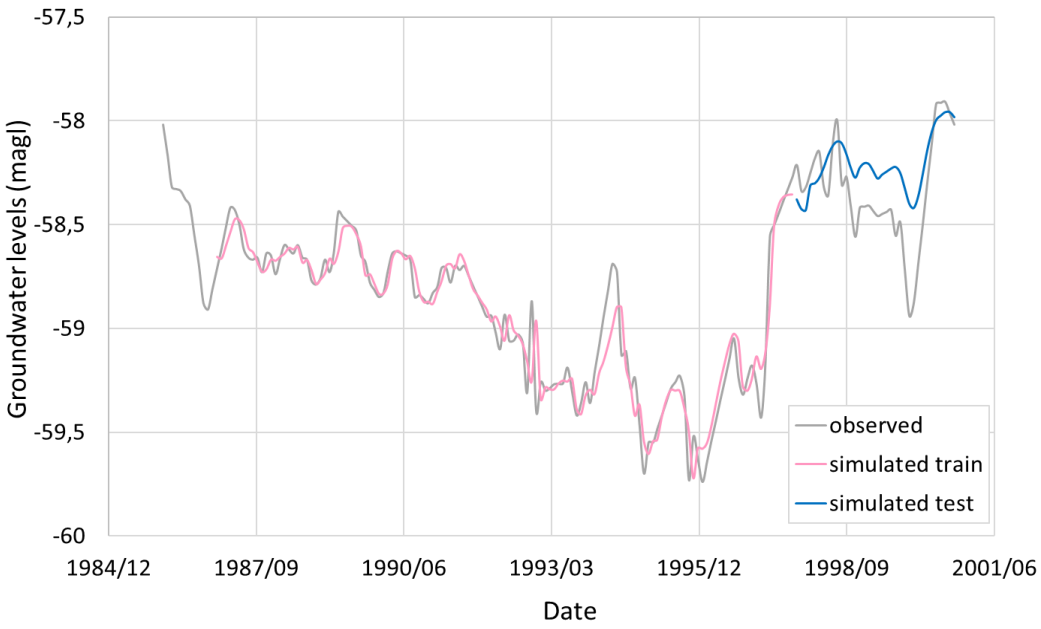
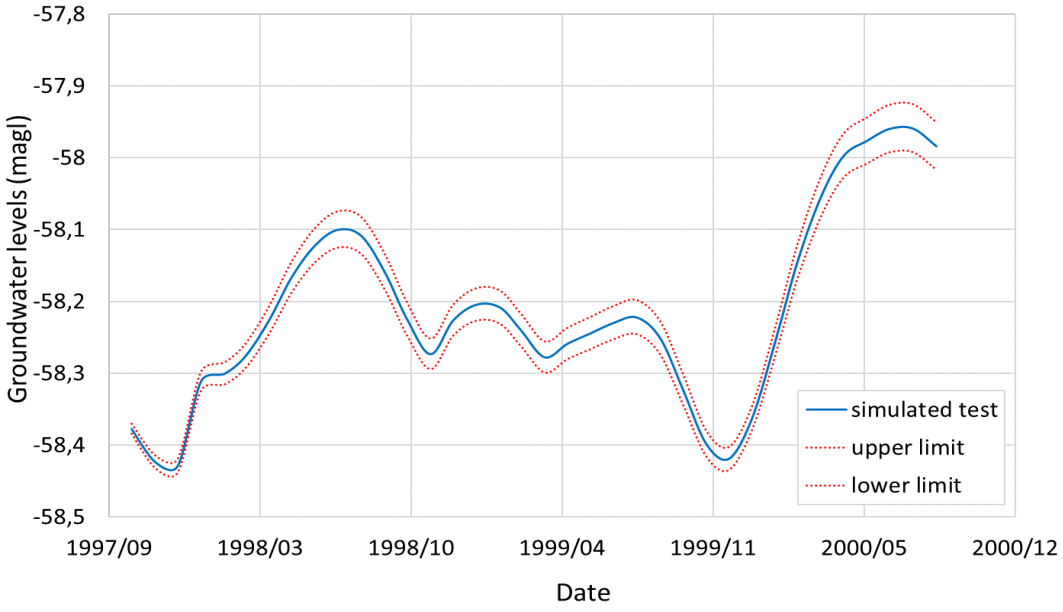
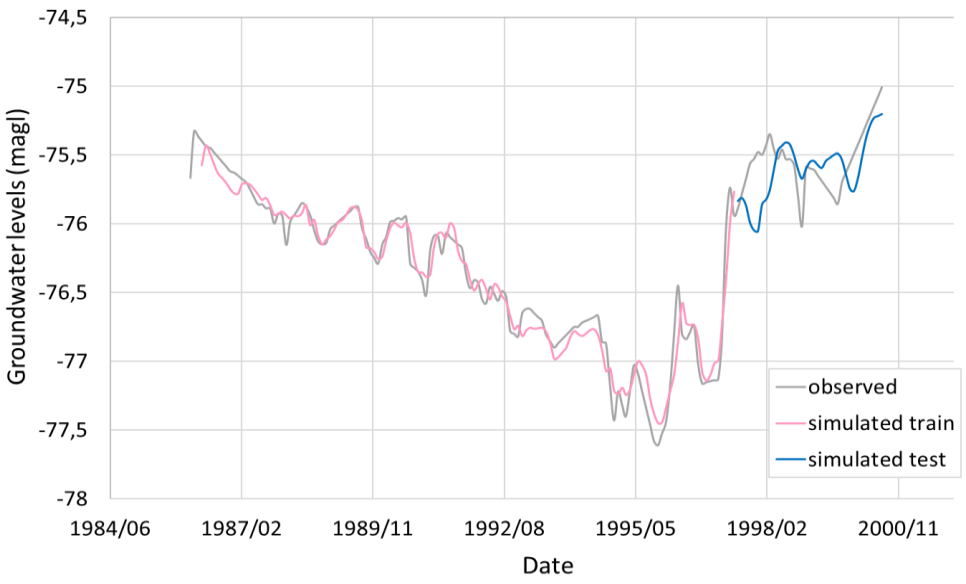
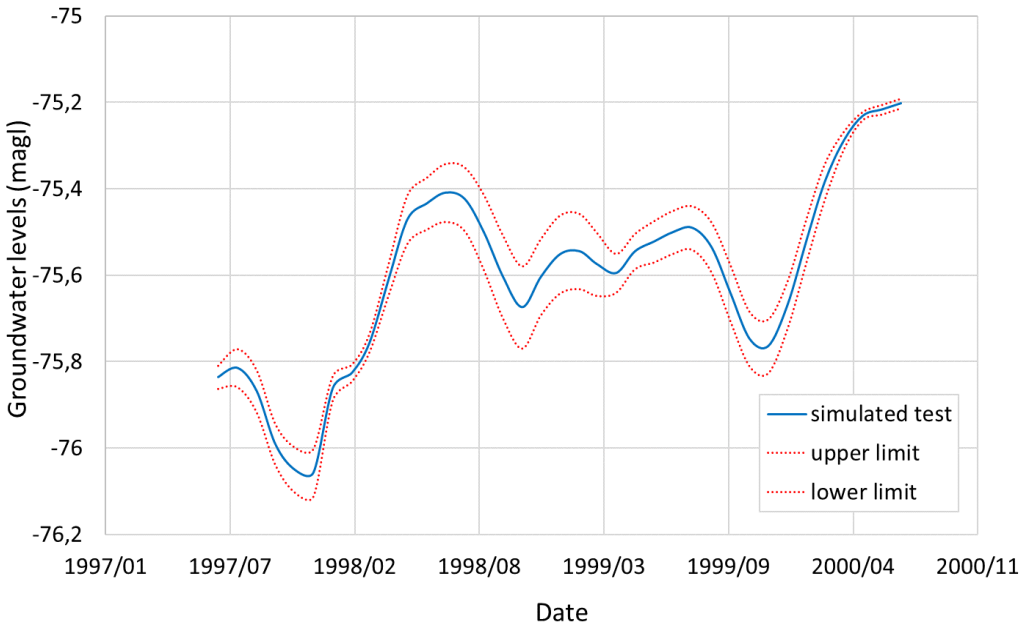
Metrics	A2N0565	
R^2	-1.70	
MSE	0.05	
RMSE	0.22	
MAE	0.18	
sd	0.01	

Table A-12: The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from borehole A2N0569 in the Steenkoppies compartment. The standard deviation for the model predictions across the ten model runs is also presented.

Metrics	A2N0569	
R^2	-0.31	
MSE	0.06	
RMSE	0.25	
MAE	0.20	
sd	0.02	

APPENDIX B

Code for Forecasting of a Groundwater Level Timeseries in Steenkoppies using Recurrent Neural Network (NARX) in R:

Source all external functions in folder functions

```
purrr::map(
  list.files(
    path = './scripts/functions',
    pattern = "\\.R$",
    full.names = TRUE,
    recursive = TRUE
  ),
  source
)
## list()
```

Load Packages

```
library(import)
import::from(readxl, read_excel)
import::from(janitor, clean_names, get_dupes)
# import::from(forecast, nnetar, forecast, accuracy)
library(plotly)
library(forecast)
# library(Metrics)
# library(doParallel)
# library(doFuture)
library(timetk)
library(modeltime)
library(doParallel)
library(tidymodels)
library(tune)
library(tidyverse)
library(lubridate)
```

1 Data Import

```
data <- read_excel("data/NARX_A2N0612_Kirsty.xlsx")
```

2 Minor Preparations

Clean column names

```
data <- data %>%
  clean_names()
```

Cast date column as date

```
data <- data %>%
  mutate(date = as.Date(date))
```

3 Preprocessing

```
data <- data %>%
  pivot_longer(cols = -date)
```

Plot the timeseries

```
data %>%
  ggplot(aes(x = date,
             y = value)) +
  geom_line(colour = "steelblue") +
  facet_wrap(~name, scales = "free_y", ncol = 1) +
  theme_minimal()
```

```
data <- data %>%
  pivot_wider()
```

4 Modelling

4.1 Data Splitting with rsample

```
train_test_splits <- data %>%
  initial_time_split(prop = 0.9)

data_train <-
  train_test_splits %>%
  training()

data_test <-
  train_test_splits %>%
  testing()
```

```
data_train %>%
  mutate(split = "training") %>%
  bind_rows(data_test) %>%
  replace_na(list(split = "testing")) %>%
  mutate(split = as.factor(split)) %>%
  ggplot(aes(date, gwl)) +
  geom_line(aes(colour = split)) +
  theme_minimal() +
  scale_color_manual(values = c("training" = "#F8766D", "testing" = "#00BFC4")) +
  labs(colour = "Split") +
  scale_x_date(breaks = "2 years",
               labels = lubridate::year)
```

4.2 Resampling with rsample

```
n_samples_train <-
  data_train %>%
  nrow()
```

```
n_initial <-
  (n_samples_train * 0.5) %>%
  floor()
```

```
n_slices <- 5

n_slice <-
  ((n_samples_train - n_initial) / n_slices) %>%
  floor()
```

```
resampling_strategy_cv5fold <-
  data_train %>%
  time_series_cv(
    initial = n_initial,
    assess = n_slice,
    skip = n_slice,
    cumulative = TRUE
  )
```

```
resampling_strategy_cv5fold %>%
  tk_time_series_cv_plan() %>%
  ggplot(aes(date, gwl)) +
  geom_line(aes(colour = .key)) +
  facet_wrap(~.id, ncol = 1) +
  theme_minimal() +
  labs(colour = "Split") +
  scale_x_date(breaks = "2 years",
               labels = lubridate::year)
```

4.3 Preprocessing with recipe

```
steenkopies_recipe <-
  recipe(gwl ~ ., data = data_train) %>%
  # update_role(date, new_role = "ID") %>%
  step_normalize(all_numeric(), -date)
  # step_normalize(all_predictors(), -date)
```

4.4 Defining a Learner with parsnip and modeltime

Now we define a machine learning algorithm. We use a recurrent neural net (RNN) `nnetar()` from the `forecast` package which is implemented as `parsnip` model in the `modeltime` package. This specific RNN is similar to the NARX model in Matlab.

```
tune_nnetar_model <-
  nnetar_reg(
    seasonal_period = 12,
    non_seasonal_ar = tune(),
    seasonal_ar = tune(),
    hidden_units = tune(),
    num_networks = 20,
    penalty = tune(),
    epochs = tune()
  ) %>%
  set_engine("nnetar",
             scale.inputs = FALSE) %>%
  set_mode("regression")
```

4.5 Tuning

The setting of `n_levels` determines the number of values that are tried out for each hyperparameter. This causes an exponential growth of the number of models that need to be fitted. `n_levels` = 3 for these five hyperparameters took approximately 30 minutes to calculate on an average PC. So this is a trade-off situation between number of tuned hyperparameter, computing time and number of hyperparameter combinations.

```
n_levels <- 30

tune_grid <- grid_regular(
  non_seasonal_ar(range = c(1L, 5L)),
  seasonal_ar(range = c(1L, 5L)),
  hidden_units(),
  # num_networks(),
  penalty(),
  epochs(),
  levels = n_levels
)
```

Number of parameter combinations and consequently model runs:

```
tune_grid %>%
  nrow()
## [1] 225000
```

4.5.1 Workflow

```
nnetar_workflow <-
  workflow() %>%
  add_model(tune_nnetar_model) %>%
  add_recipe(steenkopies_recipe)
```

4.6 Parallelize Tuning

```
all_cores <- parallel::detectCores(logical = FALSE)

cluster <- makePSOCKcluster(all_cores)
registerDoParallel(cluster)
```

4.6.1 Model Fitting with Resamples

```
nnetar_resampling <-
  nnetar_workflow %>%
  tune_grid(
    resamples = resampling_strategy_cv5fold,
    grid = tune_grid,
    metrics = metric_set(rmse, mae))

stopCluster(cluster)
```

```
chosen_metric <- "rmse"
```

4.7 Model Evaluation

```
nnetar_resampling %>%

  collect_metrics() %>%

  pivot_longer(cols = c("non_seasonal_ar", "seasonal_ar", "hidden_units", "penal
ty", "epochs")) %>%

  filter(.metric == chosen_metric) %>%

  group_by(name, value) %>%

  summarise(mean = mean(mean)) %>%

  ggplot(aes(value, mean)) +
```

```
  geom_line(size = 1.5,

            alpha = 0.6,

            colour = "#00BFC4") +

  geom_point(size = 2,

             colour = "#00BFC4") +

  facet_wrap(~ name, scales = "free", ncol = 1) +

  labs(y = "mean rmse") +

  theme_minimal()
```

```
nnetar_resampling %>%
  show_best(metric = chosen_metric)
```

```
best_model <-
  nnetar_resampling %>%
  select_best(metric = chosen_metric)

final_workflow <-
  nnetar_workflow %>%
  finalize_workflow(best_model)
```

```
final_model <-
  final_workflow %>%
  fit(data = data_train)
```

```
models_tibble <-
  final_model %>%
  modeltime_table()
```

```
calibration_tibble <-
  models_tibble %>%
  modeltime_calibrate(new_data = data_test)
```

```
best_model_forecast <-
  calibration_tibble %>%
  modeltime_forecast(
    new_data = data_test,
    actual_data = data
  )
```

```
best_model_forecast %>%
  plot_modeltime_forecast(
    .legend_max_width = 25 # For mobile screens
  )

actual_vs_prediction_data <-
  final_model$fit$fit$fit$data %>%
    select(date, .actual) %>%
    rename(value = .actual) %>%
  mutate(name = "values_observed_train") %>%
  bind_rows(final_model$fit$fit$fit$data %>%
    select(date, .fitted) %>%
    rename(value = .fitted) %>%
    mutate(name = "values_fitted_train")) %>%
  bind_rows(best_model_forecast %>%
    get_dupes(.index) %>%
    select(.index, .key, .value) %>%
    rename(
      date = .index,
      name = .key,
      value = .value
    ) %>%
    mutate(name = if_else(name == "actual", "values_observed_test", "values_pred
icted_test")))

prediction_plot <-
  actual_vs_prediction_data %>%
  ggplot(aes(date, value)) +
  geom_line(data = filter(actual_vs_prediction_data,
                          name == "values_observed_train"),
            linetype = "twodash",
            colour = "gray") +
  geom_line(data = filter(actual_vs_prediction_data,
                          name == "values_fitted_train"),
            colour = "pink") +
  geom_line(data = filter(actual_vs_prediction_data,
                          name == "values_observed_test"),
            linetype = "twodash",
            colour = "gray") +
  geom_line(data = filter(actual_vs_prediction_data,
                          name == "values_predicted_test"),
            colour = "steelblue") +
  labs(y = "Groundwater Level [m below surface]") +
  theme_minimal()
```

```
prediction_plot %>%
  ggplotly()
```

Custom function to inverse normalization

```
extract_step_item <-
  function(recipe, step, item, enframe = TRUE) {
    d <- recipe$steps[[which(purrr::map_chr(recipe$steps, ~ class(.)[1]) == step
)]][[item]]
    if (enframe) {
      tibble::enframe(d) %>% tidyr::spread(key = 1, value = 2)
    } else {
      d
    }
  }

unnormalize <-
  function(x, rec, var) {
    var_sd <- extract_step_item(rec, "step_normalize", "sds") %>% dplyr::pull(va
r)
    var_mean <- extract_step_item(rec, "step_normalize", "means") %>% dplyr::pul
l(var)

    (x * var_sd) + var_mean
  }
```

We save the best model in `models_tibble`

```
models_tibble %>%
  write_rds(
    str_c(
      "data_processed/models_tibble_best_model_",
      str_replace_all(Sys.time(),
                      "[:punct:]| ",
                      "-"),
      ".Rds")
    )
```

We plot the actual vs. predicted values of the test data split as scatter plot to get another impression of the model performance.

```
actual_vs_prediction_data <-
  actual_vs_prediction_data %>%
  pivot_wider(names_from = "name",
              values_from = "value") %>%
  mutate(across(where(is.numeric),
                unnormalize,
                prep(steenkopies_recipe),
                "gwl"))

axis_limits <-
  actual_vs_prediction_data %>%
  pivot_longer(cols = -"date") %>%
  filter(str_detect(name, "test")) %>%
  pull(value) %>%
  range(na.rm = TRUE)

actual_vs_prediction_data %>%
  ggplot(aes(values_observed_test, values_predicted_test)) +
  geom_point() +
  coord_equal() +
  geom_abline(linetype = "dashed",
              colour = "grey",
              alpha = .6,
              size = 1) +
  xlim(axis_limits) +
  ylim(axis_limits) +
  theme_minimal()
```

```
rmse_training <- actual_vs_prediction_data %>%
  rmse("values_observed_train", "values_fitted_train", na_rm = TRUE)

rmse_testing <- actual_vs_prediction_data %>%
  rmse("values_observed_test", "values_predicted_test", na_rm = TRUE)
```