MACHINE LEARNING MODELS FOR GROUNDWATER AVAILABILITY INCORPORATING A FRAMEWORK FOR A SUSTAINABLE GROUNDWATER STRATEGY

H. Seyler, K. Gibson, Y. Kanyama, K. Witthüser

Machine Learning Models for Groundwater Availability

Incorporating a Framework for a Sustainable Groundwater Strategy

Report
to the Water Research Commission
by

H. Seyler[1], K. Gibson[2], Y. Kanyama[3], K. Witthüser[4],

[1]Delta-H Groundwater Systems (Pty) Ltd
[2]University of Free State
[3]University of Witwatersrand
[4]Delta-H Water Systems Modelling (Pty) Ltd

March 2021

This report emanates from the Water Research Commission Project K5/2879 "Groundwater Secure Transboundary Systems".

This report forms part of a series of four reports. The other reports are:
- *Imagining Solutions for Extracting Further Value from Existing Datasets on Surface and Groundwater Resources in Southern Africa* (WRC Report no. TT 842/20)
- *Localizing Transboundary Data Sets in Southern Africa: A Case Study Approach* (WRC Report no. TT 843)*.
- *Data Analytics and Transboundary Water Collaboration. Theme 1: Consolidation of Data and Application of Big Data Tools to Enhance National and Transboundary Data Sets in Southern Africa that Support Decision-Making for Security of Water Resources* (WRC Report no. TT 844/20)

Executive Summary

The collaboration

This report is the outcome of one of four projects carried out under an initiative called "Big Data Analytics and Transboundary Water Collaboration in Southern Africa". This initiative is funded by USAID, the South African Department of Science and Technology (DST), and the SADC Groundwater Management Institute (GMI), managed primarily by the Water Research Commission (WRC), and with technical support from the US Geological Survey (USGS) and the IBM Research Africa Lab in South Africa. Overall programme coordination is by the USAID appointed programme coordinator, the Sustainable Water Partnership (SWP).

The overall aim of the collaboration is to research how and where big data analytics or new approaches can contribute to improved groundwater management in Transboundary Aquifers (TBAs).

Research Overview

This project (theme 4 of the 4 projects) developed a **strategy for sustainable groundwater use (SGS)** which is an approach for achieving best practice groundwater management. The SGS sets out the benchmark for an approach to achieve sustainable groundwater use. The SGS is essentially a list of actions necessary for achieving sustainable groundwater use and is applicable to any aquifer or group of aquifers (a groundwater basin). It is recommended that the approach be implemented particularly in heavily used aquifers, in aquifers with sensitive receptors, and the approach would support improved groundwater management in TBAs.

The actions within the strategy can be met using traditional hydrogeological approaches. The four research projects under this collaboration tested how and where big data analytics or new approaches can support improved groundwater management, and each tool tested or developed in the collaboration represents a new approach to meet part of the strategy. The full sustainable groundwater strategy was not implemented in full in one test case, due to the scope and structure for these projects, and also because for the SGS to be successful it needs to be implemented by the responsible authority for that basin.

The outcome of the programme is therefore a set of new approaches (tools) tested in various case studies, that all support improved sustainable groundwater management and therefore meet some of the actions required in the SGS. Specifically, the aims of theme 4 were to:

- develop a sustainable groundwater strategy, reflecting a best practice approach for a sustainable groundwater use
- pilot "big data" or new approaches for meeting parts of the sustainable groundwater strategy, in a case study site

Sustainable groundwater use *(of groundwater from an aquifer or group of connected aquifers in a basin) is defined here as:*

Abstraction of a yield that causes or is expected to cause in future, impacts on groundwater level discharge, recharge, groundwater quality, saline intrusion, and subsidence, which are considered socially, economically and environmentally acceptable.
Achieving sustainable groundwater use requires that the expected impacts of abstraction are quantified, acceptability determined with relevant stakeholders, and associated thresholds for indicators which describe expected impacts established.

Framework for a Sustainable Groundwater Strategy

Whilst there are many standards and best practice guidelines for groundwater development at scheme level, there are very few for regional aquifer resources management. "Groundwater Sustainability Plans" (GSPs) are generated for major aquifers in California under the new Sustainable Groundwater Management Act, (California State, 2014), and have a relatively standard content and approach. The national modelling project in the UK similarly provides groundwater yield assessments for major aquifers in the UK, which are updated and inform management. However, the SADC countries do not have the same: there is no (routinely updated) aquifer scale groundwater resources assessment, which informs management plan and water use strategy, in place for most major aquifers in the SADC countries, nor the TBAs.

Aquifers, under natural conditions, are in a state of dynamic equilibrium. When pumping is initiated, abstracted water is met from aquifer storage and water levels decline (Bredehoeft and Durbin, 2009; Sophocleous, 2000; Alley and Leake, 2004). As the system reaches a new dynamic equilibrium the water level decline diminishes and abstracted water is met by a reduction in natural discharge, and potentially an increase in recharge (Bredehoeft et al, 1982, Bredehoeft, 2002, Devlin & Sophocleous, 2005). The time it takes the aquifer to reach this new dynamic equilibrium is termed the response time (Sophocleous, 2000; Alley and Leake, 2004). Water withdrawn artificially from an aquifer is therefore derived from a decrease in storage in the aquifer, a reduction in the previous discharge from the aquifer, an increase in the recharge, or a combination of these changes (Theis, 1940, cited in Bennett et al, 1988). The sum of the increase in recharge and decrease in discharge brought about by pumping was referred to as capture by Lohman (1972).

It follows that an assessment of the sustainability of groundwater abstraction, or quantification of a sustainable yield for abstraction, would quantify what these changes in the flow regime related to that yield are (or will be in future, and when they will be realised), and determine whether the changes and their associated impacts are considered socially, economically and environmentally acceptable by the relevant stakeholders which needs to include the responsible authority, groundwater users, and representation of the environment . If the estimated changes are acceptable, the total yield to be abstracted for the predicted changes can be considered sustainable groundwater use (Sophocleous, 2000, Alley and Leake, 2004). If the abstraction rate cannot be met by capture, because it exceeds the pre-abstraction discharge plus any potential addition from enhanced recharge, groundwater storage continues to be the source of abstracted water. A new dynamic equilibrium is not achieved, and groundwater levels continue to decline.

The SGS requires quantification of the current and expected changes in the groundwater balance or flows in the basin related to planned or potential groundwater use, which must include changes in storage, inflows and discharges. The relationship between abstraction and aquifer fluxes must be established. These predictions of future impact of abstraction are relied upon to set sustainable management criteria, which requires close collaboration with stakeholders in the basin (including the regulators or responsible authority). In line with CDWR (2014), the overall aim of the SGS is to ensure groundwater is managed to avoid the following six undesirable results:

1. Chronic lowering of groundwater levels indicating a significant and unreasonable depletion of supply if continued over the planning and implementation horizon.
2. Depletions of interconnected surface water that have significant and unreasonable adverse impacts on beneficial uses of the surface water.
3. Significant and unreasonable degraded water quality, including the migration of contaminant plumes that impair water supplies.
4. Significant and unreasonable land subsidence that substantially interferes with surface land uses.
5. Significant and unreasonable seawater intrusion.
6. Significant and unreasonable reduction of groundwater storage.

Each of the six undesirable results are related to an equivalent sustainability indicator, which are the "six effects caused by groundwater conditions occurring throughout the basin that, when [the effects are] significant and unreasonable, represent undesirable results" (CDWR, 2017, pg. 26). The six indicators are:
1. Lowering of groundwater levels
2. Surface water depletion
3. Degraded groundwater quality
4. Land subsidence
5. Seawater intrusion
6. Reduction of storage

The framework for the SGS outlines the steps required to establish the expected impacts of abstraction, set thresholds for the indicators, establish the necessary monitoring protocol to ensure the thresholds are not exceeded, and implement and update the strategy.

The authority mandated as responsible for management of water resources in the area would be the appropriate body to implement the strategy, i.e., the Department of Water and Sanitation (or the Catchment Management Agencies) for South Africa, or the Department of Water Affairs in Botswana.

Overview of model testing

Fundamental to achieving sustainable groundwater use, and a central part of the SGS, is establishing the relationship between abstraction and aquifer fluxes. Given the complex nature of groundwater and the interdependent responses of the system to change, consideration of the long-term implications of groundwater abstraction on these systems is virtually impossible without the use of numerical models. However, numerical models require time and appropriate expertise to develop, and have large data requirements to adequately parameterise the physical system. It has therefore become appealing to investigate data-driven & machine learning methods that may be able to predict aquifer behaviour without deep knowledge of the underlying physical parameters (Sahoo et al, 2017). The research therefore tested the viability of machine learning methods to quantify the relationship between abstraction and aquifer fluxes (i.e., generating the data for establishing sustainability indictors). In piloting this aspect of the SGS, the research therefore aims to determine:

- Can we use ML methods to **forward predict** groundwater level in relation to different future stresses (recharge, abstraction)? What is the efficacy in terms of length of prediction and stresses that are not seen in training dataset?
- How can we use ML methods to **generate aquifer fluxes**? What is the efficacy?

To address these questions ML models were developed to first simulate historical groundwater level (as a proxy for storage) and aquifer fluxes over time (discharge to surface water) and if possible, investigate demonstrate causal / effect links. This requires a minimum a ML model able to recognise pattern between groundwater level, discharge, groundwater use, rainfall. Subsequently whether the developed ML model could project discharge forward for a different abstraction and rainfall was considered.

Case Study Selection

Although the Ramotswa TBA became the case study focus for theme 1, various other case study sites were used to develop and test the tools developed under the programme, selecting a site that leant itself to this task. A case study area was required where groundwater is heavily used such that the relationship between (and impacts of) groundwater use and reduced discharge and/or storage are detectable in datasets. Areas with a detectable decline in groundwater levels as the basin transitions to a new dynamic equilibrium in response to abstraction were required. The abstraction and monitoring record must be sufficient compared to the aquifer response time to demonstrate these impacts. The relationship between abstraction and groundwater flow regime or impacts is established over the aquifer or basin scale, and so this assignment also implicitly dictated that we consider a regional aquifer – scale assessment (rather than wellfield or response to pumping in one borehole). The dolomite terrains of RSA had the best potential to meet these requirements.

Data from all dolomite compartments was collated to identify areas with the greatest density of boreholes with groundwater level readings; areas where the groundwater level records were the longest and had least data gaps; areas where weather stations were in close proximity to groundwater data; and where flow gauge data was available from groundwater fed springs in the same or hydrogeologically-connected compartment. Based on these requirements, two dolomite compartments were identified as meeting the criteria:
1. Steenkoppies where a reduction in discharge has been related to groundwater use (Seyler et al, 2016).
2. Molopo/ Grootfontein which also has been reported as experiencing a reduction in groundwater storage related to abstraction (Cobbing, 2018).

Machine Learning Models overview

Machine learning models do not include the physical principles of groundwater flow or mass conservation. The models instead learn the relationships that may exist between the target variable and the other datasets during the training and calibration stages and uses these relationships to predict the target variable as accurately as possible.

The groundwater level at the current point in time is influenced by and related to the groundwater level at previous points in time. The potential exists therefore to achieve better predictions of groundwater level by understanding what the (predicted or observed) groundwater level was at previous times. Recurrent neural network models (RNN) were therefore used for modelling groundwater behaviour rather than Feed Forward Neural Network models (both types of artificial neural network models). Two types of RNN models were tested: the Long Short-term Memory (LSTM) and the Neural Network Autoregression (NNAR).

The LSTM was first introduced by Hochreiter and Schmidhuber in 1997 to address the accumulation of error gradients during the update process which is a drawback of the RNN, by adding a unique set of memory cells that replace the hidden layer neurons of the RNN (Le et al., 2019). LSTMs are capable of avoiding long-term dependence problems that occur with the traditional RNN due to the added memory cells. Instead of a single neural network like the standard RNN, the LSTM model is capable of filtering information through a gate structure to maintain and update the state of the memory cells. Conversely, the NNAR's previous timestamps are not stored in a hidden state but given as another input to the model (Izady et al., 2013). The model iteratively makes predictions. To predict one step, the model simply uses the available historical inputs. For forecasting two steps, the model simply uses the one-step prediction as an input, along

with the historical data. This process proceeds until all the required predictions are computed. The NNAR does the same for the input variables used to help predict the target variable.

Model Setup

The groundwater level at all boreholes that had adequate data in each case study area were used as prediction targets. Input variables included discharge rates, temperature, precipitation and abstraction rates. Precipitation acts as the primary source of recharge for the aquifer. Initial model results showed a lack of influence of precipitation on groundwater levels hence the rainfall data set was smoothed to improve model performance. Time series decomposition was used to de-noise the data and extract the rainfall data trend, which was then used as model input, rather than the actual data. Information Gain Ranking was performed to test the correlation between variables and understand what would have most influence when carrying out the ML models. In both case studies cross validation was performed, models were run 10 times for repeatability, and model results evaluated using R^2 and RMSE. In both cases, 80% of the data was used for training, and 20% for testing.

Training and Test Results from the Long Short-term Memory (LSTM) model in Grootfontien

Generally speaking, the observed and modelled groundwater levels are closely matched, with key features of the observed datasets replicated in the modelled data. As would be expected, the training dataset the modelled groundwater level matches observed much more closely than the test, because the observed groundwater levels are used in training, i.e., groundwater level at the previous time steps largely dictates the predicted output. During test, the model only "sees" the modelled groundwater level rather than observed. The LSTM model achieves an average MAE value of 0.25m and a maximum of 0.67m. Most of the R^2 results are above 0.5. The average R^2 value is 0.61. The results demonstrate that the model generally performs better the greater number of points in the training dataset, and the more variability it sees during the training phase and performs better in test when the variability in the test is similar to that seen in training.

The LSTM model is often unable to capture the minimum and extreme groundwater level peaks whenever they occur in a short timeframe. In all cases the predicted groundwater level in test data is lagged behind the observed groundwater level. This is common in RNN model results and is related to the fact that the model is "looking" at the model output groundwater levels for the previous X number of time steps (X is user defined).

Training and Test Results from the NNAR in Steenkoppies

Similar to the results from the LSTM, the NNAR model closely replicates the observed groundwater levels, with key features of the observed datasets replicated in the modelled data. Again, as would be expected, the modelled groundwater level matches observed much more closely in training rather than the test. The NNAR model achieves an average MAE value of 0.61m (achieving a median of 0.26m, skewed by a few boreholes that fit poorly) and a maximum of 4.03m. Eight (of 18) of the R^2 results are above 0.4, and the median R^2 value is 0.2. In some cases, the standard deviation across the ten model runs increases with time due to the accumulated error in model prediction with time, and similarly to the LSTM the predicted output of the model is lagged compared to the observed dataset. As per the LSTM, the results demonstrate that the model generally performs better the greater number of points and variability in the training dataset. In addition, the variability in model performance is also related to how closely the input variables relate to the target variable of groundwater level.

Scenario projection results

The model results show that both the LSTM and the NNAR models are able to replicate groundwater levels relatively well for most boreholes for ~5-year periods with ~20 years of training data (i.e., an 80/20 training test split with monthly data points). Whilst this is useful for applications such as filling historical groundwater level data gaps, the key aim (and requirement for managing sustainable groundwater use) was to test the use of ML methods to forward predict

groundwater level in relation to different future stresses (recharge, abstraction). The scenario test results revealed that the models were able to pick up the influence of a certain change in a variable, by generating predicted water level values that differed from the base case. For example, halving recharge in both models generates a decline in groundwater levels. The LSTM model particularly was able to show greater changes in groundwater levels than the NNAR model.

The scenario test results highlighted, however, several shortcomings in the use of ML models for forward prediction and for quantifying aquifer fluxes (quantifying the relationship between groundwater use and aquifer fluxes; this critical part of sustainable groundwater management), some of which are applicable perhaps only to these particular case studies and some of which would be applicable in any attempt to use ML models. These are summarised as:

- For the ML model to provide the impact on discharge of increasing groundwater use (or changed climate), a pattern must be detectable between groundwater use and discharge, which of course requires data. It is rare to have a dataset(s) for groundwater discharge from an aquifer. Furthermore, where this data is available to establish the models, groundwater discharge is a product of groundwater levels and so wouldn't be known in a future scenario test in which the models are required to predict the impact of changed abstraction or rainfall on groundwater levels (storage) and on discharge. Those scenarios demonstrated here used averaged discharge for future scenarios which unnaturally controls the resulting groundwater level. Taking discharge out of the scenarios as a parameter and essentially driving the prediction based on rainfall and abstraction only would be appropriate (and required in settings where there is no discharge dataset) but would reduce the accuracy. A work-around this would be to use the pattern between rainfall and groundwater levels to model groundwater levels (and predict them into the future), and the pattern between groundwater levels at several boreholes and discharge at the spring, to then predict the resulting impact on discharge of reduced rainfall. This approach was tested with the NNAR and is feasible, however, accuracy reduces significantly as parameters are excluded in the modelling. The benefit of numerical modelling, therefore, is that even when there is no time-series dataset for discharge, (and ideally at least a few measurements for calibration), the resulting aquifer discharge can be calculated based on modelled groundwater levels.
- Using ML models to predict groundwater levels may be less feasible in an aquifer setting where there is the relationship between groundwater levels and rainfall is weaker or much more muted (i.e., confined aquifer settings, or aquifers in arid areas with highly episodic recharge).
- The main challenge in using ML models in scenario analysis is establishing a relationship between aquifer-scale groundwater use and groundwater levels (storage), and thereby groundwater discharge. The research aims require assessment of aquifer-scale processes (widespread abstraction causing a gradual and consistent lowering of groundwater levels) and therefore cumulative groundwater abstraction across the whole aquifer was incorporated in the models. The research aims also dictated that long-term datasets spanning many years were used in order to detect the reducing trend in groundwater levels i.e., datasets ideally longer than the aquifer response time are required. The Grootfontein case study results do detect the impact of changing groundwater use in future scenarios on groundwater levels (which could, in turn, be used to predict future discharge for those groundwater levels), however, the Steenkoppies case study was (more understandably) unable to detect a strong correlation between use and groundwater levels, and as such changing use in the scenario predictions had little impact.

In summary, it is seemingly unlikely that ML models can be used to quantify the relationship between increased abstraction and aquifer fluxes. Being able to do so would rely on a relationship being detected between a curve of increasing groundwater use (or even sustained use at one abstraction rate) and groundwater levels and aquifer fluxes, already detected in the training dataset. ML models are therefore an inapplicable approach for under-utilised aquifers or

where abstraction has not yet commenced. It cannot, therefore, be used, for example, for mining assessments at greenfield sites where numerical modelling is also relied upon to give the same outputs.

Related to this is the challenge of using point-data (groundwater levels) to provide indications of aquifer-scale processes. In cases where there is sufficient spread of boreholes across an aquifer, the groundwater level response at every borehole could be predicted for future scenarios, and the predictions interpolated to generate a piezometric surface, and used to calculate aquifer storage.

Summary of applicability of ML models for achieving SGS

The research showed that ML models have application in groundwater for aspects such as assessment of the causal of patterns in datasets, predicting the short-term impact of climate variability, and filling data gaps. However, their applicability to predicting the impacts of abstraction to inform thresholds of acceptable impact (the core of the SGS) is limited.

Acknowledgements

Reference Group Member Name	Organization	Reference Group Member Name	Organization
Shafick Adams	WRC	James Sauramba	SADC-GMI
Shanna Nienaber	WRC	Yannick Nuapia	WITS
Wandile Nomquphu	WRC	Andrew Gemmell	Umvoto Africa
John Dini	WRC	Ebrahiem Abrahams	Umvoto Africa
Clara Bocchino	SWP	Helen Seyler	Delta H GWS
Rodolfo Camacho	SWP	Kai Witthüser	Delta H WSM / UFS
Fhedzisani Ramusiya	DWS SA	Kirsty Gibson	UFS
Mxolisi Mukhawana	DWS SA	Sibusisiwe Makhanya	IBM
Henry Roman	DSI	Claudia Ruz-Vargus	UN IRGRAC
Inga Jacobs-Mata	IWMI	Zaheed Gaffoor	UWC
Karen Villholth	IWMI	Yolanda Kanyama	WITS
Nebo Jovanovic	CSIR	Tyrel Flugel	Umvoto
Kevin Pietersen	UWC	Bochengedu Somolekae	DWS Botswana
Hlanganani Tutu	WITS	Brent Wells	USAID-GDL
Antoine Bagula	UWC	Francois Davel	DSI
Mark Anderson	USGS	Mmapitso Mokotedi	DSI
Graham Paul	USAID	Kgomotso Seikaneng	DSS

Table of Contents

List of Figures and Tables

Abbreviations

ANN	Artificial neural network
ARC	Agricultural Research Council
CDWR	California Department of Water Resources
DWS	Department of water and sanitation (South Africa)
FFNN	Feed forward neural network
GSP	Groundwater Sustainability Plan
GMA	Groundwater management area
GMU	Groundwater management unit
LSTM	Long short-term memory
magl	Metres above ground level
mamsl	Metres above mean sea level
MI	Mutual information
ML	Machine learning
NGA	National Groundwater Archive
NARX	Nonlinear Autoregressive with Exogenous Inputs
NNAR	Neural Network Autoregression
RNN	Recurrent neural network
SAWS	South African Weather Service
SGMA	Sustainable Groundwater Management Act
SGS	Sustainable groundwater strategy
TBA	Transboundary aquifer
WARMS	Water Authorisation Registration Management System

1 INTRODUCTION AND BACKGROUND

1.1 THE BIG DATA ANALYTICS AND TRANSBOUNDARY WATER COLLABORATION FOR SOUTHERN AFRICA

This research project, managed by the Water Research Commission of South Africa, is part of a series of four projects under the Big Data Analytics and Transboundary Water Collaboration for Southern Africa, bringing together key stakeholders in Water and Big Data sectors.

The Collaboration was first conceptualised in 2014 during the African Leaders Forum in Washington D.C., between United States Agency for International Development (USAID) Global Development Lab and IBM Africa Research, which had opened its first hub in Nairobi (Kenya) in 2013, followed by the Johannesburg Lab in 2015. Since the early 2000s, the regional USAID mission for Southern Africa had been intensifying its regional support for transboundary water systems with both the Ramotswa Aquifer Project, involving Botswana and South Africa and the Resilience in the Limpopo River Basin Program (currently in its second phase with the Resilient Waters Programme, covering the entire Southern Africa region, with a focus on the Limpopo and Okavango River Systems). As part of this process, USAID had also been engaging with Southern African Development (SADC): Groundwater Management Institute and the Department of Science and Innovation of South Africa to support knowledge and technological advancement in the region. The focus of this multi-agency collaboration was agreed as Big Data Analytics and Transboundary Water. On 3 April 2017, the partners met with a multi-stakeholder regional group in a dynamic "Idea Jam" hosted by the IBM Africa Research Lab in Johannesburg. The objective was twofold:

- to answer the broad question "how best can big data analytics be used to enhance transboundary water management", and
- to identity the research questions, which would have guided the projects.

Requiring the collaboration of at least five high profile government agencies and private institutions, it took over one year to move from the Idea Jam to the launch of the Call for Proposals in August 2018, and the awarding of the four research projects in January 2019.

Funding partners	Supporting partners	Managing partners
Department of Science and Innovation	IBM Africa Lab	Water Research Commission
USAID GDL & Water	USAID Southern Africa Mission	Sustainable Water Partnership
SADC-GMI	USGS	
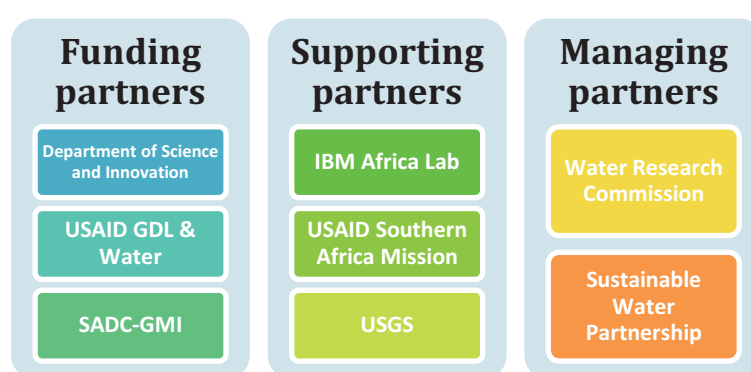
Figure 1-1   Collaboration partners and functions

The Collaboration: its partners and objectives

Currently, the Collaboration has seven partners, with a joint function for USAID Global Development Lab, Water Office, and Southern African Mission. The partners each contributed to the development of the research projects based on own technical and funding capacity, see Figure 1-1. The total funds provided by the Funding Partners to research directly amount to USD $ 500,000 (40%, 40%, 20%). IBM Africa contributed with the provision of the venue in Johannesburg, ad hoc, but more importantly, by sponsoring the internship programme to the five candidates from the research projects.

The Water Research Commission (WRC) is primarily tasked to oversee the financial and implementation management of the four research projects, as well as final reporting. The Sustainable Water Partnership (SWP) was called in by USAID in 2018 to act as the overarching Programme Coordinator, tasked with providing relation management, overall objective achievement, direction and positioning for the Collaboration in the region, and the fostering of a Community of Practice.

The United States Geological Survey (USGS), IBM Research and SWP provided three sets of online training on issues pertaining to the focal topics of the Collaboration, which are now available on the Collaboration YouTube channel.

The Collaboration partners defined the objectives for this first phase of action, see Table 1-1. However, the long-term vision is to create a Community of Practice for research and innovation on Big Data for Water Security, building on the multi-donor environment which has proven successful.

Table 1-1   Collaboration goals and objectives

Goals	Objectives
Enhance current understanding of shared groundwater resources	Improve transboundary groundwater management and collaboration
Provide big data skills development, capacity building and networking opportunities for Southern African researchers and their students	To foster multi-agency collaborative funding opportunities
To promote innovative thinking and application of Big Data Analytics to the Transboundary Water sector for integrated decision-making	To plant the seed for a growing community of pioneers in the use of Big Data Analytics for the study and management of Transboundary Water Aquifers

Research projects: funding and training

The four projects were awarded between December 2018 and January 2019, and they all focus on a secondary river basin in the region: the Ramotswa, part of the Limpopo River Basin, spanning Botswana and South Africa. All the lead institutions of the project teams have partnered, see Figure 1-2, with Botswana government and private institutions, as well as other leaders in previous water programmes in the area, such as UN-IGRAC[1] (partner of Team 1) and IWMI[2], implementers of the Ramotswa 2 USAID Project.

Despite working independently to address own project topics, the four research teams have progressively worked together to provide better integration for their outcomes. This process was led by the SWP in respect of providing a communication forum for the team leaders but was enhanced by the Internship Programme. The IBM mentors created a dedicated team and engaged the interns as individuals, as well as a group to help each other resolve new questions in coding and Machine Learning.

---

[1] International Groundwater Resources Assessment Centre of the United Nations
[2] International Water Management Institute

T1: Consolidation of data and application of big data tools to enhance national and transboundary data sets in Southern Africa that support decision-making for security of water resources.
- Umvoto Africa, University of Botswana, other global

T2: Consolidation of data and application of big data tools to enhance national and transboundary data sets in Southern Africa that support decision-making for security of water resources.
- Witwatersrand University, Geological Services of Botswana, DWS

T3: Localizing transboundary data sets in Southern African: A case study approach
- University of the Western Cape, CSIR, L2K2 Consultants

T4: Groundwater secure transboundary systems
- Delta-H Groundwater Systems Modelling and Institute for Groundwater Studies

Figure 1-2   Title of the four thematic areas and projects

The future prospects

As the current phase is coming to an end with the closing of the four research projects, the Collaboration partners are already identifying new opportunities to build on the lessons learnt and address the gaps recognised in this preliminary work, enhance the partnership to include national and regional government stakeholders, as well as new funding partners.

The focus of the Collaboration will remain the nexus between Big Data Analytics and (Transboundary) Water Security, recognising the inter-relatedness of successful water management in both national and shared aquifers to both human development and environmental goals.

1.2   PROJECT MOTIVATION

Around 70% of the SADC region is rural in nature, and more than 60% of communities in the region rely on groundwater (GWC, 2001). In addition to providing for domestic needs, groundwater, as a local resource, has the potential to catalyse further development by providing for irrigation schemes, and providing for industrial use. Altchenko and Villholth (2013) have identified 29 regional scale aquifers that extend across, and are therefore shared across, international borders (Transboundary aquifers; TBAs). It has been noted that "proper development and management systems need to be in place in order to jointly manage and harness this resource [TBAs] in an economically, environmentally and socially sustainable manner" (GWC, 2001).

But there are significant challenges to realising the potential for groundwater use. In many rural areas the capacity of the relevant local authorities is inadequate (Cobbing et al, 2015). The necessary finance is often absent for the maintenance of groundwater infrastructure. Rural groundwater supply schemes (often developed through funding sources external to the local authority i.e., international development or emergency funds) are left to fall into disrepair, and the scheme fails (for example schemes developed under the SADC Sustainable Groundwater and Drought Management Project in the Limpopo Basin TBA; SADC-GMI, 2017). Cooperation is required between users and between relevant authorities. These requirements are particularly challenging in rural areas that cross international boundaries, where limited datasets are in different formats, and where activities in one country may impact on groundwater resources in another. TBAs are at risk of over-exploitation related to various shortcomings including knowledge gaps on the situation in the neighbouring country, and lack of appropriate organisational structures.

Responding to these challenges for TBAs requires a variety of interventions. Some research has focused on harvesting data from various international sources, and making this data available internationally (for example, the SADC Groundwater Information Portal (SADC-GIP)[3], and the Ramotswa Information Management System, which is Phase 1 of the Ramotswa Project[4]). Theme 1, 2, and 3 of this proposal address the need to merge data from disparate databases or structures, and researching new ways of generating, analysing and visualising groundwater data.

But access to data is only the start of sustainable groundwater management. Fundamental to the management of any aquifer is knowledge of the groundwater flow regime, and knowledge of the long-term change in groundwater levels, recharge, and discharge to surface water when it is pumped. Knowledge of contamination threats and potential responses is also key. This kind of hydrogeological assessment forms the foundation for the determination of limits of acceptable impact and provides a benchmark against which to compare future monitoring datasets and take groundwater management decisions.

To establish groundwater-secure aquifer systems (and TBAs) requires that this kind of baseline hydrogeological assessment is in place, is routinely updated, and that it is used to inform groundwater management decisions. It is this element of sustainable groundwater management that this project intends to address, through i) the development of a best practice approach for a sustainable groundwater use and ii) the testing of new tools and approaches feasible from big data for sustainable groundwater use.

Whilst there are many standards and best practice guidelines for groundwater development at scheme level (for example GWC, 2001), there are very few for regional aquifer resources management. "Groundwater Sustainability Plans" (GSPs) are generated for major aquifers in California under the new Sustainable Groundwater Management Act, (California State, 2014), and have a relatively standard content and approach. The national modelling project in the UK similarly provides groundwater yield assessments for major aquifers in the UK, which are updated and inform management. However, the SADC counties do not have the same: there is no (routinely updated) aquifer scale groundwater resources assessment, which informs management plan and water use strategy, in place for most major aquifers in the SADC countries, nor the TBAs.

1.3   RESEARCH AIMS, OBJECTIVES, AND OUTCOMES

Whilst the original terms of references and proposal suggested that a sustainable groundwater strategy be developed and implemented for one prioritised TBA (i.e., completing an assessment of groundwater availability, groundwater quantity and quality protection, and developing a water supply strategy), this approach was amended during the inception phase of the study with the input from the reference group. It was deemed more important to trial new approaches and develop new tools than carry out routine hydrogeology. This project therefore aimed to:

- develop a sustainable groundwater strategy (SGS), reflecting a best practice approach for achieving sustainable groundwater use (as defined in Box 1-1)
- pilot "big data" or new approaches for meeting parts of the sustainable groundwater strategy, in a case study site

The SGS that has been developed is essentially a list of actions necessary for achieving sustainable groundwater use and would be applied to an aquifer or group of aquifers (a groundwater basin). It is recommended that the approach be implemented particularly in heavily used aquifers, in aquifers with sensitive receptors, and the approach would support improved groundwater management in TBAs. The strategy is explained in full in this report. The actions within the

---

[3] https://apps.geodan.nl/igrac/ggis-viewer/viewer/sadcgip/public/default
[4] http://ramotswa.iwmi.org/home

strategy can be met using traditional hydrogeological approaches. The research under this collaboration tested how and where big data analytics or new approaches can support improved groundwater management, and hence each tool developed by each project in the collaboration represents a new approach to meet part of the strategy.

The full sustainable groundwater strategy was not implemented in its entirety in one test case, due to the scope and structure for these projects and for the SGS to be successful it needs to be implemented by the responsible authority for that basin. The authority mandated as responsible for management of water resources in the area would be the appropriate body to implement the strategy, i.e., the Department of Water and Sanitation (or the Catchment Management Agencies) for South Africa, or the Department of Water Affairs in Botswana. Whilst the Ramotswa TBA became the focus for theme 1 (section 1.5), various other case study sites were used to develop and test the tools developed under the programme, selecting a site that leant themselves to this task.

The overall outcome of the programme is, therefore, a set of new approaches (tools) tested in various case studies, that all support improved sustainable groundwater management and therefore represent tools to meet actions required in the sustainable groundwater strategy. The collaboration did not develop one single tool for achieving improved groundwater management because there is no single tool conceivable that could enable a decision maker to achieve sustainable groundwater use. The decisions required to achieve sustainable groundwater use cannot be automated; they need to be based on hydrogeological information therein based on hydrogeological analysis of data and model results. Hydrogeologists are required to do the necessary analysis and provide information to decision-makers. This process is illustrated by the SGS and described in this report.

Sustainable groundwater use (of groundwater from an aquifer or group of connected aquifers in a basin) is defined here as:

Abstraction of a yield that causes or is expected to cause in future, impacts on groundwater level discharge, recharge, groundwater quality, saline intrusion, and subsidence, which are considered socially, economically and environmentally acceptable.
Achieving sustainable groundwater use requires that the expected impacts of abstraction are quantified, acceptability determined with relevant stakeholders, and used to set limits on these impact areas. The limits become thresholds for each impact area which becomes an indicator.

Box 1-1   Definition of sustainable groundwater use

1.4 PROJECT TEAM

The project was led by Helen Seyler of Delta-H Groundwater Systems, who conceptualised the project, managed and directed the machine learning modelling from a hydrogeological perspective, and contributed to the final report. Project researchers Kirsty Gibson and Yolanda Kanyama completed the modelling for the project, which has contributed to the project and also to their Masters' theses. The students were identified during project inception phase and due to the inter-disciplinary nature of the project, students from different backgrounds were selected: Kirsty Gibson is a hydrogeologist whose Masters' thesis is registered with the University of the Free State at the institute of Groundwater Studies; and Yolanda Kanyama is a computer scientist whose Masters' thesis is registered with the University of Witwatersrand at the Computer Sciences Institute.

Prof Kai Witthüser of the University of the Free State and of Delta-h Water Systems Modelling contributed as a project advisor and student supervisor to Kirsty Gibson. Prof Ritesh Ajoodha of the University of the Witwatersrand is the student

supervisor to Yolanda Kanyama, and also contributed to the overall project by reviewing and advising on machine learning modelling carried out by Kirsty Gibson.

As part of the collaboration, internship-style training was provided by IBM for the two project researchers. The IBM supervisors contributed to brainstorming solutions to model challenges encountered along the way. Ndivhuwo Makondo provided technical support with the model codes.

1.5   CASE STUDY SITE SELECTION

During the initial reference group meeting for the programme, it was proposed that all projects apply the same case study, due to the overlap in aims, the intention for data sharing between projects, and the intention for the projects to integrate the tools generated towards one overall aim. The Ramotswa TBA was selected as this case study, because the Ramotswa TBA had been identified during proposal stage by Theme 1 as the most appropriate TBA for theme 1 to achieve their aims. The Ramotswa TBA is described in detail in the theme 1 report.

However, during mid to end 2019 delays were experienced by theme 1 in the collection of data from Ramotswa and a preliminary analysis of data availability suggested a paucity of time-series groundwater level data. In order to avoid delays for Theme 4, and in order to ensure sufficient data was available for the application of machine learning models (selected as the big data approach to pilot, section 3.2), this team considered the dolomite terrains across RSA for case study sites, as agreed with the reference group.

Where and how actions required in the SGS can be met by new approaches is outlined in section 3.2. The viability of using machine learning (ML) models for quantifying the relationship between abstraction and impacts on the groundwater flow regime as part of a groundwater availability assessment was selected for piloting within this study (our focus area, section 3.2). A case study area was required where groundwater is heavily used such that the relationship between (or impacts of) groundwater use on reduced discharge and/or storage are detectable in datasets. Areas with a detectable decline in groundwater levels as the basin transitions to a new dynamic equilibrium in response to abstraction were required. The abstraction and monitoring record must be sufficient compared to the aquifer response time to demonstrate these impacts. The relationship between abstraction and groundwater flow regime or impacts is established over the aquifer or basin scale, and so this assignment also implicitly dictated that we consider a regional aquifer – scale assessment (rather than wellfield or response to pumping in one borehole). The dolomite terrains of RSA had the best potential to meet these requirements.

Among South Africa's most important aquifers are the dolomites of the North West and Gauteng Province, which are found in the northern part of the country (Figure 1-3). The dolomite aquifers are highly productive aquifers containing good quality water that supplies significant domestic and agricultural water needs particularly in the North West. Several towns and settlements such as, Itsoseng, Lichtenburg, Mahikeng largely rely on groundwater from the dolomites as their main source of water. The dolomite aquifers have been naturally compartmentalised by volcanic rocks. These rocks intrude upwards through the dolomite sequence in long sheets referred to as dykes. Dykes are less permeable than dolomite hence they act as a barrier to groundwater flow across them. Figure 1-3 shows the major compartments that the dolomite terrains have been subdivided into.

Data from all dolomite compartments was collated to identify areas: with the greatest density of boreholes with groundwater level readings; where the groundwater level records were the longest and had least data gaps; where weather stations were in close proximity to groundwater data; and where flow gauge data was available from groundwater fed springs in the same or hydrogeologically-connected compartment. Based on these requirements, two dolomite compartments were identified as meeting the criteria:

Machine Learning Models for Groundwater Availability   20

1. Steenkoppies where a reduction in discharge has been related to groundwater use (Seyler et al, 2016).
2. Molopo/ Grootfontein which also has been reported as experiencing a reduction in groundwater storage related to abstraction (Cobbing, 2018).

Delineation of these case study area boundaries is based on the groundwater management areas (GMAs) delineated by DWA (2009), based on the compartment boundaries defined therein. This study seeks to utilise the best available dataset to test the viability of a new approach and does not seek to provide detailed hydrogeological analysis of the case study sites. The references cited should be referred to in order to provide more detailed hydrogeological information on the Molopo/ Grootfontein and Steenkoppies GMAs.
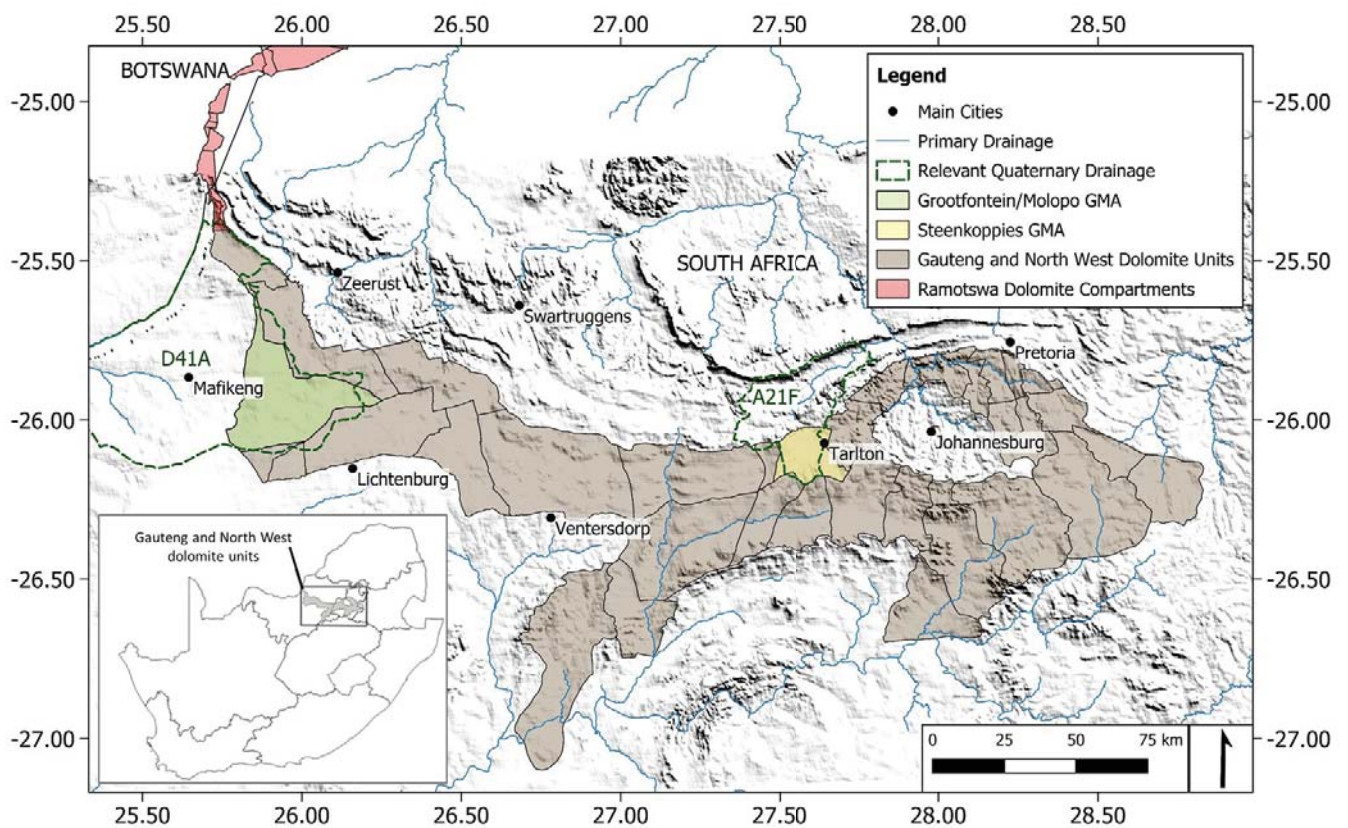
Figure 1-3 **Map showing Dolomite GMAs of South Africa with Molopo/ Grootfontein/ Molopo GMA and Steenkoppies GMA highlighted (compartment boundaries from DWA 2009)**

2 SUSTAINABLE GROUNDWATER STRATEGY

2.1 INTRODUCTION TO SUSTAINABLE GROUNDWATER USE

Supporting groundwater security in TBAs is the focus of the project and what the sustainable groundwater strategy (SGS) needs to address. Whist the theory of groundwater flow has not changed since the original work of by Theis (Theis, 1940, cited in Bennett et al, 1988), the concept of sustainable groundwater use has been more explicitly defined in recent years.

There are unavoidable consequences to abstracting groundwater, related to the source of the water abstracted. Aquifers, under natural conditions, are in a state of dynamic equilibrium. When pumping is initiated, abstracted water is met from aquifer storage and water levels decline (Bredehoeft and Durbin, 2009, Sophocleous, 2000; Alley and Leake, 2004). As the system reaches a new dynamic equilibrium the water level decline diminishes and abstracted water is met by a reduction in natural discharge, and potentially an increase in recharge (Bredehoeft et al, 1982, Bredehoeft, 2002, Devlin & Sophocleous, 2005). The time it takes the aquifer to reach this new dynamic equilibrium is termed the response time (Sophocleous, 2000; Alley and Leake, 2004). Water withdrawn artificially from an aquifer is therefore derived from a decrease in storage in the aquifer, a reduction in the previous discharge from the aquifer, an increase in the recharge, or a combination of these changes (Theis, 1940, cited in Bennett et al, 1988). The sum of the increase in recharge and decrease in discharge brought about by pumping was referred to as capture by Lohman (1972). It follows that an assessment of the sustainability of groundwater abstraction, or quantification of a sustainable yield for abstraction, would quantify what these changes in the flow regime related to that yield are (or will be in future, and when they will be realised), and determine whether the changes and their associated impacts are considered socially, economically and environmentally acceptable by the relevant stakeholders which needs to include the responsible authority, groundwater users, and representation of the environment . If the estimated changes are acceptable, the pumping (or total yield to be abstracted for the predicted changes) can be considered sustainable groundwater use (Sophocleous, 2000, Alley and Leake, 2004). Seyler et al (2016) label this approach as the capture principle approach to sustainable groundwater use. If the abstraction rate cannot be met by capture, because it exceeds the pre-abstraction discharge r plus any potential addition from enhanced recharge, groundwater storage continues to be the source of abstracted water. A new dynamic equilibrium is not achieved, and groundwater levels continue to decline. This is referred to as groundwater mining, and can be considered socially, economically and environmentally acceptable in specific cases (such as the Great Man-made River project in Libya, (Hiscock, 2005)).

Understanding these physical processes is fundamental to understanding what we are trying to achieve in testing the viability of using machine learning (ML) models for quantifying the relationship between abstraction and impacts on the groundwater flow regime as part of a groundwater availability assessment. A brief illustration of these physical processes is given in Box 2-1, and an illustration quantifying abstraction impacts on the flow regime is shown in Box 2-2. The references cited above should be examined to for fuller illustration of these processes.

Box 2-1   Introduction to Sustainable Groundwater Use – illustration of physical process

In their natural state, assuming the climate is stable, aquifers are in a state of dynamic equilibrium meaning that wet and dry years balance in the long term, and groundwater levels are relatively stable. Pumping disrupts this equilibrium. The graph below shows data from the West Coast District Municipality wellfield at Langebaan Road, on the West Coast (Woodford, 2005). The graph shows the groundwater level from two boreholes at the wellfield in the blue and green series, which are labelled "lower" because they are pumped from the lower of two aquifers. These boreholes are pumped at the combined rate shown ("LRA Wellfield"). Abstraction from this wellfield commenced in November 1999, prior to which there was minimal abstraction from this aquifer in this vicinity. The total abstraction rate varies slightly and in months with much less abstraction (i.e. October and November 2000, and May and June 2002) the groundwater level recovers.

> **"Initially** all abstracted water is met from aquifer storage and water levels decline. As the system reaches a new dynamic equilibrium the water level decline diminishes and abstracted water is met by a reduction in natural discharge, and potentially an increase in recharge."

Short term variability aside, the groundwater levels in the graph illustrate this physical process perfectly: initially the groundwater level decline is rapid; dropping around 8m in the first year. The abstracted water is being sourced from stored groundwater around the boreholes. Subsequently the rate of change in groundwater level slows significantly and in the subsequent 3 years the change is less than 2m. The aquifer system is approaching a new dynamic equilibrium in response to abstraction: a similar and relatively constant yield is being abstracted yet groundwater levels no longer significantly decline, because the yield is being sourced (or balanced) from sources other than storage. Discharge has been reduced and/ or recharge increased. In this case the response time is somewhat over 4 years.
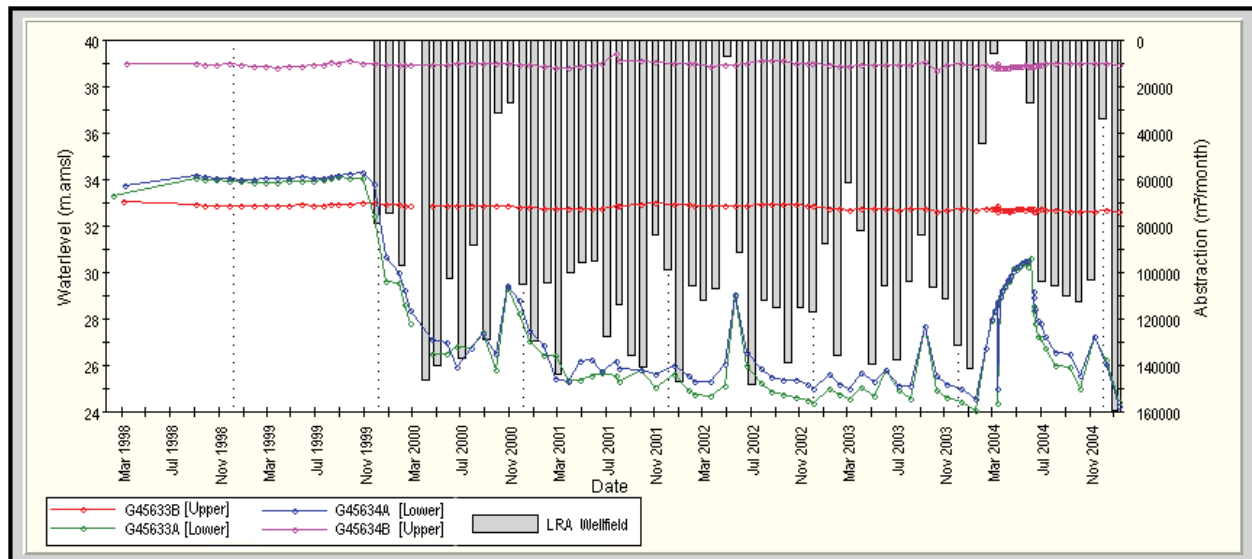
Both the expected impact on groundwater levels (at some stage potentially in future, after dynamic equilibrium is reached) and the expected impact on the flow regime need to be quantified in order to explore what impacts can be considered acceptable. The groundwater level gradient drives the rate of groundwater flow and discharge, however whilst groundwater levels are straightforward to measure, the corresponding flow and discharge rate is less so. At the aquifer or basin scale, quantifying impacts on the flow regime over time as abstraction progresses, and at the new dynamic equilibrium, and estimating the future groundwater level under dynamic equilibrium, relies on numerical models. Numerical models are calibrated generally against observed groundwater levels, and the model provides the resulting aquifer fluxes for the modelled groundwater level distribution. Models are generally developed that replicate the current situation (base case), and the response to other hypothetical pumping regimes is estimated. In a time-varying or transient simulation the numerical model solves the groundwater flow equation (thus providing groundwater level and aquifer fluxes) for each time step, and a steady-state model represents the basin's response at a new dynamic equilibrium.

The first output required from numerical models to enable stakeholders to take a decision over acceptability of impact of abstraction, is to run steady state models with increasing abstraction rates, and plot the resulting aquifer fluxes (recommended by Henriksen et al, 2008, and Seyler et al, 2016). An example of this graph is given below which shows fluxes out of the aquifer as negative and influxes as positive. The graph illustrates hypothetical abstraction scenarios from a model developed for the West Coast District Municipality wellfield at Langebaan Road, on the West Coast (Seyler et al, 2016). As abstraction increases, the discharge from the aquifer to the Berg River, Saldanha Bay and the ocean decreases. Whilst overall the river continues to gain from groundwater i.e. negative, abstraction greater than 6 million m³/a causes the Berg River to become a losing river over a short reach (increasing recharge by causing indirect recharge). In this case study, indirect recharge would be an unacceptable impact of abstraction, since the Lower Berg River has poorer water quality than groundwater in the area. The second output required is plot of each flux on the y-axis with time on x axis from time-varying simulations, and determine the response time based on when changes in fluxes minimise.
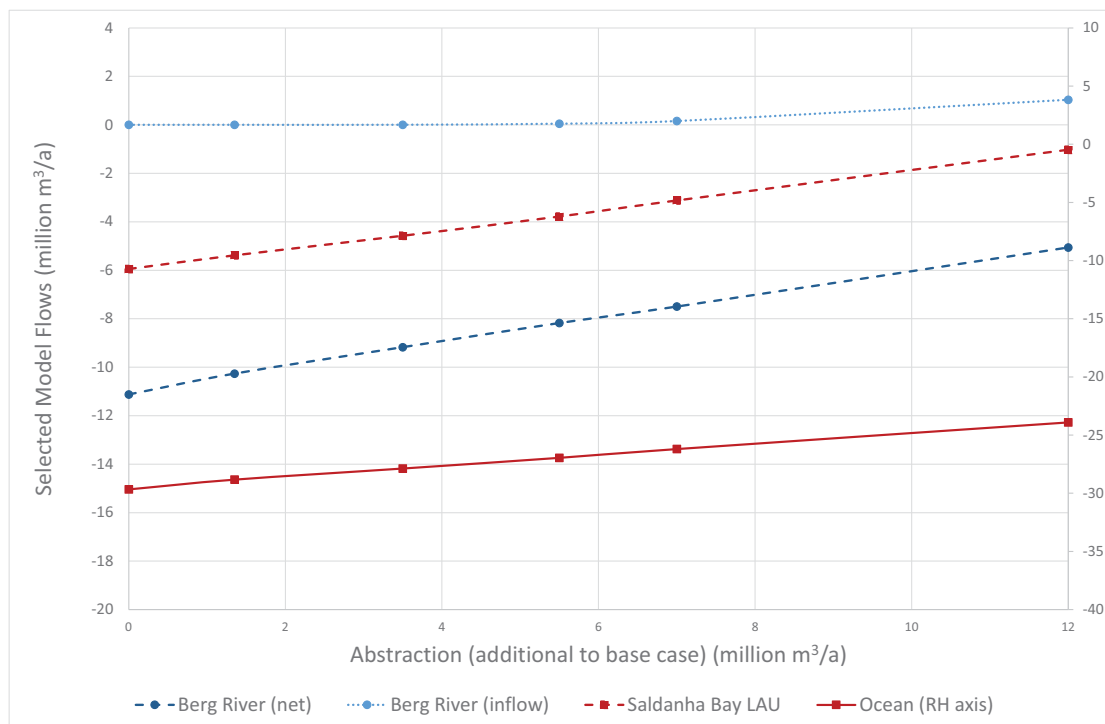
2.2   EXISTING APPLICABLE STRATEGIES OR APPROACHES

The capture principle approach to sustainable groundwater use is enshrined within the Groundwater Sustainability Plans (GSPs) required by the California Department of Water Resources (CDWR, 2014). The development of GSPs became a legal requirement for prioritised basins in California State related to the passing of the Sustainable Groundwater Management Act (SGMA) in 2014. The GSPs require that the impacts of abstraction are estimated, and acceptable limits on impact set by relevant stakeholders, dictating the available groundwater yield. Furthermore, the GSPs require that sustainability indicators are established, with thresholds related to the acceptable level of impact, in order to monitor the responses to abstraction against expected and acceptable impact.

This approach is not however mainstreamed in groundwater resources assessments in SADC countries. The fact that the changes in flow regime are hard to monitor in real time and that the estimation of future impact carries uncertainties, forcing the need for adaptive management, has led to the application of over-simplified static water balance type approaches that relate groundwater availability directly to some portion of recharge, and consider groundwater use to be sustainable if it is less than recharge (Seyler et al, 2016). Water level declines (alone) are also often mis-interpreted as indicative of unsustainable groundwater use, without quantification of the groundwater decline that _will_ occur as the aquifer system transitions to a new dynamic equilibrium. Seyler et al (2016) aimed to promote the capture principle approach to sustainable groundwater use with a focus on implementation in South Africa and proposed a decision framework (flow chart with set of required actions) for implementing sustainable groundwater use. The decision framework for sustainable groundwater use and the guidance document for the GSPs outlining the standard approach (CDWR, 2016a) were developed at the same time and are aligned technically; both implement the capture principle approach to sustainable groundwater use and contain sustainability indicators which overlap.

At the core of the decision framework is the upfront estimation of the future impact of abstraction on the flow regime (source of abstracted water), taking a socio-economic-environmental decision over acceptability, and monitoring to verify and update the estimation of future impact through the use of a set of sustainability indicators. The sustainability indicators derived by Seyler et al (2016) include changes in discharge, recharge, and aquifer storage and are intended to track the progress of the aquifer towards the new dynamic equilibrium under abstraction, in order to update the initial estimation of these conditions. The SGMA (and associated GSPs) incorporates a broader definition of sustainable groundwater use by explicitly including other impacts of abstraction including water quality, subsidence, and saline intrusion. These two resources (Seyler et al (2016) and the SGMA) form the basis for the SGS.

2.3   FRAMEWORK FOR A SUSTAINABLE GROUNDWATER STRATEGY

The proposed framework for a SGS is outlined in Table 2-1 and summarised in Figure 2-1, and is based closely on the steps listed in the California Code of Regulations for the Groundwater sustainability plans (CDWR, 2014), and the steps in the Decision Framework of Seyler et al, 2016.

SGS section 1, and subsections 2.1 to 2.4 represent relatively routine requirements. A groundwater availability assessment (termed groundwater balance in the GSPs and Californian regulations) is completed under **SGS subsection 2.5**, the requirements for which are in line with the capture principle approach to sustainable groundwater management. Subsection 2.5 specifically requires quantification of current and annual future groundwater balance or flows in the basin, which must include:

> i)   Current and predicted future groundwater inflows and outflows and resulting groundwater storage. Inflows to be considered include subsurface groundwater inflow and infiltration of precipitation, applied water, and surface water systems, such as lakes, streams, rivers, canals, springs and conveyance systems. Outflows from the groundwater system to be quantified include evapotranspiration, groundwater extraction, and groundwater discharge to surface water. Future groundwater inflows and outflows are to be estimated for

potential changes in abstraction regimes across the basin and known changes in surface water abstraction and/ or recharge. The potential for managed aquifer recharge should be incorporated in the future aquifer fluxes. It is recommended that the relationship between abstraction and groundwater fluxes is presented in a graph as illustrated in Box 2-2.

    ii)   Estimate of the resulting groundwater conditions (inflows, outflows, groundwater levels) when the planned or proposed yield is abstracted.

    iii)   Estimate of the response time.

The outputs of the groundwater availability assessment (specifically the predicted future impacts of abstraction) are relied upon to set sustainable management criteria in **SGS section 3**, which requires close collaboration with stakeholders in the basin (including the regulators or responsible authority). In line with CDWR (2014), the overall aim of the activities in **SGS section 3** is to ensure groundwater is managed to avoid the following six undesirable results:

1. Chronic lowering of groundwater levels indicating a significant and unreasonable depletion of supply if continued over the planning and implementation horizon.
2. Depletions of interconnected surface water that have significant and unreasonable adverse impacts on beneficial uses of the surface water.
3. Significant and unreasonable degraded water quality, including the migration of contaminant plumes that impair water supplies.
4. Significant and unreasonable land subsidence that substantially interferes with surface land uses.
5. Significant and unreasonable seawater intrusion.
6. Significant and unreasonable reduction of groundwater storage.

Each of the six undesirable results are related to an equivalent sustainability indicator, which are the "six effects caused by groundwater conditions occurring throughout the basin that, when [the effects are] significant and unreasonable, represent undesirable results" (CDWR, 2017, pg. 26). For example, the sustainability indicator for the first undesirable result, chronic lowering of groundwater level, is lowering groundwater levels. The six indicators are below and shown in (Figure 2-2):

1. Lowering of groundwater levels
2. Surface water depletion
3. Degraded groundwater quality
4. Land subsidence
5. Seawater intrusion
6. Reduction of storage

As a first step (**SGS sub-section 3.1**), a list of goals must be set for the basin, in the form of statements detailing the acceptable level impact. The sustainability indicators that are relevant to the basin are selected, and statements developed for each. Once acceptable impacts are agreed to by stakeholders, within **SGS subsection 3.2**, the equivalent sustainable aquifer yield can be established (using the outputs from **SGS subsection 2.5** that detail the impact of various levels of abstraction). Actual measurement of abstraction yields (and comparison of this to the established sustainable aquifer yield) is not included as a sustainability indicator (in line with both Seyler et al 2016 and CDWR 2014). The sustainable aquifer yield cannot be considered a fixed value as the factors influencing it may change over time, and the focus of the SGS is on the impact of abstracting the planned abstraction, rather than the yield itself:

> "Basin-wide pumping within the sustainable yield estimate is neither a measure of, nor proof of, sustainability. Sustainability under SGMA is only demonstrated by avoiding undesirable results for the six sustainability indicators" (CDWR, 2017, pg. 32).

Within **SGS subsection 3.2**, the recommended uses for groundwater are provided by matching groundwater availability with water requirements. Where groundwater is not being fully utilised, recommended uses of groundwater can be provided including decentralised use, centralised use for domestic purposes, industrial and agricultural uses. Where groundwater use already exceeds the sustainable aquifer yield, the associated thresholds will either be already breached or be breached in future (related to the response time), the SGS may inform the need for re-allocation of authorisations to abstract.

For each sustainability indicator, a measurable objective is set in **SGS subsection 3.3**, for example an objective for indicator 1 would be "maintain groundwater level above a specified level". A minimum threshold is established for this objective in **SGS subsection 3.4**, as illustrated in Figure 2-2, and Figure 2-3 using examples from CDWR (2017).

For protection groundwater quantity, the expected impacts when abstracting the sustainable aquifer yield is what informs the thresholds set for indicators on groundwater level, surface water depletion; an approach that therefore reflects pro-active protection of groundwater availability. Groundwater quality is a prescribed indicator and will therefore have thresholds set for concentration of relevant parameters. The thresholds would be selected based on an acceptable upper limit for variability from natural background, in order to protect against contamination of groundwater *resources*. There is a need in addition to establish a proactive approach to the protection of the quality of groundwater *sources*; through identification of threats (completed within section 2) and the delineation of source protection zones (SPZ), and the control of activities within the SPZs (for a summary on the definition of SPZ, and the approach to their delineation see Seyler et al, 2019). There is overlap in responsibilities between the authority who would implement the SGS (likely to be national government or catchment management agency), concerned with overall resource protection, and local government concerned with supplying clean water from boreholes. Delineation of SPZs could be the responsibility of either national government or local government, but the land use planning that the SPZ intends to inform is generally the responsibility of local gov. It therefore makes sense that local government also delineate SPZs at domestic supply boreholes; however, the SGS should at least support this process (in **SGS subsection 3.5**).

In **subsection 3.6** a monitoring network must be established, capable of collecting sufficient data to demonstrate the groundwater condition and whether sustainability indicators and their thresholds are being met thus evaluating SGS implementation and demonstrate progress towards achieving goals. The reasons for site selections should be given, along with monitoring protocols including methods, reporting requirements and standards.

In **SGS section 4** any additional or special measures that may be relevant and required for the area are handled, such as measures to manage saline intrusion, or remediation requirements, in addition to implementing the sustainability indicators and their thresholds.

Prior to implementation of the SGS there may be specific projects that need to be launched in order to support achieving the sustainability goals; for example, the establishment of a monitoring network and drilling of boreholes. **SGS section 5** allows for these projects to be planned, along with their costs, timelines and legislative requirements. As part of launching the SGS, implementation measures including how the SGS is updated, and reported on, are provided as part of **SGS section 6**.

2.4   APPLICABILITY, IMPLEMENTATION, AND ALIGNMENT WITH EXISTING LEGISLATION

Who is responsible to implement the SGS?
Development of the SGS would fall with the authority responsible for management and allocation of groundwater resources, which is generally national government (i.e., the Department of Water and Sanitation for RSA, or the Department of Water Affairs in Botswana). In many cases this responsibility is delegated to local catchment management

authorities (as is the case in RSA, similarly to the basin management units in California). In cases where there is significant use of groundwater for public water supply, and especially where this represents the major use of groundwater in the basin, there is overlap of responsibilities with local government, who would need to understand the impact of their planned abstraction (in order to receive a licence from the responsible authority and establish SPZs).

Is the approach applicable to TBAs?
The GSPs are required for all basins in California, with their development and implementation prioritised first in basins deemed as high and medium priority, based on the groundwater condition in those areas. Specifically, if long-term average abstraction exceeds the long-term average supply of water to the basin over a 10-year period, including enhanced recharge, such that abstraction is not maintainable (not met by a new dynamic equilibrium, section 2.1), then the basin is deemed to be in conditions of 'overdraft', and it is considered high- priority. The Decision Framework developed by Seyler et al 2016 was recommended therein for heavily used basins (defined as basins where groundwater use is greater than 65% of pre-abstraction recharge over the same area).

The need to carefully manage abstraction impacts with thresholds and indicators is arguably more necessary where there are significant impacts and therefore more likely where groundwater use is high compared to groundwater water supply. However, even where use is relatively low, unacceptable impacts may occur if abstraction occurs close to sensitive receptors (since distance to receptor and hydraulic diffusivity dictate the propagation of impact of abstraction, not abstraction yield).

The SGS represents an overarching approach applicable to any groundwater basin. Not establishing the relationship between abstraction yield and impact upfront prior to abstraction (i.e., not implementing a groundwater availability assessment as outlined in subsection 2.5) is equivalent to allowing an undefined future dynamic equilibrium to establish (Seyler et al 2016). Planned increases in groundwater abstraction often cause stakeholder concern over potential impact and can lead to misunderstanding over groundwater level declines and what these mean. The SGS approach enables expected impacts to be quantified up-front, prior to abstraction commencing, such that acceptable impacts can be pro-actively managed towards. This increases confidence, trust and buy-in from stakeholders, and the approach is considered especially necessary in TBAs where a groundwater basin is shared across international borders. Applying the SGS would generate the technical information on which to base discussions and necessary agreements over abstraction in one country which may cause impacts in another. The SGS therefore represents the idealised best-practice approach to sustainable groundwater management that these projects are recommending is implemented for TBAs.

How can the SGS be implemented?
The framework for a sustainable groundwater strategy presented in section 2.3 does not have the same level of detail as the literature for California, which includes legislation outlining the necessity for sustainable groundwater management and outlining that this be achieving through the development of GSPs (SGMA, CDWR 2014). The legislation controls the content of the GSPs, and additional guidance and best practice approach documents provide details to each of the GSP steps in legislation. It is acknowledged that full implementation of the SGS would require legislation to be amended, and more detailed guidance documents. Therefore, similarly to California, whilst the actions in the SGS are necessary in TBAs, they are unlikely to be fulfilled in their entirety until legislation is amended;

> "SGMA requires actions that have been necessary for many years or, in some cases, decades, but have not been politically feasible without a state mandate." (CDWR 2017)

The SGS does not address the governance aspects required for achieving sustainable groundwater management in TBAs but focuses on the technical steps required for supporting sustainable groundwater management. The SGS was developed at the start of the project process and is presented as the point of integration for the 4 project themes under

the collaboration. As all themes have sustainable groundwater management as the ultimate aim of any new tools or approaches being developed or tested, the outputs of each theme therefore meet part of the SGS.

Does the SGS align to existing plans / legislation?

There are some similarities between the indicators, measurable objectives and thresholds required in SGS, and the requirements of the Water Resources Classification System (WRCS) prescribed by the National Water Act in South Africa, and currently under development in Kenya. The WRCS also requires that thresholds are set to protect groundwater quality and groundwater availability, through the setting of Resource Quality Objectives (RQOs). However, most studies that have been completed have implemented a static groundwater balance type approach and dictated a groundwater yield as part of the RQOs, rather than focussing on undesirable consequences predicted based on the abstraction of an acceptable yield (reduced storage or reduced discharge). It is also a challenge to incorporate all the necessary elements of a sustainable groundwater management plan within the prescribed constraints of a WRCS study, where the methodology is based on surface water resources (Riemann, 2013; DWS, 2017). Nevertheless, as there is overlap at least between the RQOs that may be set for groundwater quality limits and the indicators required as part of the SGS, these should be in alignment.

Wherever the SGS is developed it should be used to feed information to the other processes that are mandated in those countries, for example there is clear overlap with any existing allocation planning processes that exist.

Table 2-1   **Outline of SGS showing required sections and required content**

SGS Section	SGS Sub-sections	Details / required content
1. Water requirements	1.1 Water requirements	Current and future water requirements of the groundwater basin and surrounding region (areas the basin may supply currently or in future), including current users and & potential future users, including domestic, agricultural and industrial requirements, and including seasonal patterns and peak demands. Required as basis to provide recommended uses of groundwater.
	1.2 Current water supply sources and yields of other (non-groundwater) sources.	Description of (all) available water resources to meet requirements and their yields. Required in order to understand the potential current and future demand from groundwater.
2. SGS area and aquifer setting	2.1 Description of SGS area	Describe area covered by SWS including locally relevant jurisdiction aspects, existing groundwater monitoring / management programmes and land use.
	2.2 Hydrogeological conceptual model	Graphical and narrative description of the groundwater flow regime (sources, sinks, flow direction) in the basin. To include information on and analysis of; topography, geology, soils, recharge areas and rates, groundwater level information and flow directions, discharge areas and rates, surface water bodies, identification of interconnected surface water systems and groundwater dependent ecosystems (including related factors such as environmental flow requirements), natural and impacted groundwater quality.
	2.3 Groundwater conditions	Current and historical trends in groundwater levels, groundwater quality, saline intrusion conditions, identification of subsidence and saline intrusion conditions.
	2.4 Groundwater quality threats	Assessment of aquifer vulnerability, and potential contamination sources based on analysis of land use, potentially polluting activities in the basin, and planned developments.
	2.5 Groundwater availability assessment	Quantification of current and annual future groundwater balance in the basin. To include: i) current groundwater inflows and outflows and predicted future flows and storage, based on potential changes in abstraction (quantification of relationship between abstraction yield and groundwater flows) ii) estimate of resulting conditions when the planned yield is abstracted iii) estimate of the response time.
3. Sustainable management criteria	3.1 Goals for sustainability indicators	Develop a list of sustainability goals for the basin, and select sustainability indicators (i.e., those that are relevant of the 6 prescribed indicators).
	3.2 Recommend groundwater uses	Based on sustainability goals, and related acceptable impact, establish sustainable aquifer yield (through liaison with stakeholders & authorities). Match groundwater availability with water requirements. Where groundwater is not being fully utilised recommended uses of groundwater can be provided. Includes decentralised use, centralised use for domestic purposes, industrial and agricultural uses. Where groundwater use already exceeds the sustainable aquifer yield the associated thresholds will either be already breached or be breached in future (related to the response time), and a re-allocation may be triggered.
	3.3 Measurable objectives	A measurable objective is set for each sustainability indicator.

SGS Section	SGS Sub-sections	Details / required content
	3.4 Thresholds	A minimum threshold is set for each measurable objective based on impacts that are acceptable.
	3.5 Source Protection Zones	Along with the identification of groundwater quality threats (in subsection 2.4), support the delineation of SPZs and the establishment of control of activities in recharge and SPZ areas.
	3.6 Monitoring network	Description of a monitoring network capable of collecting sufficient data to demonstrate groundwater condition, evaluate SGS implementation, demonstrate progress towards achieving goals. Reasons for site selections. Monitoring protocol including methods, reporting requirements and standards
4. Additional SGS elements	4.1 Any required additional measures	Development of any additional or special measures, such as measures to manage saline intrusion, or remediation requirements.
5. Management actions	5.1 Any projects or actions implemented to achieve sustainability goal	Planning, along with costs, timelines and legislative requirements, for any specific projects that need to be launched to support achieving the sustainability goals.
6. Plan implementation	6.1 Implementation measures	To include a roadmap for implementation including costs and schedule, mechanism for updates, mechanisms for control and reporting.

Figure 2-1 Flowchart showing the steps taken in the development of a SGS

Groundwater availability assessment
- Current and annual future groundwater balance in basin i.e. quantification of relationship between aquifer flows and abstraction.
- Estimate of response time

Groundwater quality threats
- Aquifer vulnerability assessment
- Identification of contamination sources

Groundwater conditions
- GW level and quality trends
- Seawater intrusion, subsidence

Hydrogeological conceptual model
- Topography, geology, soils
- Description, delineation, quantification of recharge & discharge
- ID interconnected SW systems and ecology

SGS area & aquifer setting

Description of SGS area
- Area covered by SWSS
- Landuse, existing programmes
- Relationship to other aquifers and other jurisdictional areas for water

Current water supply sources and yields
- Description of available water resources to meet requirements

Water requirements of the aquifer users
- Current & future domestic, agricultural, industrial requirements

Water requirements ①

②

③ **Sustainable management criteria**

Goals
- Describe goals for the basin in terms of sustainability indicators
- Related to decision on available yield and acceptable impact

Recommend GW uses
- Match availability with requirements
- Informs re-allocation or promoting increased use

Measureable objectives
- Describe the measurements to be taken to inform each sustainability indicator
- Describe path to achieve and maintain goal including interim milestones

Minimum thresholds
- Determine the acceptable threshold for each measureable objective
- How each will be measured, how the thresholds may affect interests of users or land uses

Source protection zones
- Delineate or support delineation of SPZ for supply wells prioritised based on identified threats
- Support the control of landuse activities in recharge zones and SPZ

Implementation measures
- Roadmap for implementation including costs and schedule
- Planned mechanism for update
- Planned mechanism for control and reporting

⑥ **SGS Implementation**

Project and action description
- Description of projects and actions implemented to achieve goals
- Costs, timelines, legal requirements

⑤ **Management actions**

Additional elements
- Any required additional or special measures for the area i.e. saline intrusion control, remediation requirements

④ **Additional elements**

Monitoring network
- Description of a monitoring network capable of collecting sufficient data to demonstrate groundwater condition, evaluate SGS implementation, demonstrate progress towards achieving goals
- Reasons for site selections
- Monitoring protocol including methods, reporting requirements and standards

Figure 2-2 Relationship between sustainability indicators, minimum thresholds, and undesirable results. CDWR, 2017.

Figure 2-3 Example of minimum threshold, for groundwater level. CDWR, 2017.

3 BIG DATA ANALYTICS

3.1 OVERVIEW

The overall aim of the collaboration is to test the implementation of big data analytics to achieve sustainable transboundary groundwater management. An overview of big data analytics is provided in the report for Theme 1.

3.2 APPLICATION OF NEW ("BIG DATA") APPROACHES WITHIN THE SGS

Konikow and Bredehoeft (2019) provide a thorough discussion of the source of abstracted groundwater based on a definition of sustainable groundwater use that is in line with that provided in section 2.1. Furthermore, Konikow and Bredehoeft (2019) summarise available approaches and technologies for measuring elements of the aquifer flow regime on pumping i.e., change in storage, change in recharge and change in discharge. Many of these approaches are traditional to groundwater science, and some are new and emerging technologies related to "big data approaches" (machine and deep learning). Examples from literature where "big data approaches" have been applied for meeting the requirements of the SGS are listed in Table 3-1 based on known examples. Machine learning has been applied in groundwater assessments for the analysis of high-resolution imagery, for monitoring changes in land use (the USGS LCMAP programme; Zhe & Woodstock 2014), and streamflow (the USGS PROSPER programme). The research carried out in parallel to this project, under the collaboration, includes:

1) the development of a master database for groundwater data, including the development of codes to assemble, verify and harmonise data from disparate datasets (theme 1)
2) the generation of new data sources using machine learning modelling techniques to generate groundwater quality data from citizen science sources (theme 2)
3) the generation of new data sources based on the downscaling of data from remote sensing data sources (theme 3)

Each of these research projects essentially provide new approaches that support the monitoring of groundwater, data from which informs part 1 and 2 of the SGS, and data from which is used to confirm aquifer performance compared to the established sustainability indicators (section 3 of the SGS). There appear to be opportunities for new approaches *for the monitoring* of groundwater and of the sustainability indicators established within the SGS, and there appears less application of new (big data) approaches for the setting of the groundwater availability related sustainability indicators.

The foundation to achieving sustainable groundwater use is to quantify the relationship between abstraction and aquifer fluxes, which generally requires numerical models (Box 2-2). Once this relationship is known, an acceptable level of impact and thereby thresholds for sustainability indicators can be established. The reliance on numerical models is recognised in the GSP guidelines and legislation which state that "GSPs **must use a numerical groundwater model** or an equally effective method, tool, or analytical model **to evaluate** and quantify **projected** [**groundwater availability and groundwater levels**] and to quantify surface water depletion (Cal. Code of Regulations §354.18 and §354.28(c)(6)). A **model is required to determine the equivalent groundwater yield available for minimum thresholds such as groundwater level**, or the amount of artificial recharge required to meet minimum threshold. Given the complex nature of groundwater and the interdependent responses of the system to change, consideration of the long-term implications of different management actions on these systems is virtually impossible without the use of models (Moran, 2016).

However, numerical models require time and appropriate expertise to develop, and have large data requirements to adequately parameterise the physical system. Furthermore, errors in the conceptual understanding of the aquifer lead to uncertainty in the numerical simulation. Our reliance on numerical models and their shortcomings has made it "appealing to consider data-driven & machine learning methods based on nonlinear interdependencies that may be able to predict groundwater level change without deep knowledge of the underlying physical parameters" (Sahoo et al, 2017,

pg. 3878). Machine learning methods have been widely applied to date in examples such as replicating monitored (historical) groundwater level change in order investigate the aquifer behaviour, for example determining whether climate or groundwater use is the major control on groundwater levels, and for changes in pumping (Ngoie, 2017). Research has also been carried out predicting the future groundwater level response to climate variability (Wunsch et al, 2017).

This project was to pilot various aspects of the SGS that can be supported by "big data approaches", within the case study site. Given the focus of this theme on supporting sustainable groundwater use, it is appropriate to concentrate specifically on testing the viability of machine learning methods to quantify the relationship between abstraction and aquifer fluxes (providing the data for establishing sustainability indictors). In piloting this aspect of the SGS, the research therefore aims to address the following questions (a subset of aim 2 in section 1.3):

- Can we use ML methods to **forward predict** groundwater level in relation to different future stresses (recharge, abstraction)? What is the efficacy in terms of length of prediction and stresses that are not seen in training dataset?
- How can we use ML methods to **generate aquifer fluxes**? What is the efficacy?

Whilst there is research on the use of ML models to predict the impact of rainfall changes on groundwater levels, and the use of ML models to understand the influence of factors on groundwater level change, ML models have not (to the knowledge of the team) been used for assessing aquifer wide groundwater fluxes and the impact of abstraction on these fluxes.

The approach applied to the machine learning modelling to address these research questions is as follows:
1. Simulate historical groundwater level [storage] and aquifer fluxes over time [discharge to surface water, groundwater abstraction] and if possible, investigate demonstrate causal / effect links. This requires at minimum a ML model able to recognise pattern between groundwater level, discharge, use, rainfall.
2. Use the developed ML model to project groundwater level (storage) forward for a different abstraction or rainfall input.
3. Use the developed ML model to project discharge forward for a different minimum groundwater level (storage) and different abstraction.

Table 3-1 **Selected elements of SGS where new technologies (big data approaches) have been applied in literature**

SGS Sub-sections	Traditional tools	Selected new tools (literature)
2.1 Description of SGS area	Spatial data processing, literature review	1) Airborne electromagnetic methods to delineate aquifer in 3D and changes in storage detectible (USGS; Valseth / Valder). 2) Artificial neural networks (deep learning) for predicting and understanding groundwater level responses (Ngoie, 2017; Wunsch et al, 2017)
2.2 Hydrogeological conceptual model	Spatial and point data processing and analysis, literature review.	
2.3 Groundwater conditions	Mapping and graphing of water levels and quality.	
2.5 Groundwater availability assessment	Borehole pump testing for borehole yields. Numerical groundwater models to determine aquifer yield in relation to acceptable impacts of abstraction.	**Changes in storage:** 1) Numerical models calibrated to groundwater level change provide change in storage as output (Faunt et al, 2009). 2) Head changes monitored and integrated with estimates of storage coefficient (McGuire, 2017). 3) Microgravity: Repeat surveys for monitoring changes in groundwater storage (if GWL and porosity is known) (USGS: Koth & Long 2012) **Changes in discharge:** Monitoring changes in streamflow (and could be correlated to changes in GW contribution to baseflow) via high resolution satellite imagery (images processed with machine learning techniques to recognise and quantify flow). USGS PROSPER. Recharge and changes in recharge: Monitoring land use changes (and predicting land use changes) in order to assess impact on water demand (agricultural changes) or recharge (hardening or surfaces). (USGS's LCMAP: Zhe & Woodstock 2014. (Use of algorithm for continual automated analysis of Landsat images)

3.3 MACHINE LEARNING

3.3.1 Introduction to machine learning

Machine learning is an evolved subfield of artificial intelligence (Marinósdóttir, 2019). Mohri *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 has become a powerful tool for predictive analysis (Ye, 2015). Machine learning is a broad and growing field. Figure 3-1 summarises the various categories and types of machine learning models applicable to this study.

Machine learning algorithms are generally grouped into two categories: supervised and unsupervised algorithms (Table 3-2). 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 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 (Alloghani et al., 2020). In unsupervised learning, the algorithm infers patterns from the dataset with no reference to known or labelled outcomes. The goal is to discover the underlying structure of the data. Unlike in supervised machine learning, unsupervised machine learning methods cannot be directly applied to regression or classification problems because the model does not have a reference for what the output data should be, making it impossible to conduct model training (Alloghani et al., 2020). To extract meaningful information, the model explores the structure of the data. There are a wide variety of machine learning techniques to model regression data using supervised learning. One of the more popular techniques is the artificial neural network (ANN) (Maier and Dandy, 2000).

Table 3-2   **Comparison of supervised and unsupervised machine learning algorithms**

| Basis for Comparisons	Supervised	Unsupervised
Requirements	Input and output variables are clearly labelled	Only input data is given with no reference to output data
Goal	To determine the function so well that when new input data is given, the model can predict output	To model the hidden patterns or underlying structure in the given input data to learn about new data
Example	Classification - predicting a discrete class label output (i.e., 0 or 1) Regression – predicting a continuous quantity output (i.e., time series data)	Clustering - dividing the data points into several groups such that same trait will be together in the form of cluster Association - finding associations amongst items within large commercial databases
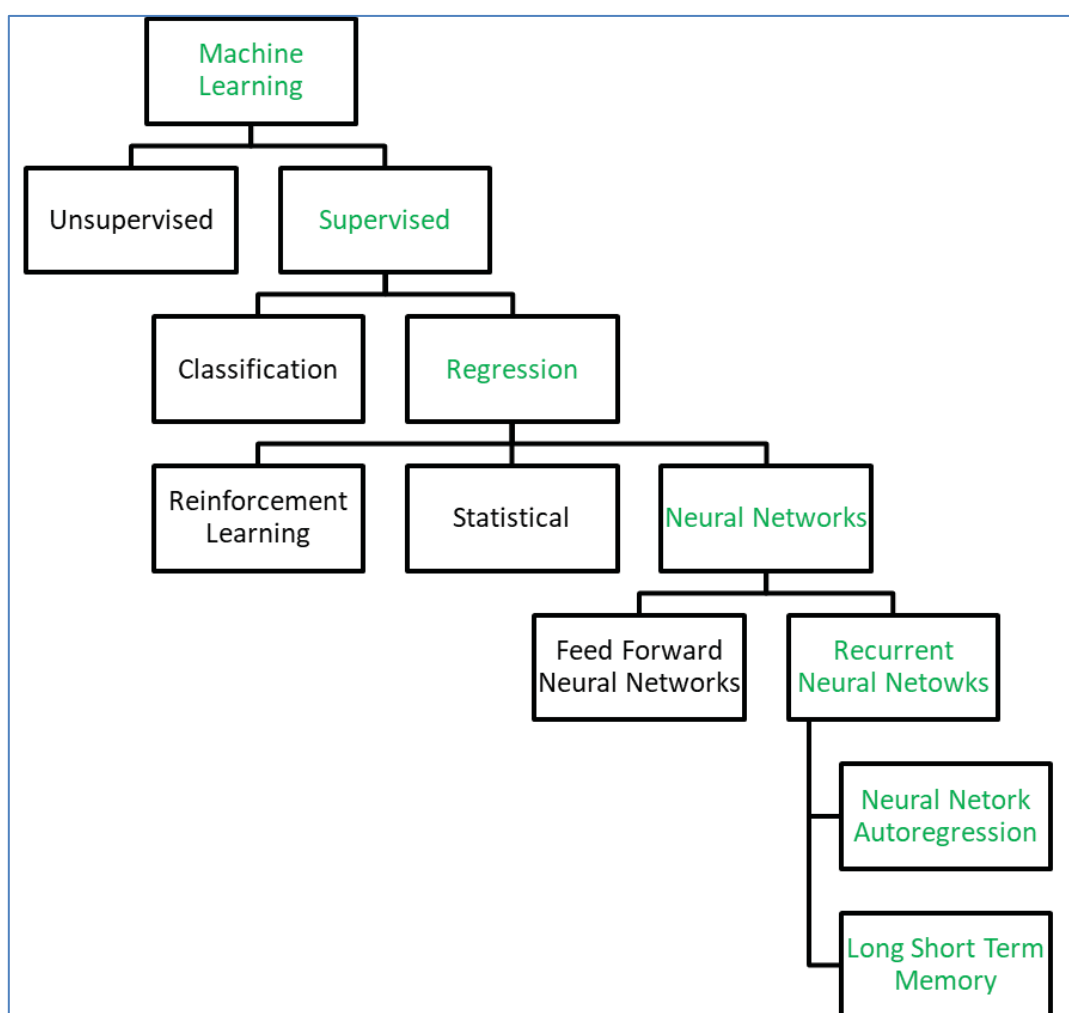
Figure 3-1   **Flow chart summarising the classification of some machine learning techniques. The green text highlights those applicable to this study.**

3.3.2   Artificial neural networks (ANNs)

A wide variety of different machine learning techniques are available to model groundwater levels. In this work, supervised ANNs, specifically the RNNs the long short-term memory and the neural network auto-regression models were applied to predict groundwater levels in the Grootfontein and Steenkoppies compartments respectively.

ANNs are a type of machine learning algorithm designed to simulate the way the human brain analyses and processes information. The human brain can memorize, learn, recognize patterns and yet still generalise in a wide variety of circumstances. The ability of the brain to do this is the driving factor behind the development of neural networks which attempt to mimic biological neural systems (Engelbrecht, 2007). ANNs are one of the most well-known machine learning algorithms and was first developed by McCulloch and Pitts in 1943 (Engelbrecht, 2007). 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). In hydrology, ANNs have been widely used as a supervised algorithm, specifically as a regression problem for time series forecasting. Neural networks have been favoured due to their ability to capture and represent input and output relationships, whether linear or nonlinear.

The architecture of an ANN consists of three main parts namely; the input layer, where data is received from the external environment; hidden layers, 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 (Da Silva *et al.*, 2017). The hidden layer can be considered similarly to a "black box" which is used in modelling. Each layer of an ANN comprises of nodes that are linked by weighted connections. 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). Depending on how the connections between the neurons are organised, an ANN can either be classified as a feedforward neural network (FFNN) or a recurrent neural network (RNN).

If the connections between the neurons are in one direction, from inputs to outputs, the ANN is classified as an FFNN (Figure 3-2) (Agatonovic-Kustrin and Beresford, 2000). Nodes in one layer are connected to nodes in the next layer but not nodes in the same or previous layer. FFNNs have been used in modelling hydrological events and for groundwater levels in aquifer systems (Coulibaly and Baldwin 2000). However, FFNNs can only accurately predict future values one step ahead (for a short period) hence they are often described as being static (Giles et al. 1997; Chiang et al. 2004). The only input FFNN uses for predictions is the current data point fed into the model. FFNN are not able to "remember" the past data but instead remembers the formative moments of training (Giles *et al.* 1997; Haykin 1999). They "remember" the relationships between data, but to make a prediction using only the information at that specific time step. Therefore, while FFNNs are more prevalent in modelling hydrological systems as stated by Maier and Dandy (2000), they lack the feedback connections necessary to model dynamic systems making them less applicable in the modelling of time-dependent patterns (Brezak *et al.* 2011).

If the connection between the neurons is in both a forward and backwards direction, the ANN is classified as an RNN (Figure 3-2) (Da Silva *et al.*, 2017). RNNs are different from FFNNs as the networks final output is directed 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 for time-variant systems, such as time series prediction, system identification and optimisation, process control etc. (Brezak et al., 2012). The chronological information stored in the RNNs hidden state extents several time stamps and cascades forward to influencing the predictions going forward (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 RNN to determine time dependant patterns more accurately compared to the FFNNs (Brezak *et al.* 2011). RNN have the potential, therefore, to be successfully applied to model hydrogeological processes and produce reasonable solutions even when there is limited information about the physical properties of the problem area. There are many variants of RNNs, each having a slightly

different architecture. 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 (Le *et al.*, 2019).
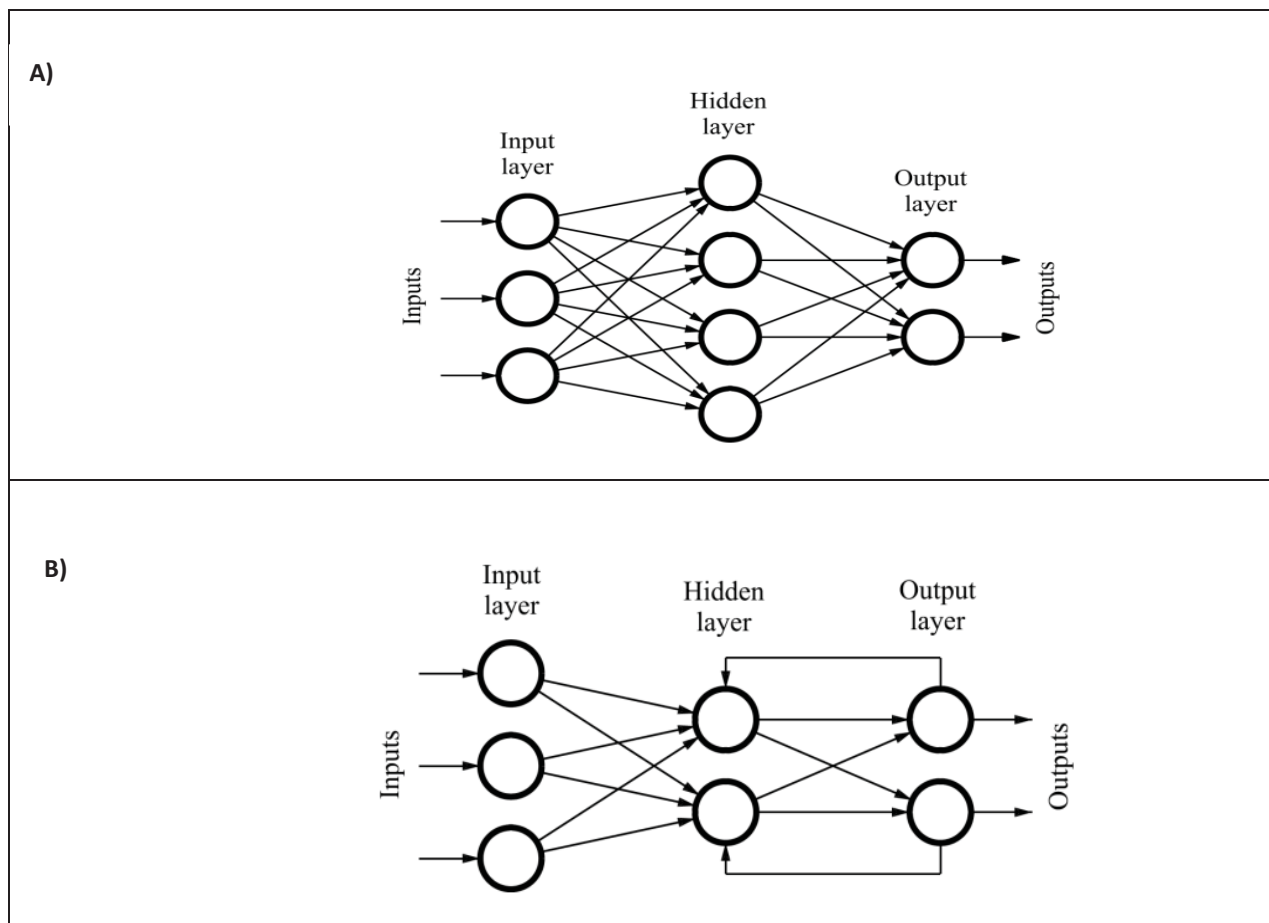
Figure 3-2 **The architecture of Artificial Neural Networks: A) Feed Forward Neural Network where the connections between the nodes are in a forwards direction and B) Recurrent Neural Network where the connections between the nodes are in a forward and backwards direction, and the output is directed back into the input of the same or previous layer and stored in a hidden state (Source: Quiza and Davim, 2011; Ruiz et al., 2016)**

3.3.3 Modelling Time-series 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, 2016). In this study, the groundwater levels are the forecasted target variable, using the x variables of rainfall, spring discharge, temperature and groundwater usage.

How to utilise the time series at hand to create the model and assess its performance is referred to 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). Time series order must be maintained when split to ensure that the model picks up the patterns in the data (Agrawal and Adhikari, 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, 2016). During training, the model makes use of 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, referred to as hyper-parameter tuning. Although one can set the model parameters manually, hyper-parameter tuning is typically carried out by the ANN adaptively based on the data fed to the model (Wickham, 2016).

The hyper-parameter tuning ability of the ANN can be optimised through several methods, one of which 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 section 4.2.2.1. Once the model training is complete, and the hyper-parameters 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 under-fitting. 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, 2016). Under-fitting is when the model is unable to simulate the training data or predict new data (Wickham, 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 (Agrawal and Adhikari, 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.

3.3.4 Use of machine learning models 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 has been explored in research (Kenda *et al.,* 2018). Some of the machine learning techniques successfully used to model groundwater levels include: ANNs first used in the field of hydrogeology 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).

Maier and Dandy (2000) assessed 43 papers on neural networks for the prediction and forecasting of water resources. It was found that most authors using ANNs for hydrogeological applications make use of FFNN, however recently there has been more investigations on the use of RNN, and the Neural Network Autoregression (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. Monthly groundwater levels were collected between 2000-2006 from 22 wells in the study area. The ANNs were developed by using feedforward neural network and trained with the Levenberg-Marquart algorithm. The ANNs predicted groundwater with an accuracy (an R^2) of 0.93. This study concluded that satisfactory groundwater level predictions could be made using limited groundwater level 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 up to four months. However, the model

performance deteriorates after two months prediction because the error in the model starts to accumulate and shows in the predictions.

Shamsuddin *et al.* (2017) and other authors have used neural networks to forecast groundwater levels on a daily temporal resolution. Shamsuddin *et al.* (2017) showed that neural networks 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. The FFNN was trained using the back-propagation algorithm. It was concluded that neural networks models were useful to predict groundwater level fluctuations even when highly variable anthropogenic factors are considered. The model was successfully used to predict groundwater levels with acceptable errors. However, a slightly higher error was encountered in the wells where there was a more significant anthropogenic influence on the groundwater.

Daliakopoulos and Coulibaly (2005) made use of FFNN, RNN and the Radial basis function network (RBF) to predict groundwater level fluctuations 18 months ahead. 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 backpropagation (GDX) algorithm.

Chang *et al.* (2016), is one of the few studies where groundwater level change at an aquifer scale is modelled at a monthly temporal resolution. This study used a combination of a Self-Organised Map (SOM) and Nonlinear Autoregressive with Exogenous Inputs (NARX) network (a type of 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 a NNAR for large scale environmental systems.

Guzman *et al*. (2017) used a 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 0.0013 m between measured and predicted values. Comparable to Guzman *et al.,* (2017), Wunsch *et al*., (2018) also made use of a NNAR model, the NARX, to model groundwater level as 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. The study demonstrated that the NARX can be used in these three aquifer types with simple readily available input variables. The wells that were influenced by pumping performed worse than those uninfluenced by pumping. Overall, the NARX showed promising results on groundwater predictions with the only two input variables.

Table 3-3 Summary of studies using ANNs in hydro(geo)logical systems. FFNN: Feed Forward Neural Network, RNN: Recurrent Neural Network, NARX: Non-linear Auto Regressive model with Exogenous Outputs, LSTM: Long Short-Term Memory

Author	Model algorithm	Target variable and Input variables	Temporal resolution	Total sample size	Prediction length	Accuracy
Sreekanth *et al.* (2009)	FFNN	**Target:** Groundwater levels **Input:** Rainfall, evapotranspiration, temperature, humidity, river discharge	Monthly	72	24 months	$R^2 = 0.93$ RMSE= 4.50 m
Nayak *et al.* (2006)	FFNN	**Target:** Groundwater levels **Input:** Rainfall, canal releases	Monthly	96	4 months	RMSE=0.32-0.58 m
Shamsuddin *et al.,* (2017)	FFNN	**Target:** Groundwater levels **Input:** Rainfall, temperature, river stage, stream flow rate	Daily	406	121 days	$R^2 = 0.55$-0.75 MSE=0.03-0.1 m RMSE= 0.01m
Lee *et al.* 2018	FFNN	**Target:** Groundwater levels **Input:** Rainfall, groundwater pumping, artificial recharge	Hourly	8712	2520 hours (105 days)	$R^2 = >0.9$ NSE= >0.75 *RMSE=3-6 cm*
Daliakopoulos and Coulibaly (2005)	FFNN RNN RBF	**Target:** Groundwater levels **Input:** Rainfall, temperature	Monthly	180	18 months	**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 months	$R^2 = 0.68$-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 dependent)	1 day – 6 months	$R^2 = 0.18$-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	Up to 3 months	$R^2 = 0.83$-0.92 MSE=0.001-1.002 m

Author	Model algorithm	Target variable and Input variables	Temporal resolution	Total sample size	Prediction length	Accuracy
						NSE=0.84-0.96 m
Le *et al.*, (2019)	LSTM	**Target:** Discharge **Input:** discharge, rainfall	Daily	Station dependant	1 – 3 days	*RMSE = 152.4 – 594m* *NSE = 86.2 – 99.1%*
Supreetha *et al.*, (2019)	LSTM FFNN	**Target:** Groundwater levels **Input: P**revious groundwater levels, rainfall	Monthly	Not highlighted	12 months	*LSTM: RMSE= 2.4* *FFNN: RMSE =3.1*
Zhang *et al.*, (2018)	LSTM FFNN	**Target:** Groundwater levels **Input:** water diversion, evapotranspiration, temperature, precipitation, time	Monthly	168	12months	*LSTM: R^2 = 0.789 – 0.952* *FFNN: R^2 = 0.004 – 0.495*
Bowes et al., (2019)	LSTM RNN	**Target:** Groundwater levels **Input**: groundwater levels, rainfall, sea level	Hourly	8 years	1 – 18 hours	*LSTM: RMSE = 0.09m* *RNN: RMSE = 0.14*

3.3.5 Long Short-term memory (LSTM)

The LSTM (a type of RNN) was first introduced by Hochreiter and Schmidhuber in 1997 to address problems of the drawbacks of modelling using an RNN (the accumulation of error with predictions further in time) by adding a unique set of memory cells that replace the hidden layer neurons of the RNN (Le et al., 2019). LSTMs are capable of avoiding long-term dependence problems that occur with the traditional RNN due to the added memory cells. Instead of a single neural network like the standard RNN, the LSTM model is capable of filtering information through a gate structure to maintain and update the state of the memory cells. Figure 3-4 shows the structure of the LSTM neural network.
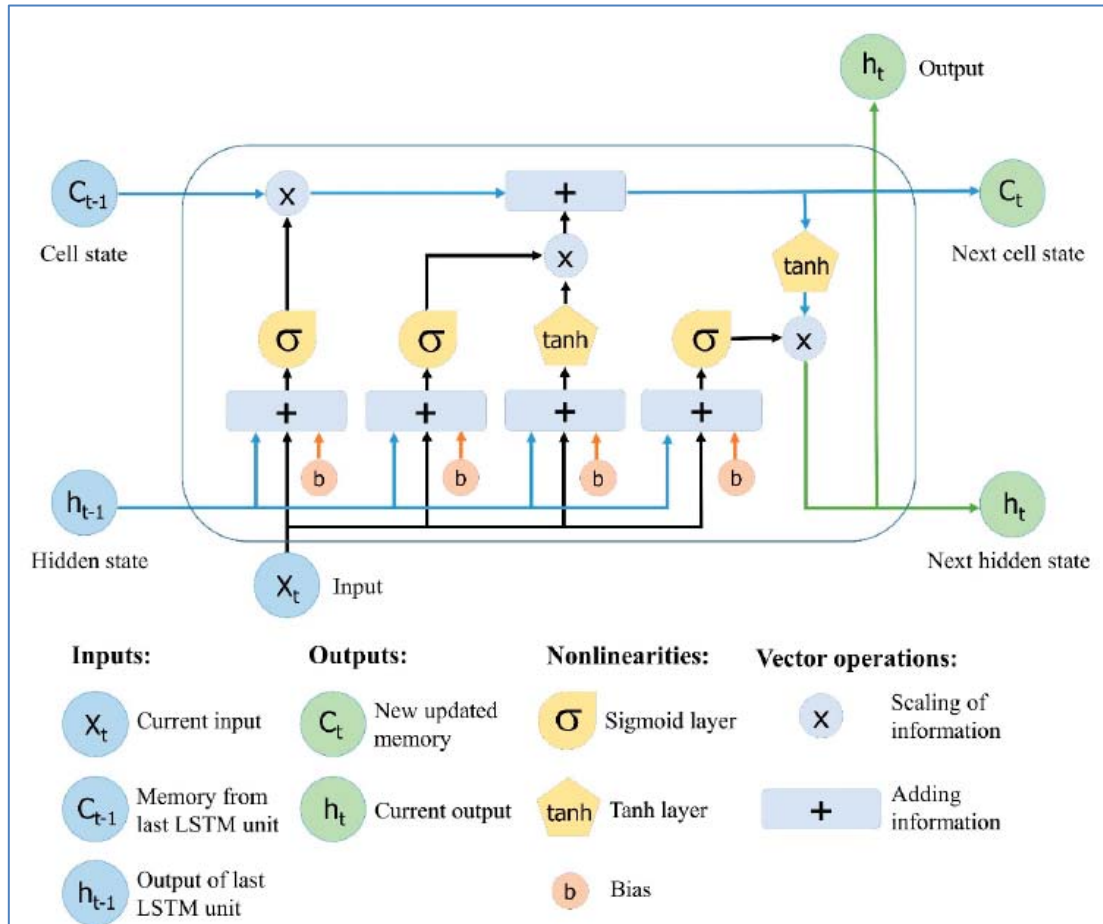
Figure 3-3   Structure of the LSTM neural network. Reproduced from Yan (2018)

As shown in Figure 3-4, the hidden state and the cell state are transferred to the next cell. The cell state is the main chain of data flow, allowing data to flow forward unchanged. Sigmoid functions are used to add or remove data from the cell state. Gates are similar to a series or layer of matrix operations which contain different individual weights. By using the gate structure to control the memorizing process, LSTMs can avoid the long-term dependency problem. The LSTM network has three gates: input, forgotten and output gates (Le et al., 2018).

The forgotten gate determines which cell state information is discarded from the model. This process of excluding and identifying data is decided by the sigmoid function. The function accepts the output $h_{(t-1)}$ at time $t − 1$ and the current input (X_t) at time t as inputs and combines them in a long vector $[h_{t-1}, X_t]$ through σ transformation to become

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f)$$

<div align="right">3-1</div>

Where W_f is the weight of the matrix, b_f is the bias of the forget gate and σ is the sigmoid function. The gate will output a value between 0 and 1 where 1 indicates complete reservation of information and 0 is complete discarding.

The input gate determines how much of the current time network input (X_t) is reserved into the cell state, which prevents insignificant information from entering the memory state. The input gate has two functions: the sigmoid function which decides whether the new information should be ignored or updated (0 or 1) and the tanh function which decides the level of importance (-1 to 1) by giving weights to the values. The two values are then multiplied to update the cell state, and this new memory is added to old memory C_{t-1} resulting in C_t.

$$i_t = \sigma(W_f[h_{t-1}, X_t] + b_f)$$

<div align="right">3-2</div>

$$N_t = tanh(W_n[h_{t-1}, X_t] + b_n)$$

<div align="right">3-3</div>

$$C_t = C_{t-1}f_t + N_t i_t$$

<div align="right">3-4</div>

Here, C_{t-1} and C_t are the cell states at time t - 1 and t, while W is the weight matrices and b is the bias of the cell state.

The output gate controls how much of the current state is discarded. The sigmoid layer determines which parts of the cell state make it to the output layer, and then the cell sate is processed by the tanh layer and multiplied by the output of the sigmoid function (O_t) to obtain the final portion.

$$O_t = \sigma(W_O[h_{t-1}, X_t] + b_O)$$

<div align="right">3-5</div>

The final output value of the cell is defined as:

$$h_t = O_t \tanh(C_t)$$

<div align="right">3-6</div>

3.3.6 Neural network auto-regression (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 and the LSTM), the NNAR's previous timestamps are not stored in a hidden state but given as another input to the model (Figure 3-5) (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 (3-7 and Equation 3-8) (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):

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

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):

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

3-8

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 different predictions each time the model is run (Scardapane and Wang, 2017). To obtain a robust evaluation of the model's 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 below 3-9 (Ruiz *et al.*, 2016):

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

3-9

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 noise

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.
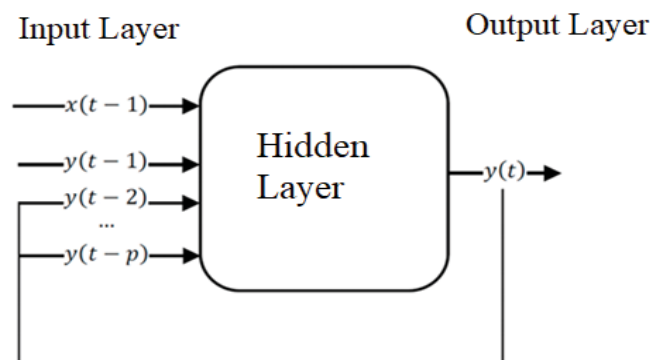
Figure 3-4 **Basic architecture of the Neural Network Autoregression (NNAR) where the connections between the nodes are in a forward and backwards direction, and the outputs are fed as a new input to the model (Source: Ruiz et al., 2016).**

4 CASE STUDY 1: GROOTFONTEIN

4.1 DESCRIPTION OF STUDY AREA (CONCEPTUAL MODEL)

4.1.1 Location and setting

The Grootfontein dolomitic compartment (Barnard, 2000; Meyer, 2014), formally delineated by DWA (2009) as the Molopo/ Grootfontein water management area (GMA) (section 1.5), is Among South Africa's most important aquifers. The compartment is located southeast of Mafikeng in the North West Province (Figure 4-1). The aquifer contains good quality water that supplies the domestic and agricultural water needs of people in Mafikeng region with 20% of Mafikeng's domestic water being supplied by the aquifer. Other towns such as Ventersdorp, Zeerust, Lichtenburg, Itsoseng and Ottoshoop also rely on the dolomitic aquifer as a source of water (Holland and Wiegmans, 2009).
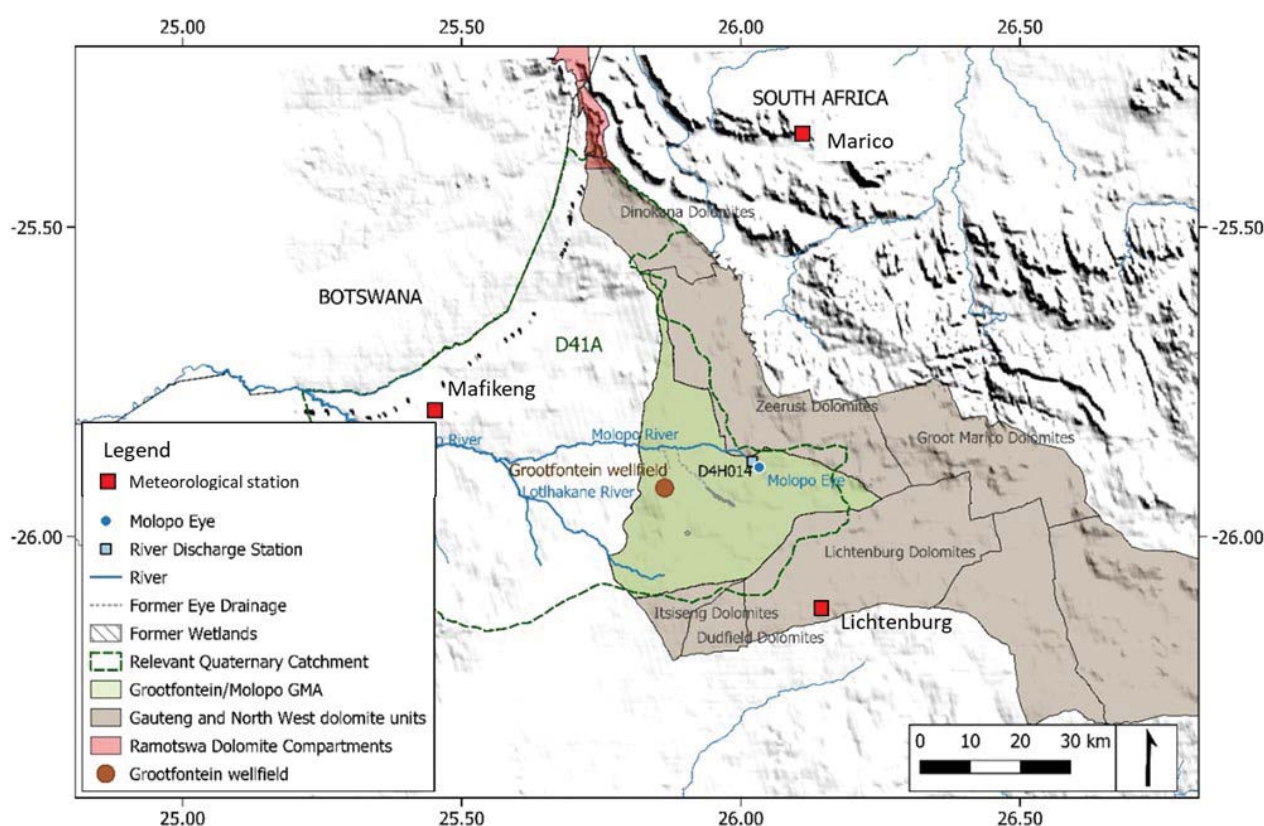
Figure 4-1   Grootfontein/ Molopo GMA location map.

4.1.2 Topography

The Grootfontein aquifer is a typical ancient karst landscape with a very flat land surface (Van Tonder et al., 1986), and the GMA covers an area of approximately 139 km². The flat, weathered land surface slopes very gently towards the north with a gradient of approximately 50 m in 16 km (Van Tonder et al., 1986).

4.1.3 Climate

Forming the southern part of the Kalahari Desert, the North West Province offers almost year-round sunshine. The summer months (from August to March) bring brief thunderstorms. The province has an above (national) average rainfall of 200 to 350 mm annually. Summer temperatures range between 22 and 27º C. The average winter (from May to July) temperature is 16ºC but can range from 2 to 20º C in a single day.

Three South African Weather Service (SAWS) weather stations surround the compartment; one within the Mafikeng region (North West of the compartment), one in Marico (North of the compartment), and another in Lichtenburg (South of the compartment), shown in Figure 4-1. The three stations were selected due to the availability of longer historical data and their positions relative to the compartment boundaries to provide an indication of the rainfall and temperature variability across the compartment. Monthly rainfall and temperature data were obtained from the stations (because groundwater levels were available monthly), over the timeframe shown in Table 4-1. The minimum, maximum and mean annual rainfall of each station is shown in Table 4-1. Figure 4-2 and Figure 4-3 illustrate the monthly rainfall and temperature values for each station over a common period.

The station located south-east of the compartment (Lichtenburg) appears to have higher rainfall between the period 1984 to 2009. Rainfall stations in the northern part of the compartment (Mafikeng and Marico) appear to have lower rainfall as compared to the southern area. High rainfall events usually occur in summer (from August to March). Moreover, the Marico area experiences higher temperatures as compared to Mafikeng and Lichtenburg areas.

Table 4-1   **Mean, minimum and maximum annual rainfall for stations around Grootfontein.**

Station	Data Availability	Minimum Annual Rainfall (mm), year	Maximum Annual rainfall (mm), year	Mean Annual Rainfall (mm)
04722780_Mafikeng	1994-2018	85.6 (1994)	777.4 (2010)	524.51
05080470_Litchenburg	1984-2018	152.7 (1984)	873.8 (1997)	507.86
05466303_Marico	1950-2019	159.8 (1994)	902.5 (1956)	609.43
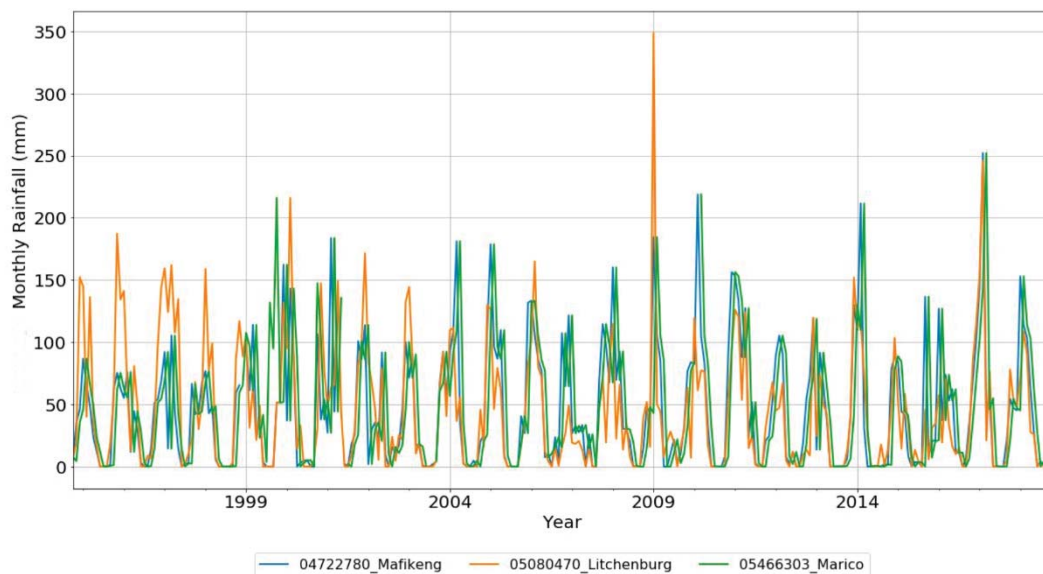
Figure 4-2   **Monthly rainfall from three weather stations around Grootfontein.**
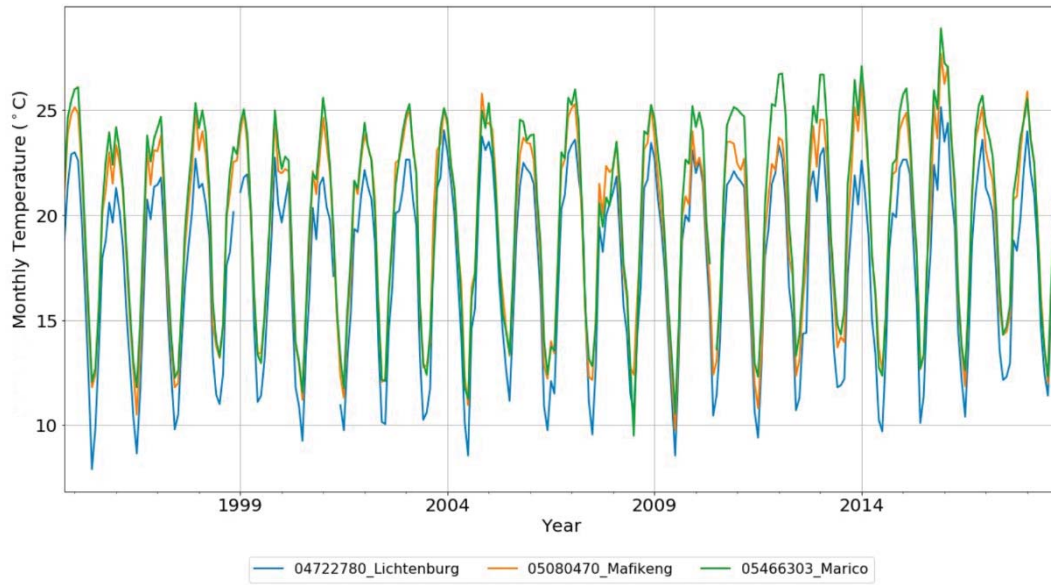
Figure 4-3   Monthly temperature from three weather stations around Grootfontein.

4.1.4   Hydrology and spring flow

Under natural conditions, recharge to the dolomitic aquifers is largely through rainfall, and they discharge to springs, seeps and wetlands (Cobbing, 2018). Spring discharge datasets were therefore required to meet the aims of the ML modelling (section 3.2).

The Grootfontein compartment falls within quaternary drainage D41A. The course of the Molopo River cuts across the compartment as shown in Figure 4-1. Historical data for two flow-gauge stations along the Molopo River recorded at monthly intervals was sourced from the DWS, namely: D4H013 (1964 – 2017) and D4H014 (1982 – 2019), shown in Figure 4-4. Gauge D4H014 is located at Molopo eye, which is the source of the Molopo River. Gauge D4H013 is located further downstream (west of D4H014) on the Molopo River (see Figure 4-1). The Grootfontein eye is located more centrally in the Grootfontein compartment. The spring discharge dried in 1981, which Cobbing (2018) attributed to the drilling of the Grootfontein wellfield boreholes in the vicinity of the spring, and to agricultural abstraction.

The Groundwater Management Area (GMA) of the Molopo/ Grootfontein aquifer incorporates sub-compartments. Cobbing (2018) uses the Kliplaagte (or D41A-03) Groundwater Management Unit delineated by DWA (2018) as the "Grootfontein aquifer" study area and considered the groundwater in this sub-compartment to discharge to the Grootfontien Eye. The study area selected here is the larger Groundwater Management Area (GMA) of the Molopo/ Grootfontein aquifer. The GMAs boundaries are considered by DWA (2018) to be aquifer boundaries in that recharge and discharge across that area is related. The spring discharge at Molopo is also considered (somewhat) hydraulically connected to groundwater across the Molopo/ Grootfontein GMA and is therefore used as a representation of long-term groundwater discharge from the GMA. The results of Information Gain Ranking assessment demonstrate there is indeed a strong correlation between groundwater levels across the Molopo/ Grootfontein Groundwater Management Area (GMA) and the Molopo Eye. Whilst this correlation does not prove or disprove causality (i.e., the central Kliplaagte Groundwater Management Unit may be less strongly hydraulically linked to the Molopo Eye than the Grootfontein Eye), it at least suggests that the Molopo Eye discharge data can be used as a proxy for discharge across the wider area.
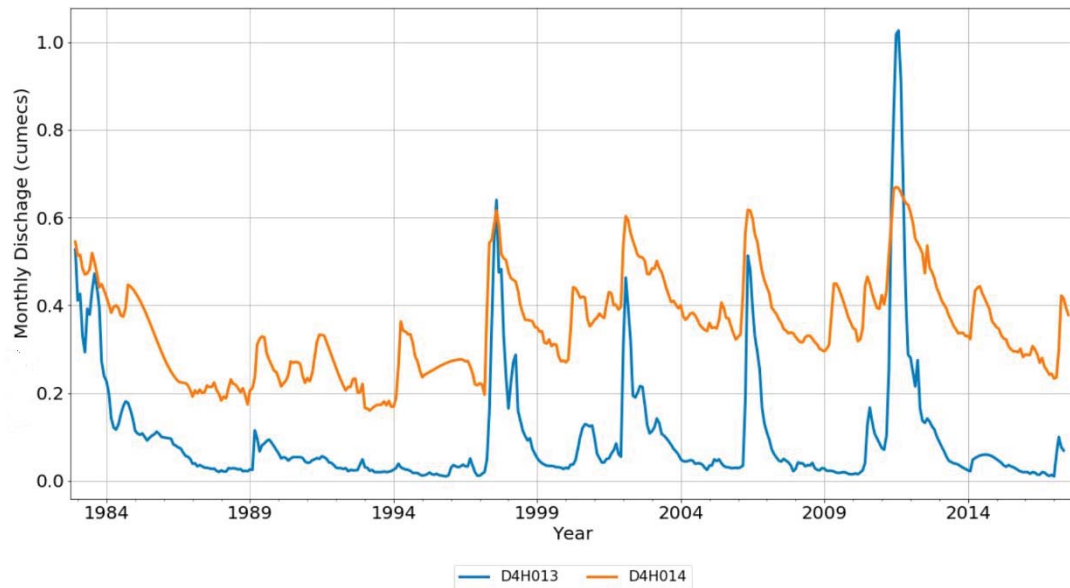
Figure 4-4   **Monthly discharge from two flow gauges on the Molopo River.**

4.1.5   Geology, aquifer systems

Formed around 2.7 billion years ago, dolomites of the Chuniespoort Group have been tectonically deformed and faulted and are intruded by igneous dykes (Tinker et al., 2002). Often weathering to a surface 'elephant skin' texture, the dolomite outcrops are moderately hard and greyish to brown in colour (Cobbing, 2016). The North West dolomites are divided into formations based partly on the chert content which comprise of the Eccles, Lyttleton, Monte Christo and Oaktree Formations, that together comprise the Malmani Subgroup of the Chuniespoort Group, Table 4-2 (Johnson et al., 2006). The area of the Grootfontein compartment is partly covered by Quaternary alluvium and soils with good agricultural potential (Stephens and Bredenkamp, 2002).

Table 4-2   **Major lithology's in the Grootfontein (area Barnard, 2000; CGS, 1991; Johnson et al, 2006).**

Appx. Age (Ma)	Supergroup / Period	Group	Formation		Lithology
0–2.6	Quaternary / Tertiary	Kalahari	Gordonia		Sands, gravels, alluvium, calcrete
65–570	Karoo Supergroup	Intrusives	Karoo dolerite		Intrusive dolerite dykes, sills
		Ecca	Volksrust, Vryheid		Mudstones, shales, sandstones
		Dwyka			Tillite, shale, mudstone, sandstone
2050–2700	Transvaal Supergroup	Pretoria	Magaliesberg, Daspoort, Hekpoort, Timeball Hill		Quartzite, shale, andesite
		Chuniespoort	Penge		Dolomite, chert
			Malmani Subgroup	Frisco	Stromatolitic dolomite, chert-poor dolomite, shale
				Eccles	Interbanded dolomite and chert
				Lyttleton	Chert-poor dolomite, shale, quartzite
				Monte Christo	M1 – Chert-rich dolomite; M2 – Interbanded chert & dolomite; M3 – Oolitic chert & dolomite
				Oaktree	Chert-poor dolomite, carbonaceous shale
			Black Reef		Quartzite, conglomerate
2700–2800	Ventersdorp Supergroup		Andesite, quartz porphyry, conglomerate, calcareous shale, quartzite, lava		
3090+	Basement Complex		Granite, gneiss		

4.1.6   Groundwater levels and flow directions

Groundwater level data was obtained from the DWS databases HYDSTRA and National Groundwater Archive (NGA) for a total of 105 boreholes in the compartment. The groundwater level is measured in meters above ground level (magl) and as such is represented as a negative value. Boreholes that had less than 15 years' worth of data and more than 50% missing data were excluded from analyses resulting in a subset of 28 groundwater level monitoring stations across the compartment for modelling. All 28 stations are either currently active or have monitoring records that extend over a 20-year period. Irrigation is known to have started before the 1960s in the Grootfontein compartment, however records showing water levels before the early 1970's in the Grootfontein area is scarce (Cobbing, 2016). Records at all 28 stations began between 1973 and 1986, during which time irrigation increased. All boreholes in the compartment have a gap in data from late 2005 to early 2008 (see Figure 4-5), and few water level measurements are available after 2009. In order to generate a continuous time-series dataset, which is required for ML modelling, linear interpolation was performed to replace missing data with an estimated value (section 4.2.2.3). Only small data gaps (up to a few months) were filled in this way, and large data gaps (2005-2019) were not used in modelling.
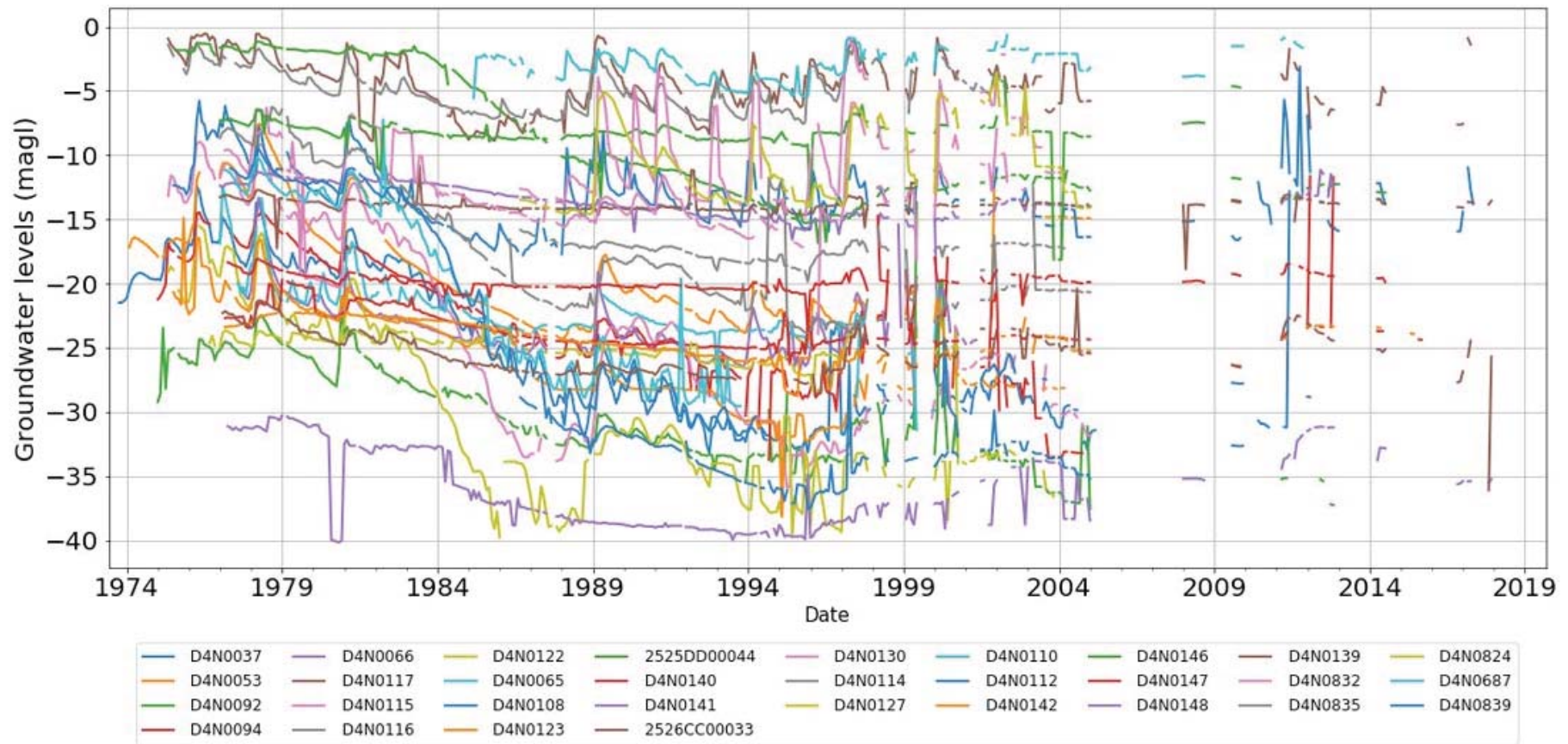
Figure 4-5 Groundwater level in boreholes selected for modelling in Grootfontein.
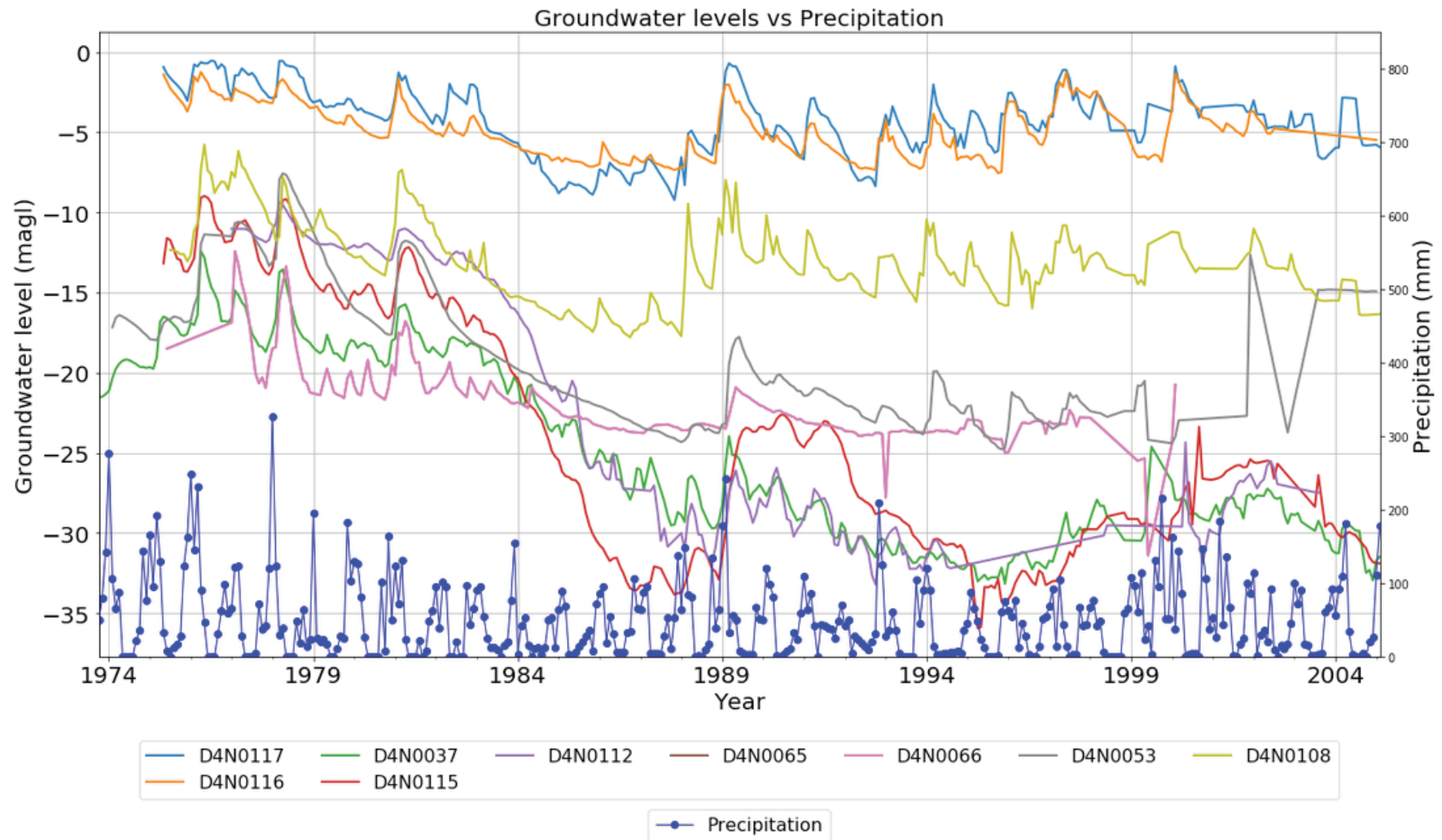
Figure 4-6 Rainfall and groundwater level in the Grootfontein compartment over a common period.

Cobbing (2018) analysed groundwater levels in Grootfontein and reports a long-term decline in groundwater levels over the period of available data. Several of the boreholes used by Cobbing (2018) are included in the 28 selected here, and the same analysis has been performed, showing that 19 of the 28 boreholes selected show a decline over the record, with an average gradient (or rate of decline in meters per month) of 0.005905; equivalent to a fall in water level of 2.2 m per decade. The rate of decline varies by an order of magnitude across the boreholes from 0.0002 for D4N0108 and 0.002 for D4N0115, expected due to the heterogeneity of the karst aquifer and the potential local effects of abstraction or recharge. The decline in groundwater levels across the region has been related to "over-abstraction", and the aquifer is considered to be in a "dewatered state" (Cobbing, 2018).

A subset of these boreholes is shown in Figure 4-6, along with a 10-year moving average rainfall. This shows a general declining trend in rainfall particularly evident from 1977 to 1989, and across the entire dataset (shown in Table 4-3). The declining trend in rainfall mirrors the trend in groundwater level closely; both datasets show a significant decline from 1977 to 1989 with some levelling off between 1989 and 1999. Whilst correlation alone doesn't signify causation, this pattern has not received attention in the literature that attributes the declining groundwater level trend to abstraction alone. Groundwater levels will decline on the onset of abstraction until a new dynamic equilibrium is reached (section 2.1), which will impact natural discharge rates, however it appears likely that declining rainfall and hence recharge has also driven the long-term decline in groundwater levels.

Figure 4-7   Location of boreholes used in modelling.

Table 4-3   Analysis of linear trend in groundwater level for boreholes Grootfontein, along with rainfall.

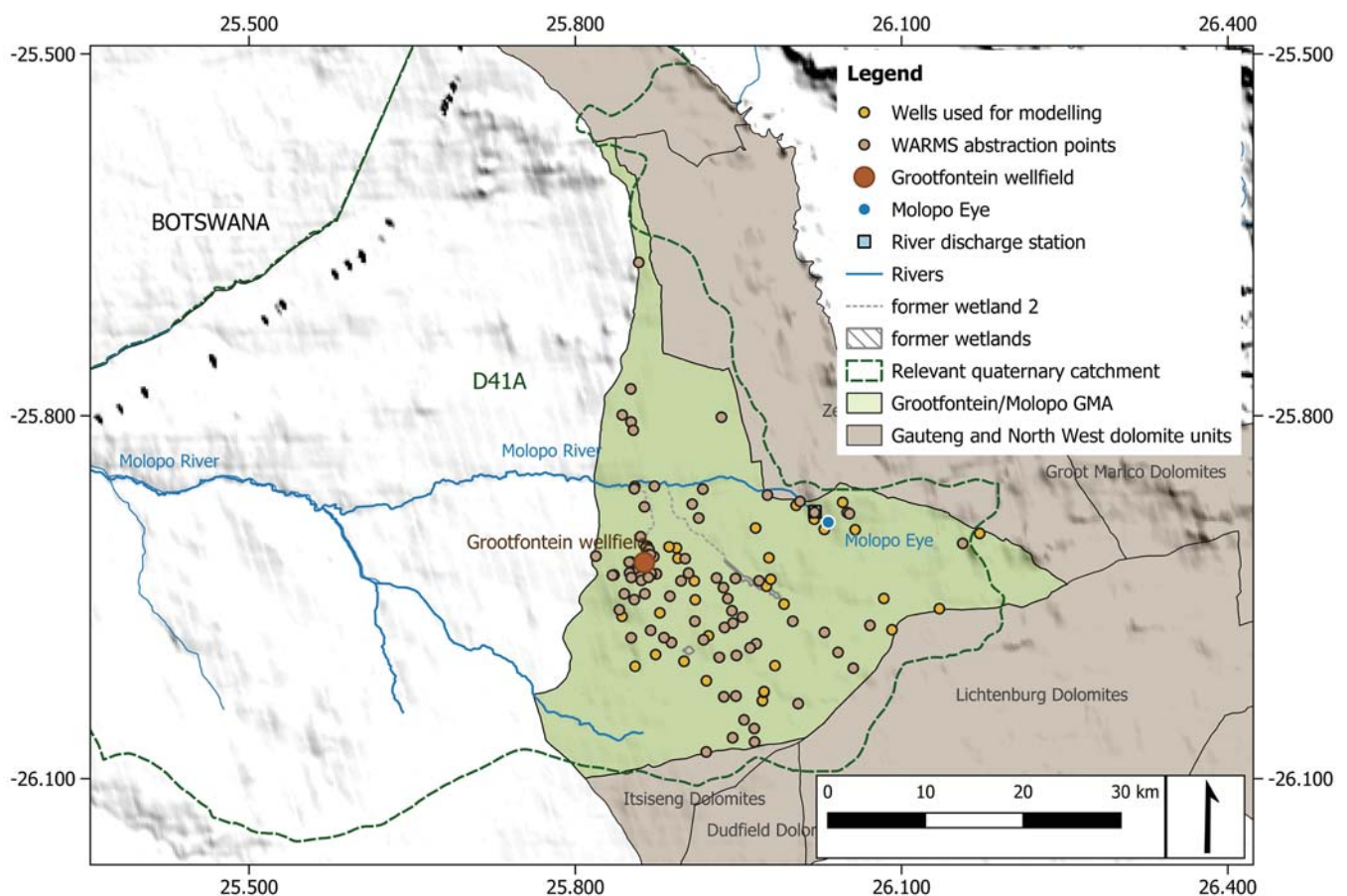| Borehole	Latitude	Longitude	Start date	Start water level	End date	End water level	Years Active	Trend	Trend Gradient	Annual drop
D4N0037	-25.98200	25.92253	1973/08/01	20.62	2012/02/01	28.85	39	Decline	0.00120	0.211
D4N0053	-25.90942	25.89306	1974/02/01	17.18	2005/01/01	14.92	31	Incline	0.00070	-0.073
D4N0065	-26.00306	25.90000	1975/06/01	18.50	2000/02/01	20.76	25	Decline	0.00070	0.090
D4N0066	-26.01917	25.92056	1975/03/01	23.62	2002/04/01	24.50	27	Decline	0.00040	0.033
D4N0092	-26.03570	25.97200	1975/01/01	29.22	2013/04/01	36.67	38	Decline	0.00090	0.196
D4N0094	-26.03556	25.97222	1975/01/01	21.24	2004/12/01	33.21	29	Decline	0.00130	0.413
D4N0108	-25.90844	25.88625	1975/07/01	12.35	2019/06/01	14.42	44	Decline	0.00020	0.047
D4N0110	-25.95222	25.91028	1976/11/01	11.58	2000/01/01	29.90	24	Decline	0.00320	0.764
D4N0112	-25.93667	25.90944	1977/01/01	11.02	2003/08/01	27.48	26	Decline	0.00250	0.633
D4N0115	-25.91797	25.89435	1975/05/01	13.18	2005/03/01	31.89	30	Decline	0.00200	0.624
D4N0116	-26.00722	25.85500	1975/05/01	1.41	2005/01/01	5.08	30	Decline	0.00010	0.122
D4N0117	-25.96600	25.84297	1975/05/01	0.93	2019/04/01	6.27	44	Decline	0.00020	0.121
D4N0123	-26.02806	25.97361	1975/07/01	20.68	2005/01/01	28.14	30	Decline	0.00100	0.249
D4N0126	-25.96422	25.98111	1976/08/01	17.50	2018/04/01	33.70	42	Decline	0.00120	0.386
D4N0127	-25.87417	26.00306	1976/09/01	24.64	2005/01/01	25.41	29	Decline	0.00030	0.027
D4N0130	-25.91750	25.97806	1976/04/01	6.30	2002/12/01	12.70	26	Decline	0.00020	0.246
D4N0139	-25.95114	26.08372	1977/02/01	22.31	2019/02/01	26.87	42	Decline	0.00008	0.109
D4N0140	-25.97700	26.09120	1976/01/01	22.95	2017/09/01	24.46	41	Decline	0.00020	0.037
D4N0141	-25.95958	26.13483	1976/02/01	10.00	2014/05/01	13.46	38	Decline	0.00020	0.091
D4N0142	-25.89728	26.17186	1977/01/01	23.69	2017/08/01	24.38	40	Decline	0.00020	0.017
D4N0146	-25.88553	26.01967	1977/01/01	7.31	2011/11/01	20.00	34	Decline	0.00005	0.373
D4N0147	-25.89425	26.05733	1977/01/01	18.68	2015/06/01	10.93	38	Incline	0.00001	-0.204
D4N0687	-25.89281	25.96578	1985/03/01	5.55	2012/01/01	1.53	27	Incline	0.00010	-0.149
D4N0824	-25.93167	25.83556	1983/11/01	13.49	2005/01/01	14.10	22	Decline	0.00050	0.028
D4N0832	-25.93528	25.97972	1983/10/01	12.62	2004/03/01	11.30	21	Incline	0.00070	-0.063
D4N0835	-25.95583	25.99194	1983/10/01	15.83	2005/01/01	17.51	22	Decline	0.00009	0.076
2526CC00033	-25.89400	26.02903	1976/03/01	12.00	2019/09/01	14.04	43	Decline	0.00001	0.047
2625BB00028	-26.02881	25.97351	1975/07/01	20.68	2004/03/01	28.14	29	Decline	0.00100	0.257
Mean Precipitation	-	-	1950/01/01	95.90	2019/01/01	95.20	69	Decline	0.00060	-0.010

4.1.7 Recharge

Long-term recharge values of between 4.5% and 10% of annual rainfall have been estimated for Grootfontein (Cogho, 1982; Bredenkamp and van Rensburg, 1983; Dziembowski, 1995, cited in Cobbing 2018). Estimating groundwater recharge is particularly complex in the region due to its episodic nature, and lack of data including chloride, stable isotope, groundwater level, and local rainfall (Cobbing, 2018). Indirect recharge via irrigation return flow is also unknown.

4.1.8 Abstraction

The Grootfontein aquifer supports extensive groundwater abstractions for irrigated agriculture, water supply to towns and settlements and for industrial use (Cobbing, 2016). Significant groundwater abstraction occurs for agricultural irrigation and also for mining and other industrial-based activities. The sum of the registered abstractions per water use sector is shown in Table 4-4 based on an extract of registered abstractions (from the DWS Water Authorisation Registration Management System, WARMS) in November 2016.

Table 4-4   **Summary of registered groundwater use in Grootfontein (WARMS, DWS).**

Water use sector	Number of Registrations	Total Use (m^3/a)
Agriculture: Irrigation	261	21 220 316
Agriculture: watering livestock	11	33 708
Mining	10	38 927
Industry	5	951 614

There are other known inaccuracies with the WARMS dataset, including incomplete entries, inaccurate coordinates and duplicate entries. Furthermore, the registrations may be registered use rather than actual use and may be under-estimates for the purposes of securing a license or over-estimate in an attempt to secure a supply. Moreover, several users may not be registered (i.e., unlicensed abstraction). Duplicates in the dataset were corrected and registrations with missing registered volume amounts were discarded. Following this process, the cumulative registered abstraction within the Molopo/Grootfontein GMA boundary totals 22.2 million m^3/a. This appears in line with Cobbing's (2016) lower estimate of 18.3 million m^3/a for the Grootfontein GMU boundary.

To generate the time series of total use across the compartment shown Figure 4-8, the "registered volume start date" and "registered volume end date" was used in conjunction with the "register status". If the registered status column had a value of "active" then the close date for those boreholes was set to 01/11/2016 as this is when the data was acquired. The earliest registration start date in the compartment was used to replace the blank values that were found in the registered volume start date column.
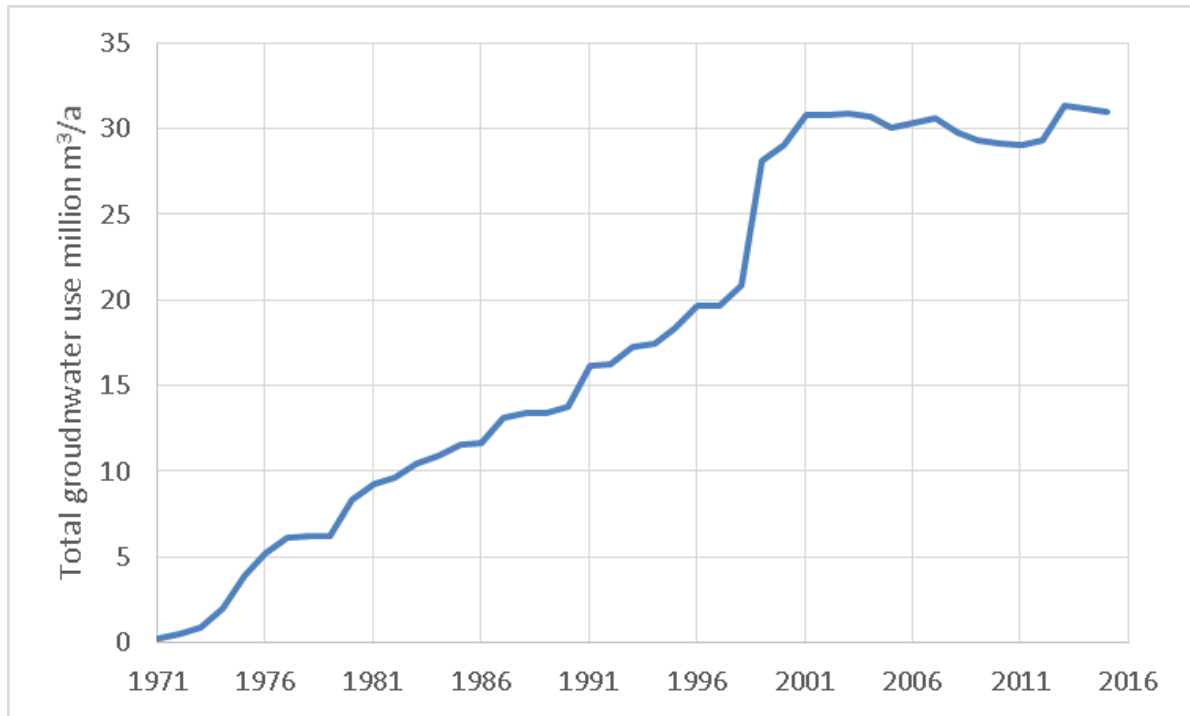
Figure 4-8   **Total groundwater use in the study area over time based on WARMS licence registrations.**

4.2.1   Model software and approach

Tensorflow (Al-Abadi, 2016) in Python (Rossum, 1995) was used to model the LSTM (McKinney, 2010). Pandas, NumPy (Van Der Walt et al., 2011) and Matplotlib (Hunter, 2007) libraries were imported for management, processing and visualisation of data. Figure 4-9 summarises the methodological approach taken to model the LSTM.
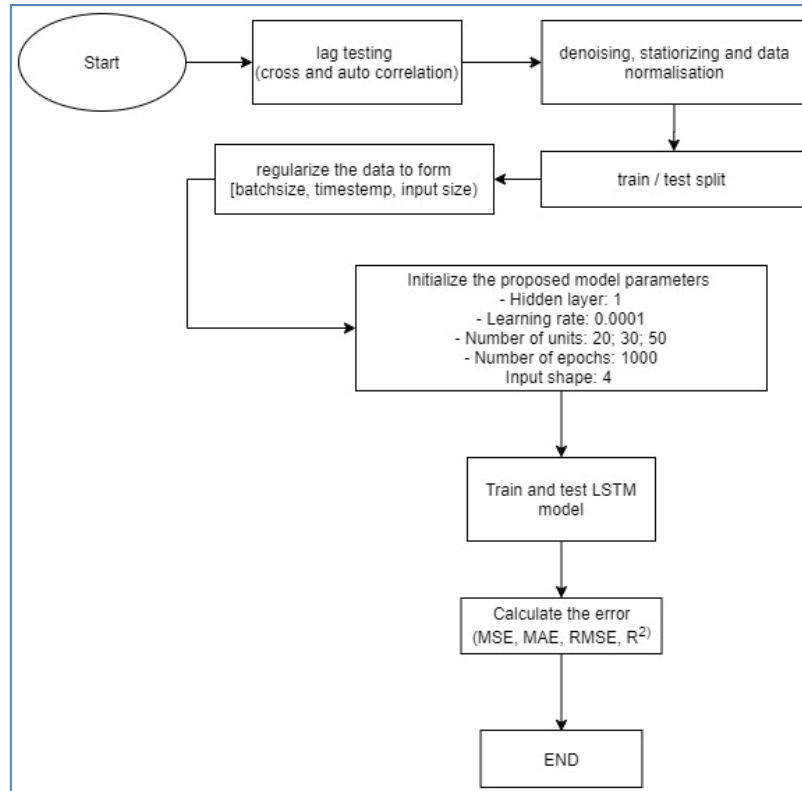
Figure 4-9   Flowchart Summarising LSTM model approach.

4.2.2   Model input variables

This study makes use of the relationship between the target variable (groundwater level), and the input variables to make predictions. Suitable input parameters were chosen based on the expected influence the input variable has on groundwater levels, and on the data available. Groundwater levels in the Grootfrontein compartment are influenced by recharge, and artificial discharge, and changes in these drive groundwater discharge from the compartment. Groundwater level time series from 28 boreholes in the Grootfontein compartment were used as prediction targets. Input variables included discharge rates, temperature, and precipitation and abstraction rates. 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.

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. The trend of the rainfall time series shown in Figure 4-10 effectively represents a smoothed rainfall time series and was used as an input parameter which achieved better results. 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):

$$Y_t = T_t + S_t + R_t$$

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.
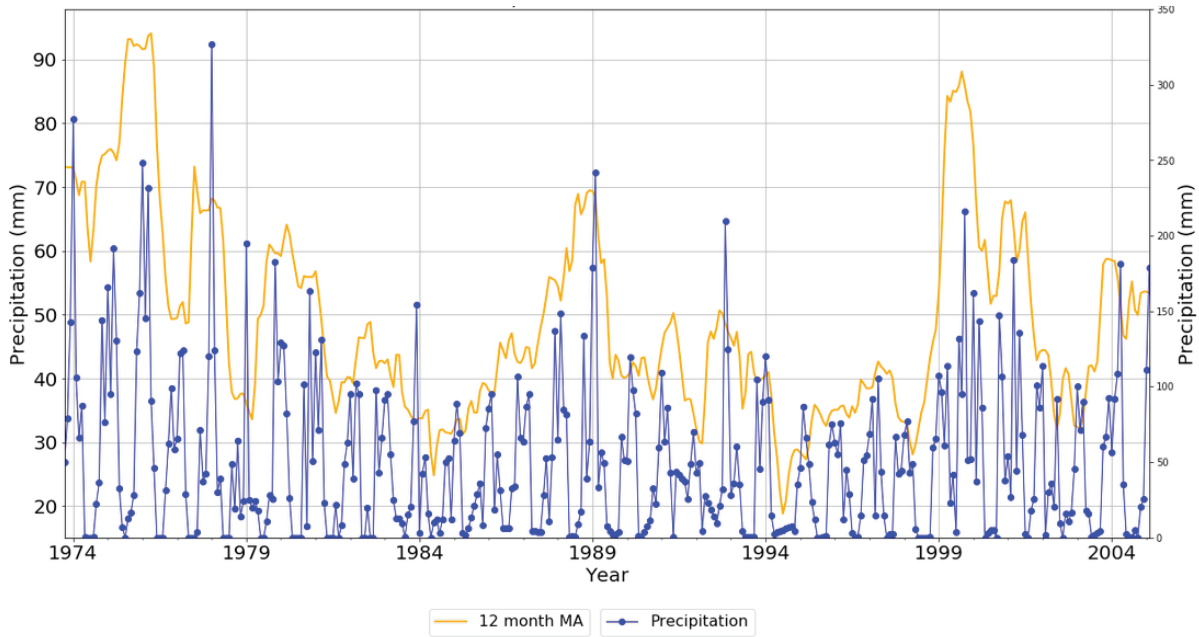
Figure 4-10   **Graphs showing original precipitation data and the de-noised dataset.**

4.2.3   LSTM Model pipeline

4.2.3.1   Lag and hyper-parameters

Lag testing was performed to determine the appropriate lag parameter between rainfall and groundwater levels. Hydrogeologically there is a time lag between the occurrence of rainfall and groundwater level response (Mcphie and Post, 2010), hence accounting for the lag can potentially increase the model's ability to pick up the correlation of rainfall on groundwater levels. The optimum time lag between rainfall and groundwater levels in the Grootfontein compartment was between 2-3 months.

It is necessary to transform the datasets to make them stationary. In the most intuitive sense, stationarity means the statistical properties of a process generating a time series does not change over time. This does not mean that the series does not change over time but instead the way it changes does not itself change overtime. Stationary data is easier to model and will very likely result in better forecasts hence the trend was removed from the features to stationarise then added back to forecasts later to return the prediction to the original scale and calculate a comparable error score. A standard way to remove a trend is by differencing the data. The observation from the previous time step *(t-1)* is subtracted from the current observation *(t)*. This removes the trend leaving a differenced series. The differenced time series is then inverted to its original scale to make forecasts.

The time series were transformed though normalisation according to 4-2. 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). Normalisation applied to the time series after split (Pedregosa *et al.,* 2011):

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

Where Z is the standard score after normalisation, x is the observed value, μ represents the mean of the sample and σ the standard deviation of the sample.

Twenty percent of the data was used as test and the remaining 80% for training. The LSTM layer expects input to be in a matrix with three dimensions: samples (the total of time steps collected or rows of data), time steps (past observation for the feature, i.e., a lag variable) and features (the data columns or number of features). Hence the input layer had to be reformatted into three-dimensional (3D) vectors to match the architecture of the LSTM model, a process referred to as regularising the data. The samples differed per borehole depending on the length of data for that borehole.

For the LSTM model to learn the dependencies more efficiently, there were several modifications related to the arrangement and format of data for the input layer. Instead of using the input data vector as only the data at a specified time step, the vector has been formatted into a sequential data with a selected sequence length of ten time-steps. This may explain the changing trend of the forecasted values which can be seen from the observed values of the ten most recent time steps. Increasing the value of the sequential length can potentially increase model accuracy since the model can capture a better trend. However, this may result in the LSTM model consuming more resources of the computer during training time since the model will have to train a complete neural network. Furthermore, when the sequence length varies from 5 to 15 or more the model demonstrates equally excellent performance (Le et al, 2018).

An important step in constructing the LSTM model is hyper-parameter selection such as learning rate, number of units, number of epochs. The recommended optimization algorithm for this study is Adam (Kingma & Ba, 2015). This algorithm has been widely applied in natural language processing and computer vision. It is an extension on the stochastic gradient descent procedure which updates its weights iteratively based on training data. A learning rate of 0.00001 was chosen instead of the default 0.001. This was done to make the training process slower and the curve of the loss function smoother. The number of units selected was 20. There is no specific reference structure for selecting the number of units in each cell hence different values were tested (20, 30, 50) and evaluated on the forecast results. Lastly, the maximum number of epochs was set to 1000 in order to record the necessary information during training and validation. One epoch in deep learning is defined as when an entire dataset is passed forward and backward through the neural network only once.

To prevent overfitting and to improve the effectiveness of the model early stopping and dropout techniques were employed. Overfitting is when the model learns the noise and detail in the training data to the extent that it negatively impacts its generalisation to new data. Early stopping is used to stop the model when performance on the validation dataset starts to decrease or remains constant. The developed LSTM model was set to stop the training process if the performance on the validation dataset does not improve after 100 epochs.

4.2.3.2 Cross-validation

Cross-validation was performed to improve the generalization error of the model and also prevent overfitting, essentially improving how accurately the model can predict the outcome value for previously unseen data. Time series cross-validation based on the rolling window approach (Hyndman and Athanasopoulos, 2014) was used. In this procedure, the last few years (depending on the dataset length) of the data is put aside as the test set. The rest of the data is used in cross validation to evaluate the predictive power of various models and fine-tune each model's hyper-parameters. This method ensures each training set only consists of observations that occurred prior to any observation in the test set. Therefore, no future information is used to train the models. Figure 4-11 illustrates how the data was partitioned.

The forecast accuracy is then computed by averaging over the test sets. Since the data is sequentially correlated, traditional k-fold cross validation techniques were omitted due to their randomized splits which would cause the time dependencies between the observations to be lost. The model run for each borehole was repeated ten times and the average prediction was taken as the final simulated result. The error metrics were calculated on the averaged values to indicate how well the model is expected to perform on unseen data on average.
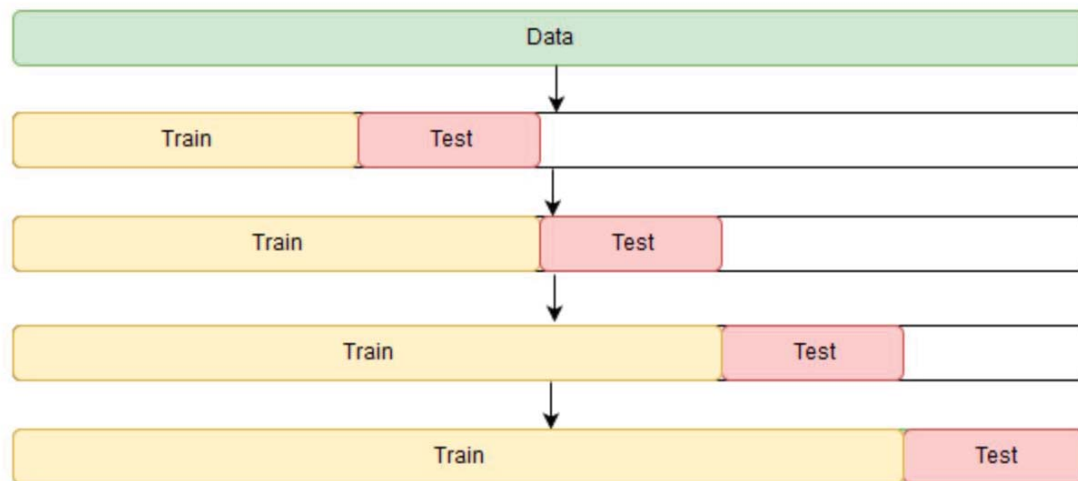
Figure 4-11   Cross- Validation Scheme for Time Series.

4.2.3.3   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 (used by authors such as Wunsch *et al.*, (2018) and Daliakopoulos *et al.*, (2005) respectively). A study by North and Livingstone (2013) compared 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 result of the study suggests that the linear interpolation method interpolates the gaps most accurately.

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 $R^2 = 0.95$).

4.2.3.4   Model evaluation criteria

The performance of a machine learning model is evaluated using a set of metrics and based on the results, improvements are made to the model until a desired accuracy is achieved (Engelbrecht, 2007). The performance of machine learning models is defined by how well the models can map the input space to the output space (i.e., fit the target variable to observed data) and how well the models can generalize to new points that were previously not seen during training (Engelbrecht, 2007).

Mean squared error (MSE) measures the average of errors squared. Error in this case means the difference between the observed values O_i and the predicted values P_i. The difference is squared so that negative and positive values do not cancel each other out.

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

<div align="right">4-3</div>

Root mean squared error (RMSE) measures the average magnitude of the error between observed (o) and predicted (p) values. RMSE is a good measure of error that can serve as a loss function to minimize (Engelbrecht, 2007).

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

<div align="right">4-4</div>

Mean absolute error (MAE) measures the average magnitude of the errors without considering their direction. It calculates the averages of the absolute differences between prediction and actual observation where all individual differences have equal weight.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

<div align="right">4-5</div>

R-squared (R^2) explains how much variability of one factor can be caused by its relationship to another factor (co-linearity between the observed and predicted data). R^2 values range from 0 to 1, with higher values indicating less error variance (i.e., the closer the value is to 1, the better the fit, or relationship, between the two factors) (Krause, 2005).

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

<div align="right">4-6</div>

In addition to the above metrics, the fit between observed and modelled groundwater levels was assessed qualitatively in order to establish that the models that made sense.

4.2.4   Information Gain Ranking

The input variables (x) are not equally informative in predicting the target variable (y), as some variables would be noisier or have less of a relationship and hence correlation with the target data set. 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 (Rossum, 1995) according to equation 4-7 (MacKay, 2005):

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

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

$H\ (x\ |\ y)$ = 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 $(x; y\ ) \geq 0$. 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.

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 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.

4.2.5   Scenario testing

Once the developed models performed adequately, the established model was used to predict groundwater levels under hypothetical scenarios with adjusted rainfall and groundwater abstraction. To simulate groundwater predictions for scenarios one to three (Table 4-5) the models were trained using the original training dataset (80% of the time series). The test dataset for the input parameters were then altered accordingly in alignment with the selected scenario. For example, when groundwater level predictions were generated for scenario 1a (Table 4-5), 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 the rainfall test dataset was generated by subtracting 100mm for any months with > 100 mm of rainfall during that month. The model will then use the input variable altered to mimic the scenario with the other input variables altered, based on what the model has learnt during training to make groundwater level predictions.

The scenarios tested are summarised in Table 4-5 showing the shortened ID, name and description assigned to each scenario tested. Scenario 1 aims to simulate changes in rainfall intensity and overall reduction in the magnitude of rainfall.

Scenario 2 simulated an increase in groundwater abstraction. Scenario 3 simulates both a decrease in rainfall intensity and increase in abstraction.

Scenario 4 tests the ability of the model to make long-term predictions. Scenario 4a aimed to test if the 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. 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 4b 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 average for each month of the year was then repeated until a 30-year time series was created for each input variable.

Table 4-5   Description of the scenarios applied to the developed models for the Steenkoppies aquifer

Scenario ID	Name	Description
1a	Decrease rainfall intensity	If monthly rainfall > 100mm then minus 100mm. No alteration to spring discharge, temperature and groundwater usage input variables.
1b	Increase rainfall intensity	If monthly rainfall > 100mm then add 200mm. No alteration to spring discharge, temperature and groundwater usage input variables.
1c	Decrease rainfall	Half the monthly rainfall. 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.
3	Simulate a worst-case scenario	Combination of both scenario 1a and 2. No alteration to spring discharge and temperature variables.
4a	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.
4b	Long term prediction 30 years	100% data for training, predict groundwater levels for 30 years using long term averages of input variables.

4.3   MODEL RESULTS

4.3.1   Information gain ranking

The contribution of each feature against the target variables (groundwater levels), following the method in section 4.2.3, is shown in Table 4-6. The first column indicates the borehole ID (BHID), and the remaining 5 columns indicate the feature names and the corresponding entropy scores to the target variable (groundwater level). Table 4-6 shows that total abstraction rate correlates with water levels the strongest in each borehole (entropy ranges from 0.48 to 0.69). Again, correlation does not imply causality, and this essentially shows that the method detects a correlation between the curve of increasing groundwater use over time with the groundwater levels at each borehole. The result is surprising as the groundwater levels are significantly noisier than the groundwater use.

Table 4-6   A ranking by entropy of each input feature against each target variable

	BHID	Temperature	Rainfall trend	Actual rainfall	River Discharge	Groundwater Abstraction
0	D4N0037	0.01	0.36	0.05	0.45	0.61
1	D4N0053	0.02	0.35	0.02	0.40	0.56
2	D4N0065	0.01	0.29	0.05	0.38	0.57
3	D4N0066	0.01	0.32	0.01	0.39	0.68
4	D4N0092	0.04	0.39	0.02	0.43	0.53
5	D4N0094	0.02	0.38	0.01	0.49	0.55
6	D4N0108	0.05	0.33	0.04	0.37	0.58
7	D4N0110	0.03	0.39	0.03	0.43	0.58
8	D4N0112	0.02	0.37	0.05	0.38	0.57
9	D4N0115	0.01	0.36	0.05	0.47	0.69
10	D4N0116	0.04	0.33	0.03	0.32	0.48
11	D4N0117	0.03	0.37	0.04	0.36	0.60
12	D4N0123	0.03	0.37	0.05	0.38	0.67
13	D4N0126	0.02	0.39	0.04	0.48	0.62
14	D4N0127	0.01	0.32	0.05	0.35	0.69
15	D4N0130	0.02	0.38	0.02	0.45	0.65
16	D4N0139	0.02	0.32	0.01	0.49	0.61
17	D4N0140	0.02	0.32	0.03	0.35	0.66
18	D4N0141	0.04	0.37	0.06	0.32	0.65
19	D4N0142	0.03	0.36	0.06	0.32	0.61
20	D4N0146	0.00	0.40	0.06	0.45	0.60
21	D4N0147	0.05	0.35	0.02	0.46	0.62
22	D4N0687	0.06	0.37	0.06	0.52	0.60
23	D4N0824	0.06	0.35	0.05	0.45	0.59
24	D4N0832	0.06	0.29	0.03	0.42	0.54
25	D4N0835	0.02	0.36	0.03	0.46	0.63
26	2526CC00033	0.05	0.44	0.03	0.49	0.64
27	2625BB00028	0.04	0.34	0.03	0.41	0.61

The second most correlating feature with groundwater levels is discharge with an entropy ranging from 0.32 to 0.52. A strong relationship is expected, given that the hydraulic gradient between the surrounding aquifer and the eye is what drives discharge hence changes in groundwater level will directly translate to changes in discharge rate. The rainfall trend (described in section 4.2) is the feature with 3rd strongest correlation (0.29 to 0.45). The actual observed rainfall exhibits very low entropy scores (ranging from 0.01 to 0.06) due to the data being noisy when compared to the trend rainfall. Xu et al. (2020) looked at the decomposed signals of rainfall as inputs for forecasting rainfall and found that the decomposed signals of rainfall produced better predictions compared to actual measured rainfall. A similar observation is seen in this study. As would be expected, temperature has the lowest correlation (and therefore least influence in predicting water levels) with a maximum entropy of 0.06.

It should be noted that MI only measures the mutual dependence between two random variables by identifying how much information of one of the features can be obtained from the other features, but it does not tell us what drives the groundwater levels.
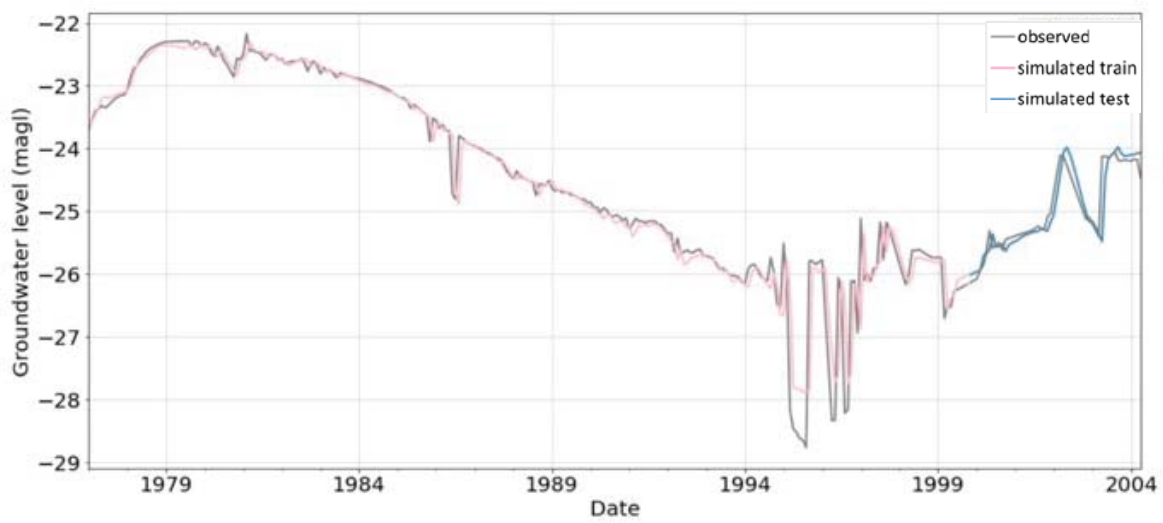
4.3.2   Model training and testing

The model results for the 28 boreholes are summarised in Table 4-7. While the model reports MAE values of less than 0.40m for most of the boreholes, it reports MAE values of around 0.4m and 0.5m for six boreholes, and for borehole D4N0147, the model recorded the highest MAE value of 0.672m. Most of the R^2 results between the predicted and the observed at all the borehole sites are above 0.5. Only five boreholes recorded R^2 values of less than 0.5 with the lowest score from borehole D4N0140 which had an R^2 of 0.341. The high MAE, MSE and RMSE scores of the predicted outcome at boreholes D4N0066, D4N0127, D4N0146, D4N0147, D4N0687 and D4N0824 demonstrate that the accuracy of the model outputs at these boreholes is questionable. Since the same structure of the LSTM model was used on all the boreholes the high errors in these boreholes might be caused by inaccurate input data or by other factors that might have a strong influence on the groundwater level fluctuations at these particular sites. A subset of the boreholes shown in Table 4-7 was selected to show a comparison of the predicted results versus the observed results (see Figure 4-12).

Table 4-7   LSTM model result metrics per borehole

|	BHID	R-Squared	MSE	RMSE	MAE	STDV
0	D4N0037	0.73	0.02	0.12	0.10	0.02
1	D4N0053	0.55	0.03	0.17	0.09	0.01
2	D4N0065	0.58	0.03	0.16	0.13	0.02
3	D4N0066	0.58	0.62	0.79	0.54	0.02
4	D4N0092	0.65	0.09	0.30	0.23	0.03
5	D4N0094	0.65	0.25	0.50	0.35	0.03
6	D4N0108	0.80	0.05	0.23	0.16	0.02
7	D4N0110	0.64	0.01	0.12	0.10	0.01
8	D4N0112	0.27	0.01	0.11	0.08	0.02
9	D4N0115	0.82	0.02	0.12	0.08	0.02
10	D4N0116	0.85	0.08	0.28	0.20	0.02
11	D4N0117	0.62	0.15	0.39	0.23	0.03
12	D4N0123	0.74	0.12	0.35	0.21	0.03
13	D4N0126	0.63	0.03	0.16	0.10	0.02
14	D4N0127	0.73	0.46	0.68	0.47	0.03
15	D4N0130	0.60	0.19	0.44	0.27	0.04
16	D4N0139	0.68	0.12	0.35	0.23	0.02
17	D4N0140	-0.34	0.04	0.19	0.18	0.03
18	D4N0141	0.82	0.05	0.22	0.18	0.01
19	D4N0142	0.81	0.03	0.18	0.13	0.01
20	D4N0146	0.64	0.57	0.75	0.47	0.03
21	D4N0147	0.42	0.63	0.79	0.67	0.03
22	D4N0687	0.79	0.58	0.76	0.48	0.03
23	D4N0824	0.52	0.37	0.60	0.54	0.06
24	D4N0832	0.86	0.10	0.31	0.26	0.02
25	D4N0835	0.32	0.08	0.23	0.21	0.10
26	2526CC00033	0.29	0.20	0.45	0.22	0.03
27	2625BB00028	0.80	0.10	0.31	0.18	0.02
	Average	0.61	0.18	0.36	0.25	0.03

D4N0142

D4N0037
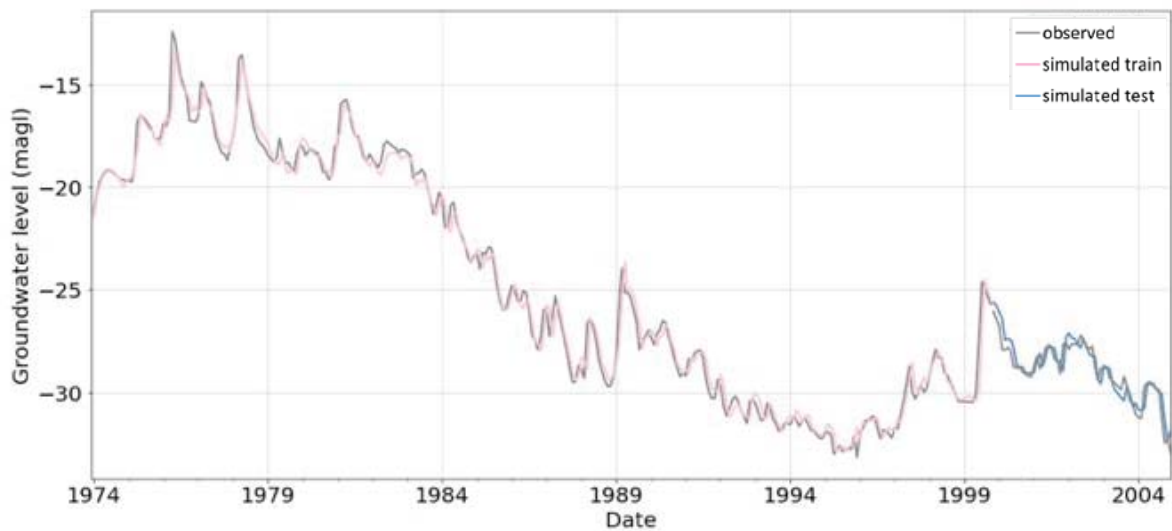
D4N0110
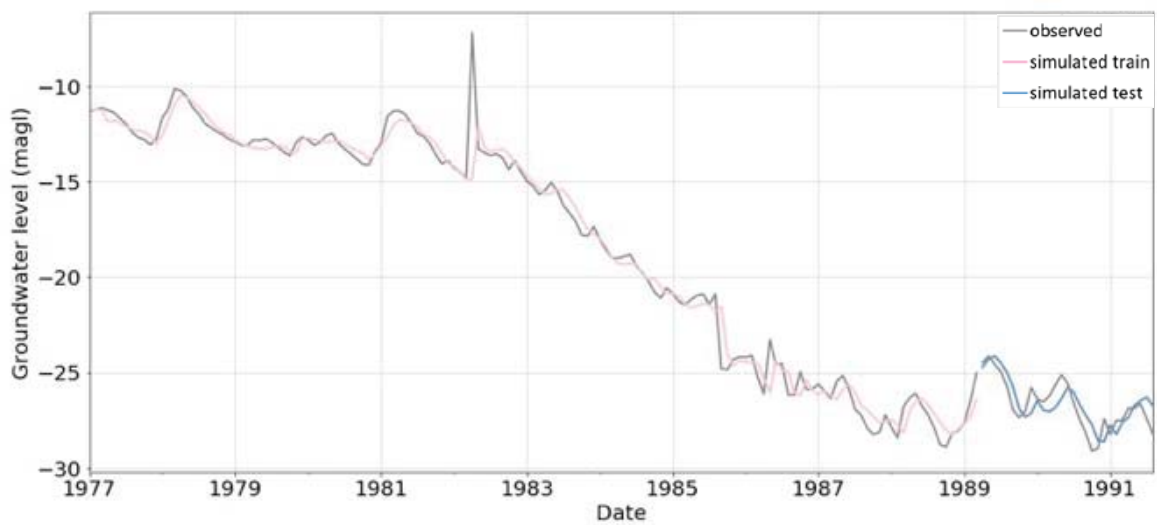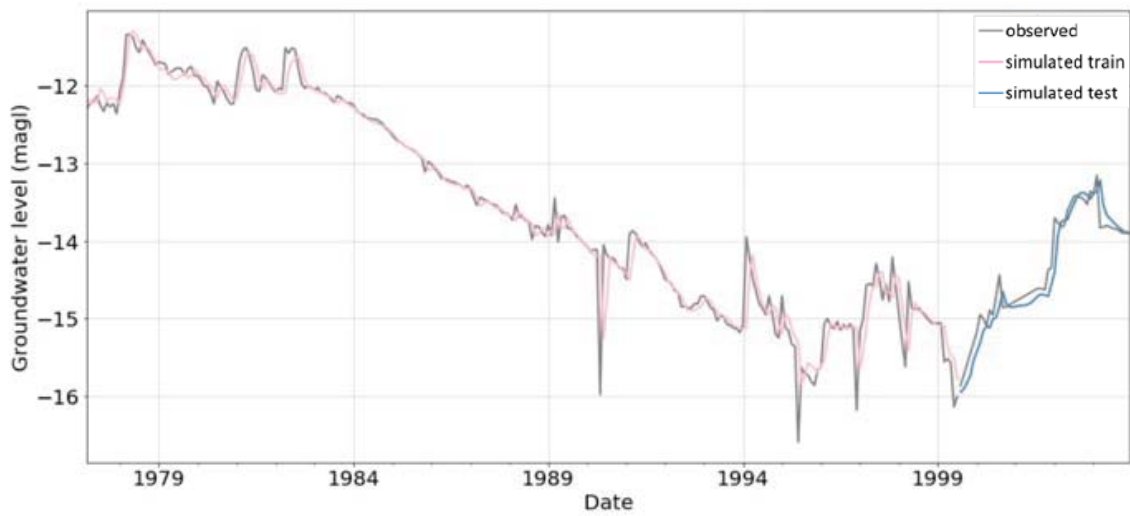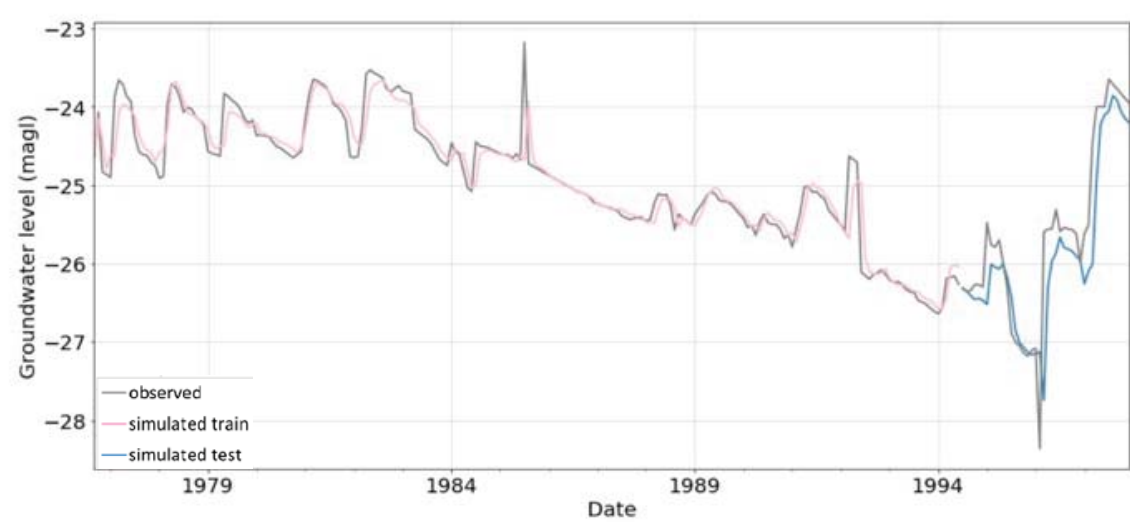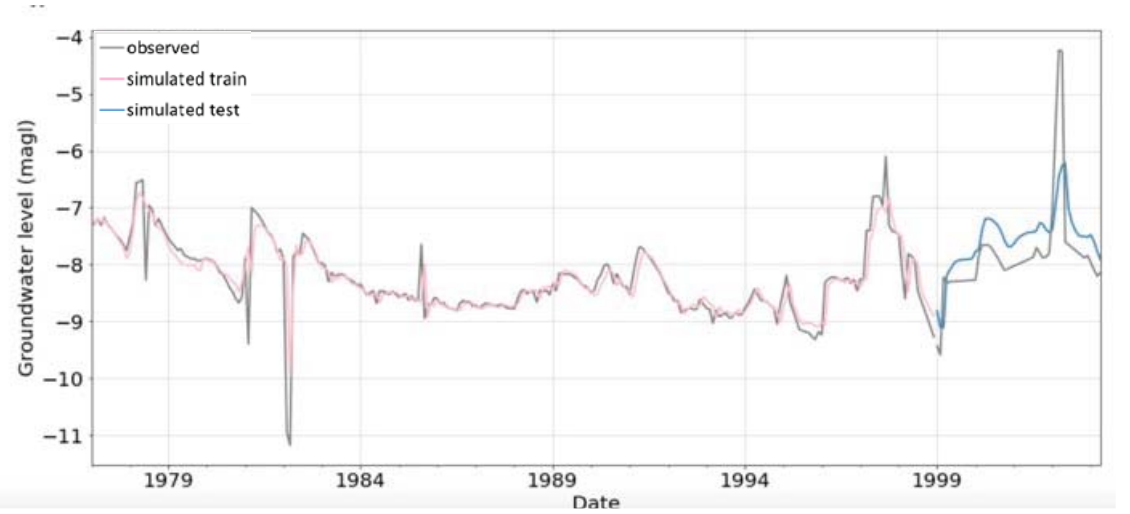
Figure 4-12   Comparison between the observed and predicted for 6 boreholes during training and testing phases.

Generally speaking, the observed and modelled groundwater levels are closely matched, with key features of the observed datasets replicated in the modelled data. As would be expected, for the training dataset the modelled groundwater level matches observed much more closely than the test, because the observed groundwater levels are used in training, i.e., groundwater level at the previous time steps largely dictates the predicted output. During test, the model only "sees" the modelled groundwater level rather than observed.

Figure 4-12 suggests a good correlation between the predicted and observed values for boreholes D4N0141 and D4N0142 which also exhibited high R^2 values (0.823, 0.821 respectively) and low RMSE values (0.224, 0.183 respectively). Boreholes D4N0037 and D4N0127 performed reasonably well. For borehole D4N0127, the model seemed to under-predict values in the training set as compared to the test set. On the other hand, borehole D4N0146 seemed to over-predict groundwater level values which also explains why the borehole exhibited high error metrics (MSE: 0.565, RMSE: 0.751 and MAE:0.456). Borehole D4N0110 seemed to under and over-predict groundwater level values throughout the test phase.

Figure 4-12 also illustrates how outliers affect model performance. The model learns patterns from data during the training phase hence patterns which deviate from patterns previously seen decrease the overall performance of the model. This can be seen with boreholes D4N0146 and D4N0127 during the test phase. During the testing phase for borehole D4N0146, a high peak occurred during 2003 that was previously not seen in the training set hence the model was unable to accurately predict the peak which resulted in decreased model performance during the testing phase. RNNs generally perform better the more variability it sees during the training phase. This is evidenced by the three troughs seen in D4N0146. The model under-predicts the first two troughs but by the time it reaches the third trough the model is able to better predict the trough. In addition, the model is unable to pick up the variability seen in the test set for borehole D4N0127, because the training data is less representative of the test data (variability happening in test set is different from train set). Due to this high variability model errors also increase significantly with RMSE of 0.75 and MAE of 0.45 and the STDV is also significant (0.032 and 0.028 respectively) showing there is high variability in the test data.

Boreholes with less variability perform sufficiently well evidenced by the low RMSE values for boreholes D4N0037 (0.124) and D4N0110 (0.119). Since the model was better able to learn and generalise with high variability data in the training phase for boreholes D4N0141 and D4N0142 the model was able to better predict the test sets which also why the STDV values for the two boreholes are quite low (0.014, 0.012 respectively).

Key observations from the results include:
- What is common across most boreholes for both the training and testing phase is the LSTM model is often unable to capture the minimum and extreme groundwater level peaks whenever they occur in a short timeframe. This is illustrated in Figure 4-12 where across all boreholes the model failed to reproduce the extreme minimums and maximum values of each borehole. For example, the minimum groundwater levels in 1990, 1995 and 1997 in D4N0141 and 1981 and 1982 in borehole D4N0146 are not captured. Similarly, extremes and short time scale fluctuations of groundwater levels are not well replicated as seen in 1982 for borehole D4N0110, 2002 for borehole D4N0146 and 1995 for borehole D4N0127.
- In all cases the predicted groundwater level in test data is lagged behind the observed groundwater level. This is common in RNN model results and is related to the fact that the model is "looking" at the model output groundwater levels for the previous X number of time steps (X is user defined). The model combines this information (perhaps a declining trend) with new input data at the current time step. If the new data at the current time-step suggests that the decline is ceasing, and a peak of rainfall &/ discharge is approaching, the model has to make a decision on the modelled groundwater level based on this and the previous modelled groundwater level. It appears that the "looking backwards" at modelled groundwater level causes the modelled result to lag behind the influence of new parameters.

- The model fit statistics improve with the number of samples or length of dataset (i.e., R^2 increases with number of samples).
- The observed groundwater level has a dominant effect on the modelled groundwater level in training, demonstrated by the measurement recording error in the observed groundwater level for 2525DD00054, where data is shifted down by ~11m between 1981 and 1984, as well as in 2526CC00255 in 1986, and data appears shifted downwards in D4N0130 in 1981 – 1982. This data was not removed or corrected prior to modelling. The model is able to reproduce this data shift which because it is a measurement error is not represented in the discharge or rainfall input variables (and therefore a physically based model would not be able to reproduce).
- Other variables (i.e., discharge and rainfall) do have some influence on modelled groundwater level, more so in the test dataset, as demonstrated by the fact that the modelled groundwater levels are able to fluctuate in the test dataset. This demonstrates that the modelled data is sufficiently influenced by the input parameters.

4.3.3 Scenario testing

Scenario testing (as per section 4.2.4) was performed for different time periods (ranging from 2 to 5 years) for the subset of boreholes mentioned in section 4.3.2.

Scenario 1a – Precipitation (-100mm)

Subtracting 100mm from rainfall events that had peaks higher than 100 resulted in a drop of groundwater levels at the specific month were precipitation was reduced. Only a few precipitation peaks above the 100mm threshold occurred during 2000 – 2004 (duration of test data for most boreholes). For this reason, the drop in groundwater levels is only seen at the particular months where rainfall was reduced. For boreholes D4N0142 and D4N0142, the decrease in water levels between year 0 and year 2 is clearly observed as rainfall events were reduced during this period. For borehole D4N0037, a decline in water levels is observed during year 2 to 3 as compared to Figure 4-12 where there is an over-predicted value of groundwater levels in 2002. Borehole D4N0127 had only a single rainfall event that was above the threshold during the testing phase hence only one negative change was observed with a drop of 0.2m at the beginning of year 3. The general trend for the rest of the years remained similar to that shown in Figure 4-12. For boreholes D4N0110 and D4N0146, a decline in water levels (0.2 to roughly 0.4m) can be observed from year 0.

Scenario 1b – Precipitation (+200mm)

Similar to scenario 1a boreholes in this scenario showed an increase in groundwater levels in particular months where the rainfall event meets the required threshold for increase. The increase in precipitation clearly generates an increase in groundwater levels, seen more clearly in boreholes D4N0037, D4N0110 and D4N0146. The magnitude of the response in these boreholes is significantly higher than boreholes D4N0142 and D4N0141. For borehole D4N0127 the groundwater levels increase for year 3 yet remained almost stable for the remainder. Generally, in this scenario the groundwater levels increased by 0.6 to 1.3m.

Scenario 1c – Precipitation (half)

By halving all precipitation values in the test set, we can see a clear decline in groundwater levels for all boreholes presented in Figure 4-15. The decline in groundwater levels is consistent throughout the years since all precipitation values were halved unlike in scenario 1a (see Figure 4-13) where only specific rainfall events were tweaked hence the water levels only declined at particular months. The LSTM model is clearly able to predict the influence of precipitation on groundwater levels in the compartment, evidenced by the constant drop in water levels for all boreholes shown in Figure 4-15.

Scenario 2 – Abstraction doubled

Doubling the regional abstraction rates in the compartment resulted in a significant decrease of groundwater levels (ranging between 0.5 to 1.7m) in the compartment. As can be seen in Figure 4-16, all boreholes had a significant decline in water levels. Although the doubled abstraction rates seemed to have a higher effect on boreholes D4N0110, D4N0141 and D4N0142, the magnitude for boreholes D4N0127, D4N0037 and D4N0146 was much smaller.

Scenario 3 - Abstraction doubled and precipitation reduced (-100mm)

The highest change in groundwater levels was observed from the third scenario where regional abstraction was doubled and rainfall was reduced by 100mm at high precipitation events. All selected boreholes show a decline in groundwater levels as indicated in Figure 4-17. Groundwater levels for boreholes D4N0141 and D4N0142 dropped by almost 1m as observed in Figure 4-17. For boreholes D4N0037, D4N0127, D4N0110 and D4N0146 groundwater levels dropped by at least 0.5m.

Scenario 4 – long term prediction

For scenario (4a), the training sample size was reduced to see how the model would perform in a long-term prediction, whilst keeping all input variables the same. The performance is clearly worse than when more data used in training (compare Figure 4-18 and Figure 4-12 for the same borehole). Generally, with RNN models specifically the LSTM model, the more data you feed the model the better the model predictions. In scenario 4a the sample size was split in half giving the model less data to train and in scenario 4b the model was given full dataset to train on, in order to predict a longer-term forecast. As can be seen when comparing the two predictions, the longer training data (30 years) produced a much better training result. In scenario 4a, the magnitude of the observed peaks and troughs are not well matched during training, and during the test period the predicted groundwater level tracks well below the observed in sections of the time series (particularly 1994 to 1999). However, the model was still able to generalise sufficiently, as post 1999 the predicted groundwater level comes closer to observed (rather than becoming progressively worse with an error accumulating), and an RMSE value of 1.075 is recorded. The input variables of precipitation and discharge appear to sufficiently influence the model prediction. The results for scenario 4a illustrate that the type of input data can at times have more influence than the quantity of input data since it was still able to generalize well with a shorter training sample.

For the longer-term forecast (scenario 4b), precipitation, abstraction, temperature and discharge data were formulated by averaging the 30 years of data for each variable and predicting 15 years into the future. The results of scenario 4b (Figure 4-18) simply demonstrate that hypothetically it is possible to predict any duration into the future, with accuracy related to the length of training data.
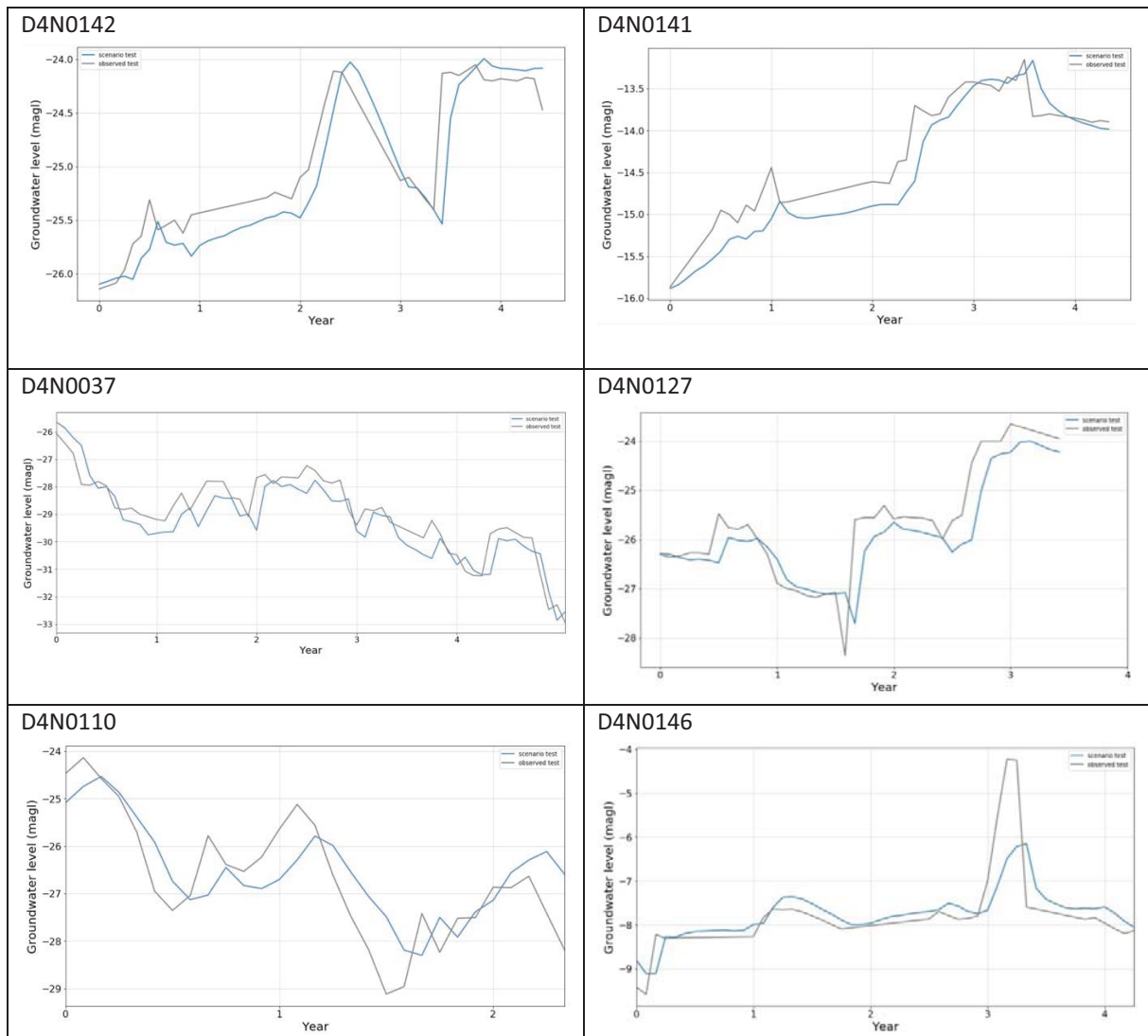
Figure 4-13 Model prediction for scenario 1a, decrease rainfall peaks.

Figure 4-14   Model prediction for scenario 1b, increase rainfall peaks.
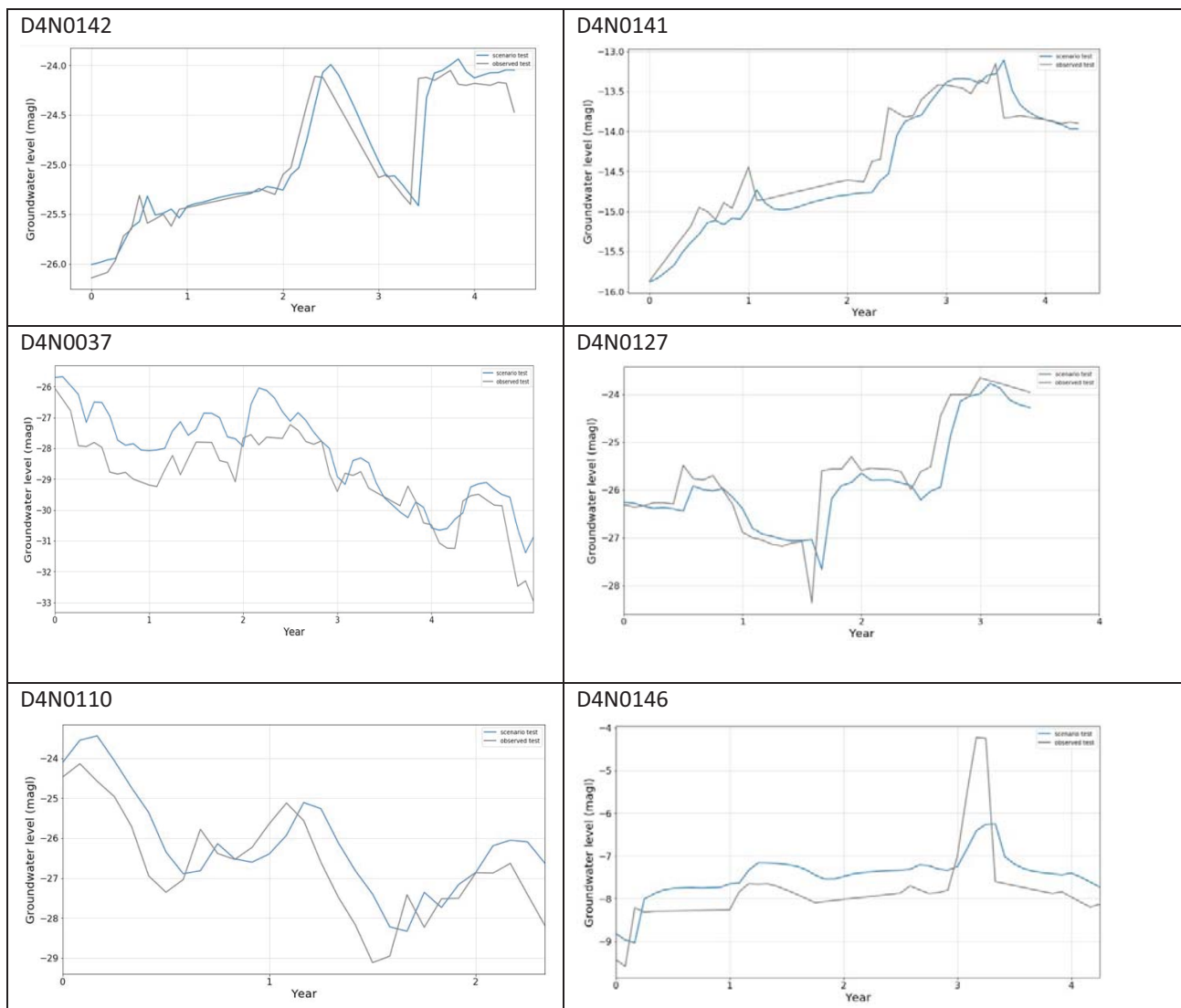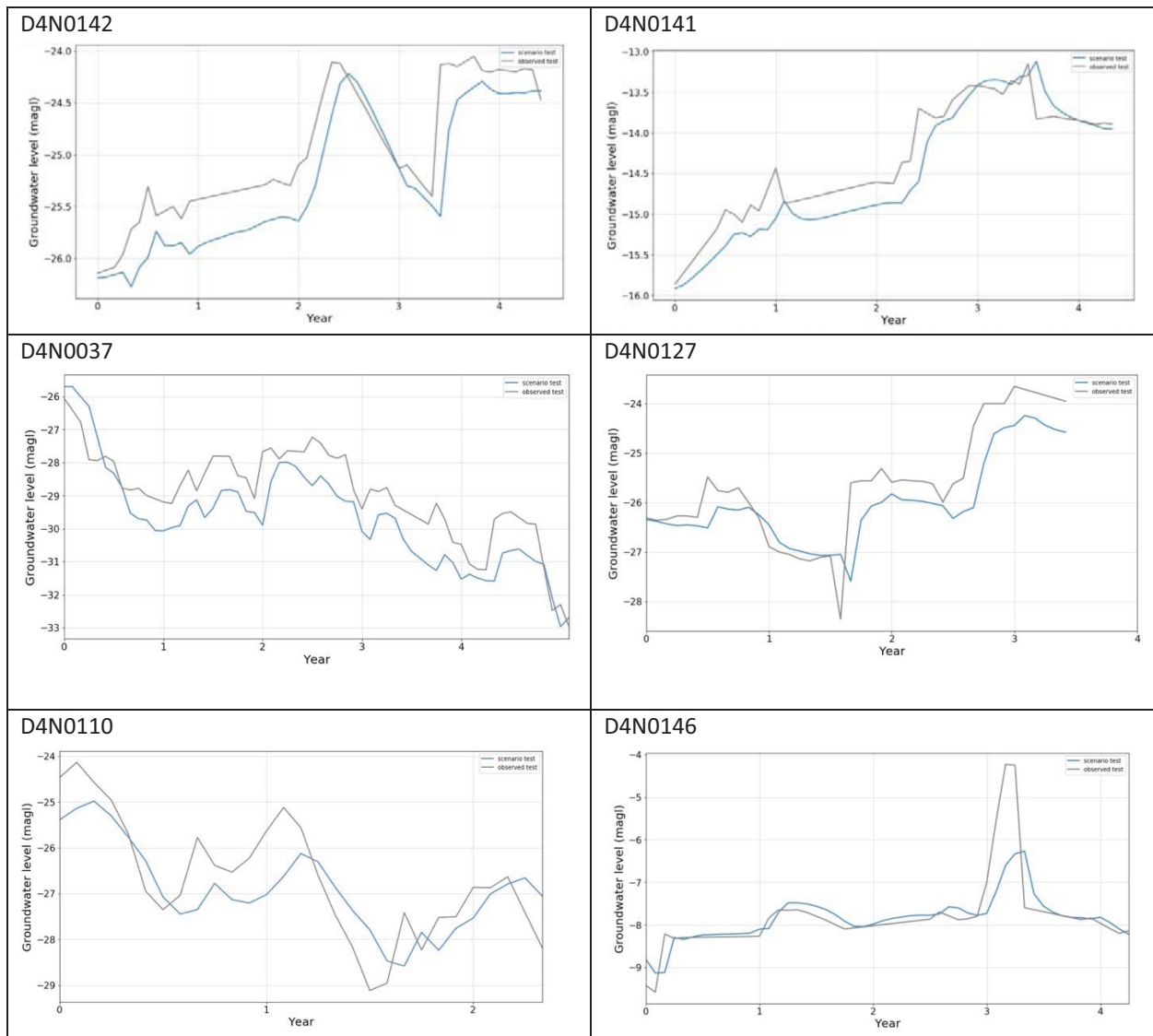
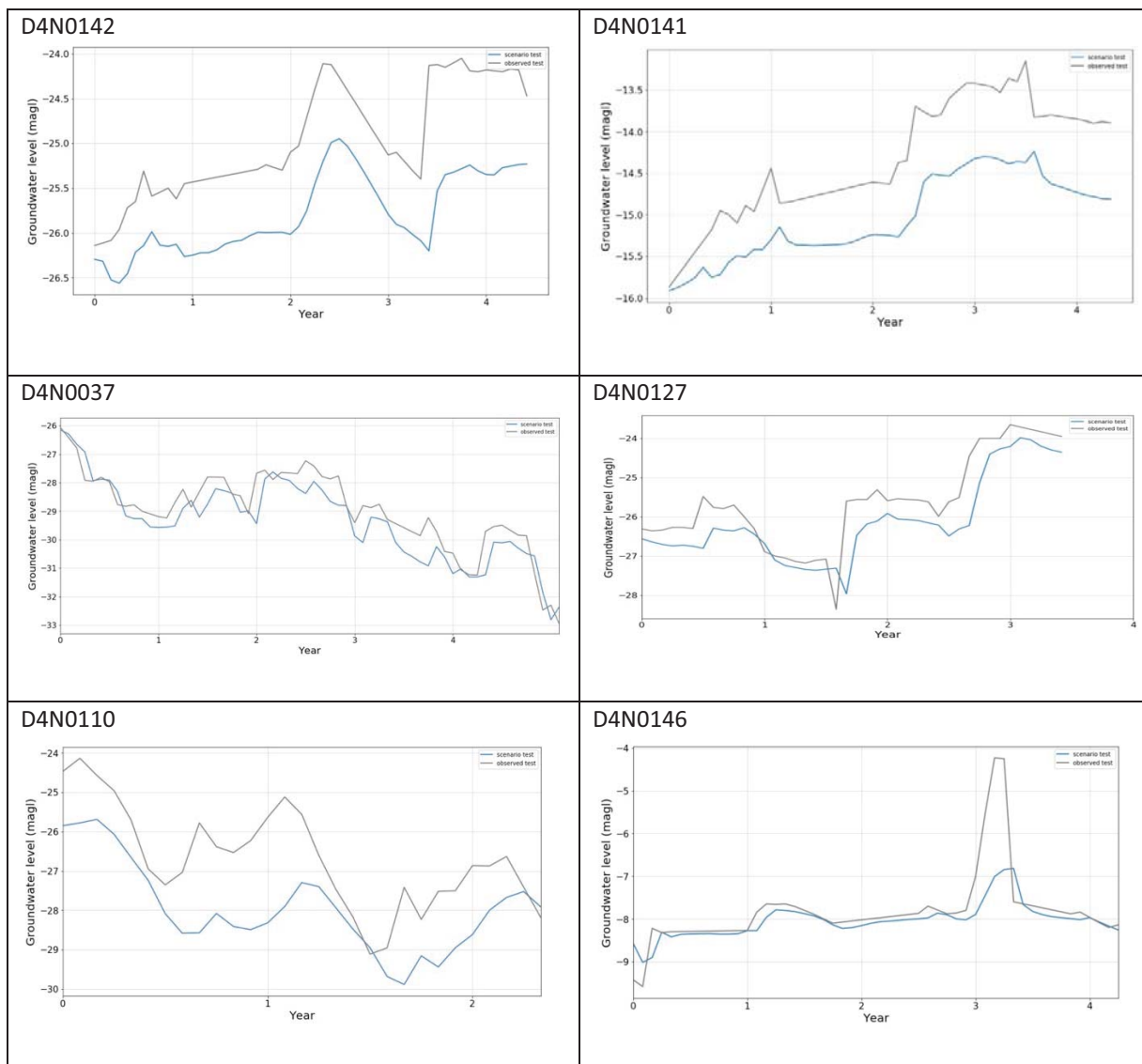Figure 4-15   Model prediction for scenario 1c, decrease rainfall.

Figure 4-16   Model prediction for scenario, increase abstraction.

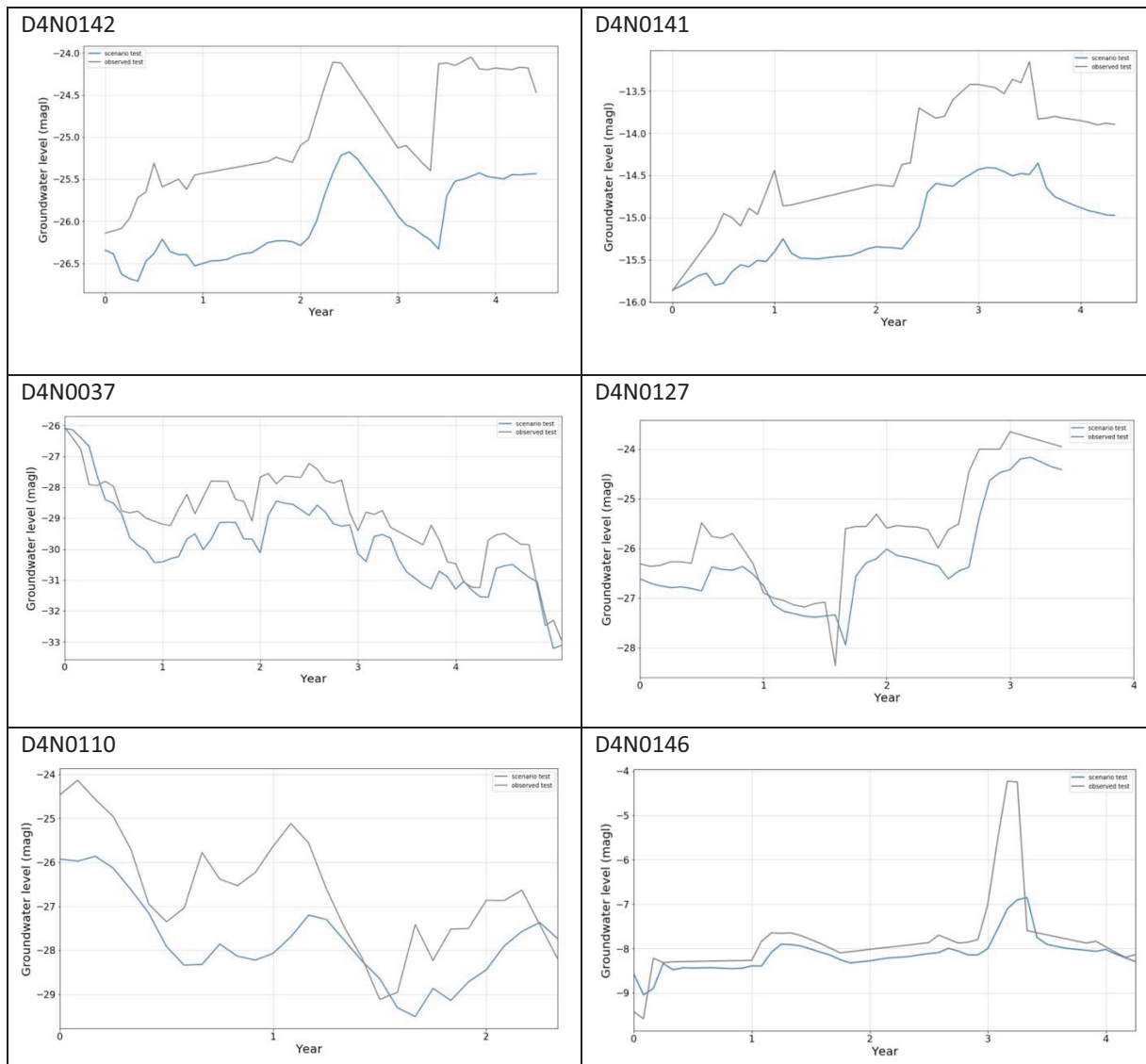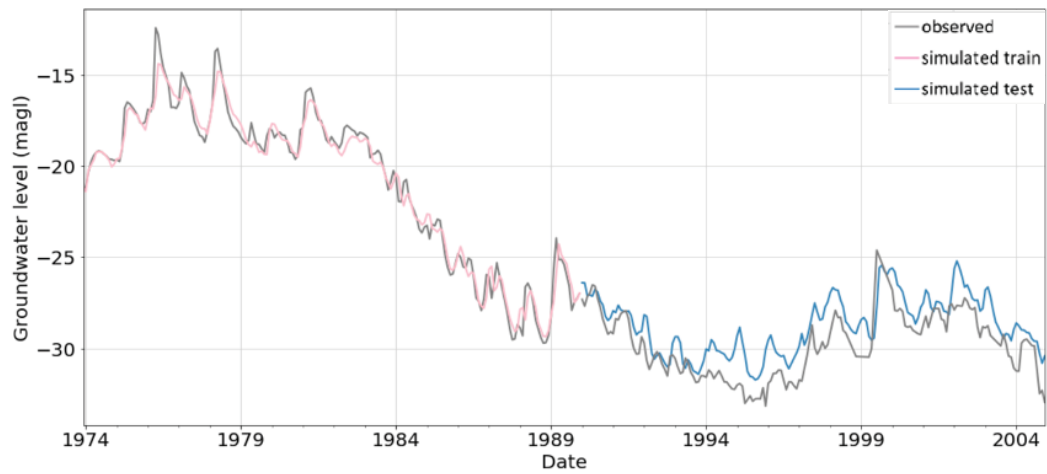**Figure 4-17   Results summary for Scenario 3.**

4a. Long term prediction with reduced train set size
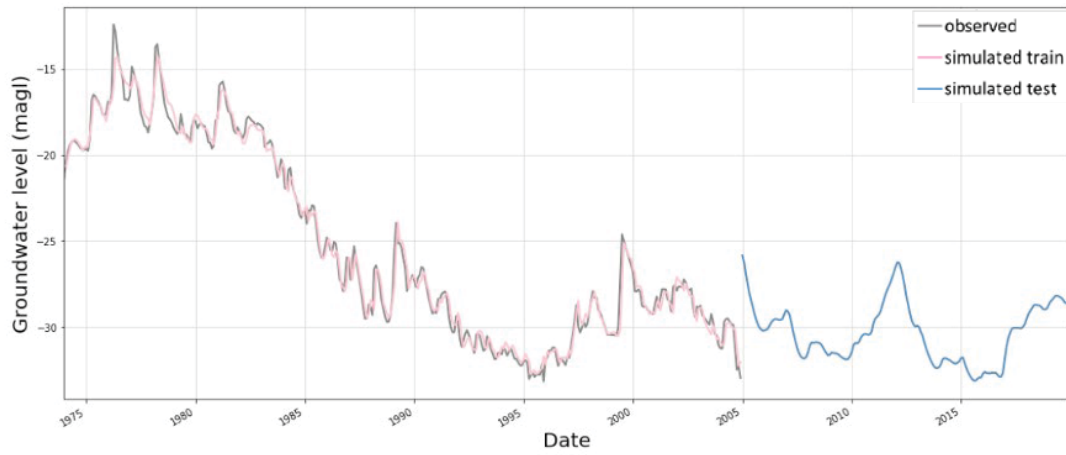
4b. long term prediction (15years)

Figure 4-18 **Model prediction for scenario 4a and 4b, long term prediction for borehole D4N0037.**

5 CASE STUDY 2: STEENKOPPPIES

5.1 DESCRIPTION OF STUDY AREA (CONCEPTUAL MODEL)

5.1.1 Location and setting

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 5-1) (Holland 2009). The aquifer provides water primarily for productive local agriculture, including the largest producer of carrots for export in South Africa (Cobbing, 2017). The farms dependant on groundwater employ over 4000 people and have significant links to other sectors of South Africa's economy (Cobbing, 2017).
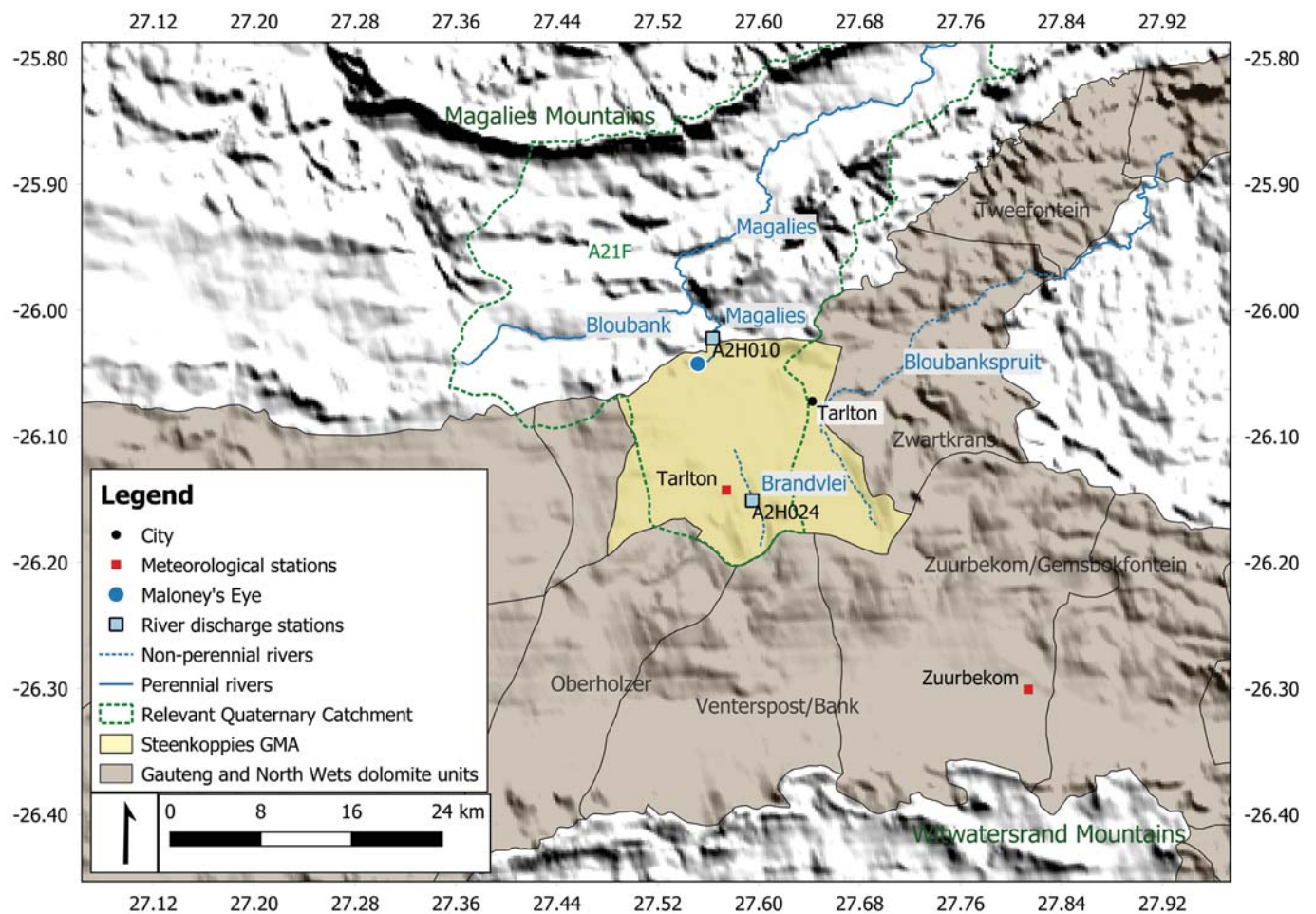
Figure 5-1 **Location of the Steenkoppies compartment boundary, relevant hydrology, meteorological stations and cities. (compartment boundaries from DWA 2009).**

5.1.2 Topography and Drainage and Land Use

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 5-1). Besides the non-perennial

Brandvlei River, there are no surface water drainage features or wetlands in the Steenkoppies compartment (Figure 5-1). This indicates that most of the rainfall infiltrates directly into the aquifer (Vahrmeijer *et al.*, 2013).

Figure 5-2 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. 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.
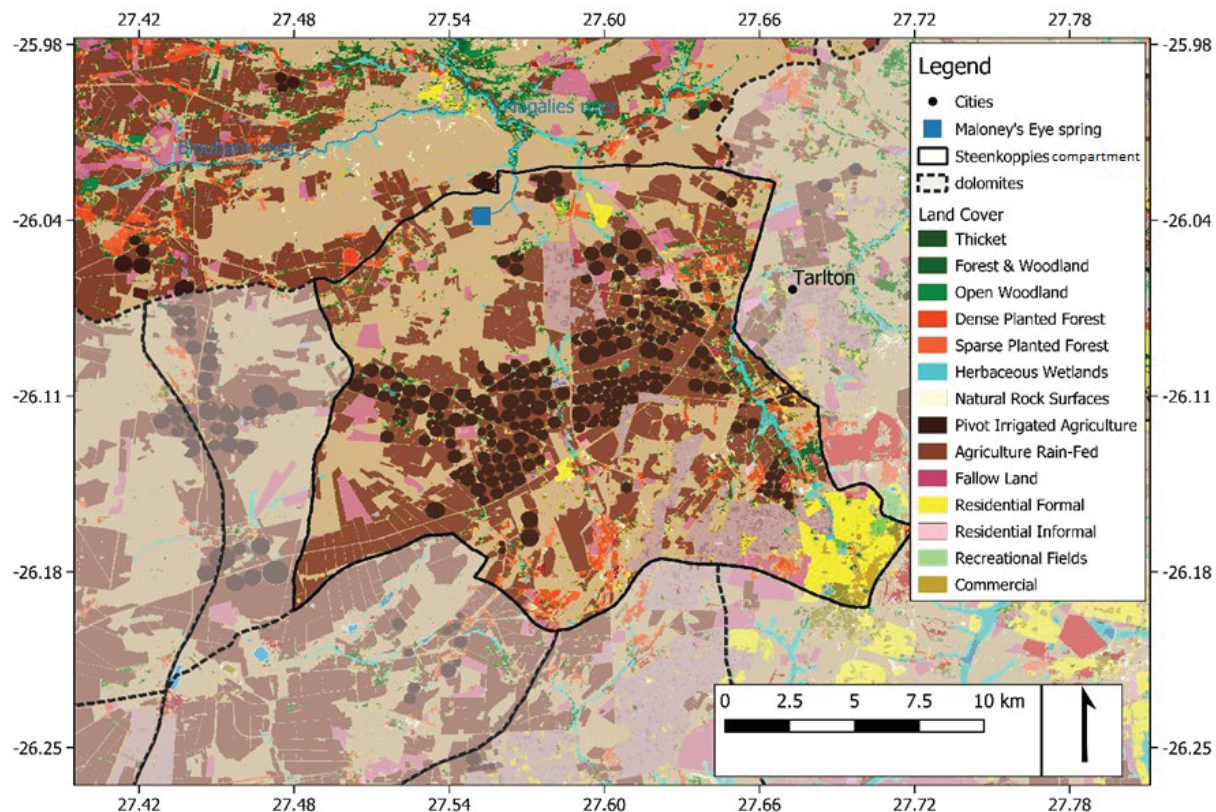
Figure 5-2   Land cover over the Steenkoppies aquifer (shapefile from the South African National Land-cover dataset of 2018 of the South African Spatial Data Infrastructure).

5.1.3   Climate

5.1.3.1   Rainfall

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.

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 5-2) to create a final rainfall data set from October 1908 to October 2019, presented in Figure 5-3. A two-year moving average of the annual rainfall (grey dotted line) and the mean annual precipitation (MAP, red dotted line) is also presented.

The annual rainfall recorded in the Steenkoppies compartment from 1909 to 2019 was the lowest in 1935 (323 mm) and the highest in 1929 (1081 mm) with a MAP of 669 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 5-2). Table 5-1 compares the MAP to the annual rainfall of four individual years showing the annual rainfall in 1994 was 87% of the MAP, while in 2004 the annual rainfall was only 73%. Table 5-1 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 5-2). Between 1966 and 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 in 1987-1989 and lower rainfall in 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, and the interval ranges from one to 15 years, with an average of 4 to 5 years, which is also illustrated in Figure 5-4, showing the monthly rainfall along with the trend of the monthly rainfall. In the last ten years, there has been a decline in the frequency of MAP above the average, and 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 previous five years (2009-2014). Groundwater levels in the Steenkoppies aquifer appear to increase when more than 100 mm rain falls in a month suggesting this is a threshold for recharge events (section 5.1.7).
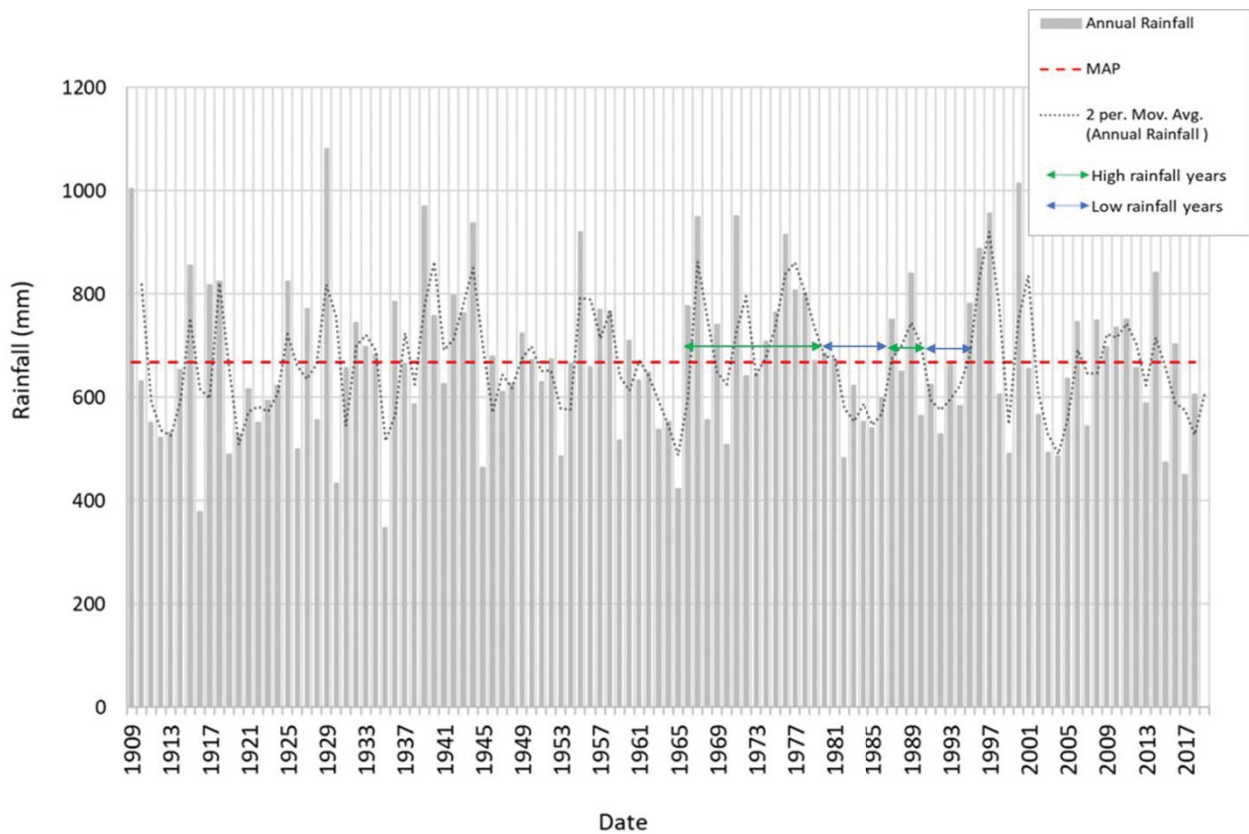
Figure 5-3   Annual rainfall in the Steenkoppies compartment (grey bars). A two-year moving average (grey dotted line), and the Mean Annual Precipitation (MAP) (the red dotted line) are also represented—An example of high and low rainfall years are represented as the green and blue arrows, respectively.

Table 5-1   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

Figure 5-4 **Monthly rainfall for the Steenkoppies aquifer (grey bars) and the trend of the monthly rainfall time series (blue line).**

5.1.3.2 Temperature

A historical temperature dataset was acquired from the SAWS and the Agricultural Research Council (ARC). The two meteorological stations recording temperature closest to the wells in the Steenkoppies aquifer was the Tarlton station on the Deodar farm situated in the Steenkoppies aquifer at 26°08'33''S and 27° 34'27''E, and the Zuurbekom station approximately 30 km southeast of the Steenkoppies aquifer at 26°18'03''S and 27° 48'49''E (Figure 5-1). The Tarlton station 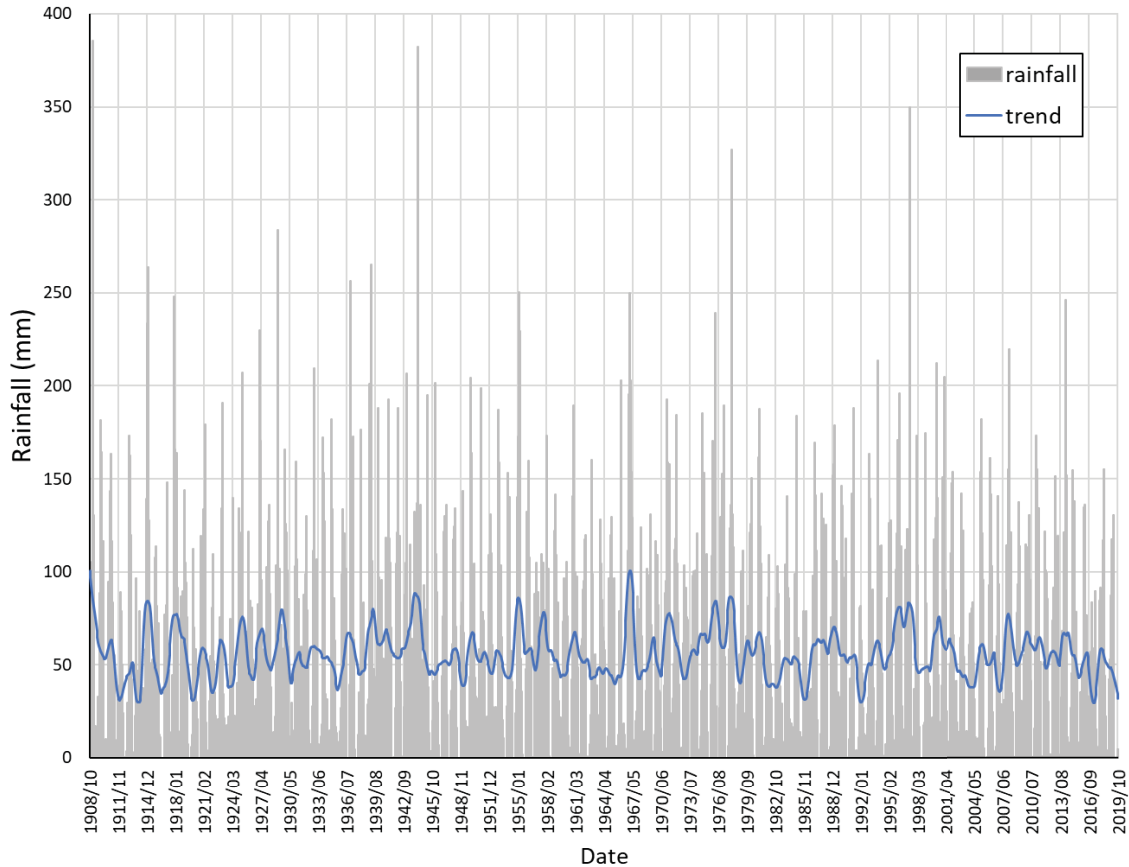recorded daily temperature from November 2003 to October 2019. The Zuurbekom station recorded daily temperature from January 1950 to August 2019.

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 datasets (November 2003 to August 2019) was calculated to assess the correlation between temperatures recorded at these two stations. Between the two datasets, there is a high linear correlation (R^2 score of 0.96). Both stations also showed a similar mean annual temperature where the deviation between the stations was less than 10%. The Zuurbekom data set can therefore be used as a realistic representation of the temperature in the Steenkoppies aquifer. The temperature in the Steenkoppies compartment has followed a consistent seasonal trend (Figure 5-5). 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 Steenkoppies compartment maintained by the ARC (Figure 5-5). The twelve-month moving average in Figure 5-5 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), which may result in less water recharging the aquifer.
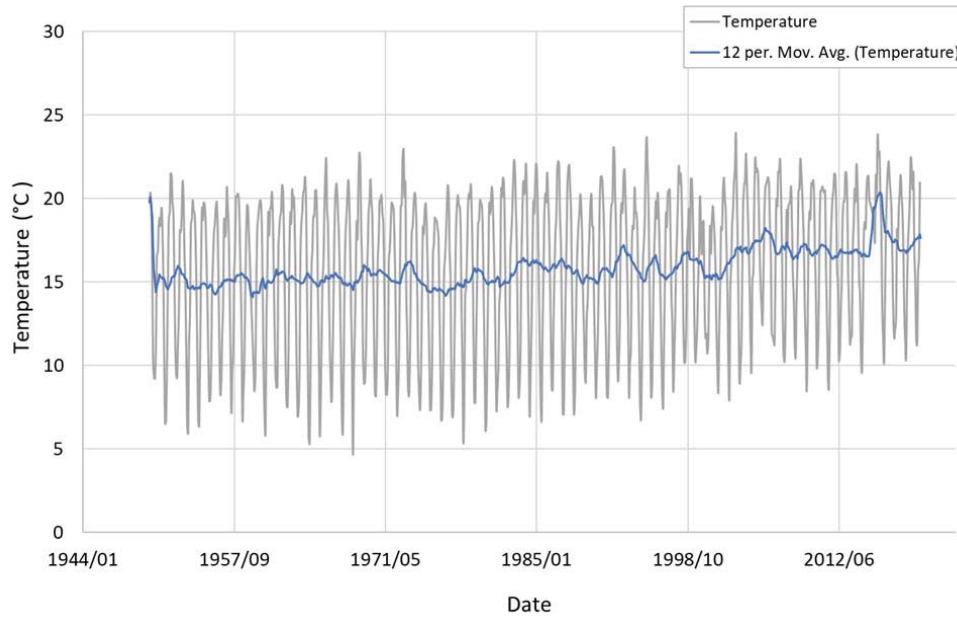
Figure 5-5   Mean monthly temperature in the Steenkoppies compartment (grey line) and the twelve-month moving average (blue line).

5.1.4   Hydrology and spring flow

The 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 mamsl (Figure 5-1). The Maloney's Eye is located at the intersection of an east-west striking fault zone and Malmani Subgroup that is overlain by of a lower permeability unit, forcing the groundwater to the surface. The Maloney's Eye serves as the only natural significant outlet for the groundwater in the Steenkoppies compartment, and is the origins of the Magalies River, that feeds the Hartbeespoort Dam.

River discharge data for the Maloney's Eye Spring was obtained from gauge stations maintained by the DWS. From 1908-2013 the monthly average discharge from Maloney's Eye spring varies from a low of 0.051 m^3/s recorded in April 2007 to a high of 1.035 m^3/s recorded in March 1979 (Figure 5-6).

There are six major peaks in spring discharge in 1919, 1945, 1979, 1997, 2001 and 2011 (peaks are seen in both Figure 5-6 and Figure 5-7). These peaks are related to the cyclical patterns described in the rainfall time series as the peaks coincide with the end of high rainfall years. 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 m^3/s) after a 15-year cycle of higher rainfall from 1966-1981 (Figure 5-7). The spring discharge peak in 2001 is much lower (0.43 m^3/s) after only one year of high rainfall recorded from 2000-2001 (Figure 5-7).

The most recent peak seen in the average spring discharge in 2011 is low (0.28 m^3/s) although it follows a seven-year cycle of high rainfall between 2005 and 2012. It is uncertain whether this peak is related another external factor acting on the spring discharge such as groundwater abstraction in the Steenkoppies aquifer. Groundwater abstraction results in a decrease in discharge (Konikow and Bredehoeft, 2019). Vahrmeijer *et al.* (2013) used the cumulative rainfall departure (CRD) method to evaluate the relationship between rainfall and discharge and confirms that there is a reasonable correlation between rainfall and spring discharge. However, 1987 the actual measured discharge was lower than the simulated discharge from the CRD, indicating that other external factors not accounted for in the model (such as abstraction) influence the discharge from around 1987.
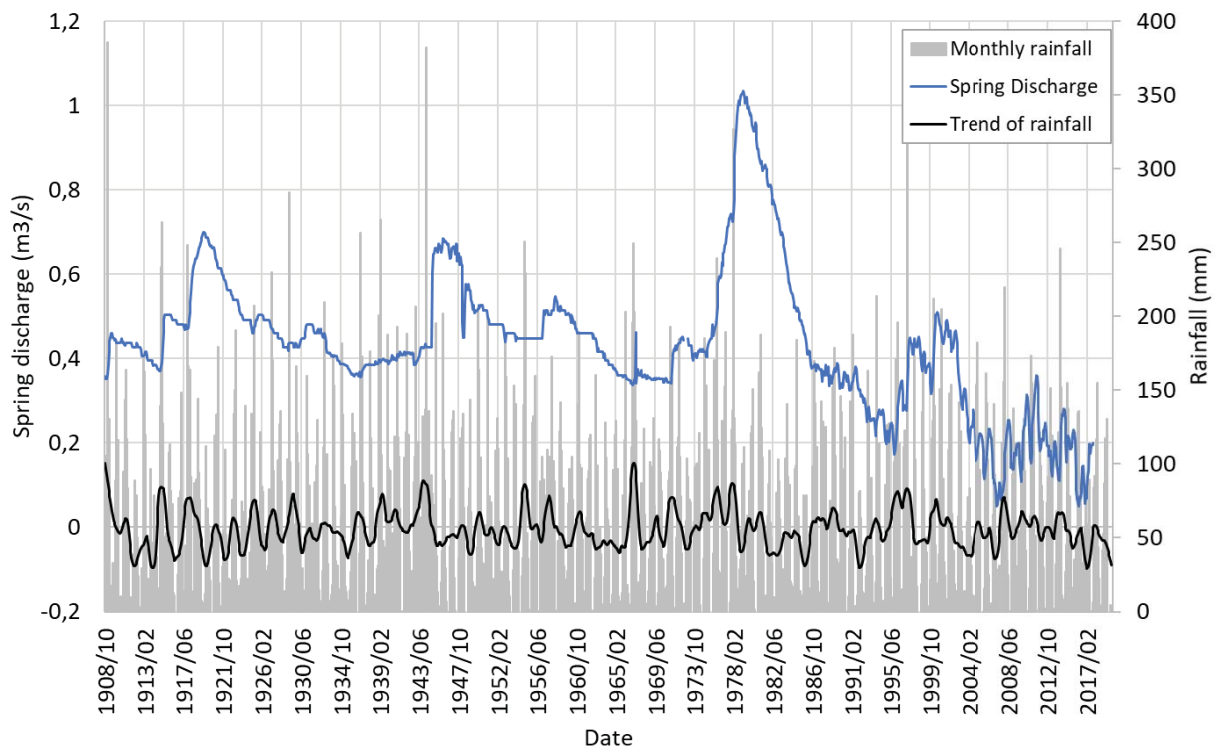
Figure 5-6 **Monthly mean discharge recorded at the Maloney's Eye Spring and monthly rainfall with the trend of the rainfall time series over time.**
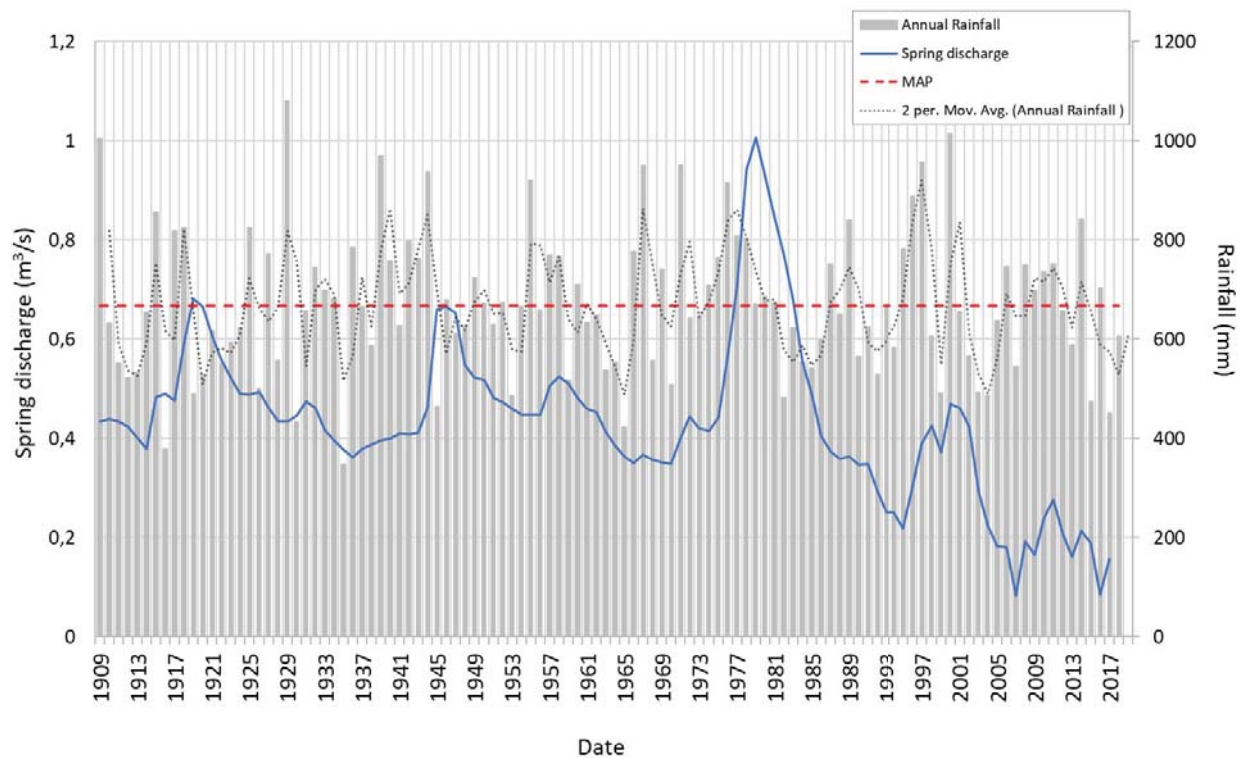
Figure 5-7 **Average annual spring discharge from the Maloney's Eye spring is represented as the blue line. Annual rainfall from nearby SAWS metrological stations is defined as the grey bars with a two-year moving average shown as the grey dotted line, and the Mean Annual Precipitation (MAP) is shown in red.**

5.1.5 Geology

The stratigraphy of the Steenkoppies compartment is outlined in Table 5-2. 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 (DWS, 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 Wolwekrans 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 5-2   Stratigraphy of the Gauteng and North West dolomites (Source: Vahrmeijer et al. 2013).

Super Group	Group	Formation	Thickness (in m)	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.
		Timeball Hill	270-660	Shale, Diamictite, Klapperkop Quartzite and ferruginous quartzite.
		Rooihoogte	10-150	Quartzite, Shale, Bevets Conglomerate Member and Breccia.
	CHUNIESPOORT	Frisco	30-158	Chert-free dolomite with some primary limestone and carbonaceous shale at the base.
		Eccles	490	Chert-rich dark dolomite with stromatolitic and oolitic bands. Chert 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
WITWATERSRAND	CENTRAL RAND	-	2 880	Aranaceous, rudaceous rocks.
	WEST RAND		5 150	Quartzite, reddish and ferruginous magnetic shales.
	DOMINION		?	Quartzite, conglomerate, shale, interbedded lava.
BASEMENT COMPLEX				

5.1.6 Groundwater levels and flow direction

For this study, historical groundwater level data was requested from the NGA and HYDSTRA databases. 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 (NGA and HYDSTRA boreholes used for modelling are shown in Figure 5-8, along with the positions of registered groundwater use from WARMS).
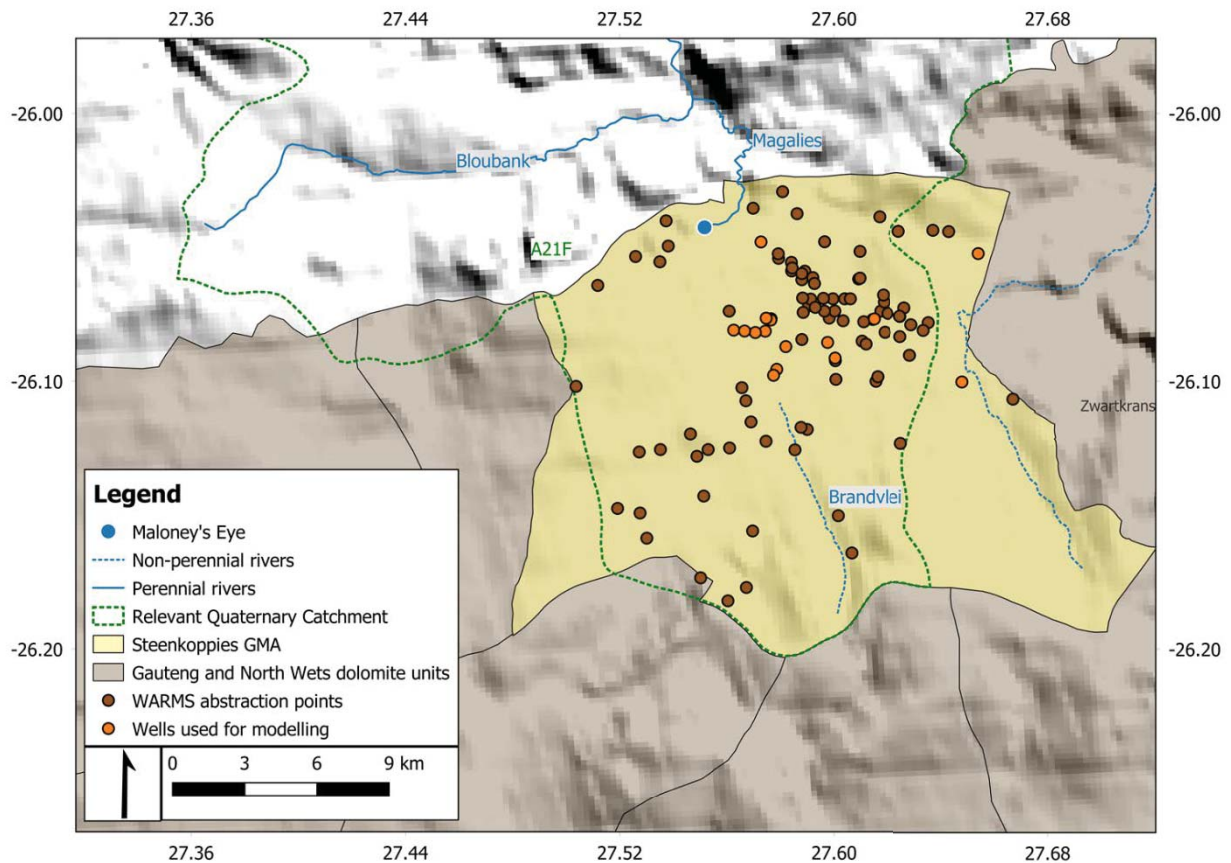
Figure 5-8 Position of NGA and HYDSTRA boreholes used for modelling in the Steenkoppies aquifer, along with abstraction points registered in WARMS.

Monthly groundwater level time series recorded from boreholes in the Steenkoppies compartment are shown in Figure 5-9. The groundwater levels in the Steenkoppies compartment range from approximately -82 magl to -29 magl. A similar trend and timing of the groundwater level increases and decreases in response to rainfall are recorded throughout the compartment (Figure 5-9), although each borehole shows a different magnitude of response.

The fluctuations in the groundwater level time series coincide with the rainfall cycles outlined in Figure 5-4. A selection of six groundwater level time series is shown with the monthly rainfall and the monthly rainfall trend (Figure 5-10). The peaks and the troughs in the groundwater levels represent a slightly delayed version of the monthly rainfall (Figure 5-10) and strongly mimic the rainfall trend.

Figure 5-9   Monthly groundwater levels recoded from boreholes in the Steenkoppies compartment.
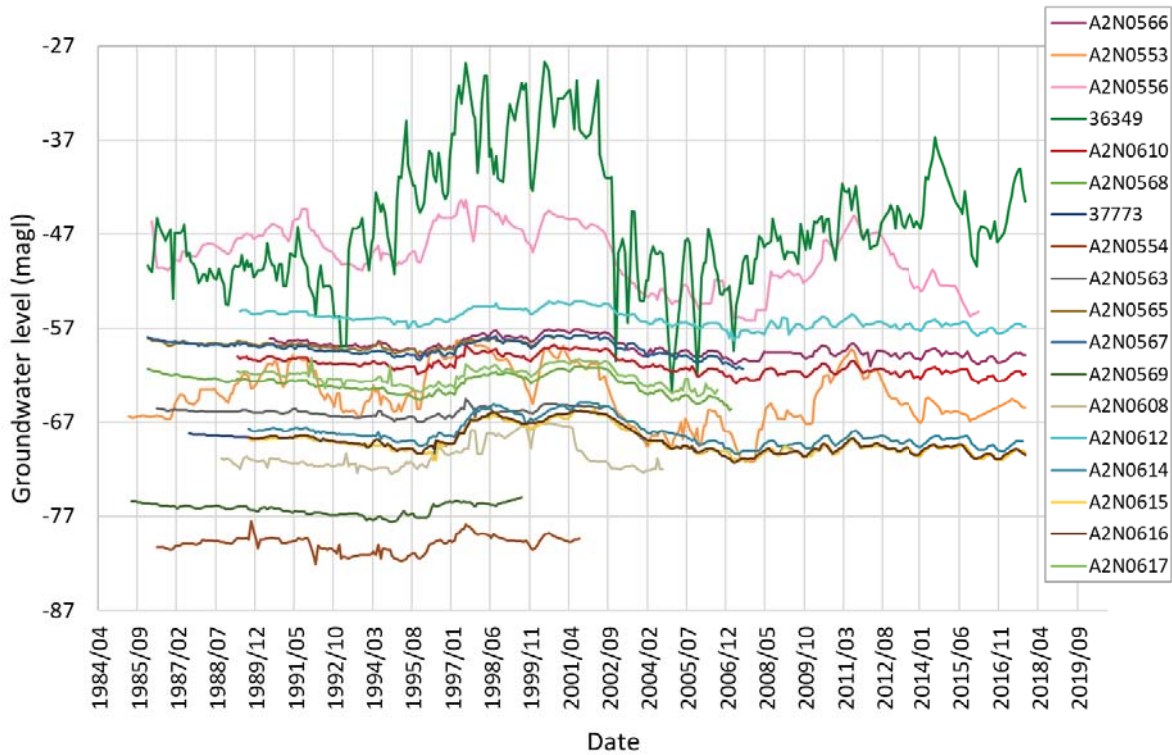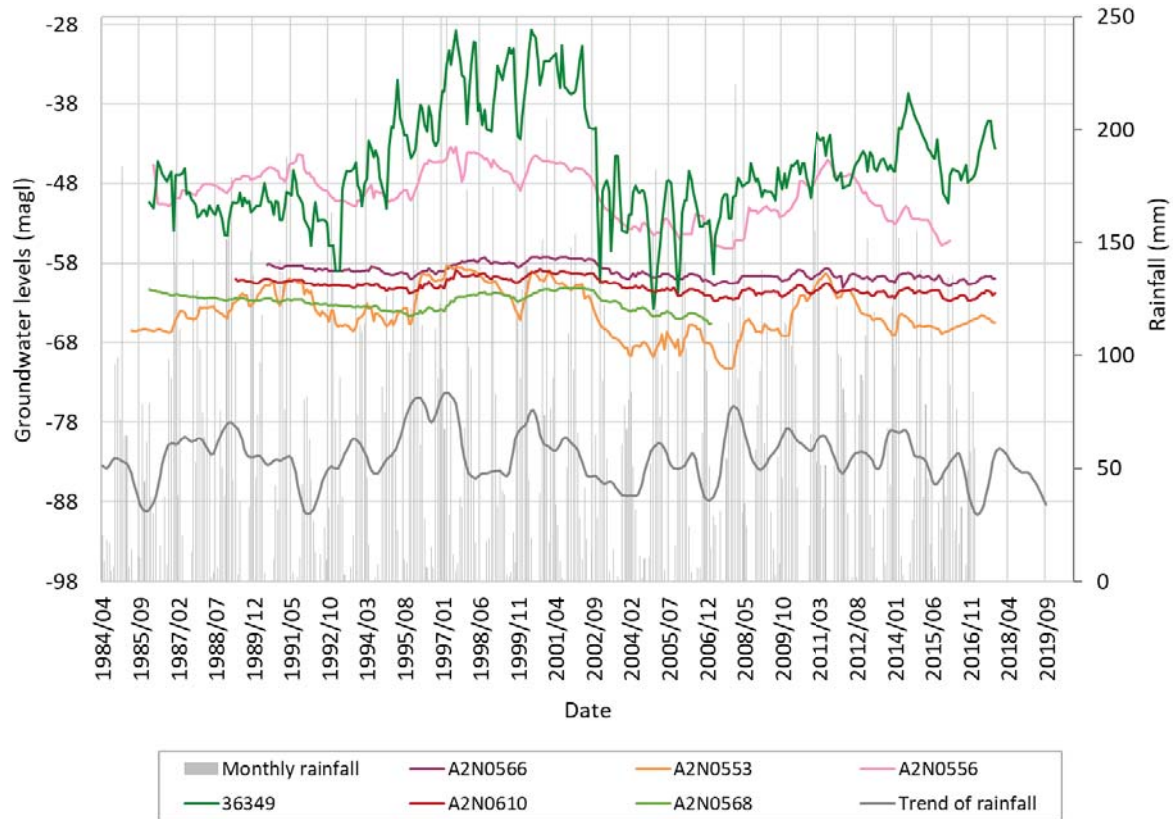

Figure 5-10   Monthly groundwater level (left axis) and monthly rainfall with the trend (right axis) over time.

Figure 5-11 shows the six groundwater level time series with the monthly mean spring discharge from the Maloney's Eye. The conceptual hydrogeological model of 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 changes will directly translate to changes in the spring discharge rate. This is also confirmed in the historical time series (Figure 5-11). The pattern of the spring discharge time series is very similar to the groundwater level time series (Figure 5-11). The spring discharge records significant peaks in 1997, 2001 and 2011, which coincide 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.
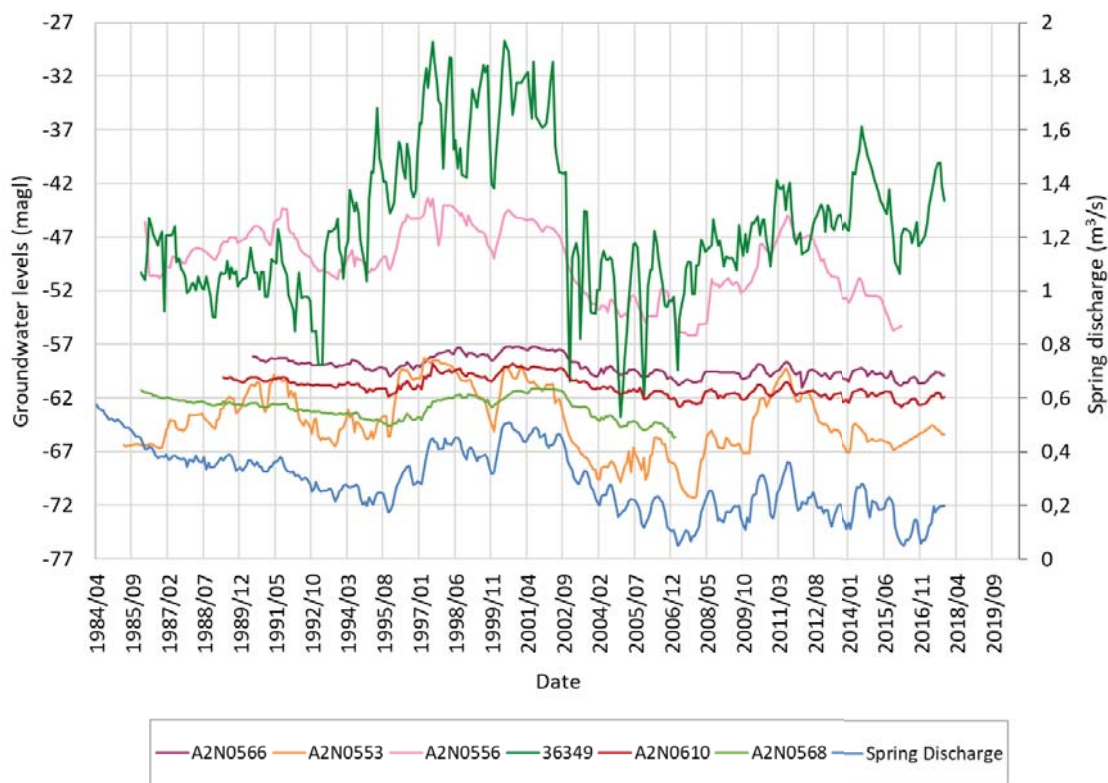
Figure 5-11   **Monthly groundwater level (left axis) and monthly spring discharge (right axis) over time.**

Table 5-3 shows a summary of the groundwater level data sets used for modelling in the Steenkoppies compartment as well as 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 trends. The declining in groundwater level and spring discharge has been related to the impact of sustained and growing abstraction across the compartment (section 5.1.8) but will also be impacted by the decline in rainfall.

Table 5-3 Summary of the groundwater level (GWL) time series in the Steenkoppies. A summary of the monthly rainfall and spring discharge time series is given. (The units for the rainfall and spring discharge time series differ from the heading and shown in the table.)

Borehole ID	Latitude	Longitude	Start date	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

5.1.7 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 of groundwater level demonstrate that groundwater levels are influenced and recharged by rainfall (e.g., Figure 5-10). 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. After sustainaed rainfell events 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). Analysis of the data showed that monthly rainfall greater than 100 mm result in significant recharge to the aquifer (increase in groundwater levels).

5.1.8 Abstraction

The Water use Authorisation and Registration Management System (WARMS) data set records a total registered groundwater use of 78 million cubic meters per annum (million m^3/a) from the Steenkoppies compartment (Table 5-4). 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 5-4). This makes up 76% of the total registered groundwater use in the Steenkoppies aquifer.

Table 5-4   **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 5-5 and Figure 5-12). The data set 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 5-12). 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. The earliest start date for registrations in the WARMS data set is 1988 for the Steenkoppies compartment, and therefore it is not a representative dataset for earlier groundwater use. The stepwise increases in both the WARMS and literature dataset are likely related only to inaccuracies in the datasets, rather than related to actual increases in groundwater use, and a more gradual increase is likely.
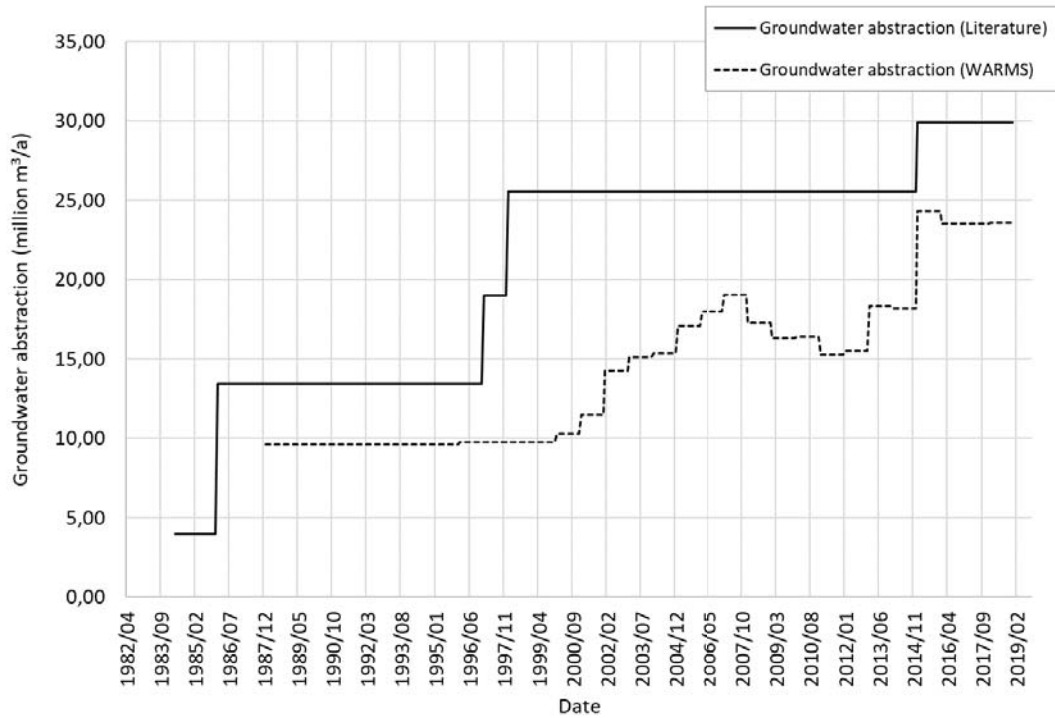
Figure 5-12 Groundwater usage in the Steenkoppies compartment. The dotted line represents registered groundwater (WARMS, 2019), the solid line represents groundwater usage from literature (Seyler et al., 2019).

Table 5-5 Groundwater usage in the Steenkoppies aquifer estimated from literature (Seyler et al., 2019).

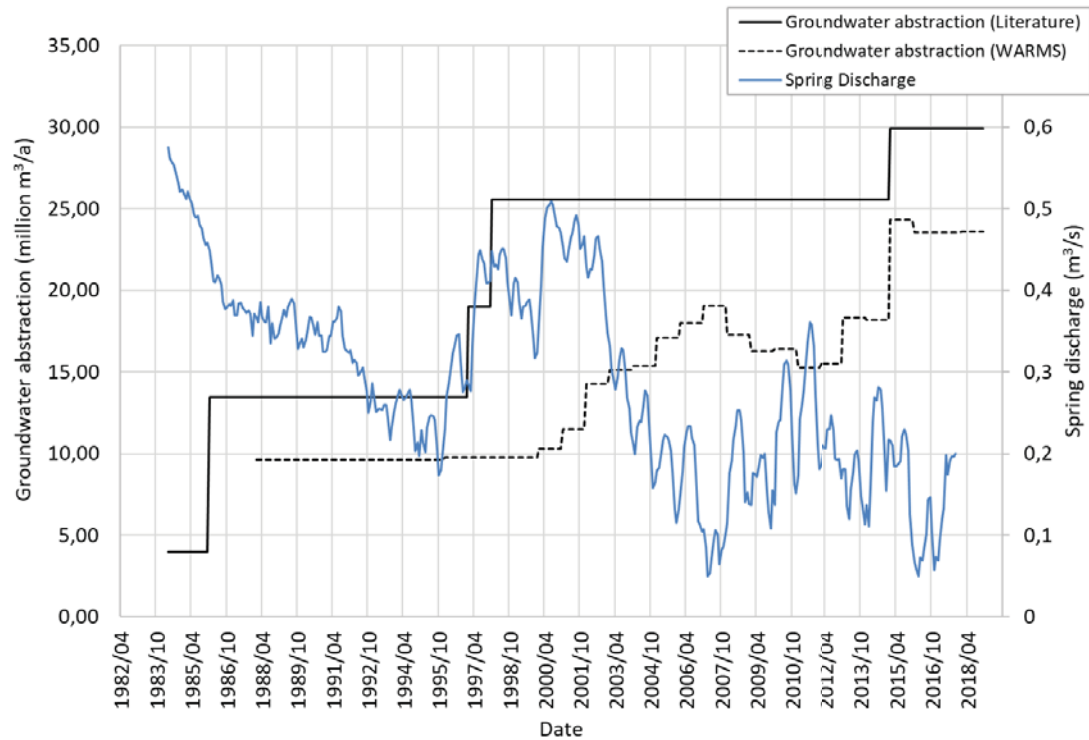| Reference	Year	Abstraction (million m³/a)
Hobbs (1980)	1980	3.95
Bredenkamp et al (1986)	1986	13.45
Barnard (1997)	1997	19
Schoeman and Partners (2016)	1998	25.55
Schoeman and Partners (2016)	2015	29.92

Figure 5-13 Monthly spring discharge (left axis) and groundwater usage in the Steenkoppies compartment (right axis).
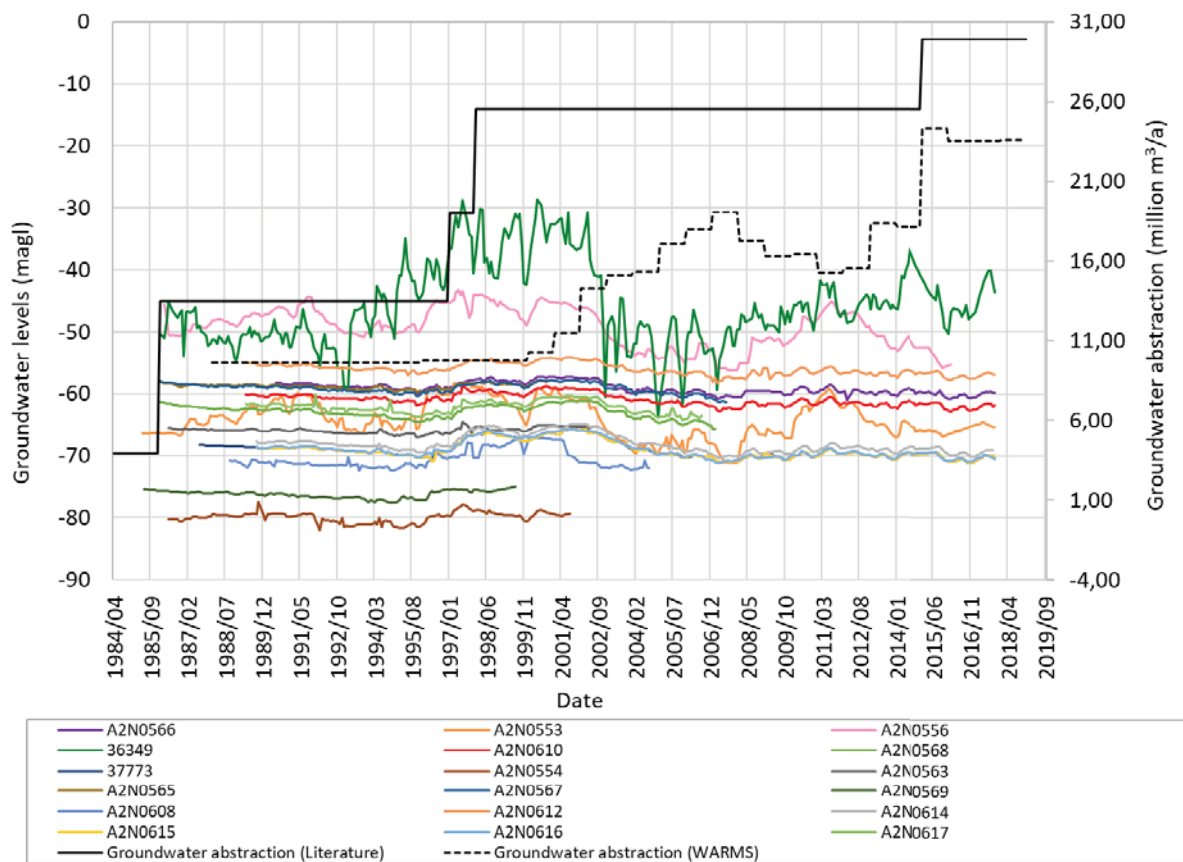

Figure 5-14 Monthly groundwater levels (left axis) and groundwater usage in the Steenkoppies compartment (right axis).

5.2   NNAR MODEL SET UP

5.2.1   Model software and approach

The nnetar function in the forecast package (Hyndman and Athanasopoulos, 2014) in R (R Core Team, 2017) is used to fit the NNAR model to the time series. The nnetar implements the NNAR model using a neural network with a sigmoid activation function, a single hidden layer and lagged inputs for forecasting the target variable. Figure 5-15 summarises the applied approach taken to model and predict groundwater levels using the NNAR. The approach is similar to that applied for the LSTM and a more detailed explanation of the approach is given in section 4.2.2.2.

The first 80% of the target and input time series is used for training and the last 20% is used for testing. During training the nnetar function automatically selected the number of lagged inputs for forecasting through hyper-parameter tuning. Once the model had been trained and the hyper-parameters optimised for the dataset, the model is ready to make predictions. There is an element of randomness in the predictions. Therefore, as mentioned in section 4.2.2.2, the model makes groundwater level predictions 10 times, and the average of these 10 predictions is calculated and the model performance is assessed using this average prediction.
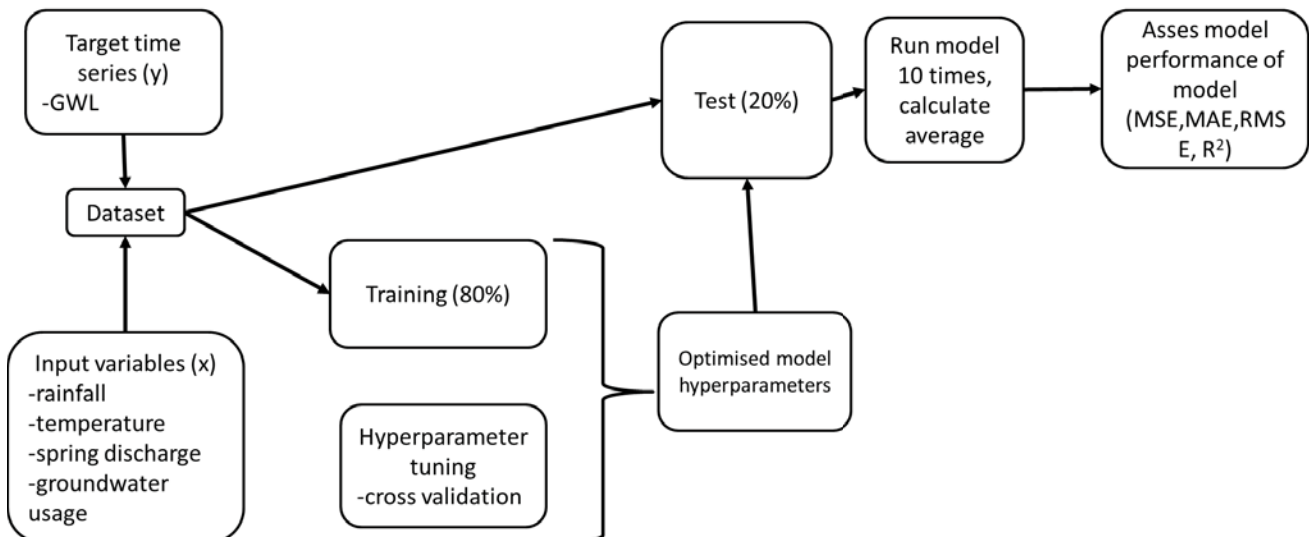

Figure 5-15   The applied methodological approach taken to model and predict groundwater levels.

5.2.2   Model input variables

In this study, the relationship between the target variable (groundwater levels) and the input variables (rainfall river discharge, temperature and groundwater usage) was used to assist predictions (as per section 4.2.3.2). Similar to section 4.2.2, the input parameters were deemed appropriate as they each have a relationship with the target groundwater levels. The relationship between the target groundwater level and each input variable in the Steenkoppies aquifer is outlined in section 5.1.

Similar to that described in section 4.2.2 for the LSTM, the NNAR model could not detect the relationship between rainfall and groundwater level in the aquifer as the raw rainfall time series dataset was too noisy. Therefore, STL time series decomposition was performed, and the trend of the rainfall was extracted and used as an input to model time series. The original time series and the trend of the time series is shown in Figure 5-16.
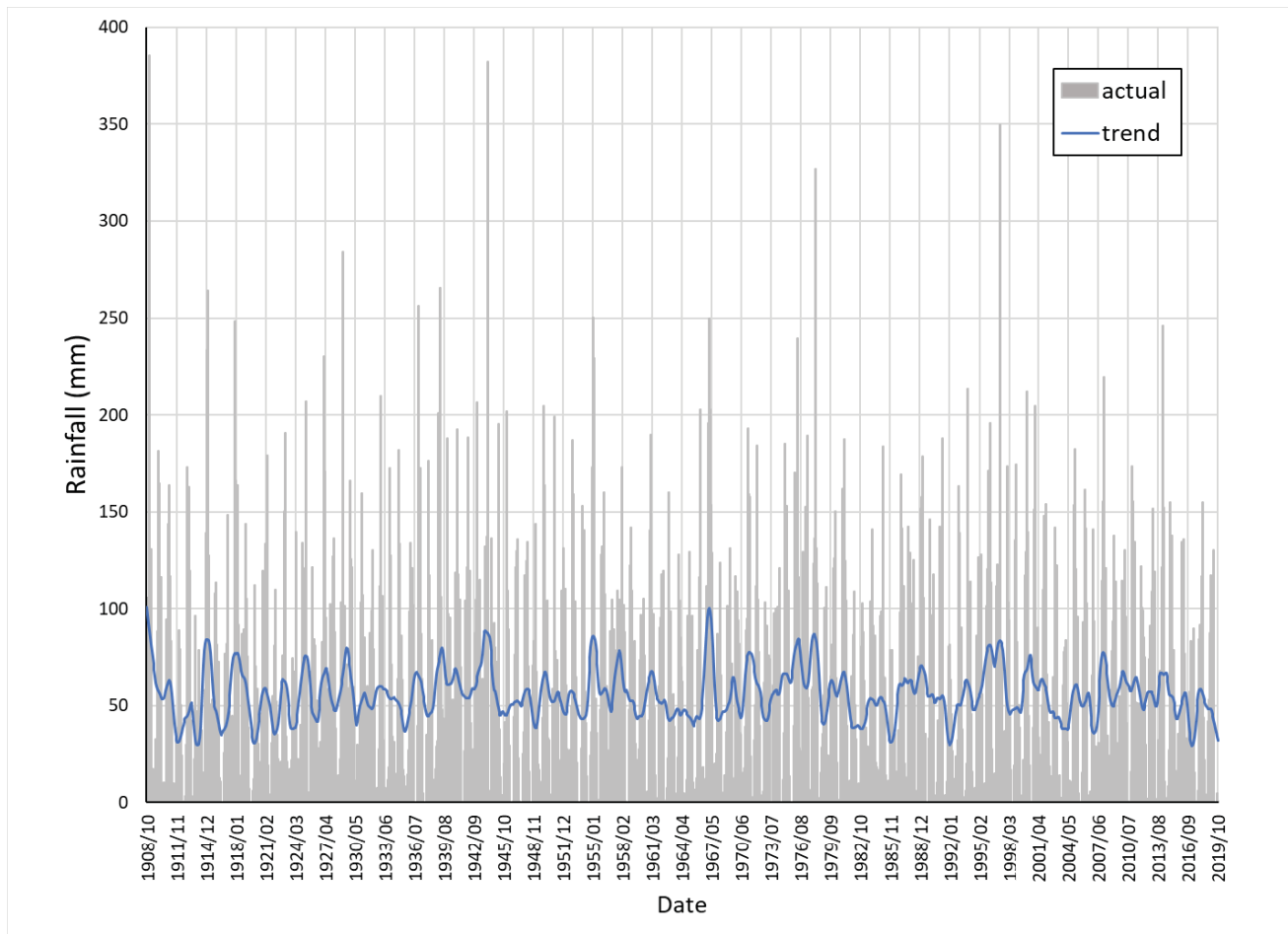
Figure 5-16   The actual rainfall and the trend of the rainfall over time.

5.2.3   NNAR model pipeline

The same cross-validation approach was implemented for modelling Steenkoppies with NNAR as for the Grootfontein case study with the LSTM, which is discussed in section 4.2.3.2. The data gaps in the input and target time series were interpolated linearly as per section 4.2.3.3. The model evaluation criteria used to assess the performance of the NNAR model in this Steenkoppies case study is the same as that used to evaluate the performance of the LSTM used in the Grootfontein case study discussed in section 4.2.3.4.

5.2.4   Information Gain Ranking

The methodological approach used to calculate the information Gain Ranking in this Steenkoppies case study is the same as that used in the Grootfontein case study discussed in section 4.2.4.

5.2.5   Scenario Testing

Scenario modelling for Steenkoppies was performed in the same way as that in Grootfontein; refer to section 4.2.4 for the methodological approach taken to perform scenario modelling.

5.3   MODEL RESULTS

5.3.1   Information Gain ranking

The measures of mutual information (MI) for the five input variables with respect to groundwater levels in the Steenkoppies compartment, with the mean (x̄) and standard deviation (sd), are presented in Table 5-6. In the Steenkoppies compartment, spring discharge recorded the highest MI with respect to groundwater levels (x̄ = MI 0.78,

sd = 0.09) 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 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 an 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 5-10). 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 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). 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. Only larger rainfall events (> 100mm per month) recharge the aquifer and correlate with groundwater level increases. Whilst smaller rainfall events (< 100 mm per month) were not informative with regards to groundwater levels (section 5.1.7). 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 (section 5.1.6), whether this pattern would be detectable in the models was uncertain before undertaking the research given the nature of the groundwater use data, and the comparatively noisy groundwater level. However, the results show 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, 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 (section5.1.3.2), a low MI recorded show that temperature (substitute for evapotranspiration in this study) was not informative for 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 evapotranspiration reduces (Gulliver *et al*., 2010). Furthermore, evapotranspiration is a complex process and involves other factors such as vegetation type, irrigation schedules, soil types etc, which contribute to there being a low MI between groundwater levels and temperature.

	Borehole ID	Temperature	Rainfall trend	Rainfall	River 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

5.3.2   Model training and testing

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 5-7. The performance of the NNAR was also assessed visually and the graphical results from six boreholes are displayed and discussed in this chapter, and the additional 12 graphs are presented in Appendix B.

The error indices indicate that the performance of the NNAR varied between the boreholes (Table 5-7). 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). The 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 5-7).

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 (better result) compared to borehole A2N0556, the MAE, MSE and RMSE recorded higher error metrics for borehole A2N0553 compared to A2N0556 (Table 5-7). In addition, boreholes with low and negative R^2 values do to replicate groundwater levels relatively well when inspecting the graphs. Wunsch et al. (2018) used NARX to model groundwater levels. Similarly, 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, 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 5-7   **Summary of the metrics used to assess NNAR performance.**

	Well ID	R^2	MSE (m)	RMSE (m)	MAE (m)	STDV across ten model runs
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 of the target groundwater level time series (Figure 5-17). Similar to the statistical performance (Table 5-7), 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 5-17).

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 5-17). 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, 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 (Agrawal and Adhikari, 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 5-17). 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, which 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 5-7 and Figure 5-17). The MI of each input feature with respect to groundwater levels for borehole A2N0566 was higher than for borehole A2N553 (Table 5-6). As a result, the model was able to make better groundwater level predictions at borehole A2N0566 compared to borehole A2N553.

The NNAR model showed overfitting at borehole 36349 (Figure 5-17). Overfitting is when a model fits extremely closely to the training data resulting in a model that has learnt to reproduce noise and peculiarities in the data and fails to learn the general predictive rule (Wickham, 2016). Overfitting causes the accuracy of the model to deteriorate when predicting anything outside the range of 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 5-17). 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 level predictions (R^2 = -1.19, MSE = 23.04, RMSE = 4.75, MAE = 4.03) (Table 5-7). Zanotti *et al.* (2019) proposed that as an alternative to cross-validation, the parameters such as the weights and bias, and 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).

The groundwater level predictions made by the NNAR were slightly lagged outputs compared to the observed groundwater levels. These lags are clearly seen in the model results of similar studies that made use of autoregressive models (e.g., Wunsch *et al.*, 2018; Guzman *et al.*, 2017), however no explanation is provided in these studies. 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 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 5-17). Using the input variables, the NNAR model was able to simulate a more realistic groundwater level for that period.
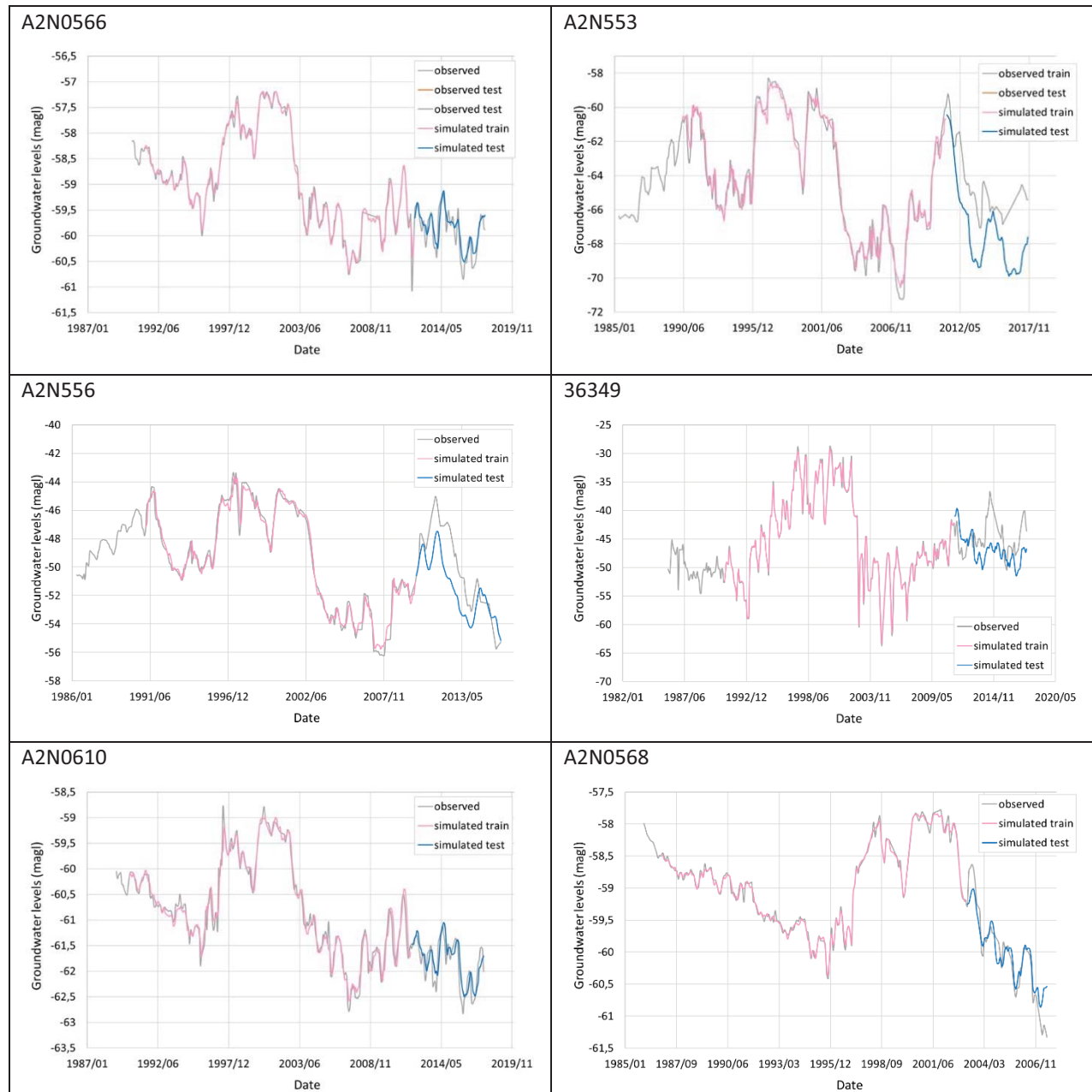

Figure 5-17   Comparison between the observed and simulated groundwater levels for six boreholes in the Steenkoppies compartment, during training and testing phases.

The Standard deviation between the ten independent runs of six models is shown graphically in Figure 5-18. 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 5-18). 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 5-18 graph A2N0610 and A2N0568), 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 5-18). This may be attributed to the fact that the model over-fits to the data during training (Figure 5-17). 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.
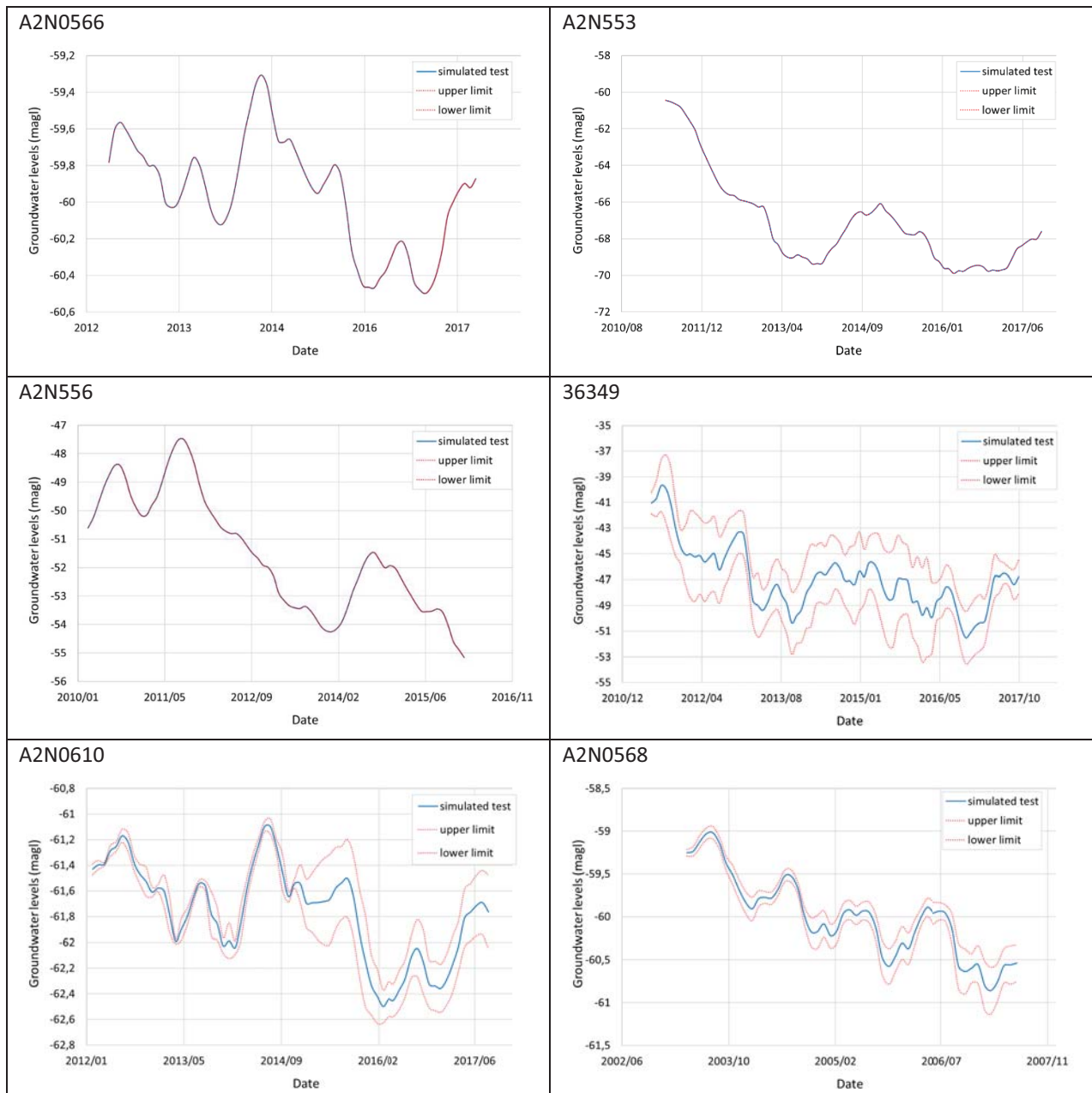
Figure 5-18 **Standard deviation error between ten independent runs of each model used to predict the groundwater levels for six boreholes.**

5.3.3 Scenario testing

Scenario 1a and c 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 1c simulated reduction in all rainfall events (monthly rainfall halved). For scenarios, 1a and c, the NNAR model simulated a decrease in groundwater levels at boreholes A2N553, A2N556 and 36349 (Figure 5-19 and Figure 5-20). At these three boreholes, 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 1c (in both Figure 5-19 and Figure 5-20). The model simulated groundwater levels similar to the observed groundwater levels suggesting rainfall has a minor influence in the modelled groundwater level.

This may be due to the very high MI recorded between spring discharge and groundwater levels at these boreholes (Table 5-6). The MI between spring discharge and groundwater levels at boreholes A2N0566, A2N0610 and A2N0568 was 1, 1 and 0.74 respectively (Table 5-6). 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 dominated the predictions.

Figure 5-21 shows the model results for scenario 1b, 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 c, 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 1b. 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 strength in the relationship between the input variable and target variable drives whether the groundwater level responds to the scenario, but also drives the magnitude of the response. For example, in borehole A2N0566 of Figure 5-21, 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. The hydraulic properties (the hydraulic diffusivity) of the aquifer are heterogeneous and impact the magnitude of the response to rainfall, along with variability in the recharge rate itself. 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. It would therefore be expected that different boreholes have different relationships with rainfall.

Figure 5-22 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 stepwise increase in abstraction with time). For boreholes A2N556, 36349, and A2N0610 (Figure 5-22) 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 5-22), the model recognises change, but it results in subduing the fluctuations in groundwater level change, since the groundwater usage data for the time period of the scenario is flat. 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 (excluding abstraction) and groundwater levels. The NNAR was clearly able to recognise there is a change and produces predictions that make sense according to hydrogeological knowledge and the knowledge of the Steenkoppies compartment (outlined in 5.1).

The long-term predictions in Figure 5-23 for scenario 4a 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 (hyper-parameter tuning) and the predictions made vary based on the data used to train the model (MacKay, 2005; Wickham, 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).

Scenario 4b used the long-term averages of the input variables to predict a long-term forecast of groundwater levels (results shown in Figure 5-23). This is merely a demonstration to show that it is possible to use the NNAR to make long-term predictions greater than 30 years. 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 to ensure reliable model predictions.
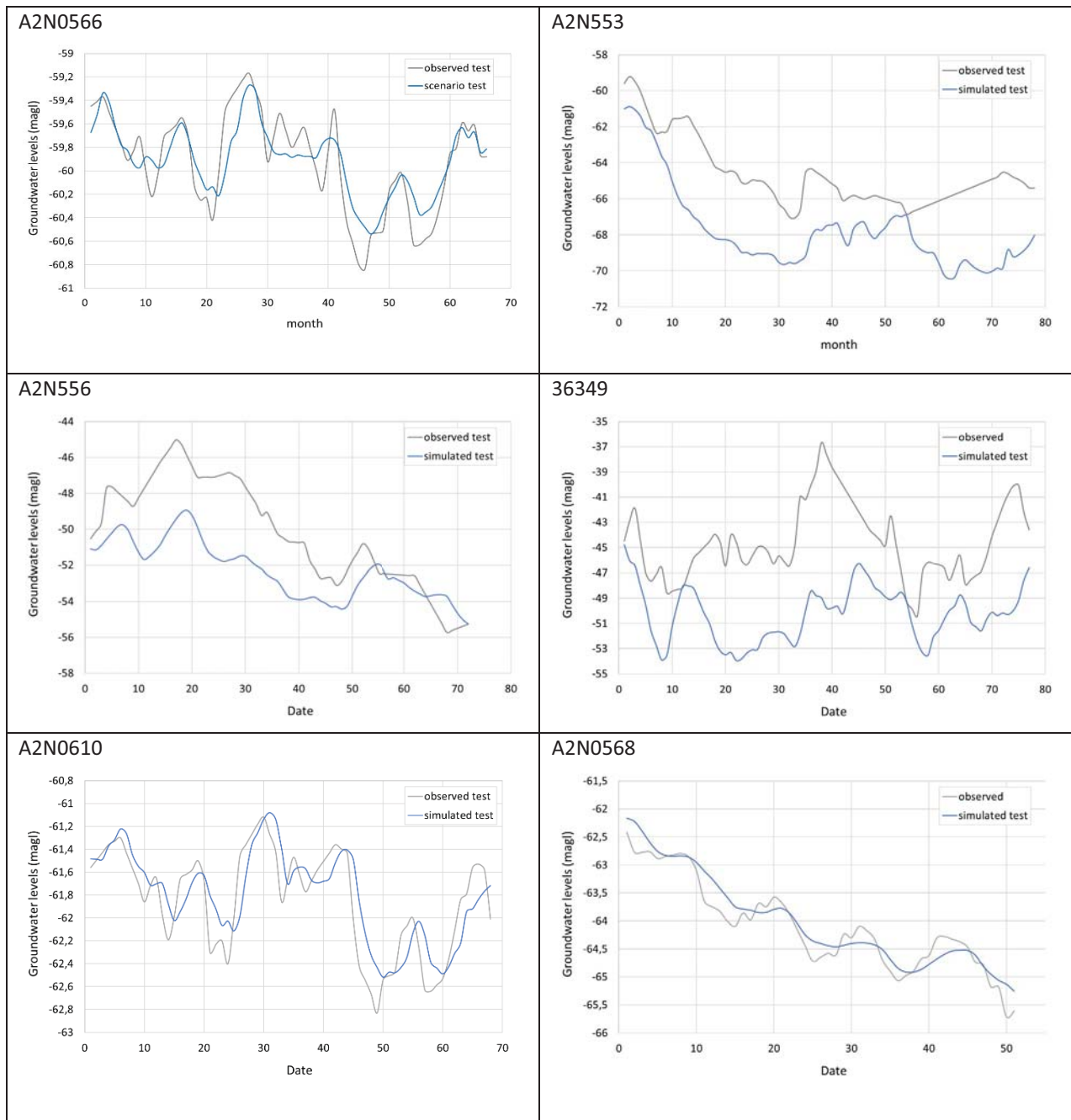
Figure 5-19 **Model prediction for scenario 1a, decrease rainfall peaks.**
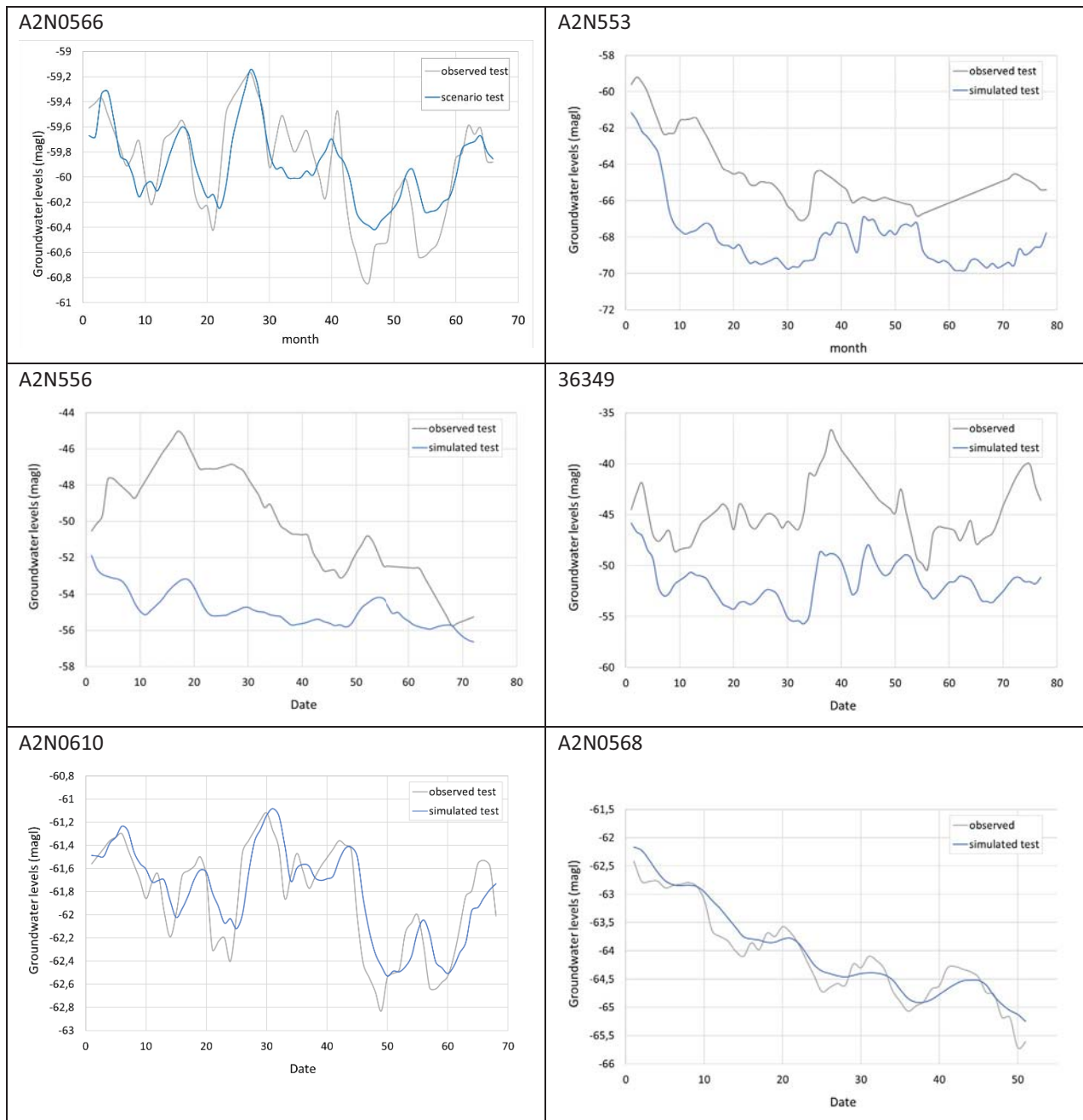
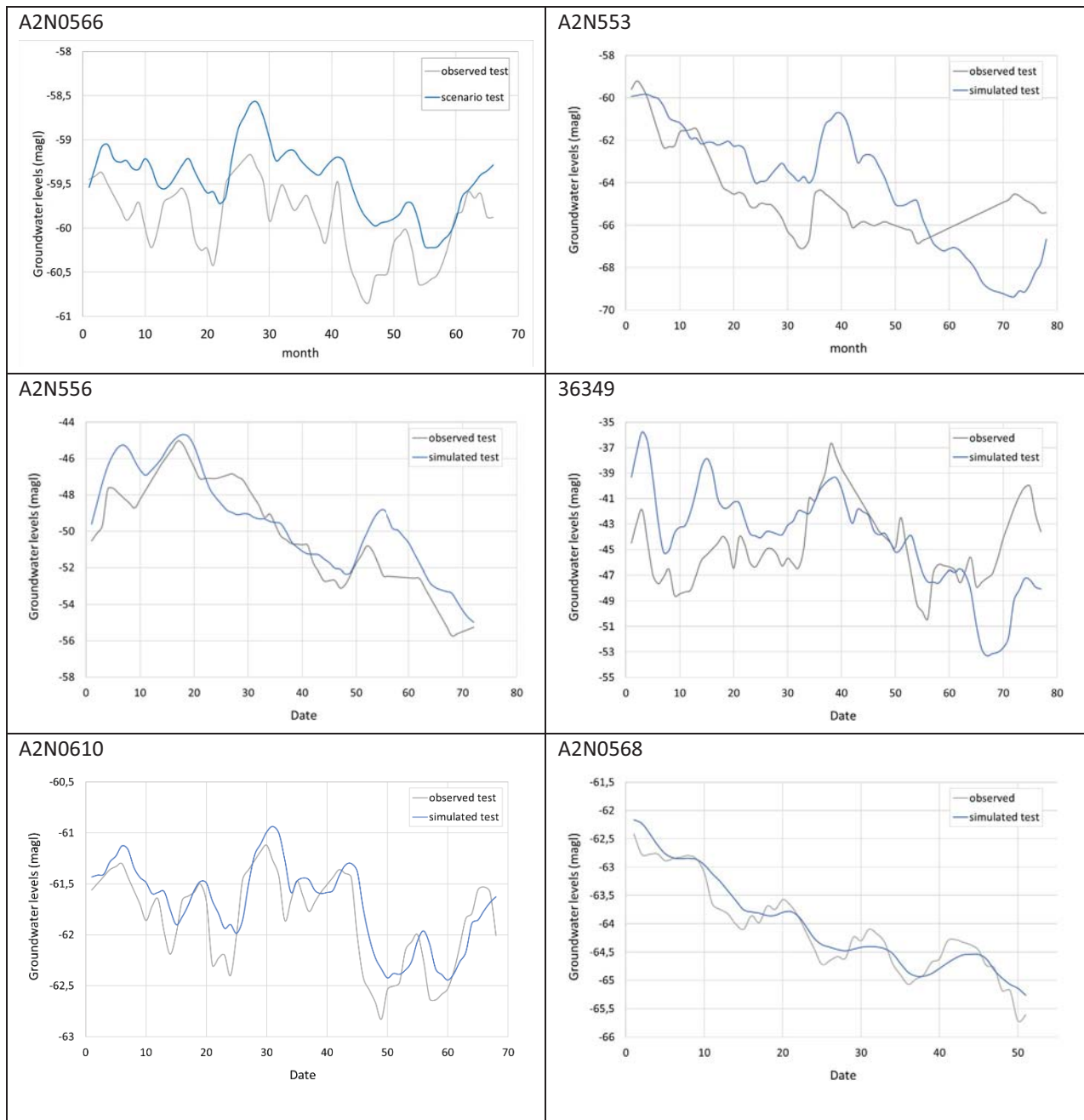**Figure 5-20   Model prediction for scenario 1c, halved monthly rainfall.**

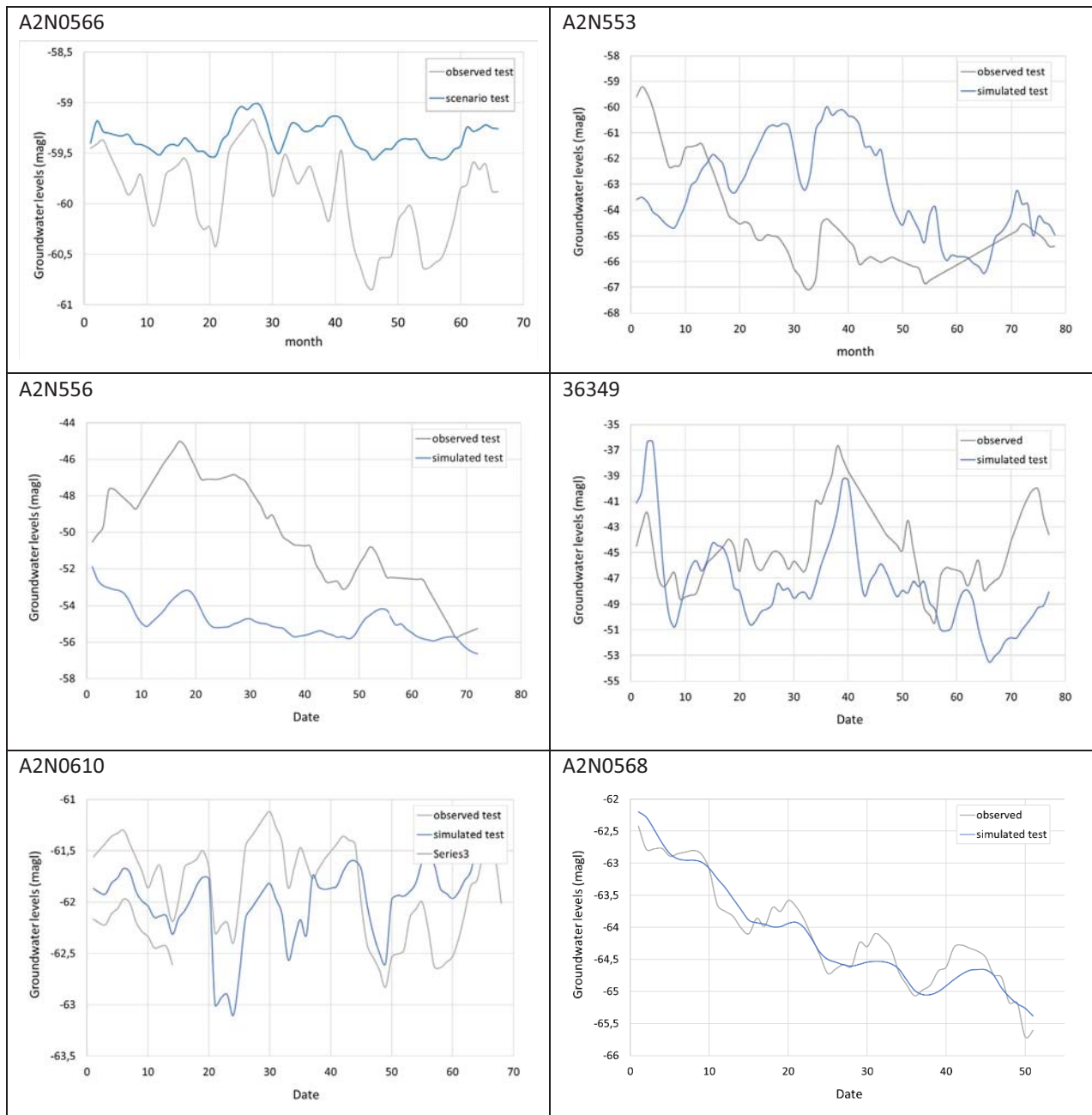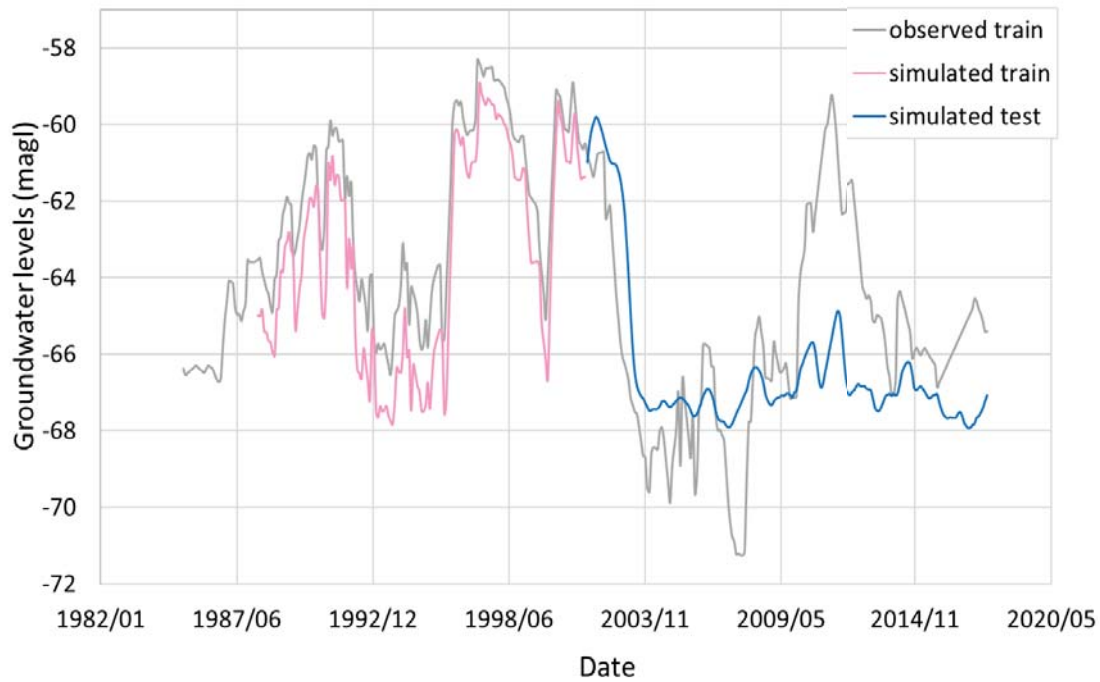**Figure 5-21   Model prediction for scenario 1b increase rainfall peaks.**

Figure 5-22 **Model prediction for scenario 2, increase groundwater abstraction.**
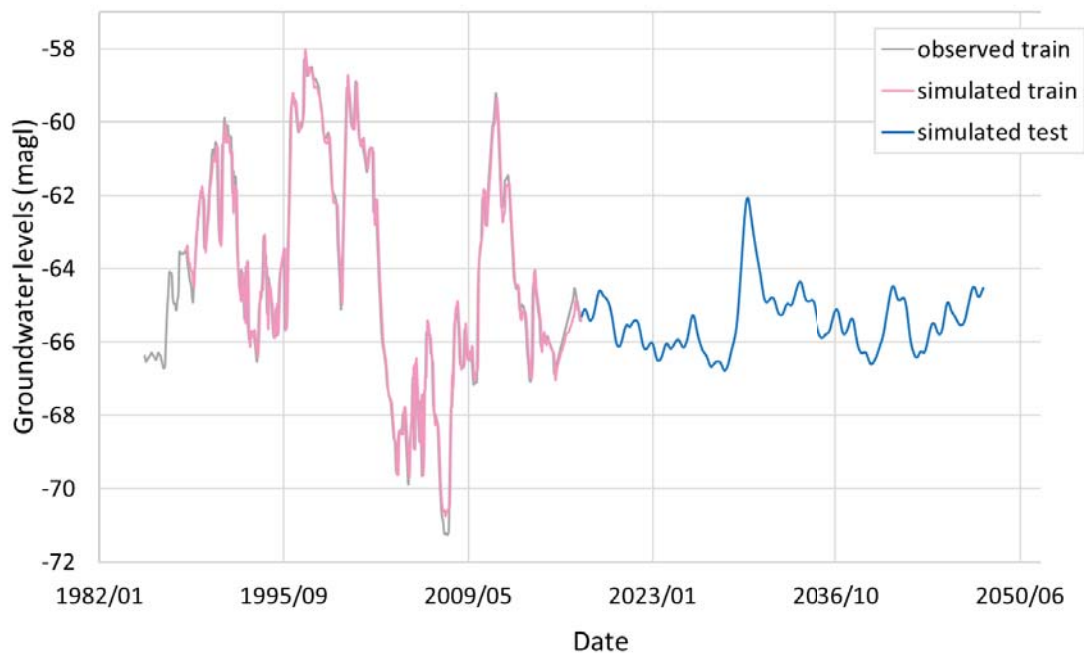
Figure 5-23 Model prediction for scenario 4a and b long term prediction at borehole A2N553.

6 GENERALISABILITY OF MODELS

The generalisability of a model is a measure of its successful application to data sets other than the one used for training and testing. Insufficient data hampered attempts to implement these ML models in Ramotswa TBA Ramotswa. Whilst there are 139 boreholes within the Ramotswa Dolomite area, the maximum length of data readings was 9 readings (compared to the ~240 used in this study). Therefore, to test generalisability, the LSTM was used to model 3 groundwater level datasets from the Steenkoppies compartment and the NNAR was used to model 3 groundwater level datasets from the Grootfontein compartment. In doing this, the generalisability of the two models was tested, albeit in very similar settings (similar rapid response to rainfall events), and with similar data structures (monthly datasets).

6.1 GENERALISABILITY OF THE LSTM TO MODEL

To test the generalizability of LSTM model the same model pipeline developed for the Grootfontein aquifer was tested with three datasets from the Steenkoppies compartment. The boreholes chosen show different characteristics as compared to the Grootfontein boreholes. For example, two of the chosen boreholes show a more muted groundwater level sequence with less sudden fluctuations as compared to the Grootfontein boreholes, and only one shows similar rapid fluctuations to those in the Grootfontein aquifer.

Table 6-1 shows a summary of the LSTM model results when compared to the NNAR model when modelling Steenkoppies data. The LSTM model produced less error rates and higher r-squared scores as compared to the NNAR model for 2 out of the 3 boreholes. The LSTM model performs better in the Grootfontein compartment than in Steenkoppies. Borehole A2N0553 had the lowest error scores and the highest r-squared of the tested boreholes. On average the 3 boreholes recorded R^2 values of 0.61 with the LSTM model and -0.06 with the NNAR model. Similarly, the boreholes also recorded MAE scores of 0.22 for the LSTM as compared to 2.13 for the NNAR model.

The NNAR model was able to model the peaks and troughs present in the Steenkoppies more closely than the LSTM model for boreholes 36349 and A2N0568 (Figure 6-1). During the training phase borehole A2N0553 performed similarly to the NNAR predictions (under-predicting or over-predicting the peaks and troughs) but during the test phase the LSTM model outperformed the NARX model as it modelled the observed more closely. Furthermore, the declining trend shown in borehole A2N0568 was unable to be modelled well by the LSTM model.

Table 6-1   Comparison between the performance of the LSTM and the NNAR to predict groundwater levels in the Steenkoppies aquifer (the NNAR model was developed for the Steenkoppies).

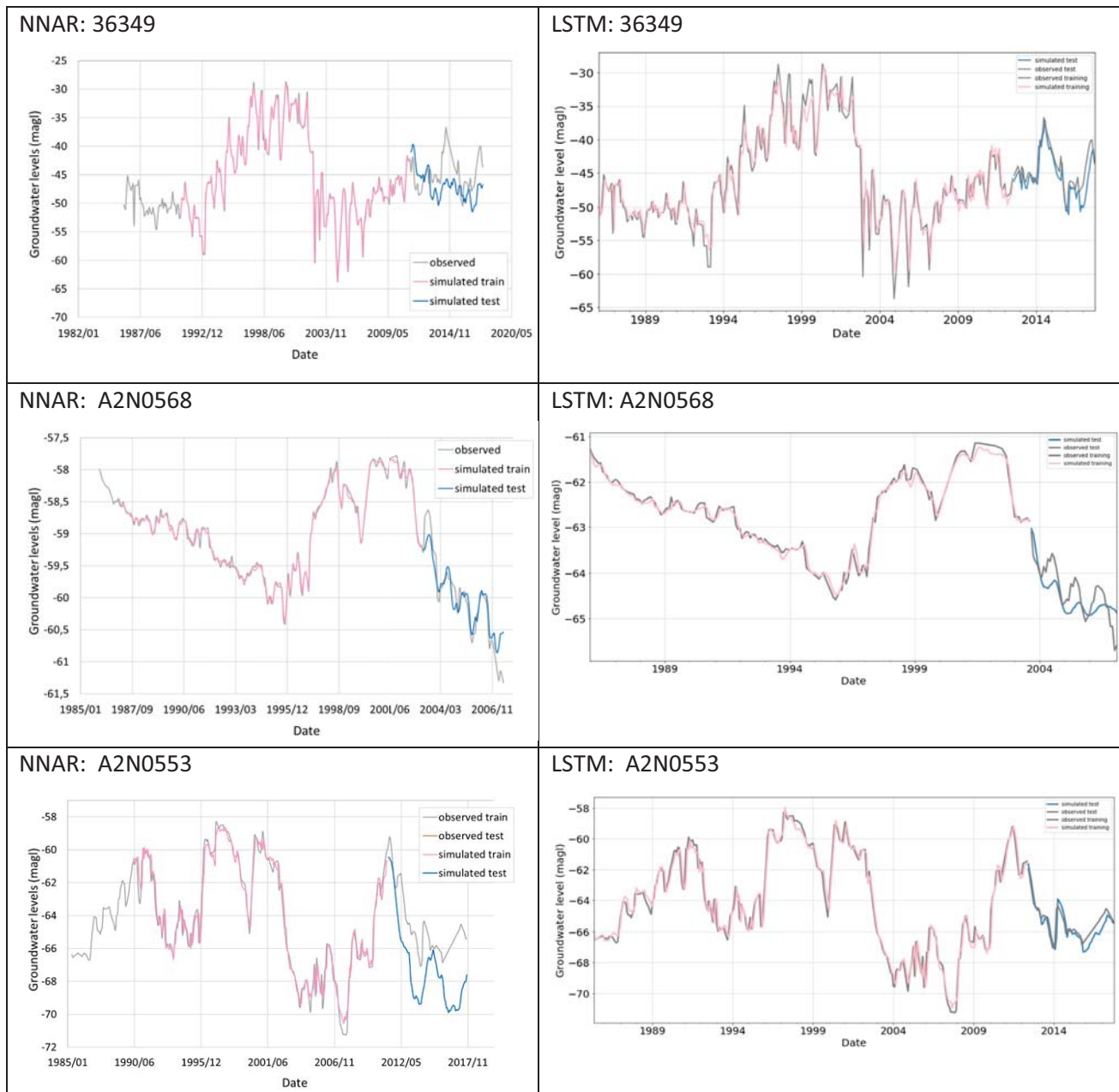|	BHID	NNAR R^2	LSTM R^2	NNAR MSE	LSTM MSE	NNAR RMSE	LSTM RMSE	NNAR MAE	LSTM MAE
1	36349	-1.19	0.63	23.04	0.08	4.75	0.28	4.03	0.21
2	A2N0568	0.81	0.45	0.09	0.14	0.30	0.37	0.23	0.31
3	A2N0553	0.20	0.74	6.50	0.03	2.54	0.17	2.14	0.14
	Average	-0.06	0.61	9.88	0.08	2.53	0.27	2.13	0.22

Figure 6-1 Comparison between the observed and predicted groundwater levels for 6 boreholes during training and testing phases.

6.2 GENERALISABILITY OF THE NNAR TO MODEL GROUNDWATER LEVELS IN THE GROOTFONTEIN COMPARTMENT

To test the generalizability of NNAR model, the same model pipeline used to model groundwater levels in the Steenkoppies compartment was used to model groundwater levels in the Grootfontein compartment. Table 6-2 compares the performance of the NNAR and the LSTM to predict groundwater levels from three boreholes in the Grootfontein compartment. The NNAR modelled groundwater levels from the Grootfontein aquifer with a greater error compared to the LSTM. Across all four performance metrics in Table 6-2, the LSTM outperforms the NNAR and is able to make better groundwater level predictions across all three boreholes. The R^2 score for groundwater level predictions made by the NNAR model is negative whereas for the LSTM the lowest R^2 recoded is 0.64 for borehole D4N0110. The error metrics (MSE, RMSE and MAE) calculated from the predictions made by the NNAR are five to ten times higher than those from the LSTM.

Figure 6-2 shows the predicted groundwater levels for 6 boreholes during training and testing phases. The predictions made by the NNAR (simulated test) do not match the observed groundwater levels and graphically confirm the poor predictions made by the model shown in Table 6-2. The NNAR is particularly unable to model the sharp peaks recorded in the groundwater level time series. This may be because the peaks are sharp and sporadic throughout the timeseries. The NNAR model requires the pattern, in this case the sharp peaks, to occur more frequently in the training data in order to "learn" and optimise the model hyper-parameters accordingly to predict the pattern in the test dataset. In all three boreholes in Figure 6-2 the groundwater level time series has infrequent sharp peaks in the training data that is not captured by the model. The predictions therefore do not capture the peaks either.

Table 6-2   Comparison between the performance of the NNAR and the LSTM to predict groundwater levels in the Grootfontein aquifer.

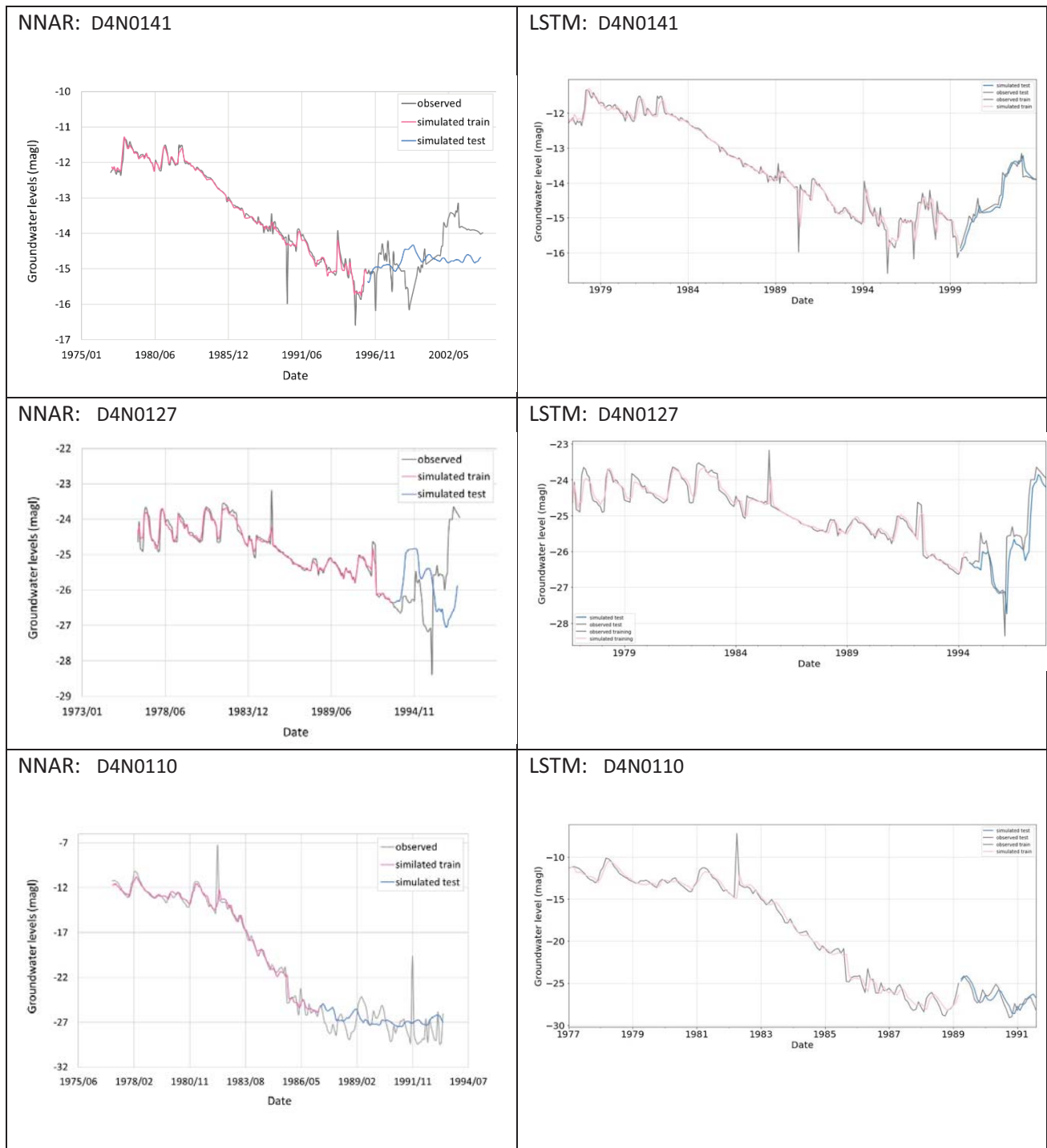|	BHID	NNAR R^2	LSTM R^2	NNAR MSE	LSTM MSE	NNAR RMSE	LSTM RMSE	NNAR MAE	LSTM MAE
1	D4N0110	-8.05	0.64	3.35	0.01	1.83	0.12	1.46	0.1
2	D4N0127	-5.41	0.73	1.83	0.46	1.35	0.68	0.88	0.47
3	D4N0141	-21.94	0.82	0.49	0.05	0.70	0.22	0.56	0.18

Figure 6-2 **Comparison between the observed and predicted groundwater levels for 6 boreholes during training and testing phases.**

6.3 CONCLUSION

The LSTM model was used to make predictions on the groundwater levels in the Steenkoppies compartment and the NNAR model was used to make groundwater level predictions in the Grootfontein compartment. Thus, each model developed was used to model a completely different dataset using the same model pipeline.

The LSTM model was able predict groundwater levels in both the Steenkoppies and Grootfontein compartment with higher accuracies than the NNAR model except for borehole A2N0568 in the Steenkoppies aquifer where the NNAR model

outperforms the LSTM. The NNAR model was unable to capture the underlying patterns of the boreholes present in the Grootfontein compartment. The reason for the LSTM to better generalise can only be related to datasets modelled and the internal functioning of the model.

The NNAR exhibited greater strength on a dataset with less peaks and troughs hence its performance on the Grootfontein boreholes was quite poor as the groundwater levels in the Grootfontein compartment show more sporadic sharp fluctuations. The model results of the NNAR in the Grootfontein compartment highlight the ability of the NNAR to model specific datasets and this could be due to the internal structure of how the model operates. On the other hand, the results of the LSTM model simulations show the ability of the model to capture the broad patterns of the Steenkoppies data so we can reasonably argue that the LSTM model is more likely to be generalizable as compared to the NNAR.

Both the LSTM and the NNAR models are sub-types of the RNN as both models use current and previous timestamps to make predictions, however the internal functioning and behaviour of the models is quite different (refer to section 3.3.4 and 3.3.5). Unlike the LSTM which stores previous time steps in the output gate, the NNAR's previous timestamps are given as another input to the model (Izady *et al.*, 2013). Hence, the internal memory of the LSTM allows the model to better "learn" and memorize the relationships in the datasets and use that to make more accurate predictions making the model more generalisable as compared to the NNAR.

7 SUMMARY AND WAY FORWARD

7.1 FRAMEWORK FOR A SUSTAINABLE GROUNDWATER STRATEGY

The approach to sustainable groundwater use that is spelled out by the Sustainable Groundwater Strategy requires quantification of the current and expected changes in the groundwater balance or flows in the basin related to planned or potential groundwater use, which must include changes in storage, inflows and discharges. The relationship between abstraction and aquifer fluxes must be established. These predictions of future impact of abstraction are relied upon to set sustainable management criteria, which requires close collaboration with stakeholders in the basin (including the regulators or responsible authority). In line with CDWR (2014), the overall aim of the SGS is to ensure groundwater is managed to avoid the following six undesirable results:

1. Chronic lowering of groundwater levels indicating a significant and unreasonable depletion of supply if continued over the planning and implementation horizon.
2. Depletions of interconnected surface water that have significant and unreasonable adverse impacts on beneficial uses of the surface water.
3. Significant and unreasonable degraded water quality, including the migration of contaminant plumes that impair water supplies.
4. Significant and unreasonable land subsidence that substantially interferes with surface land uses.
5. Significant and unreasonable seawater intrusion.
6. Significant and unreasonable reduction of groundwater storage.

Each of the six undesirable results are related to an equivalent sustainability indicator, which are the "six effects caused by groundwater conditions occurring throughout the basin that, when [the effects are] significant and unreasonable, represent undesirable results" (CDWR, 2017, pg. 26). The six indicators are:

1. Lowering of groundwater levels
2. Surface water depletion
3. Degraded groundwater quality
4. Land subsidence
5. Seawater intrusion
6. Reduction of storage

The framework for the SGS outlines the steps required to establish the expected impacts of abstraction, set thresholds for the indicators, establish the necessary monitoring protocol to ensure the thresholds are not exceeded, and implement and update the strategy. The authority mandated as responsible for management of water resources in the area would be the appropriate body to implement the strategy, i.e., the Department of Water and Sanitation (or the Catchment Management Agencies) for South Africa, or the Department of Water Affairs in Botswana.

7.2 APPLICATION OF MACHINE LEARNING MODELS IN HYDROGEOLOGY

The machine learning model aims and research questions (section 3.2) were explored through applying and testing various models and approaches. Findings on the applicability and shortcomings of various models only became clear through attempting to implement them. The models that were applicable to the task at hand and yielded useful results are presented in this report.

Numerical models are used in groundwater resources assessments to predict groundwater conditions including groundwater levels and aquifer – wide fluxes such as discharge to surface water, for adjusted stresses such as increased

abstraction, reduced recharge. The research attempted to see if ML models can do the same (see ML model aims, section 3.2).

Ability to predict groundwater level

- Contrary to requiring a variety of input data that is often unknown (or estimated) for groundwater level prediction, such hydraulic properties and aquifer geometry, the developed models use only measured data for the model input variables. Both models (NNAR and LSTM) were able to learn the dependencies between the sequential data series and demonstrated reliable performance in groundwater level forecasting in both Steenkoppies (see Table 5-8) and the Grootfontein (see Table 4-7) dolomitic compartments.

- The results presented therefore demonstrate that data driven models can successfully predict groundwater level values. However, testing showed this is only the case if the training dataset is representative of the test dataset Patterns in the test set which deviate from patterns previously seen in the training set decrease the overall performance of the model. For example, the models are seen to perform better on the peaks and troughs only when several peaks and troughs were previously seen during model training. This has implications suggesting the models would not be able to accurately replicate the groundwater level response to novel changes or system shocks, such as several peak rainfall events that are out of range of the previous data. Furthermore, aspects that are not considered in the model, such as land use changes that impact on recharge rates and impact the relationship between rainfall and groundwater level would not be replicated.

- The predictions for groundwater levels (i.e., the test datasets) are relatively successful when training datasets are longer, demonstrated particularly by the results to scenario 4a. The more data you feed the RNN the better the performance because the longer training periods allow for greater variability to be seen during the training phase which make for better predictions in the test phase. However, if the input data represents the target variable well (i.e., there is a very strong correlation between an input and target variable) then the input variables could have more influence in the accuracy of predictions, than the quantity of the input data (as observed in section 4.2.4).

- The purpose of groundwater model development is generally to test the response of the groundwater system to stresses that have not yet occurred, i.e., changing abstraction of recharge. To test the model's capabilities in predicting groundwater levels for different stresses, four scenarios were chosen, and recharge and abstraction adjusted. The results revealed that the models were able to pick up the influence of a certain change in a variable, by generating predicted water level values that differed from the base case. For example, halving recharge in both models generates a decline in groundwater levels. The LSTM model was able to show clearer declines or increases in groundwater levels (based on scenario) than the NNAR model.

- Mutual information was calculated between each feature and the target variables (see section 4.3.1 and 5.3.1). In both study areas, the discharge dataset represented a more important variable than rainfall data in predicting groundwater levels. The strong relationship between discharge and groundwater levels is the reason why the models are able to perform in the long-term forecasts without deviating significantly from the observed groundwater levels, even with a significantly smaller training dataset (see Figure 4-18a). In cases where discharge is not a measured or known parameter, groundwater level prediction would rely on rainfall only then model performance could decrease. Furthermore, for scenario modelling, the future rainfall can be estimated, but groundwater discharge is a product of groundwater levels and so would not be known in a future scenario test. So, whilst these models can predict scenarios into the future, those demonstrated here (scenario 4) used averaged discharge which un-naturally controls the result. Taking discharge out of the scenarios as a parameter and essentially driving the prediction based on rainfall and abstraction only would be appropriate and reduce the accuracy.

- The implication is also that using ML models to predict groundwater levels may be less feasible in an aquifer setting where discharge data is not available, and the relationship between groundwater levels and rainfall is

weaker or much more muted (i.e., confined aquifer settings, or aquifers in arid areas with highly episodic recharge).

- The models developed have demonstrated that it is certainly possible to accurately fill short-term groundwater level data gaps at least in dolomitic terrains where spring discharge data closely matches the groundwater levels and in areas where there is a strong correlation with rainfall. The timescale over which this would be feasible depends on the amount of other data that is available to train the model; so generalisations are not possible. In these models with ~20-30 years of training data and a strong correlation with both discharge and rainfall data gaps of up to 5 years can confidently be filled.

Generating aquifer fluxes and the translation from borehole to aquifer scale

The research aims require assessment of aquifer-scale processes (widespread causing a gradual and consistent lowering of groundwater levels in monitoring boreholes across the basin) and therefore cumulative groundwater abstraction across the whole aquifer was incorporated in the models. The models train based on a pattern between datasets, and the NNAR model was unable to "see" a pattern between the gradually increasing groundwater use, and the groundwater levels, to the extent that adjusting abstraction rates in the scenario tests generated inconsistent results in the groundwater levels. Furthermore, to be able to quantify the impacts of changed abstraction rates once dynamic equilibrium is reached, the model would need to provide the projected groundwater level in response to one constant abstraction rate (higher or lower). This type of calculation does not rely on pattern recognition but quantification of physical process, and it is unlikely that ML models could ever be able to improve upon the numerical modelling tool that is available for this.

The research aims require assessment of aquifer-scale processes, however, point data (from boreholes) must be relied on to provide indications of aquifer-scale processes. Input data for the ML models was a mixture of borehole-scale data (groundwater levels) and aquifer-scale data (rainfall assumed to be applicable across the area, abstraction across the entire compartment; discharge from the whole aquifer). A model was developed for each borehole, because the response to rainfall is different at each borehole and model trains on the data at that borehole and the "hidden layer" or "black box" of the model will memorise a pattern specific to that borehole. In cases where there is sufficient spread of boreholes across an aquifer, the groundwater level response at every borehole could be predicted for future scenarios, and the predictions interpolated to generate a piezometric surface, and used to calculate aquifer storage, thus translating from models of borehole scale to aquifer scale interpretations.

The ML models can predict anything for which there is a pattern and so if training data is directly available for aquifer fluxes, these can be predicted regardless of whether the data represents the borehole verses aquifer scale. To test this, the modelling inputs and outputs were reversed, and discharge predicted based on groundwater levels at several boreholes in both the NNAR and LSTM models. This test demonstrated it is equally feasible to predict discharge rates (refer to Kanyama et al, 2020 for a description of this test for the Grootfontein case). To predict future aquifer discharge based on adjusted rainfall (or abstraction rates) it would be feasible to use adjusted rainfall to predict the impact on future groundwater levels (with discharge removed from the predictive model), and then predict discharge based on the modelled future groundwater levels. This two-step approach allows for the fact that future groundwater levels *and* future discharge would change based on that input rainfall / use changing. However, whilst this prediction of aquifer scale processes is methodologically feasible, the ability of the ML models to predict into the future, and to predict the correct magnitude is questionable.

The Grootfontein case study results do detect the impact of changing groundwater use in future scenarios on groundwater levels (which could in turn be used to predict future discharge for those groundwater levels), however the Steenkoppies case study was (more understandably) unable to detect a strong correlation between use and groundwater levels, and as such changing use in the scenario predictions had little impact. It is seemingly unlikely that ML models can

be used to quantify the relationship between increased abstraction and aquifer fluxes demonstrated by the graph in Box 2-2. It would rely on a relationship being detected between a curve of increasing groundwater use (or even sustained use at one abstraction rate) and groundwater levels and aquifer fluxes, already detected in the training dataset; it is therefore an inapplicable approach for under-utilised aquifers or where abstraction has not yet commenced. It cannot therefore be used for example in mining greenfield sites where numerical modelling is also relied upon to give the same outputs.

It is also extremely difficult to determine (without a numerical model for verification) whether the magnitude of the prediction generated by the hidden layer is physically accurate. The magnitude is based on the training dataset, which shows (for the Grootfontein case) a range of groundwater levels and discharge rates for the historically increasing groundwater use. We do not know however whether the aquifer is at dynamic equilibrium (in groundwater levels and discharge) to the current abstraction rates or that shown by the data. Entering this data to a numerical model, and then running it forward for another 50 years, may show continued groundwater level and discharge decline for no future increase in abstraction, simply due to some of the current abstraction coming from storage. These dynamics cannot be accommodated in ML models.

One advantage of the approach in ML models is that recharge does not need to be quantified. Rainfall is simply used, and the hidden layer (black box model) can work out how the groundwater level responds to rainfall. For future scenarios we therefore do not need future recharge, only future rainfall (perhaps easier to generate, from climate models). This only applies if there is no system change / system shock such that groundwater levels start to respond differently to recharge i.e., land use changes significantly.

Hydrogeological insights

Due to the low confidence associated with the forward predictions for Grootfontein (on varying rainfall and groundwater use) and for Steenkoppies (for varying rainfall; groundwater use predictions are not feasible), it is not possible to provide new insights for hydrogeological functioning of these two case studies that can inform sustainable groundwater management in the areas. The analysis presented here shows a long-term decline in rainfall in the areas over the same time period as the decline in groundwater levels yet in both areas the decline in groundwater levels has been attributed to groundwater use. Seyler et al 2016 quantify the relationship between abstraction rate and reduced discharge at Steenkoppies, for scenarios at dynamic equilibrium (hence for long term averaged recharge), however previous work has not quantified the degree to which groundwater use verses reduced recharge has contributed to the groundwater level decline over the previous ~ 30 years (feasible with a transient numerical model scenarios). It is important for the management of these two aquifers that the relationship between groundwater use and aquifer fluxes be quantified and the relative influence of rainfall and abstraction on current groundwater levels be understood.

7.3 WAY FORWARD

Recommendations for sustainable groundwater use – implementing the SGS

The SGS represents the idealised best-practice approach to achieving sustainable groundwater management that this project recommends is implemented for TBAs. It is acknowledged that full implementation of the SGS would require legislation to be amended, and more detailed guidance documents. Therefore, similarly to California, whilst the actions in the SGS are necessary to achieve sustainable groundwater use, and could be particularly powerful for TBAs, they are unlikely to be fulfilled in their entirety until legislation is amended.

Recommendations for data management in TBAs

The partnership had initially hoped to test whether big data approaches could be used within a decision-support system for improving groundwater management in TBAs. It is not possible to automate the analyses hydrogeologists need to do

in order to provide the technical information on which to base decisions controlling sustainable groundwater use. If the SGS was implemented in its entirety in a basin, then subsequent to this, an automated system would be feasible that ingests real time monitoring data and flags to a manager when the threshold for a set sustainability indicator is breached. But it is not possible to bypass the basic requirement of an automated system; such a system relies on a hydrogeological analysis of predicted impact, and on thresholds being established, and relies entirely on a monitoring network being in place (i.e., boreholes, stream gauges, weather stations).

The programme has piloted new approaches (theme 2, 3 and 4 have tested machine learning methods in different applications) and generated prototypes (theme 1 contributed to updating the IGRAC GIP database). Known groundwater data-related shortcomings have again been highlighted through these projects, including lack of groundwater monitoring infrastructure; the lack of access to data that is available and the associated lack of mandate for sharing data; the lack of data for actual groundwater use (rather than licensed use); and the lack of a single platform for data. The research illustrates that some data challenges can be overcome i.e., data patching for missing groundwater levels, and the use of coding to automate processing data from differing formats. But, to really further develop the use of big data in SADC, the fundamental data generation and data accessibility problem must be resolved. A centralised groundwater database accessible to all is required, ideally to include all SADC countries. This is required before any automated decision-support system can be considered.

Recommendations for further testing of machine learning applications for modelling groundwater
The models developed have been optimised for the case studies applied. The generalisability of the models was tested to demonstrate that these model pipelines can be applied in different settings. It is worth testing the application of ML models in a greater number of other settings, to verify some of the model implications listed in section 7:

- Similarly structured data (and also regional behaviour) from a different aquifer system perhaps with a much more muted response to rainfall, such as aquifers in the Karoo, or confined aquifers from elsewhere.
- Data structured very differently such as 3-hourly groundwater levels from a pumped borehole, with abstraction at that same borehole. Due to the time scale the groundwater level fluctuation will be dominated by the pumping rates (which could be off or on at a relatively constant rate), with little influence from rainfall / recharge at the 3-hourly timescale, and perhaps no discharge measurement depending on the aquifer case study selected. Whether machine learning models can recognise the pattern between the pump on/off signal, and this be enough to generate a typical pumped response in the groundwater level would be a worthwhile test.

Recommendations for implementing ML models in groundwater
Based on the ability of the models to fill data gaps, it is recommended that:

- ML modelling is adopted as an approach by the groundwater community, with close guidance from computer scientists, to improve groundwater datasets
- DWS initiates a project to fill data gaps in the NGA and HYDSTRA databases, with this data earmarked as "modelled". More complete datasets would significantly benefit other researchers, and other analyses such as numerical modelling.

It is also recommended that ML modelling be considered for the following uses in groundwater studies:

- Understanding causal impacts with Bayesian models (Ajoodha and Rosman, 2020)
- Predicting the impact of rainfall variability on groundwater levels at individual boreholes. In situations where there are sufficient numbers of boreholes across an aquifer, developing a model per borehole (if each has sufficient historical data), and predicting the impact per borehole can allow this data to be interpolated and the impact on the whole aquifer groundwater table derived.

- Projecting response in a pumped borehole to changes in the pumping regime (although accuracy and magnitude of predicted response is uncertain where the change in pumped regime may be significantly different from something seen previously)

Recommendations for developing capacity in ML models for groundwater science

Expertise in machine learning modelling (specifically RNNs) for time series groundwater data is a niche area in which there is currently extremely limited expertise in South Africa. Applying ML for hydrogeology is truly multi-disciplinary, and wrong turns during the research process highlighted that there is a risk that hydrogeologists implement ML models without the necessary computer science input and apply models outside of their real purpose. Vice versa is also true with a risk of computer scientists mis-interpreting the physical meaning of the output from machine learning models.

8   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.

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.

Aller. Linda. Bennett. Truman. Lehr. J.H.. and Petty. R.J. 1985. DRASTIC—a standardized system for evaluating ground water pollution potential using hydrogeologic settings: U.S. Environmental Protection Agency. Robert S. Kerr Environmental Research Laboratory. Office of Research and Development. EPA/600/2–85/018. 163 p.

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

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.

Barnard, H.C. (1997). Geohydrological evaluation of the Maloney's Eye catchment area. M.Sc. Dissertation University of the Free State, Bloemfontein.

Bennett. R. R., R.H. Brown., H. H. Cooper. Jr., W. J. Drescher, J. G. Ferris, A. I. Johnson, S. W. Lohman, C. L. McGuinness, A. M. Piper, M. I. Rorabaugh, R. W. Stallman, and C. V. Theis. 1988 Definitions of Selected Ground-Water Terms. USGS Water-Supply Paper 1988. 21 pp.

Bredehoeft. J.D. and T.J. Durbin. 2009. Ground water development—the time to full capture problem. *Ground Water* Volume 47. no. 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 pp 51–57.

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.

California State. 2014 California Water Code sections in Division 6. Part 2.74. Sustainable Groundwater Management Act

CDWR. 2014. Title 23 of the California Code of Regulations. Division 2. Chapter 1.5. and Subchapter 2 Groundwater Sustainability Plans.

CDWR. 2016a. Groundwater Sustainability Plan (GSP) Annotated Outline. Guidance Document for the Sustainable Management of Groundwater. FINAL. December 2016.

CDWR. 2016b. Best Management Practices for Sustainable Management of Groundwater: Hydrogeologic conceptual model. December 2016. California Department of Water Resources – Sustainable Groundwater Management Program

CDWR. 2017. Best Management Practices for Sustainable Management of Groundwater: Sustainable Management Criteria. Draft. November 2017. California Department of Water Resources – Sustainable Groundwater Management Program

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*.*

Cobbing, J. E., Eales, K., Gibson, J., Lenkoe, K., & Cobbing, B. L. 2015. Operation and Maintenance (O&M) and the perceived unreliability of domestic groundwater supplies in South Africa. South African Journal of Geology. 118. 17- 32.

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

Cobbing, J. 2018. An updated water balance for the Grootfontein aquifer near Mahikeng, *Water SA* Vol. 44 No. 1 January 2018.

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.

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 Affairs. 2009. Directorate: Water Resource Planning Systems. Sub-directorate: Integrated Hydrological Planning. Geohydrology Guideline Development: Implementation of Dolomite Guideline – Phase 1. ACTIVITY 19 & 28: Desktop development of a Dolomite hydrogeological compartment map and explanation booklet (Report). by Martin Holland and Frans Wiegmans. November 2009.

Department of Water and Sanitation. South Africa. July 2017. Determination of Water Resource Classes and Associated Resource Quality Objectives in the Berg Catchment: Quantification of the Ecological Water Requirements and Changes in Ecosystem Goods. Services and Attributes. Project Number WP10987. DWS Report NO: RDM/WMA9/00/CON/CLA/0217. Prepared by Aurecon South Africa (Pty) Ltd in sub-consultancy association with Southern Waters Ecological Research and Consulting. Anchor Environmental and Delta-H Water Systems Modelling

Devlin. J. and M. Sophocleous. 2005. The persistence of the water budget myth and its relationship to sustainability. *Hydrogeology Journal*. Volume 13: 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.

Faunt, C.C., Sneed, M., Traum, J., and Brandt, J.T. 2016. Water availability and land subsidence in the Central Valley. California. USA: Hydrogeology Journal. v. 24. p. 675–684. DOI 10.1007/s10040-015-1339-x. https://link.springer.com/article/10.1007/s10040-015-1339-x

Frankfurt. Germany: Institute of Physical Geography University of Frankfurt (Main).

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.

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.

GWC. 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 (2005). Hydrogeology principles and practice. Published in 2005 by Blackwell Science Ltd. 389pp

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

Hyndman, R.J. and Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.

IGRAC (International Groundwater Resources Assessment Centre). UNESCO-IHP (UNESCO International Hydrological Programme). 2015. Transboundary Aquifers of the World [map]. Edition 2015. Scale 1: 50 000 000. Delft. Netherlands: IGRAC. 2015.

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.

Kanyama, Y., Ajoodha, R., Seyler, H., Tutu, H., (2021) Application of Artificial Neural Networks to Forecast River Discharge Rates: A Case Study of the Grootfontein Aquifer. Proceedings of the 7th Asia-Pacific Conference on Computer Science and Data Engineering, Institute of Electrical and Electronics Engineers Conference, Xplore digital library.

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. and J. Bredehoeft. 2019. GROUNDWATER RESOURCE DEVELOPMENT: EFFECTS AND SUSTAINABILITY. In The Groundwater Project. J. Cherry. *Currently in press.*

Koth. K.R. and Long. A.J. 2012. Microgravity methods for characterization of groundwater-storage changes and aquifer properties in the karstic Madison aquifer in the Black Hills of South Dakota. 2009–12: U.S. Geological Survey Scientific Investigations Report 2012–5158. 22 p.

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. (2019). *R for data science: import, tidy, transform, visualize, and model data*. CRC Press.

Le Maitre. D.C. Seyler. H. Holland. M. Smith-Adao. L. Nel. J.L. Maherry. A. and Witthüser. K. (2018)

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.

Le Maitre. D.C. Seyler. H. Holland. M. Smith-Adao. L. Nel. J.L. Maherry. A. and Witthüser. K. (2018). Identification. Delineation and Importance of the Strategic Water Source Areas of South Africa. Lesotho and Swaziland for Surface Water and Groundwater. Final Integrated Report on Project K5/2431. Water Research Commission Pretoria.

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

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.

McGuire. V.L. 2017. Water-level and recoverable water in storage changes. High Plains aquifer. predevelopment to 2015 and 2013–15: U.S. Geological Survey Scientific Investigations Report 2017–5040. 14 p. https://doi.org/10.3133/sir20175040.

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

Ngoie. S. 2017. Development of Artificial Neural Network for Mine Dewatering. PhD Thesis. submitted to Institute of Groundwater Studies. University of Free State.

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.

Pietersen, K. & Beekman, H. 2016. Position Paper: Groundwater Management in the Southern African Development Community. Prepared for the SADC-GMI.

Pietersen, K. Kellgren, N. Roos, M. & Chevallier, L. 2010. Explanatory Brochure for the South African Development Community (SADC) Hydrogeological Map & Atlas. Southern African Development Community.

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.

Riedel, C. & Döll, P. 2016. Global-scale modeling and quantification of indicators for assessing transboundary aquifers.

Riemann. K. 2013. 'Groundwater Reserve' – a critical review. Paper presented at the 13th Biennial Groundwater Conference of the Geological Society of South Africa. 17th-19th September 2013. Durban. South Africa.

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

SADC-GMI. 2017. Requests for Expressions of Interest in Consulting Services – Firms Selection Consultancy for Follow-up Impact Evaluation of Pilot Projects Implemented under the Sustainable Groundwater and Drought Management Project in the Limpopo Basin. Issuance Date: 26 June 2017

SADC-GMI. 2018a. Consultancy for capacity needs assessment to determine priority challenges for capacity development initiatives in Member States.

SADC-GMI. 2018b. Project overview SADC-GMI project CS2017/05: Capacity Building on Groundwater Data Collection and Management in SADC Member States (SADC-Groundwater DataCoM)

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*.*

Seyler, H. Witthüser, K. and Holland, M. 2016. The Capture Principle Approach to Sustainable Groundwater Use. Water Research Commission Report No. 2311/1/17. ISBN: 978-1-4312-0893-7. November 2016.

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 riverbank 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.

Sophocleous. M. 2000. From safe yield to sustainable development of water resources – the Kansas experience. *Journal of Hydrology. Volume* 235: 27–43

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.

UCS (Union of Concerned Scientists). 2017. Getting involved in Groundwater. a Guide to California's Groundwater Sustainability Plans. Available at www.ucsusa.org/CAgroundwatertoolkit

UNESCO 2016. Stampriet Transboundary Aquifer System Assessment. Paris. France: International Hydrological Programme of the United Nations Educational. Scientific and Cultural Organization.

Valder. J. G. C. Delzer. J. Carter. B. Smith. and D. Smith. Construction of a Groundwater-Flow Model for the Big Sioux Aquifer Using Airborne Electromagnetic Methods. Sioux Falls. South Dakota. USGS Fact Sheet 2016-3075. September 2016

Valseth. K.J. Delzer. G.C. and Price. C.V. 2018. Delineation of the hydrogeologic framework of the Big Sioux aquifer near Sioux Falls. South Dakota. using airborne electromagnetic data: U.S. Geological Survey Scientific Investigations Map 3393. 2 sheets. https://doi.org/10.3133/sim3393

VILLHOLTH. K. G. TØTTRUP. C. STENDEL. M. & MAHERRY. A. 2013. Integrated mapping of groundwater drought risk in the Southern African Development Community (SADC) region. *Hydrogeology Journal.* 21**.** 863-885.

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*.*

Woodford. 2005. Assessment of the response of the Langebaan Road Aquifer System to a 3-month Shutdown of the Municipal Wellfield. SRK Report No. 335975. Prepared by SRK for the West Coast District Municipality. February 2005.

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.

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.
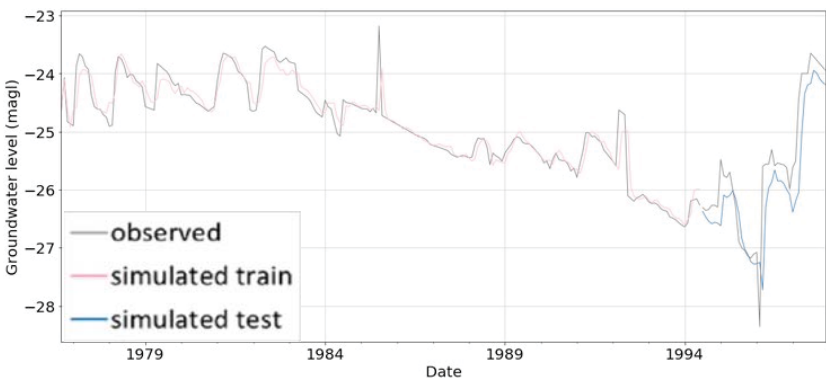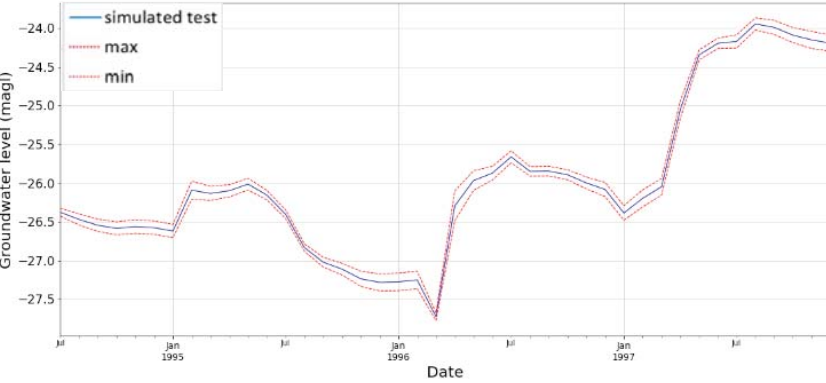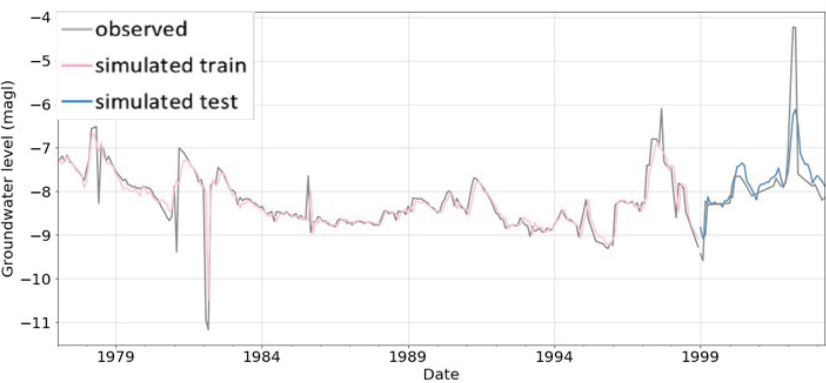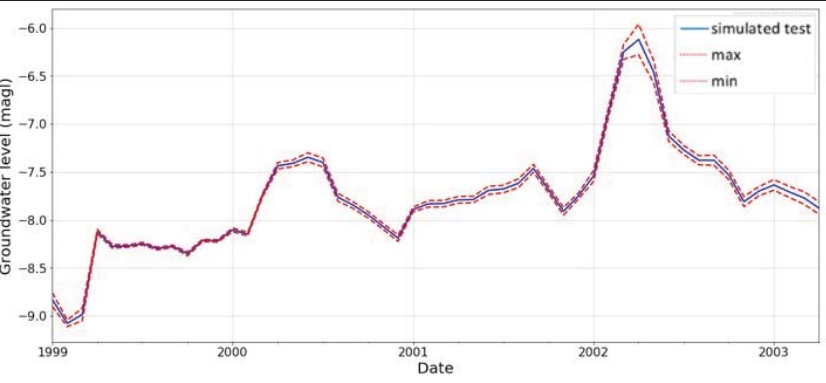
Zhang, G. and Hu, M.Y. (1998). Neural network forecasting of the British pound/US dollar exchange rate. *Omega*, *26*(4), 495-506.
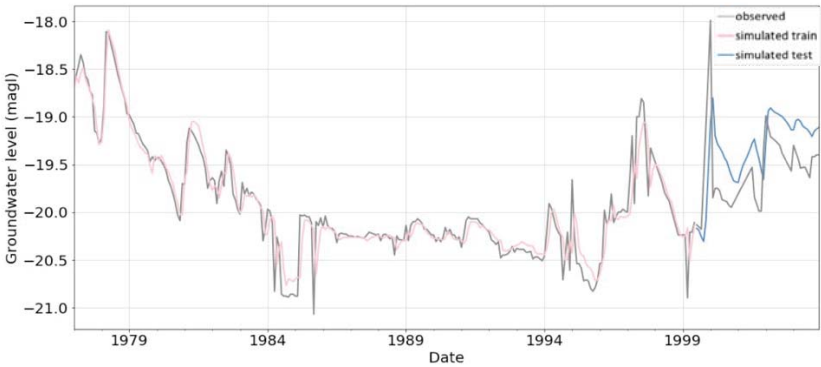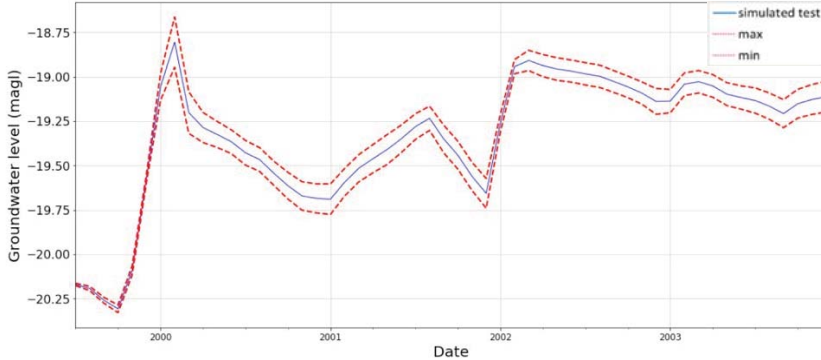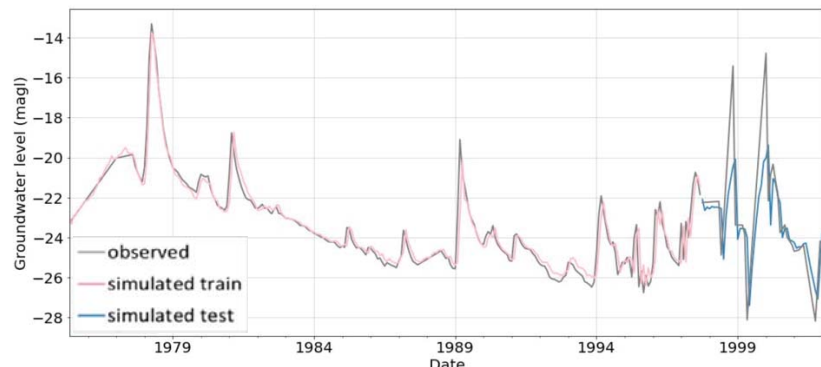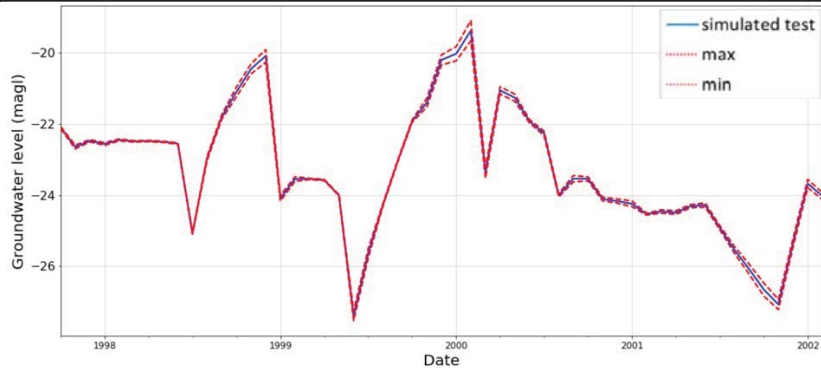
Zhu. Z. and C. Woodstock. Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment Volume 144. 25 March 2014. Pages 152-171. https://doi.org/10.1016/j.rse.2014.01.011
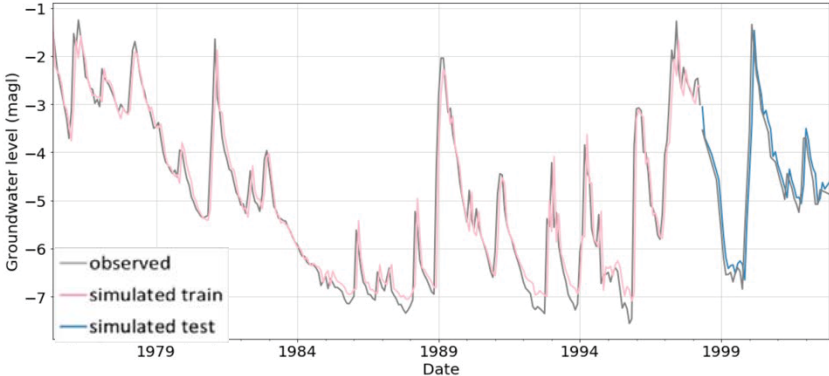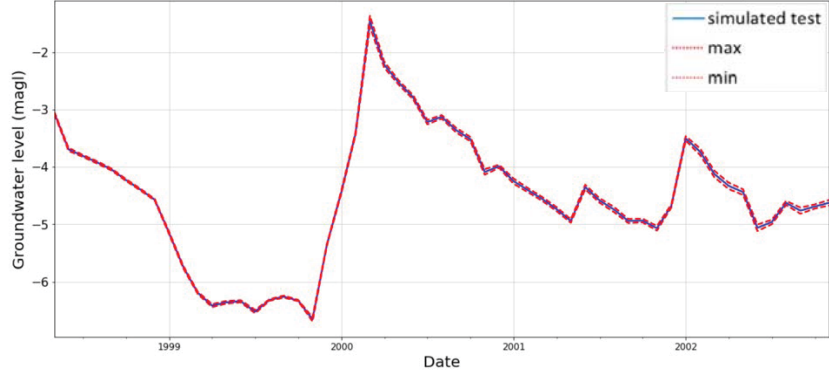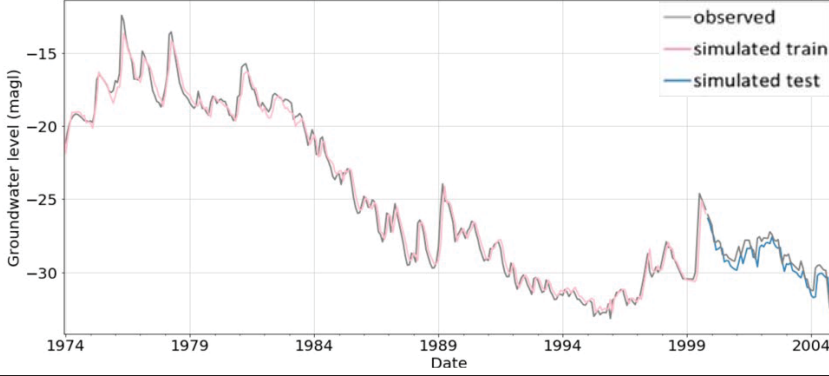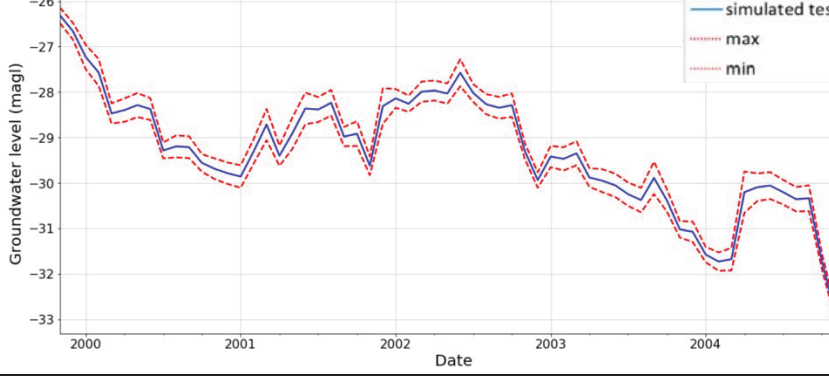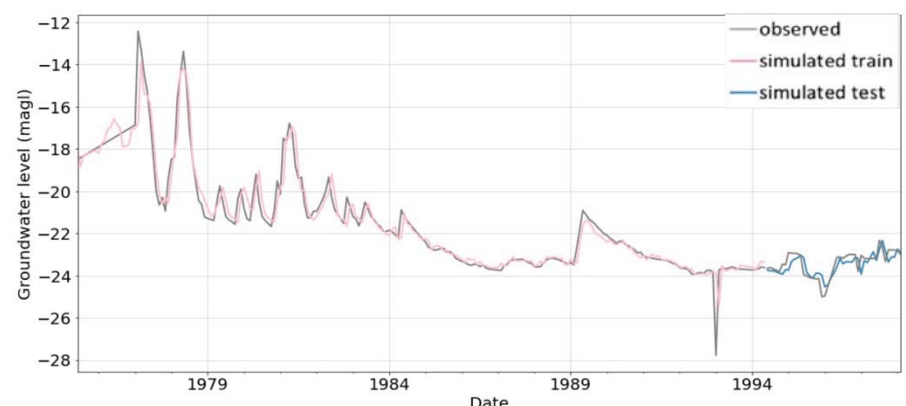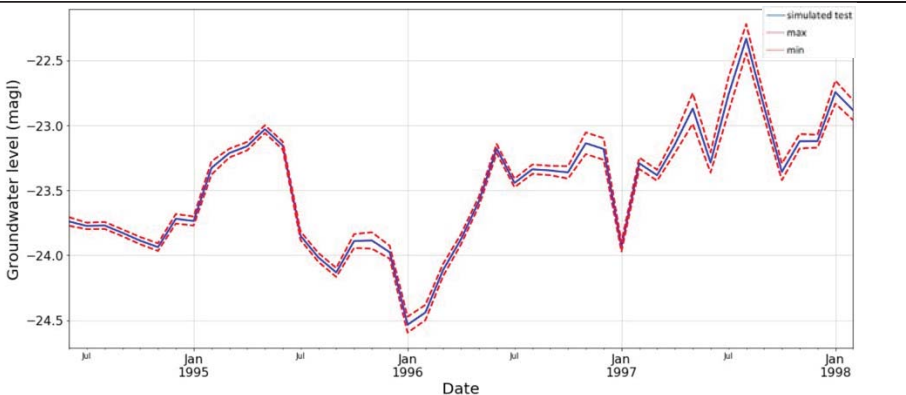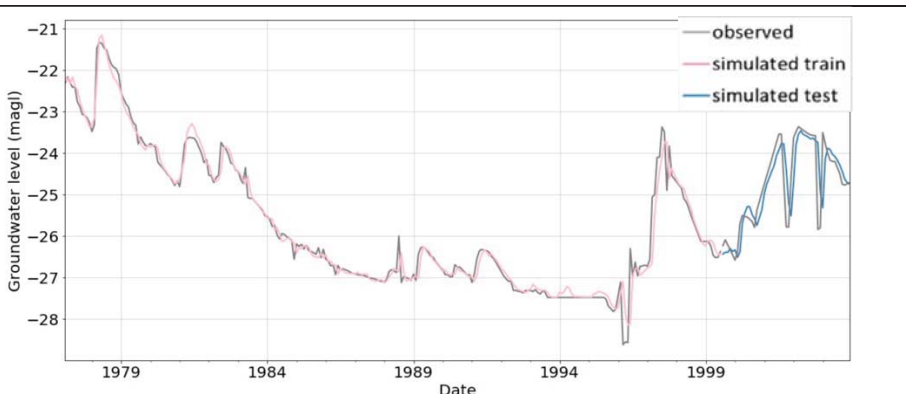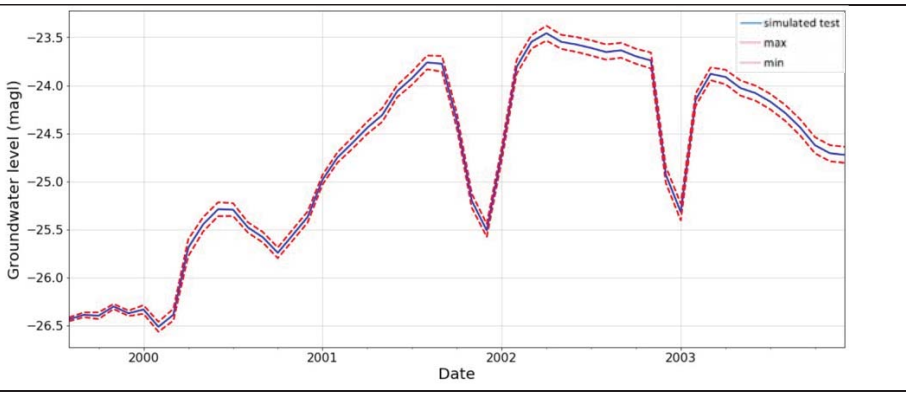
Table 10-1   The statistical and graphical results of the LSTMs performance to simulate and predict groundwater levels from each modelled borehole in the Grootfontein compartment. The standard deviation for the model predictions across the ten model runs is also presented.
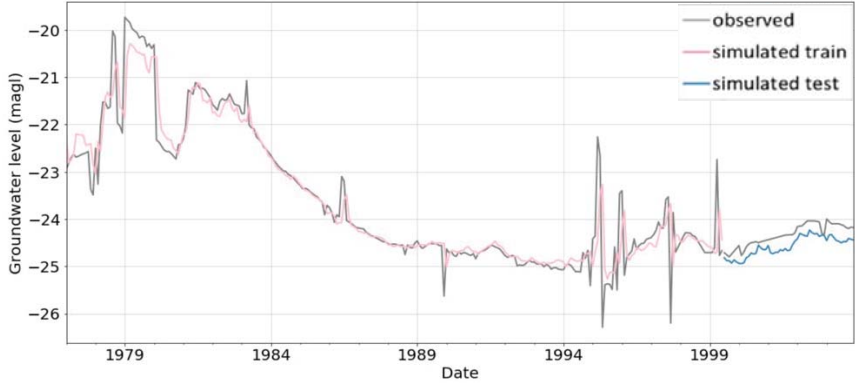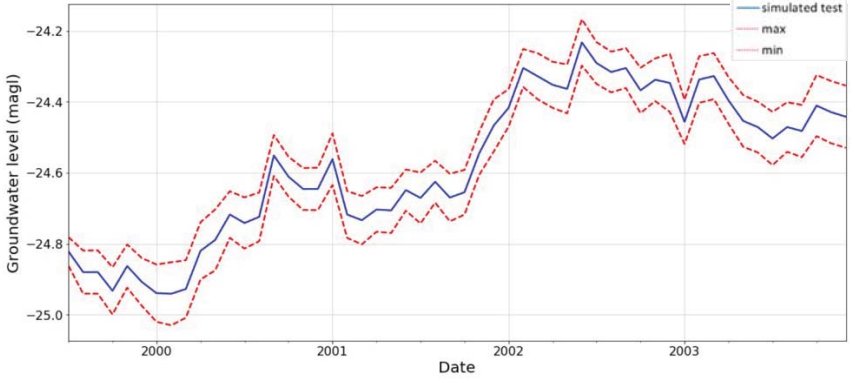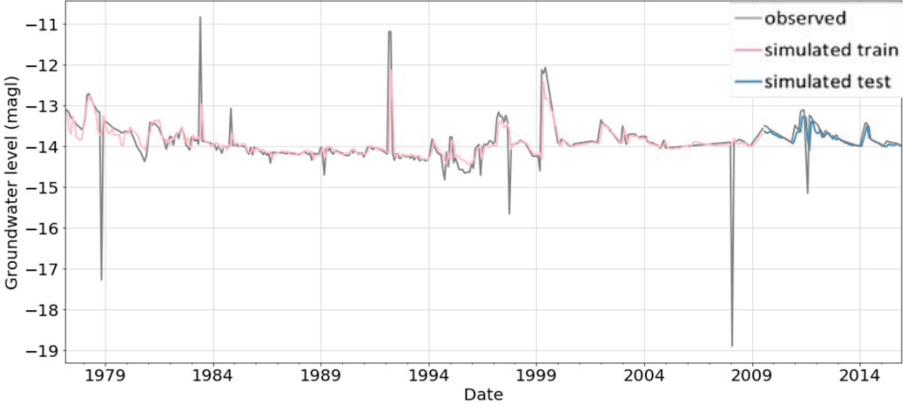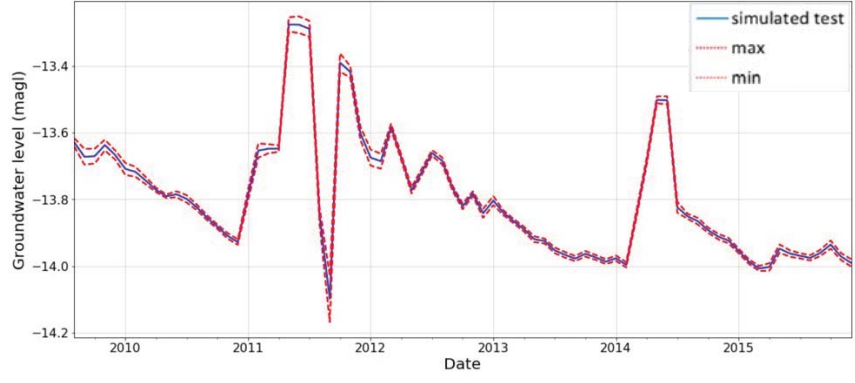
Metrics	D4N0112	
R2	0.266	
MSE	0.012	
RMSE	0.108	
MAE	0.084	
STDV	0.021	
Metrics	**D4N0117**	
R2	0.615	
MSE	0.148	
RMSE	0.385	
MAE	0.230	
STDV	0.027	
Metrics	**D4N0127**	

R2	0.725	
MSE	0.459	
RMSE	0.676	
MAE	0.473	
STDV	0.032	
Metrics	**D4N0146**	
R2	0.642	
MSE	0.565	
RMSE	0.751	
MAE	0.465	
STDV	0.028	
Metrics	**D4N0147**	
R2	0.421	
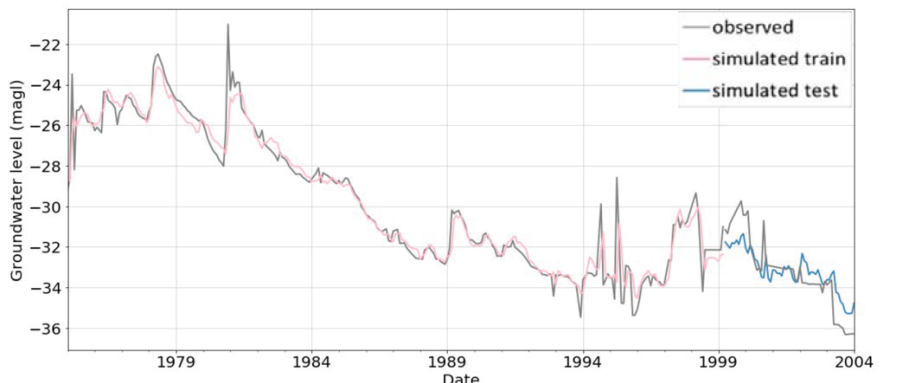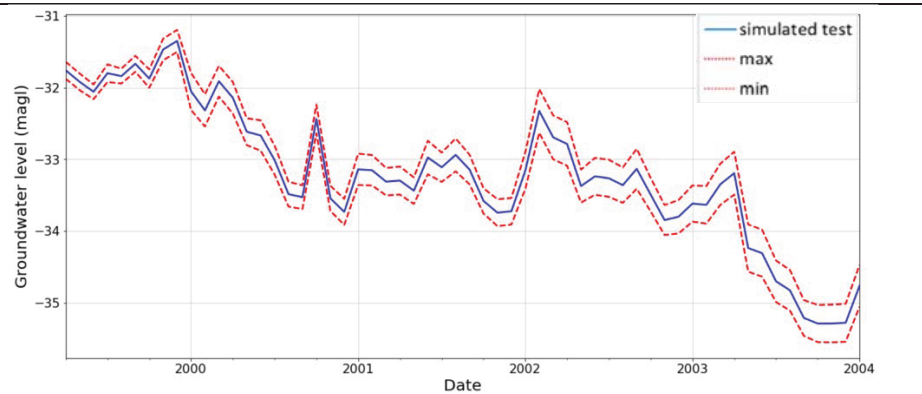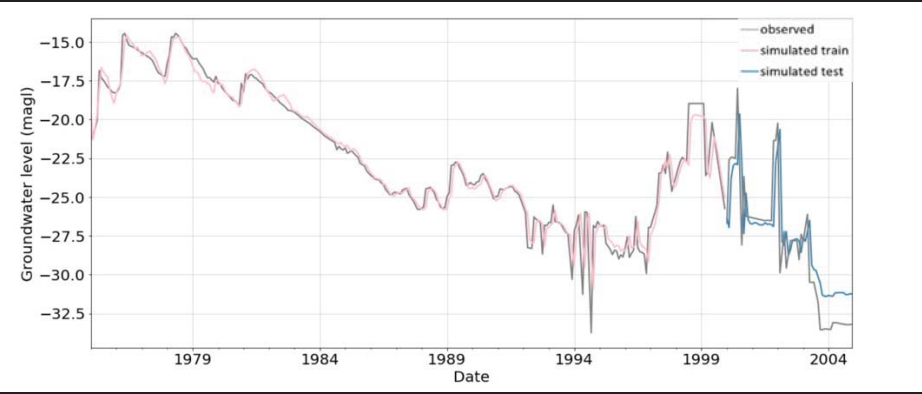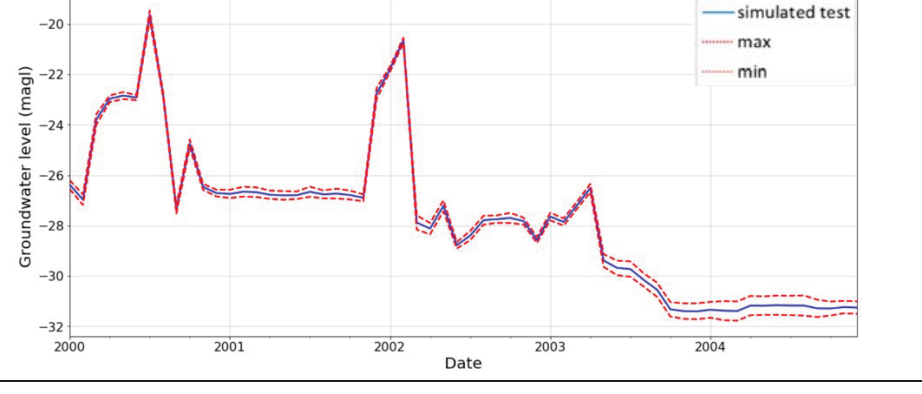
MSE	0.628	
RMSE	0.788	
MAE	0.672	
STDV	0.033	
Metrics	**D4N0066**	
R2	0.578	
MSE	0.619	
RMSE	0.787	
MAE	0.541	
STDV	0.021	
Metrics	**D4N0116**	
R2	0.848	
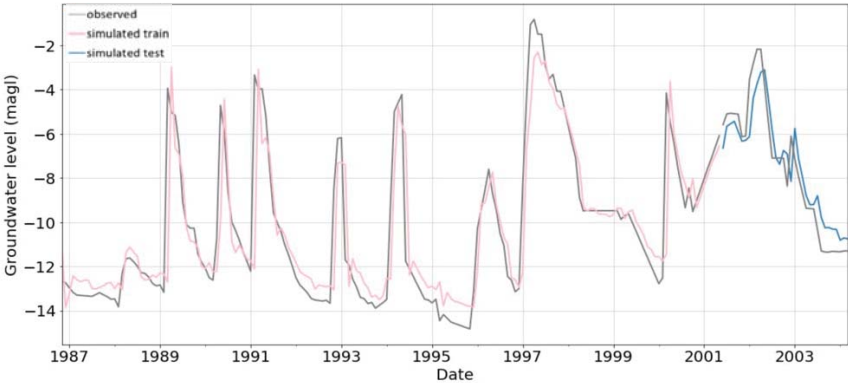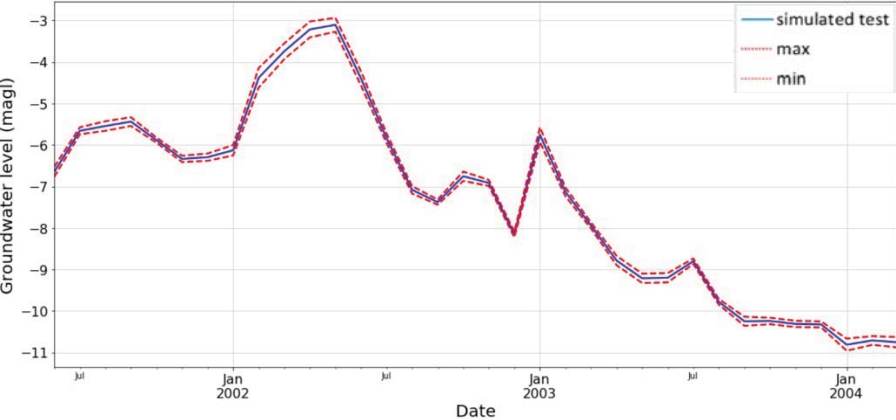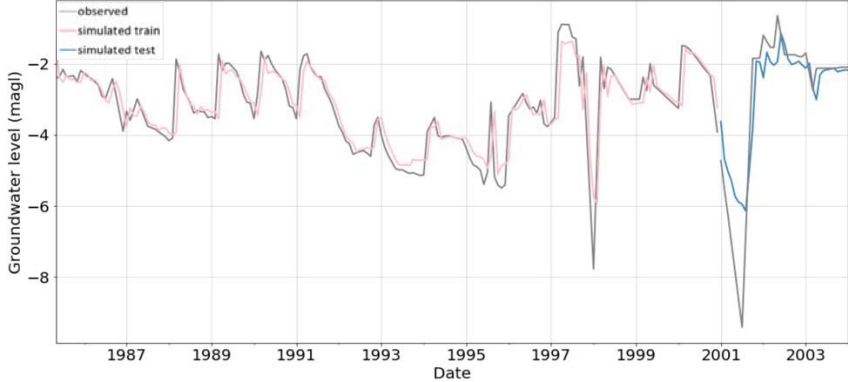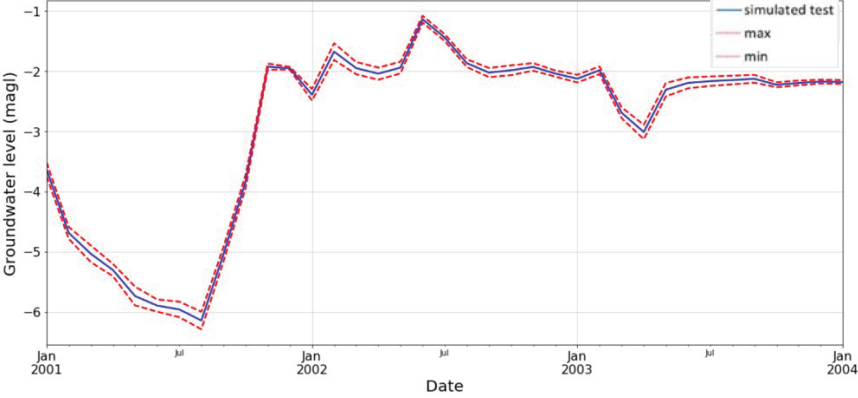
MSE	0.08	
RMSE	0.284	
MAE	0.197	
STDV	0.017	
Metrics	**D4N0037**	
R2	0.734	
MSE	0.016	
RMSE	0.124	
MAE	0.102	
STDV	0.023	
Metrics	**D4N0065**	
R2	0.578	
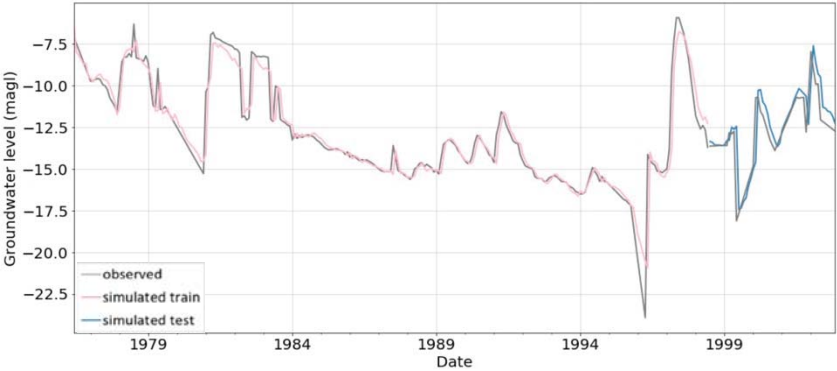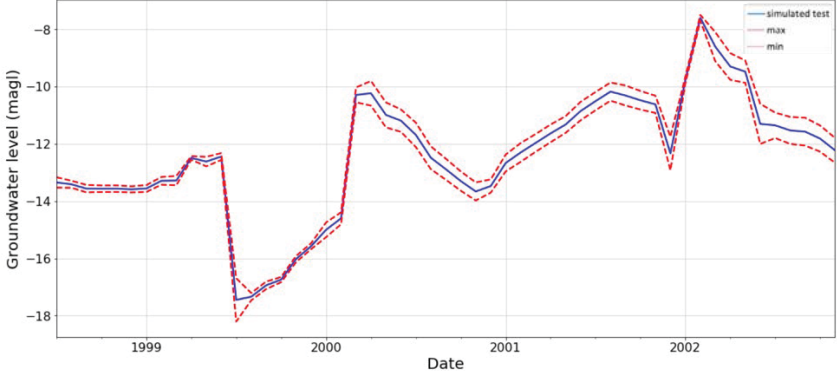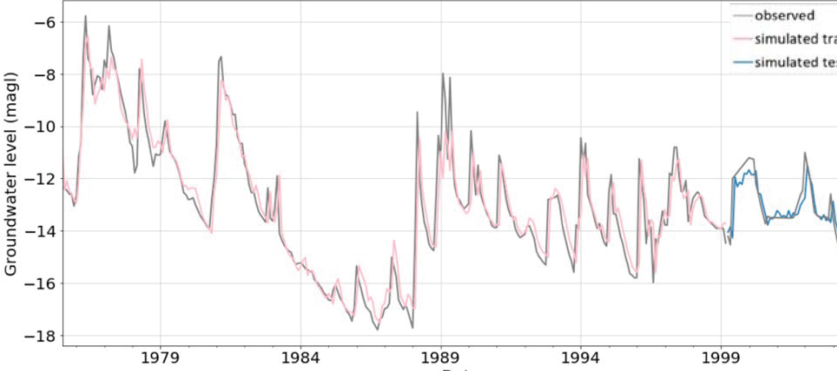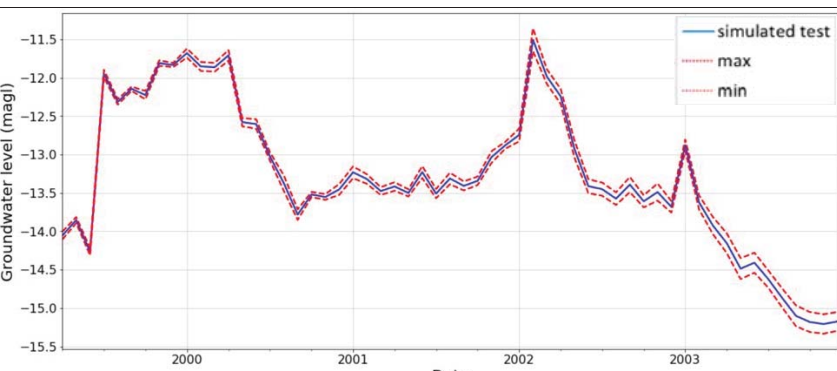
MSE	0.027	
RMSE	0.163	
MAE	0.126	
STDV	0.02	
Metrics	**D4N0139**	
R2	0.684	
MSE	0.121	
RMSE	0.347	
MAE	0.226	
STDV	0.018	
Metrics	**D4N0140**	
R2	-0.341	
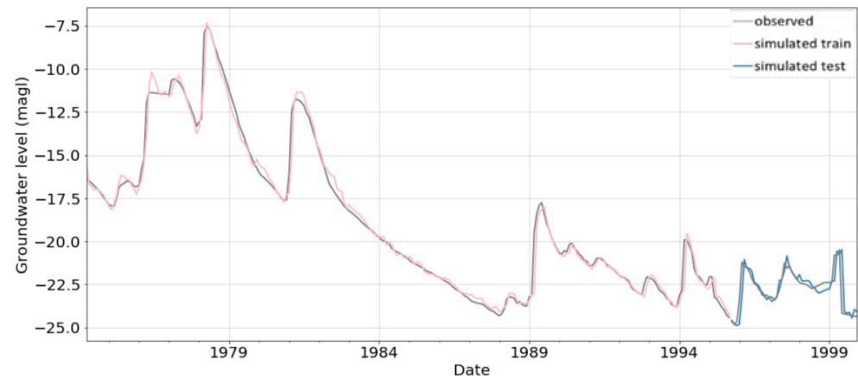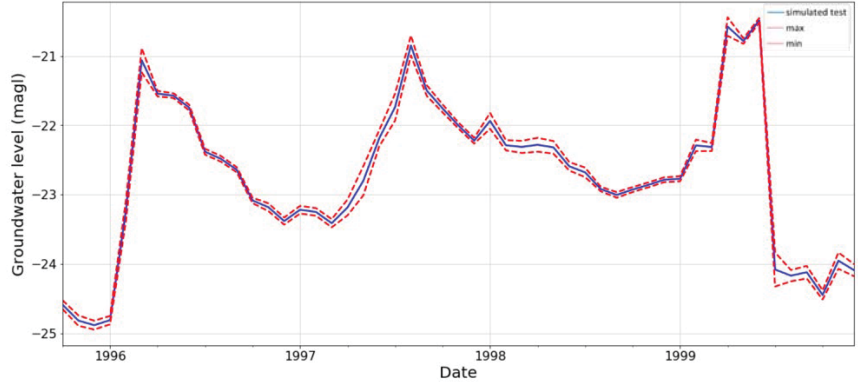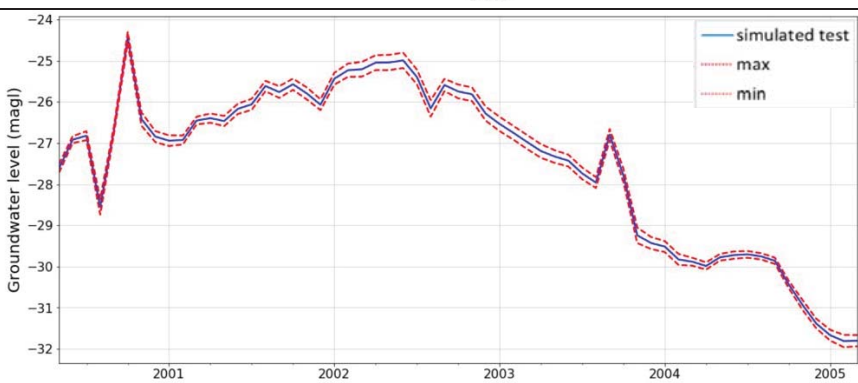
MSE	0.037	
RMSE	0.19	
MAE	0.177	
STDV	0.031	
Metrics	**2526CC00033**	
R2	0.288	
MSE	0.2	
RMSE	0.447	
MAE	0.216	
STDV	0.026	
Metrics	**2625BB00028**	
R2	0.803	

MSE	0.095	
RMSE	0.307	
MAE	0.177	
STDV	0.02	
Metrics	**D4N0141**	
R2	0.823	
MSE	0.051	
RMSE	0.224	
MAE	0.176	
STDV	0.014	
Metrics	**D4N0092**	
R2	0.653	
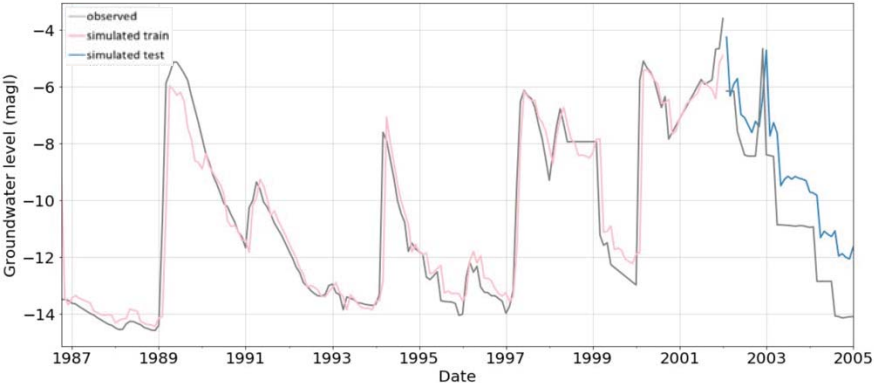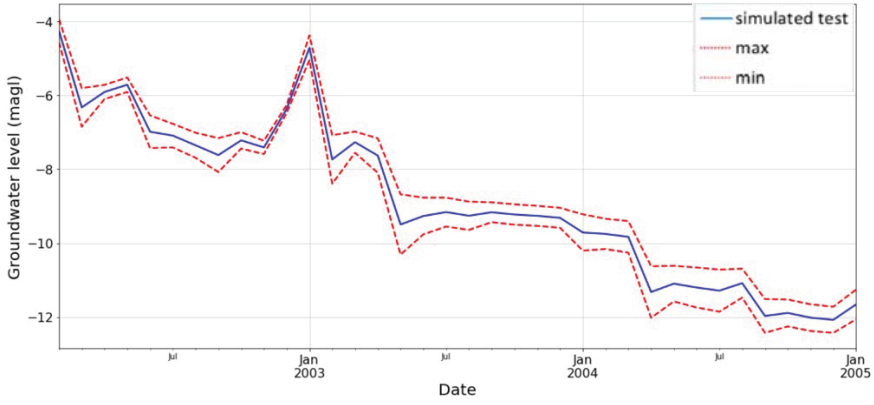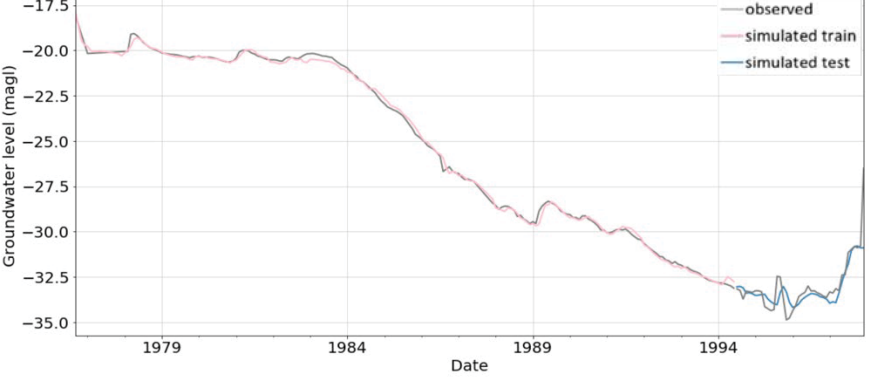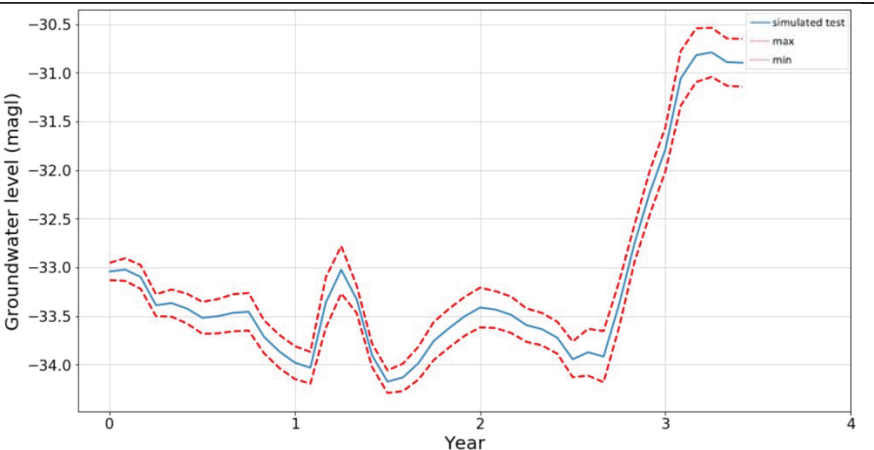
MSE	0.093	
RMSE	0.304	
MAE	0.233	
STDV	0.029	
Metrics	**D4N0094**	
R2	0.649	
MSE	0.25	
RMSE	0.499	
MAE	0.353	
STDV	0.025	
Metrics	**D4N0832**	
R2	0.862	

MSE	0.096	
RMSE	0.309	
MAE	0.258	
STDV	0. 021	
Metrics	**D4N0687**	
R2	0.789	
MSE	0.578	
RMSE	0.76	
MAE	0.477	
STDV	0.025	
Metrics	**D4N0130**	
R2	0.604	

MSE	0.19	
RMSE	0.435	
MAE	0.271	
STDV	0.04	
Metrics	**D4N0108**	
R2	0.802	
MSE	0.053	
RMSE	0.23	
MAE	0.162	
STDV	0.024	
Metrics	**D4N0053**	
R2	0.554	

MSE	0.028	
RMSE	0.167	
MAE	0.09	
STDV	0.014	
Metrics	**D4N0115**	
R2	0.819	
MSE	0.015	
RMSE	0.121	
MAE	0.082	
STDV	0.015	
Metrics	**D4N0110**	
R2	0.642	
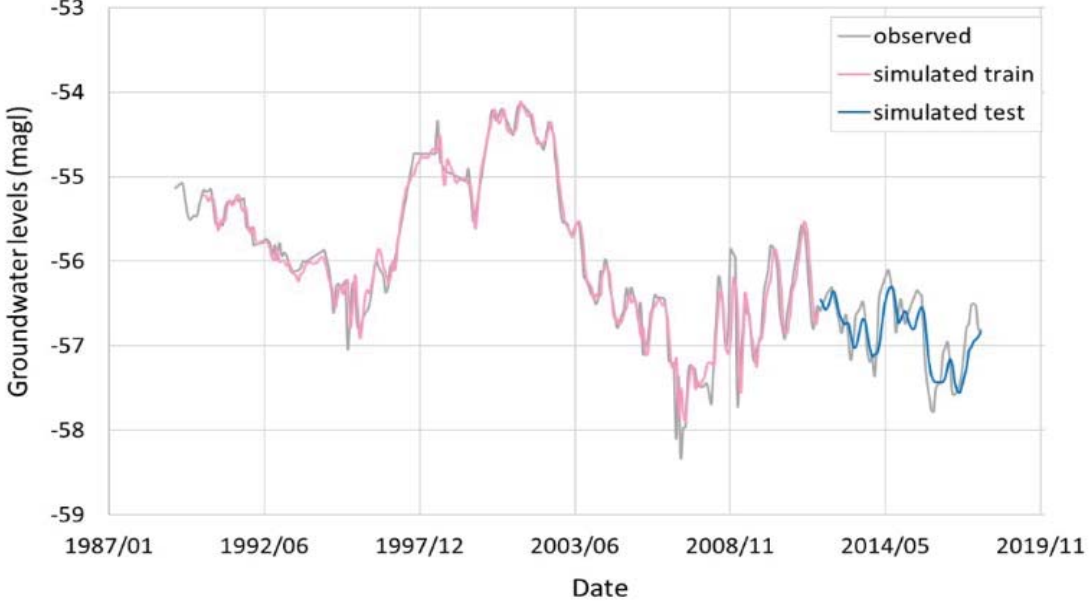
MSE	0.014	
RMSE	0.119	
MAE	0.10	
STDV	0.012	
Metrics	**D4N0123**	
R2	0.741	
MSE	0.122	
RMSE	0.346	
MAE	0.208	
STDV	0.033	
Metrics	**D4N0824**	
R2	0.515	

MSE	0.37	
RMSE	0.602	
MAE	0.539	
STDV	0.062	
Metrics	**D4N0126**	
R2	0.629	
MSE	0.026	
RMSE	0.164	
MAE	0.095	
STDV	0.016	
Metrics	**D4N0835**	
R2	0.321	

MSE	0.083	Graph
RMSE	0.225	
MAE	0.207	
STDV	0.098	
Metrics	**D4N0142**	
R2	0.812	
MSE	0.034	
RMSE	0.183	
MAE	0.133	
STDV	0.012	

10 APPENDIX B

Table 11-1   **The statistical and graphical results of the NNAR's performance to simulate and predict groundwater levels from each modelled borehole 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	

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

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

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

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

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

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

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

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

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

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

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