Hydrogeological analysis of groundwater occurrence within the Limpopo Province

RV Weidemann

orcid.org 0000-0003-2115-7544

Dissertation accepted in fulfilment of the requirements for the degree *Master of Science in Environmental Sciences with Hydrology and Geohydrology* at the North-West University

Supervisor: Dr SR Dennis

Graduation July 2020

20658249

Acknowledgements

I want to thank my heavenly father for giving me the strength, courage and perseverance to have done and finished this study. Without His grace nothing would be possible.

Secondly, I want to thank my whole family, for always believing in me and always giving me a fair chance to try anything and helped me achieve everything I have achieved up to this point.

My father and employer, managing director at VSA LEBOA CONSULTING (PTY) Ltd for allowing me to accompany him on field visits and to make me see the importance of groundwater from a very young age. Showing me the positive changes, it can bring to lives and communities especially in rural areas. He mentors me in everything about the geohydrology field of study that the books cannot do. Always helping with funding, resources and efficient time to be able to complete my MSc degree.

My fellow employees at VSA LEBOA CONSULTING (PTY) Ltd. for being patient with my studies and always being there to assist me in anything I asked from them.

Dr Rainier Dennis for helping and guiding me with the whole dissertation from the beginning right through to the end.

Ria Lubbe and the late Piet Lubbe for helping me to source relative data used for my project.

Abstract

Existing literature is reviewed to investigate what delineation methodologies on groundwater occurrence already exists and could be applied to specifically the Limpopo Province. Valuable lessons learned during this study are that a vast range of research has been done and a lot of time already spent on methodologies to reach certain outcomes, but the accumulation of new groundwater data remains of cardinal importance to be able to understand the occurrence of groundwater in different areas.

The importance of siting high yielding boreholes within the Limpopo Province and the amount of money and time that is spend on the development of sustainable groundwater resources highlights the importance of understanding the occurrence of potential high yielding water resources within the Limpopo Province. The GRIP database covering the Limpopo Province is one of the most comprehensive borehole databases in South Africa, making it the ideal dataset to test new groundwater delineation methodologies.

The aim of this study was to make use of existing borehole data and associated datasets to formulate a delineation methodology to delineate areas based on groundwater occurrence within the Limpopo Province.

An existing delineation methodology associated with Vegter's analysis method was investigated to test applicability to the Limpopo Province. A classification tree method for delineation was developed to address the short comings of the existing delineation methodology and results are compared for a previous study where the existing methodology were applied.

Finally, the newly formulated delineation methodology is applied to Vegter Region 7 as a case study. The standard Vegter analysis is also performed on Vegter Region 7 to investigate if further trends in the geohydrological data could be used to further sub-delineate areas of interest.

The developed delineation methodology is based purely on borehole data, which is a dynamic data set so that the delineation can be updated as new borehole data becomes available. Consideration is also given to delineation across study area boundaries to ensure continuity of the result.

Keywords: Groundwater; Hydrogeological; Delineation; Statistical analysis; Interpolation; Boreholes;

Table of Contents

List of Figures

List of Tables

Acronyms

AAYM	Aquifer Assured Yield Model
AD	Available Drawdown
AFYM	Aquifer Firm Yield Model
ASTER	Advanced Spaceborne Thermal Emission and Reflectance
BHN	Basic Human Need
CGS	Council for Geosciences
CIDA	Canadian International Development Agency
CSIR	Council for Scientific and Industrial Research
DWA	Department of Water Affairs
DWAF	Department of Water Affairs and Forestry
DWS	Department of Water and Sanitation
EC	Electrical Conductivity
GEP	Groundwater Exploitation Potential
GHP	Groundwater Harvest Potential
GRA	Groundwater Resource Assessment
GRA-I	Groundwater Resource Assessment Phase 1
GRA-II	Groundwater Resource Assessment Phase 2
GRDM	Groundwater Resource Directed Measures
GRIP	Groundwater Resource Information Project
GSA	Geo-Statistical Assessment
HSD	Hydrological Services Directorate
IDW	Inverse Distance Weighting
MCDA	Multi-criteria decision analysis
NGA	National Groundwater Archive
NGDB	National Groundwater Database
NWRS	National Water Resources Strategy
PC	Pearson Correlation
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SD	Standard Deviation
SRTM90	Shuttle Radar Topography Mission 90m dataset
StatsSA	Statistics South Africa
SVF	Saturated Volume Fluctuation
TDS	Total Dissolved Solids
UGEP	Utilizable groundwater exploitation potential
WRC	Water Research Commission

1 INTRODUCTION

1.1 Preamble

In 2013, the NWRS (National Water Resources Strategy) placed South Africa as the 30th driest country in the world, which highlights the fact that South Africa is a water scarce country and that our water resource management is of great importance and must be of very high quality as everyone depends on water to stay alive and it is a human right (DWAF, 2013).

The Limpopo Province of South Africa alone has approximately 5.5 million residents (StatsSA, 2014) of which only 10% are urbanized (De Cock *et al.* 2013). Almost 70% of the rural domestic water supply throughout the Province comes from groundwater (Botha *et al.* 2011). This means that groundwater exploitation is and will be an ongoing field of study for many years to come.

The NWRS (DWAF, 2013) states that an additional 165 Mm^3 of water per annum will be required for the entire Olifants water management area and the adjacent areas of Polokwane and Mogalakwena before 2035 of which 35 Mm^3/a must be sourced from groundwater.

The Utilisable Groundwater Exploitation Potential (UGEP) for South Africa was published by the Department of Water Affairs (DWAF, 2010) to be 10,343 Mm^3/a (7,500 Mm^3/a under drought conditions). A total amount of 1572 Mm^3/a can be utilised by Limpopo, Olifants, Luvuvhu and Letaba water management areas. According to this study, South Africa can only utilize between 2000 Mm^3/a and 4000 Mm^3/a, which implies that groundwater utilisation is roughly at a third of its full potential (DWAF, 2010).

Groundwater is the primary water source for both domestic and agricultural use within the Limpopo Province. The increase in potential groundwater use will lead to the spending of millions of Rands on the drilling of boreholes. According to Botha *et al*. (2011) well fields can be developed rapidly at R2/m^3. From this it is evident that a very large amount of money will be spent on groundwater resource development in the near future.

When siting for possible underground water resources various geological conditions should be taken in consideration. Holland and Witthuser (2011) determined that five factors have an influence on groundwater productivity while focusing on crystalline rocks in the Limpopo Province which includes the following:

- The geological and topographic setting
- Dykes and linear anomalies (including their orientation)
- Regional tectonics (maximum horizontal stress)
- Weathering thickness

- Proximity to surface water drainages

These different factors that influence groundwater occurrence as well as the fact that the behaviour of certain factors vary between different regions (Holland, 2012) and indicate that areas should be studied on a smaller scale rather than on a regional scale. An appropriate way of doing such a study would be to assess available data to determine geohydrological trends for areas, which can be used to predict the occurrence of underground water.

1.2 Problem statement

The importance of siting high yielding boreholes within the Limpopo Province and the amount of money and time spent on the development of sustainable groundwater resources highlight the importance of understanding the occurrence of potentially high yielding water resources within the Limpopo Province. The GRIP database covering the Limpopo Province is one of the most comprehensive borehole databases in South Africa. Making use of the GRIP database and supporting datasets, areas associated with potential high yielding boreholes can be identified by applying relevant statistical analysis on the data.

1.3 Aims and objectives

The aim of this study is to make use of existing borehole data and associated datasets to formulate a delineation methodology to delineate areas based on groundwater occurrence within the Limpopo Province.

The specific objectives of this study are outlined as follows:

- Extract required data from the GRIP database to perform relevant statistical analysis;
- Formulate a delineation methodology to express groundwater occurrence based on physical aquifer parameters to ensure continuity across boundaries irrespective of scale; and
- Apply the delineation methodology in a case study.

2 LITERATURE REVIEW

2.1 Introduction

The purpose of this literature review is to investigate what delineations on groundwater occurrence already exist specifically in the Limpopo Province. Over and above the existing datasets, relevant models and existing delineation methodologies will also be investigated.

2.2 Geohydrological maps of South Africa

Throughout the geohydrological history of South Africa, there have been numerous attempts to predict and estimate the amount of groundwater that is available to extract in a sustainable manner. With each attempt some lessons were learnt from the previous projects and the accuracy was improved. Today various data sources are available regarding the geohydrology of South Africa in the form of various maps and groundwater databases. The sections that follow provide an overview of these datasets and associated results.

2.2.1 Vegter regions

The Water Research Commission introduced a project to subdivide (delineate) South Africa into a series of Groundwater regions. This has come to be known in the field as the “Vegter Regions” after J.R. Vegter who performed the delineation of the various regions. The delineations were mainly based on the occurrence of groundwater (type of porosity), lithostratigraphy, physiographical aspects as well as climatic factors such as rainfall.

The Vegter region delineation is seen as the first attempt at a visual representation of South African groundwater resources (DWAF, 2009a). Mr J.R. Vegter from the Water Research Commission (WRC) published a report with a set of groundwater maps in 1995. He used the information from 120,000 boreholes obtained from the National Groundwater Database (NGDB) on which he performed statistical analyses. According to Vegter (1995), analysis of the yield distribution of enough randomly spaced boreholes within a certain lithological unit can be used as an indication of the groundwater supply that can be expected for that same lithology. During this study lithostratigraphic units of South Africa were divided into 16 water-bearing categories by Vegter (Rosewarne *et al.*, 2006).

Vegter (1995) also stated that the aim of delineating groundwater regions is to provide guidelines for successful and cost-effective siting of boreholes. Vegter (2000; 2001a, 2001b; 2003a; 2003b; 2006) applies his methodology to various other regions based on the same premise.

The existing methodology is primarily based on existing information derived from national groundwater maps and information/data obtained from the Department of Water and Sanitation. Vegter does, however, state:

A comprehensive description of the occurrence of groundwater requires detailed field studies in order to identify, catalogue and map not only existing sources of supply but also those particular geomorphologic and geologic features that are indicative of favourable conditions for siting boreholes. In the event of possible exploitation of the latter confirmation by geophysical surveying may be necessary.

However, he does acknowledge that this is almost impossible to do and therefore suggests the approach followed in the above-mentioned reports – which is to perform a detailed assessment of existing data in order to set guidelines for each of the groundwater regions.

The first step in this methodology is to provide a detailed description of the geology, as this plays a predominant role in the character of the groundwater found in the area. Thereafter the occurrence of groundwater is discussed based on existing knowledge – these occurrences are then related to the geology of the area. One of the major distinctions made in all of the reports is the distinction between porous (e.g. alluvial aquifers) and secondary (e.g. hard rock aquifers).

The next step is to identify and characterize springs within the region. The temperature, water quality and rate of flow are discussed in detail.

The third step is to conduct a statistical analysis of available borehole data which is the focus of this report.

Once all these assessments have been completed borehole prospects can be defined in terms of accessibility and exploitability and various sub-regions can be delineated. Guidelines are then provided for further improvement in siting boreholes and the associated production costs of a successful borehole. Suitable geophysical methods for siting boreholes within the sub-regions are recommended and discussed. Associated drilling control procedures are stipulated.

In addition to the drilling of successful boreholes, it is important to know how much water is available as this will influence the long-term success of a borehole. Therefore, recharge and storativity have to be calculated so that a water balance can be determined for each of the sub-regions. Vegter's documents provide no guidelines on how to calculate these parameters.

Finally, water quality of available boreholes is compared to the Department of Water Affairs and Forestry 1996 drinking water guidelines (DWAF, 1996). The water is classified accordingly as ideal, good, marginal, poor and unacceptable. The Vegter delineation includes 16 regions for the Limpopo Province as shown in Figure 1 and the description of each region is listed in Table 1. The sections that follow present the following Vegter parameters for the Limpopo Province:

- Depth to water level
- Accessibility and probability of drilling a successful borehole
- Groundwater recharge

- Saturated insterstices
- Groundwater quality (mean TDS) and dominant ions

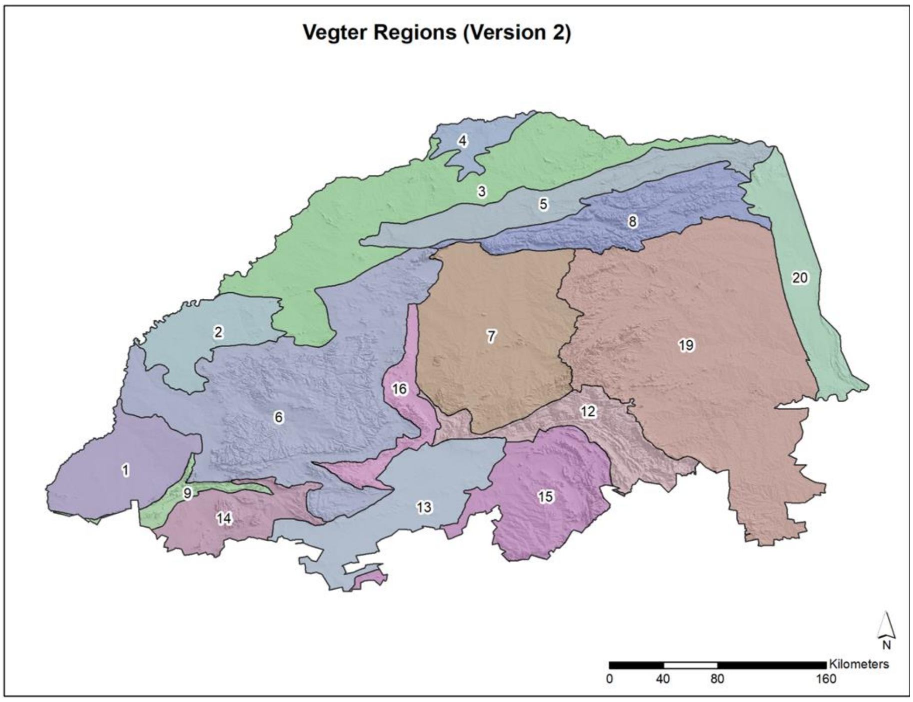


FIGURE 1 - VEGTER REGIONS OF THE LIMPOPO PROVINCE

TABLE 1 - VEGTER REGION NAMES

No	***Region Name***	***No***	***Region Name***
1	Makoppa Dome *Crystalline Igneous and metamorphic Basement rocks*	9	Western Bankeveld and Marico Bushveld
2	Waterberg Coal Basin	12	Eastern Bankeveld
3	Limpopo Granulite Gneiss Belt *Crystalline igneous and metamorphic Basement rocks*	13	Springbok Flats *Mainly extrusive rocks*
4	Limpopo Karoo Basin	14	Western Bushveld Complex *Mainly extrusive rocks*
5	Soutpansberg hinterland *Mainly extrusive rocks*	15	Eastern Bushveld Complex *Mainly extrusive rocks*
6	Waterberg Plateau	16	Northern Bushveld Complex *Mainly extrusive rocks*
7	Pietersburg Plateau *Crystalline igneous and metamorphic basement rocks*	19	Lowveld *Crystalline igneous and metamorphic basement rocks*
8	Soutpansberg	20	Northern Lebombo *Mainly extrusive rocks*

2.2.1.1 Vegter's depth to water level

The depth to water level map represents the depth to water level from the ground surface and is shown in Figure 2. This depth to water level map is an input to the Groundwater Resources of South Africa map (Seymour, 1994). In addition to the mean water level the standard deviation associated with mean value is also available as shown in Figure 3.

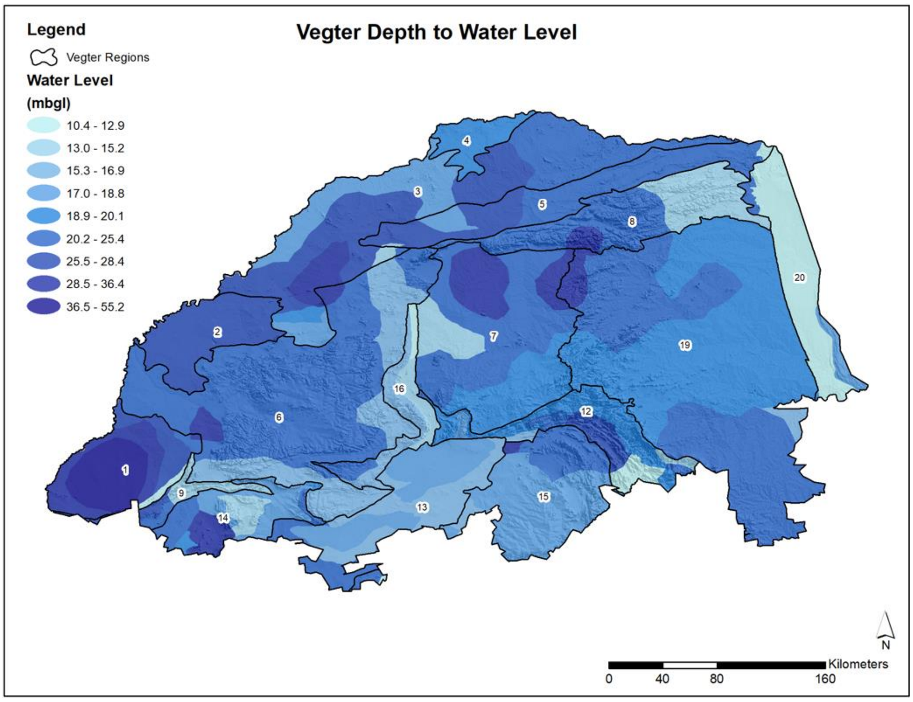


FIGURE 2 – VEGTER'S DEPTH TO WATER LEVEL FOR THE LIMPOPO PROVINCE

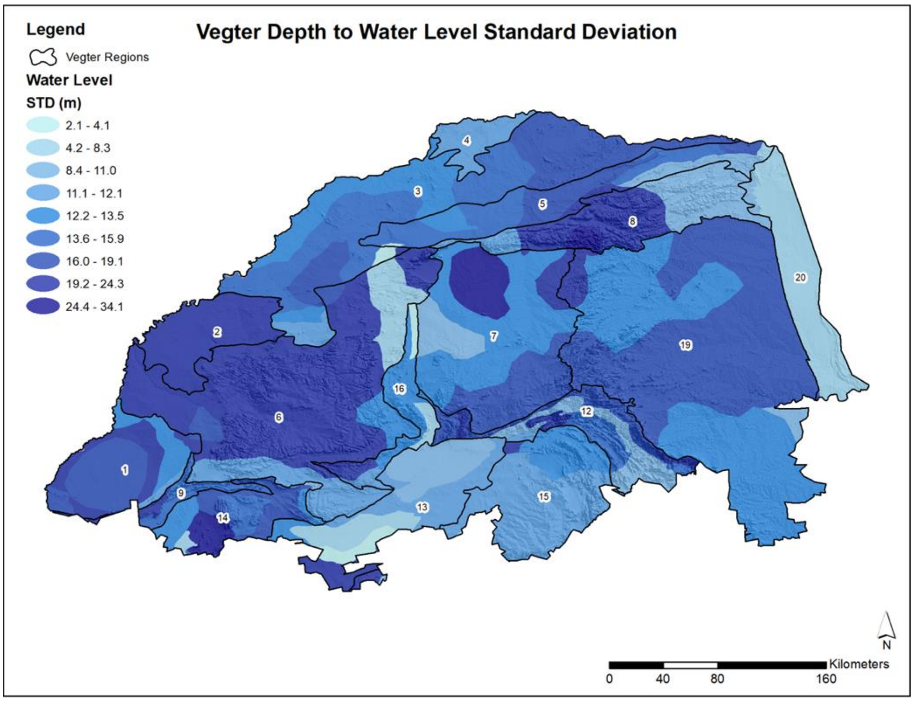


FIGURE 3 – VEGTER'S DEPTH TO WATER LEVEL STANDARD DEVIATION

2.2.1.2 Vegter's Borehole Probability Index

The probability of drilling a successful borehole is shown in Figure 4 where the probability is indicated with an index ranging between 0 and 10. The probability of drilling a successful borehole with a yield of 2 l/s and more is shown in Figure 4 – Vegter's borehole probability index of a successful borehole. The union of the aforementioned maps was the first groundwater map of South Africa (Seymore, 1994).

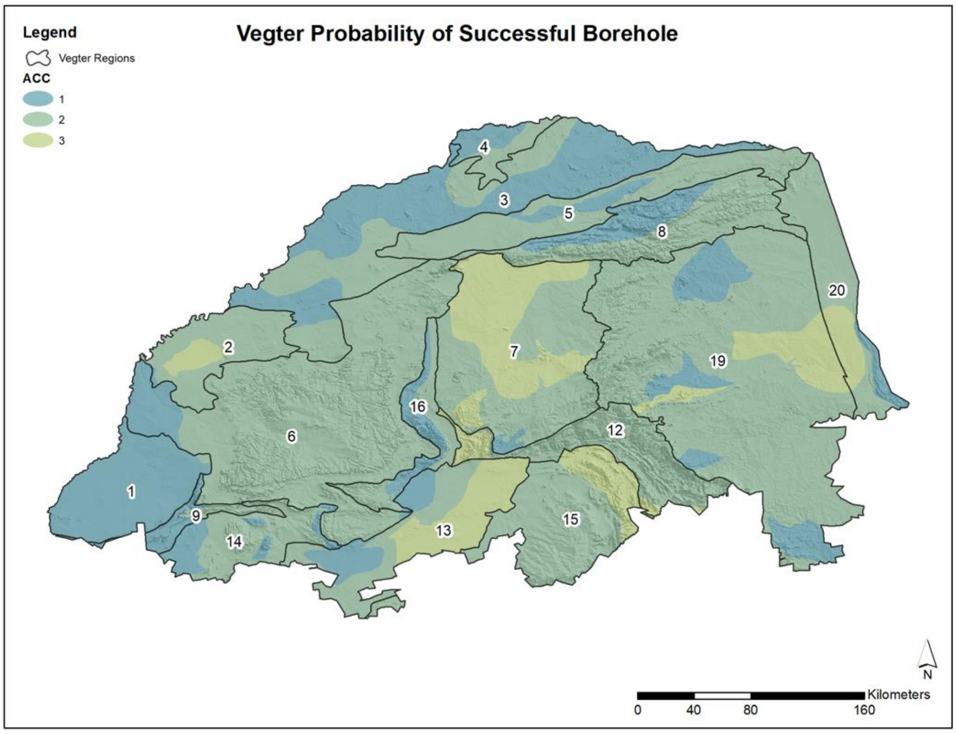


FIGURE 4 – VEGTER'S BOREHOLE PROBABILITY INDEX OF A SUCCESSFUL BOREHOLE

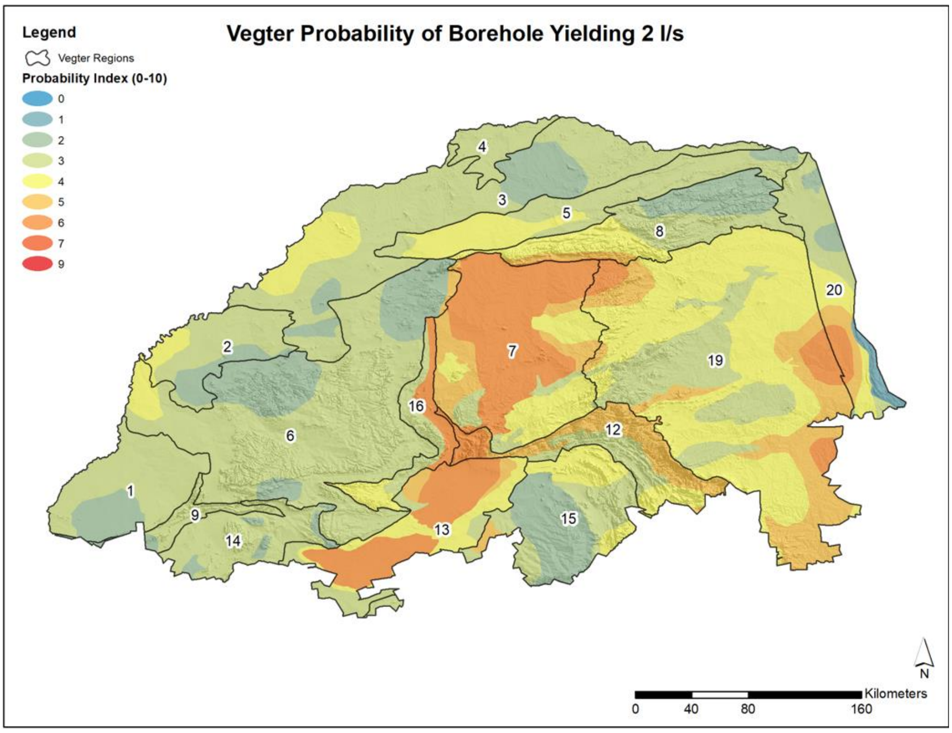


FIGURE 5 – VEGTER'S BOREHOLE PROBABILITY OF DRILLING A BOREHOLE WITH YIELD OF 2L/S AND MORE

2.2.1.3 *Vegter's recharge map*

The groundwater recharge map (groundwater recharge from rainfall) is an inset map on the Groundwater Resources of South Africa map and is shown in Figure 6. It should be noted that this is a generalised map (Seymore, 1994).

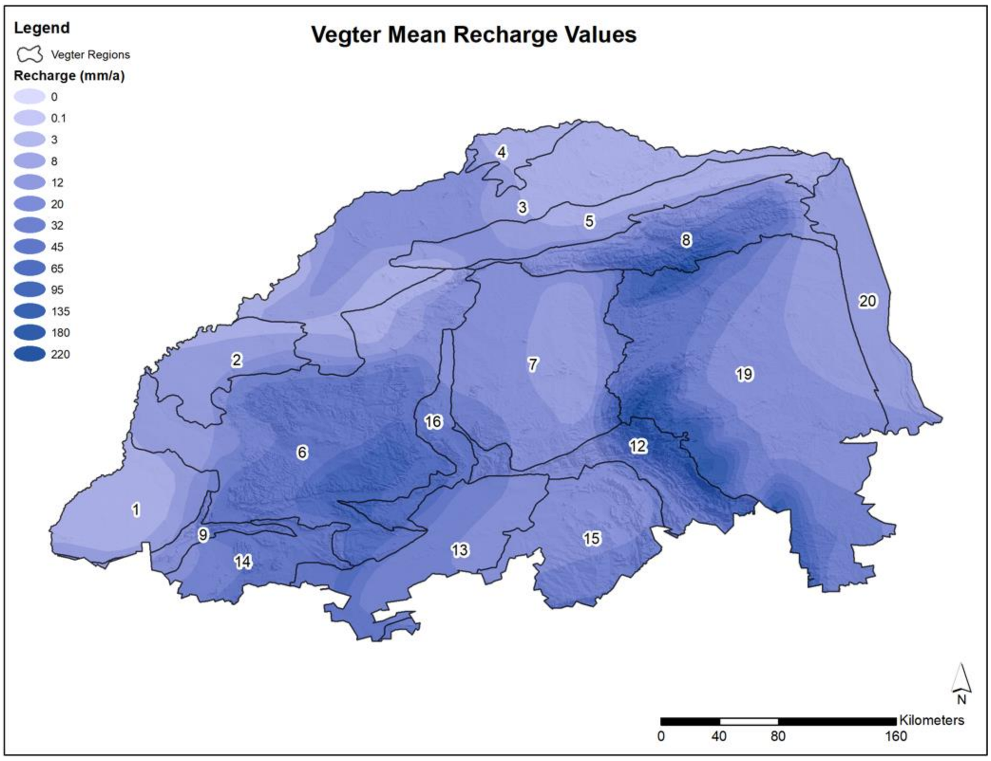


FIGURE 6 – VEGTER'S RECHARGE MAP

2.2.1.4 *Vegter's Saturated Insterstices*

The saturated interstices are an inset map on the Groundwater Resources of South Africa map. The optimal drilling depth and indication of the storage coefficient are presented in Figure 7 - Vegter's optimal drilling depth and Figure 8 respectively. Only boreholes with values greater than zero for strike depth, borehole depth, yield and water depth were used (Seymore, 1994).

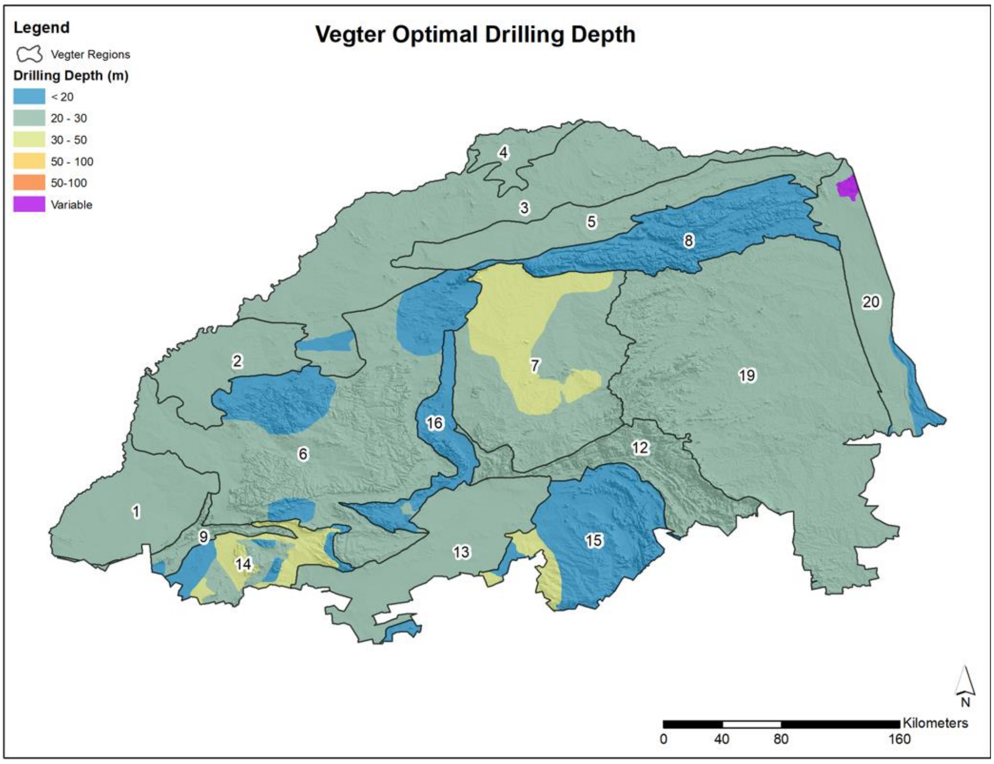


FIGURE 7 - VEGTER'S OPTIMAL DRILLING DEPTH

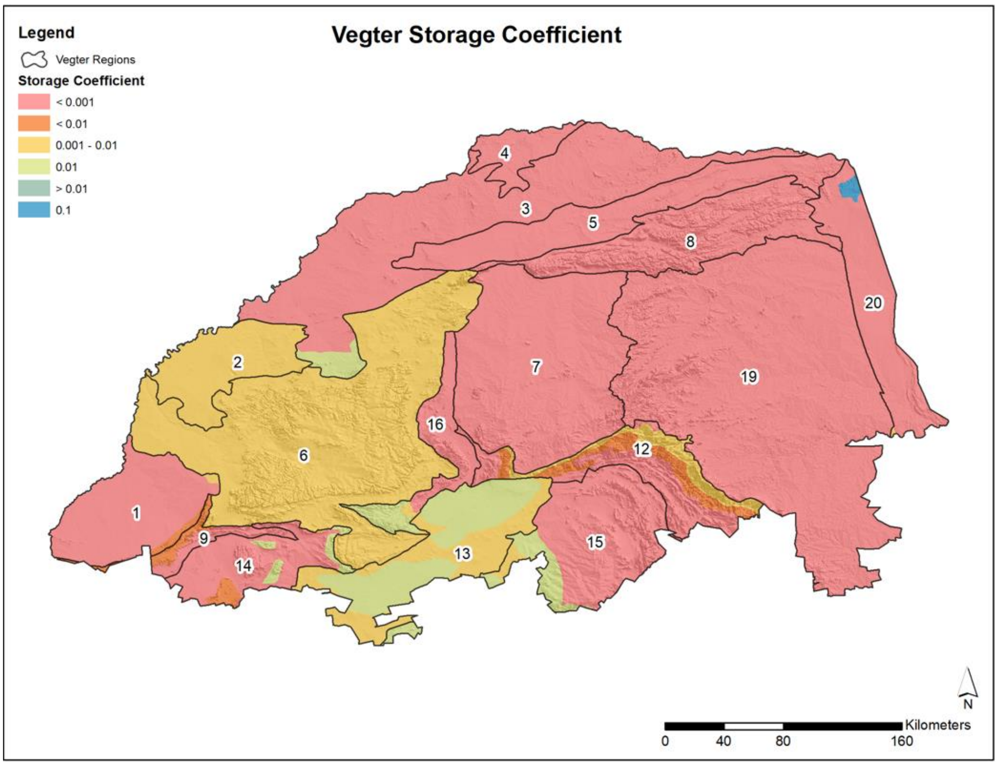


FIGURE 8 - VEGTER'S STORAGE COEFFICIENT

2.2.1.5 *Vegter's water quality maps*

The water quality map, represented by mean TDS, is presented in Figure 9 – Vegter's TDS map for Limpopo. The TDS is used as macro-indicator of water qualities.

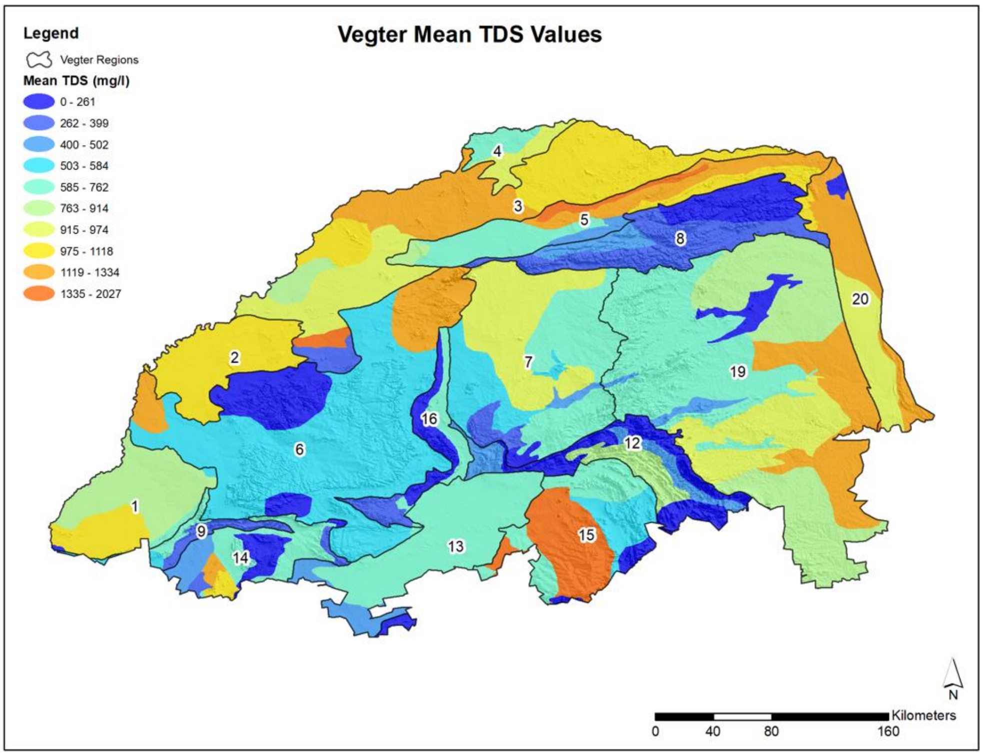


FIGURE 9 – VEGTER'S TDS MAP FOR LIMPOPO

High TDS values do not necessarily refer to polluted water as the natural background can be high as well as is the case in the Kruger National Park to name one example.

The water character as determined by the major ions is presented in Figure 10 to Figure 13. This work was done by Milo Simonic and edited by JR Vegter. The four major water character types are considered and each map indicates the percentage of the relevant water character as it relates to the spatial distribution.

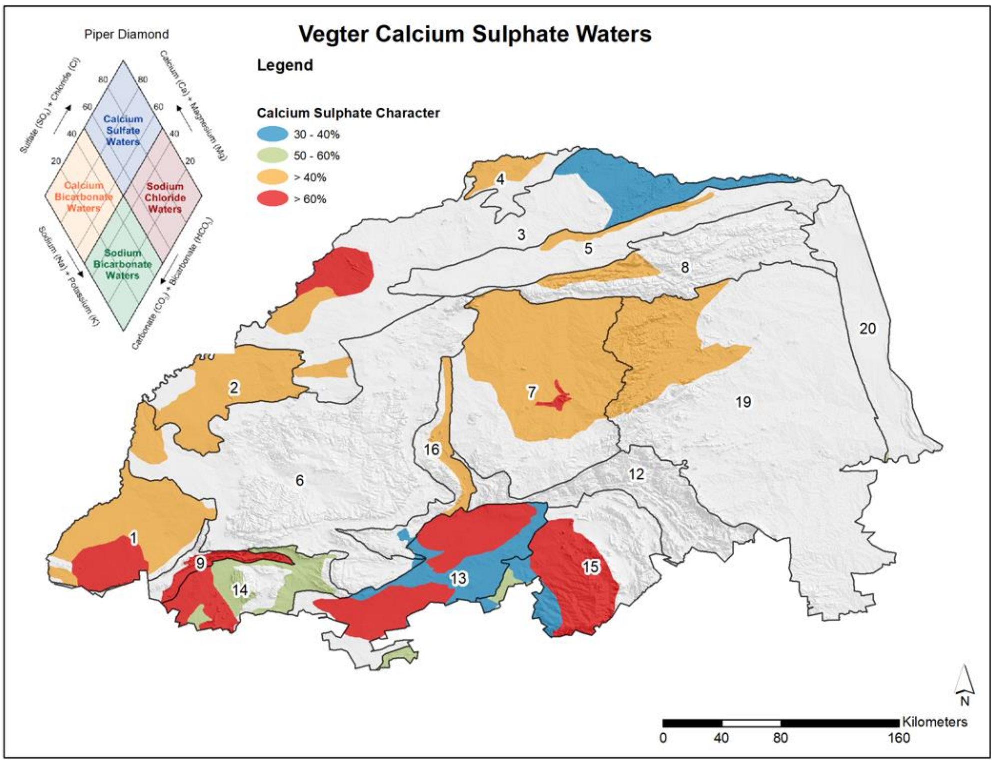


FIGURE 10 - VEGTER'S CALCIUM SULPHATE WATERS FOR LIMPOPO

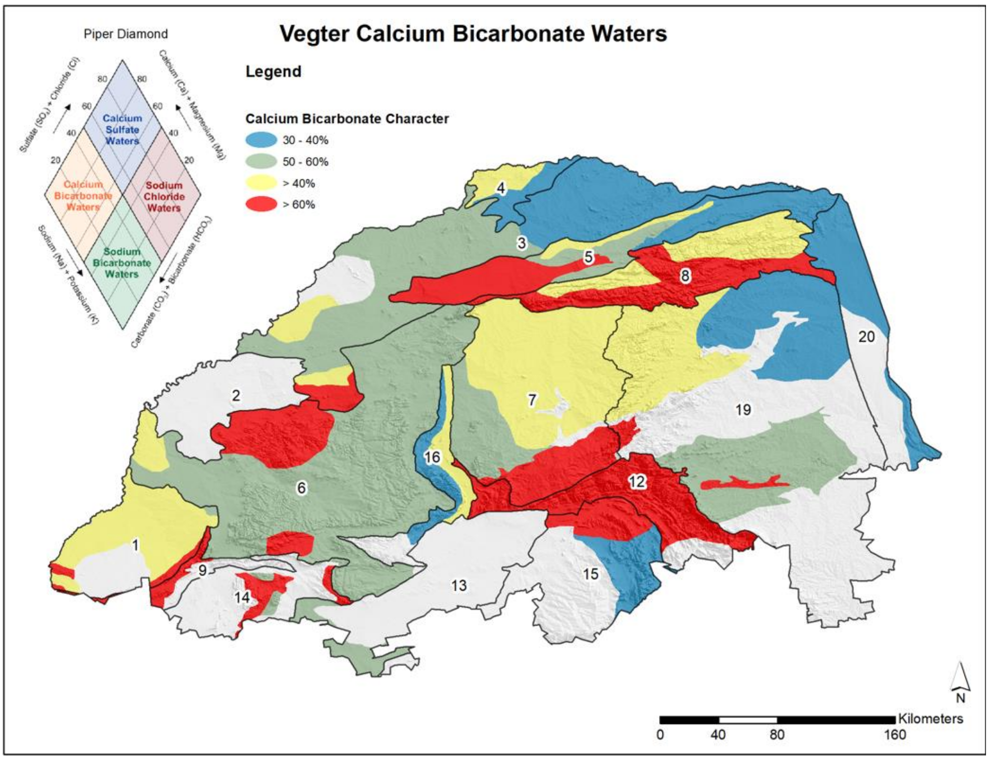


FIGURE 11 - VEGTER'S CALCIUM BICARBONATE WATERS FOR LIMPOPO

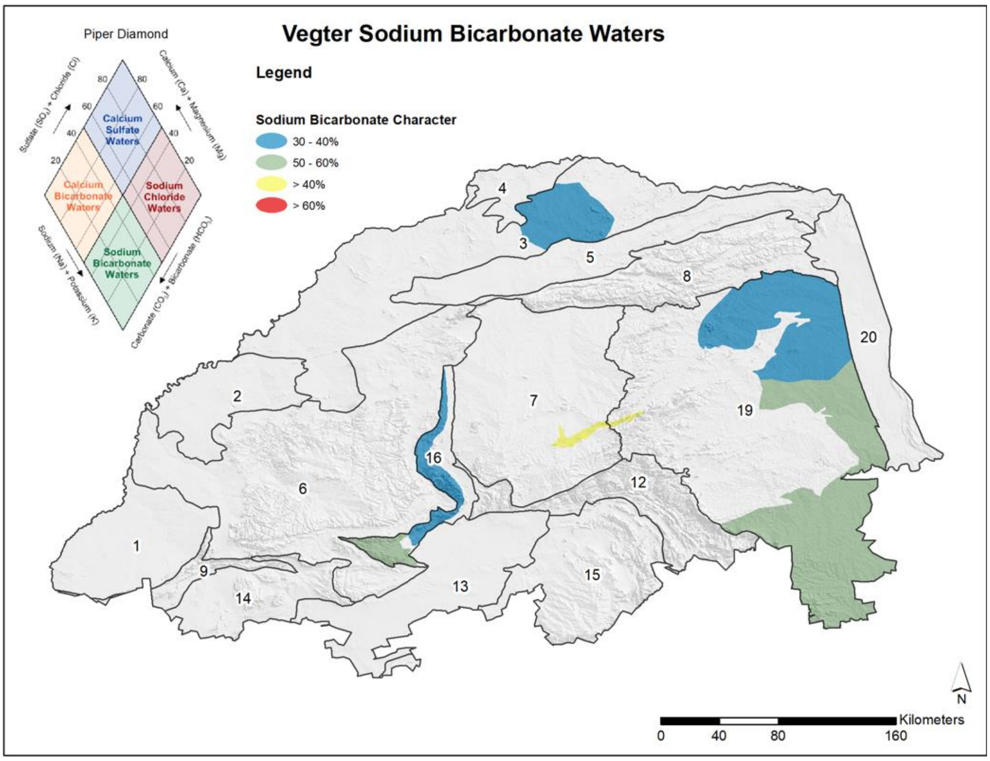


FIGURE 12 - VEGTER'S SODIUM BICARBONATE WATERS FOR LIMPOPO

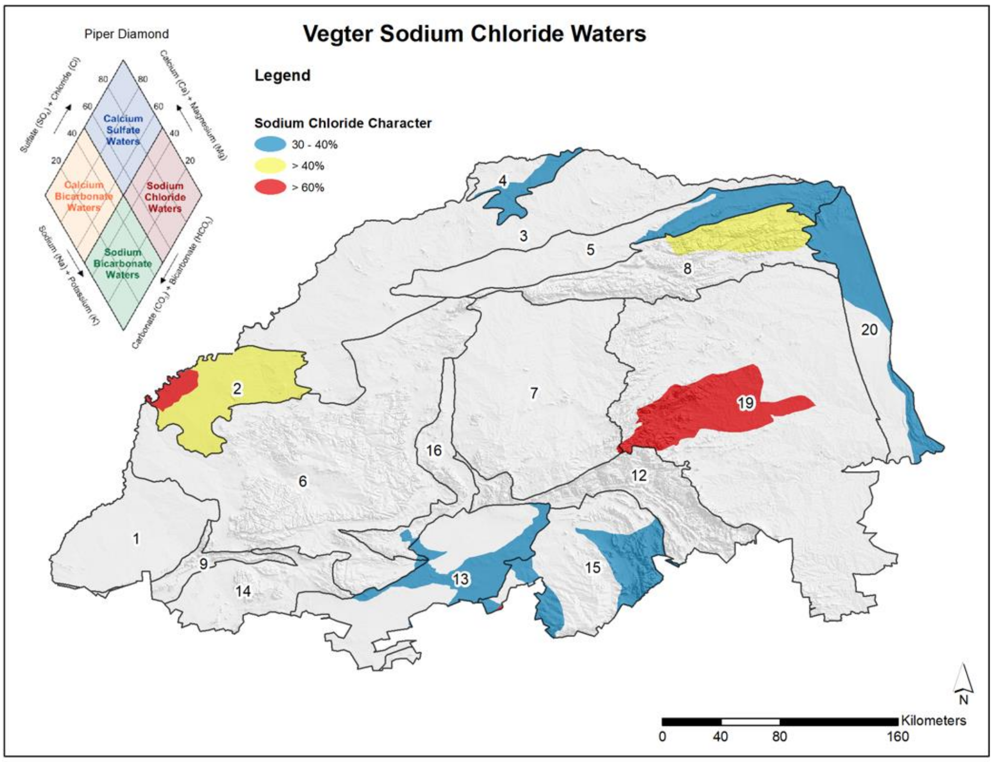


FIGURE 13 - VEGTER'S SODIUM CHLORIDE WATERS FOR LIMPOPO

2.2.2 Hydrological Map Series of South Africa

To help increase the lack of the South African groundwater knowledge, a groundwater characterisation program was initiated by Department of Water Affairs in the 1990's. A series of twenty-one hydrogeological maps were compiled at a scale of 1:500,000 with explanatory booklets to accompany each of them. The main aim was aquifer classification and the maps were completed in 2005. The hydrogeological map series includes schematic cross sections showing typical groundwater occurrence and includes four inset maps as follows (DWAF, 2009a):

- Distribution of borehole data (1:2,000,000 scale)
- Elevation above sea level (1:2,000,000 scale)
- Mean annual precipitation (1:2,000,000 scale)
- Groundwater quality (1:1,500,000 scale)

Groundwater classification on the hydrogeological map series consists of four classes:

- Intergranular (Type a)
- Fractured (Type b)
- Karstic (Type c)
- Fractured and Intergranular (Type d)

These maps should be used as a guideline only. A note that can be found on the maps states the following: "*The simplified lithology may be considered as guidelines only and the map is not to be used for the purpose of local borehole siting*" (Du Toit *et al*., 1998).

The groundwater occurrence from the geohydrological map series for the Limpopo Province is shown in Figure 14. The Vegter regions are also displayed for reference purposes.

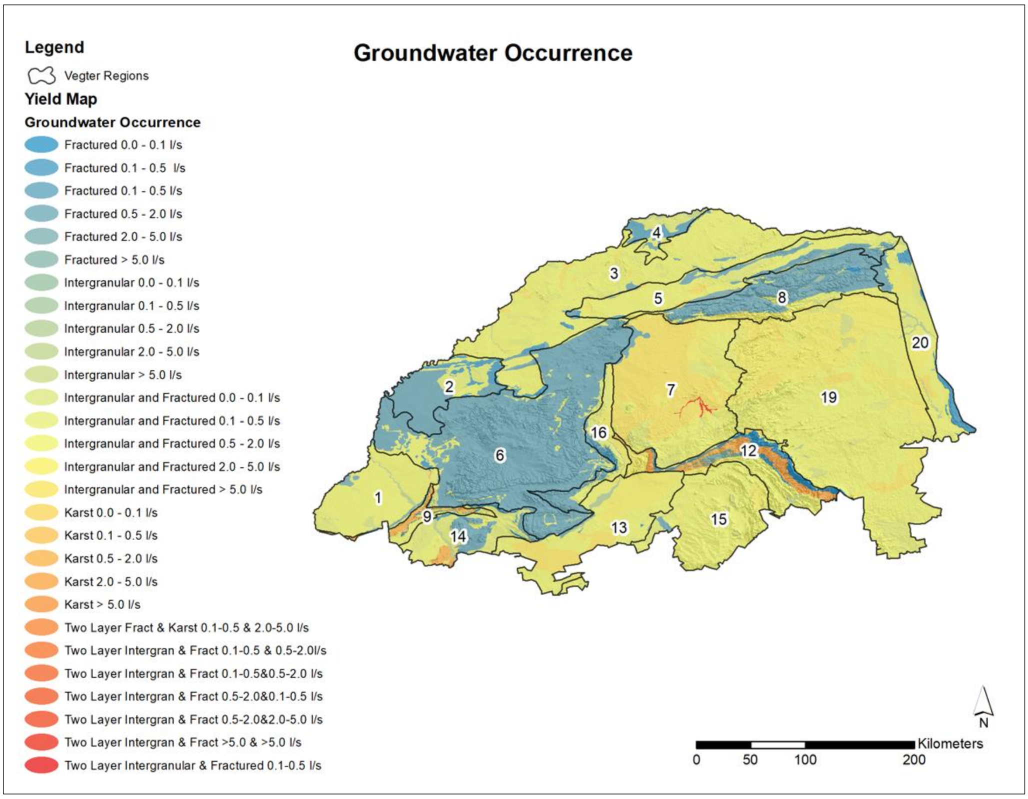


FIGURE 14 - GROUNDWATER OCCURRENCE MAP OF THE LIMPOPO PROVINCE

This study has developed into the project known as the Groundwater Resource Assessment Phase 1 (GRAI) which is discussed in the next section.

2.2.3 Groundwater Resources Assessment Project

A first attempt at the mapping of South Africa's groundwater resources was the Groundwater Resources of South Africa (GRA) project (maps and reports) which was published by the Water Research Commission (Vegter, 1995). The maps were based on information from approximately 120,000 boreholes throughout South Africa and included information such as depth of groundwater levels, mean annual groundwater recharge, groundwater quality and hydro-chemical types. At more or less the same time the Department of Water Affairs and Forestry launched an initiative to develop a series of 21 hydrogeological maps covering the whole of South Africa. This project was known as the Groundwater Resource Assessment Phase 1 (GRA-I). The GRA-I together with the national maps and the groundwater regions by Vegter (1995; 2001) raised awareness of the importance of groundwater potential (DWAF, 2009a). This awareness laid the foundations for the quality of groundwater information (maps, harvest potential and yield models) that were developed and are available today.

A criticism of GRA-I maps was that it did not present estimates on the amount of groundwater that could be utilised annually. This led to the start of the GRA-II project which commenced in 2003 with the main aim of upgrading and increasing the quality of the GRA-I deliverables. According to

Rosewarne *et al.* (2006) the main aim was to quantify South Africa's groundwater. DWAF (2009a) summarises the main GRA-II tasks as follows:

- Quantification (aquifer storage)
- Update the harvest potential map (discussed in next section)
- Groundwater-surface water interaction (recharge)
- Aquifer classification
- Groundwater use

A large amount of work was conducted during the GRA-II project and the general perspective was that good results were obtained. In a summary regarding the GRA-II, DWAF (2009a) stated that some of the methods used in the process were questionable. This was mainly because of the shortage of reliable groundwater data as well as the uneven distribution of the data across the country; this emphasizes the need for more and good quality groundwater data. GRA-I used geological and aquifer boundaries where GRA-II was based on quaternary catchment boundaries to allow easy interface with the surface water modelling systems of DWAF at the time. The problem with the afore-mentioned approach is that it does not assure the supply per aquifer or per borehole. A GRA-III project was proposed to eliminate the problems encountered within the GRA-II, but to date have not realised.

Even though the GRA-II project was quaternary based in the final presentation of the assessment, some grid files do exist. The spatial distribution of transmissivity and storativity across the Limpopo Province is shown in

Figure 15 – GRA-II transmissivity of the Limpopo Province

and Figure 16 respectively.

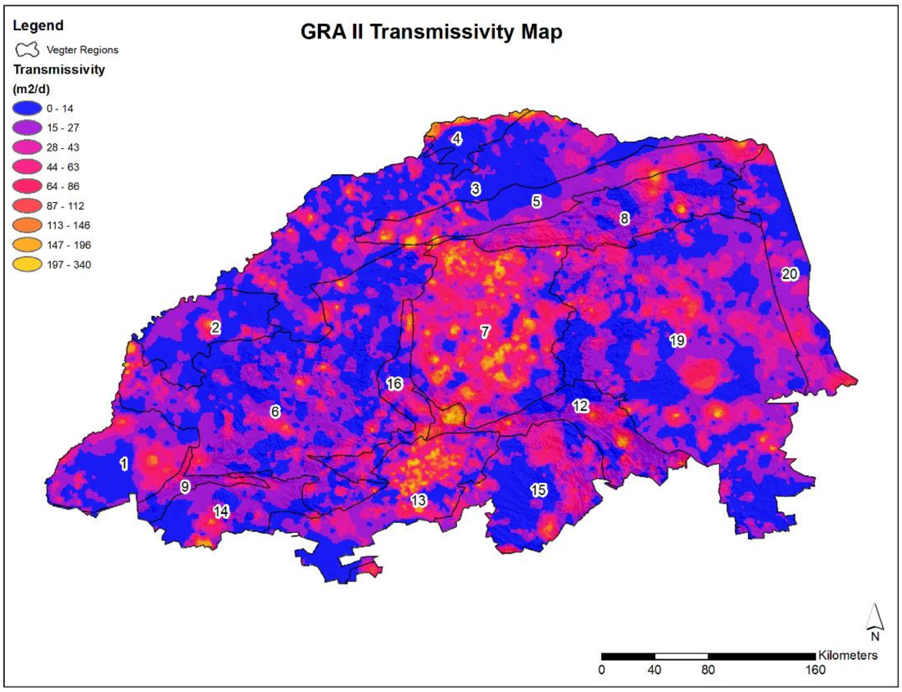


FIGURE 15 – GRA-II TRANSMISSIVITY OF THE LIMPOPO PROVINCE

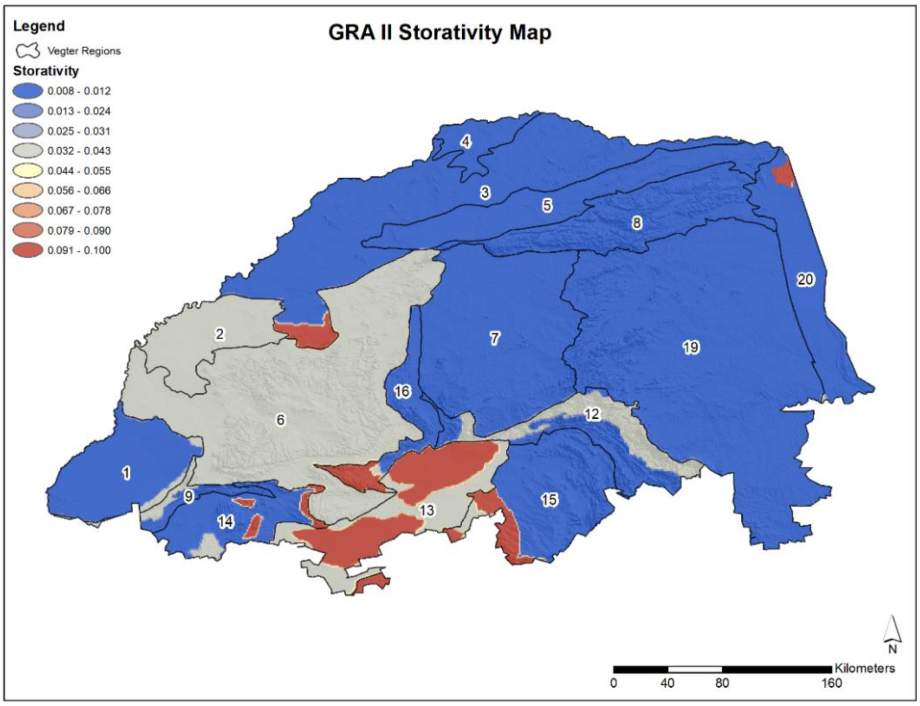


FIGURE 16 – GRA-II STORATIVITY OF THE LIMPOPO PROVINCE

Similarly, even though the final recharge and water levels from the GRA-II are reported on a quaternary scale, the source grid files are available and the spatial distribution of the recharge and groundwater levels across the Limpopo Province are shown in Figure 17 and Figure 18 respectively.

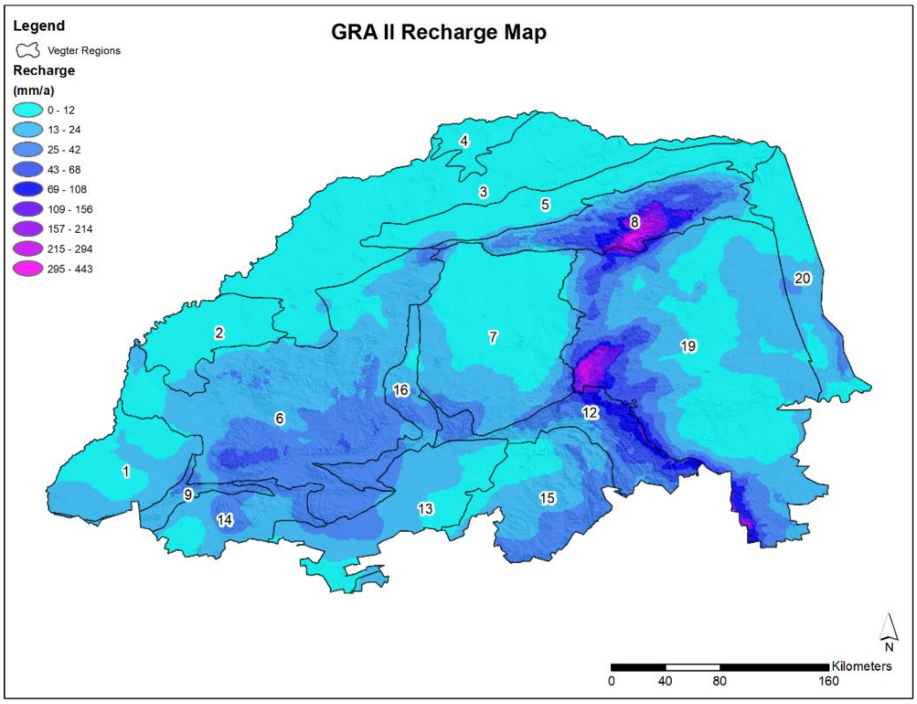


FIGURE 17 – GRA-II RECHARGE OF THE LIMPOPO PROVINCE

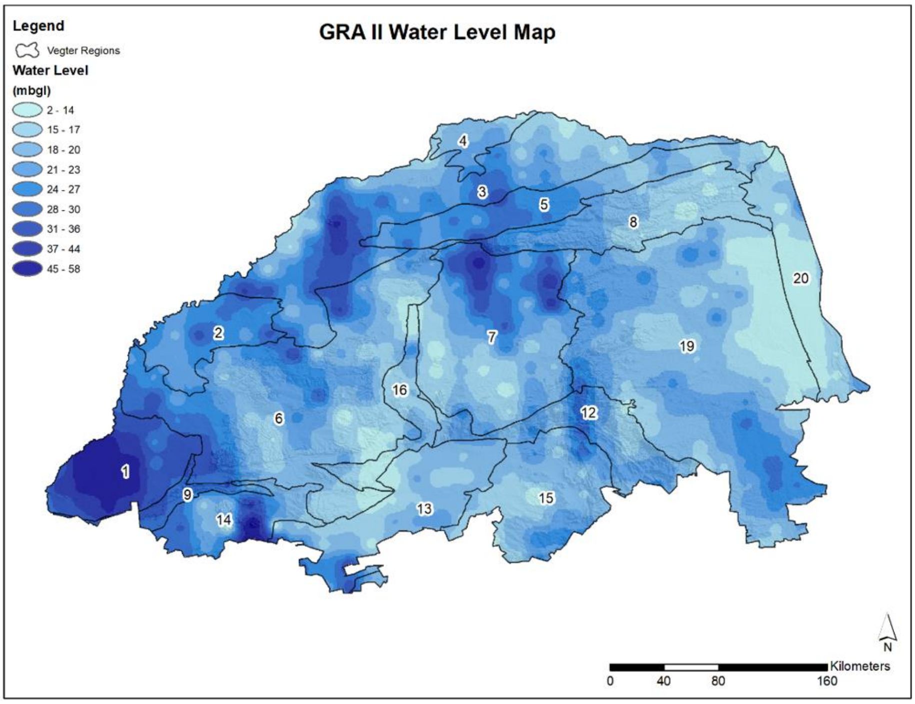


FIGURE 18 – GRA-II GROUNDWATER WATER LEVELS OF THE LIMPOPO PROVINCE

2.2.4 Groundwater Harvest Potential Map

The Groundwater Harvest Potential Map was developed from the GRA project that was published in 1995 (Baron *et al.*, 1998). This map estimated an annual sustainable volume of groundwater that could be extracted for different regions. A total of 19,000 Mm^3/a was estimated, that can be extracted from

South Africa's groundwater resources without depleting the aquifer. Although this effort only used regional estimates of aquifer recharge and storage to calculate a sustainable groundwater yield (Rosewarne *et al.*, 2006), it was definitely the best effort to map groundwater potential at the time (DWAF, 2009a).

Haupt (2001) estimated a total harvest potential of 10,000 Mm^3/a where he recognised the importance of aquifer permeability when calculating the harvest potential. The permeability as limiting factor (based on country-wide borehole yield analysis) was applied to the harvest potential maps in order to increase the accuracy of the data (Rosewarne *et al*., 2006).

Since the Groundwater Harvest Potential map was derived from the GRAII project, it is also quaternary based in scale. The harvest potential for the Limpopo Province is shown in Figure 19.

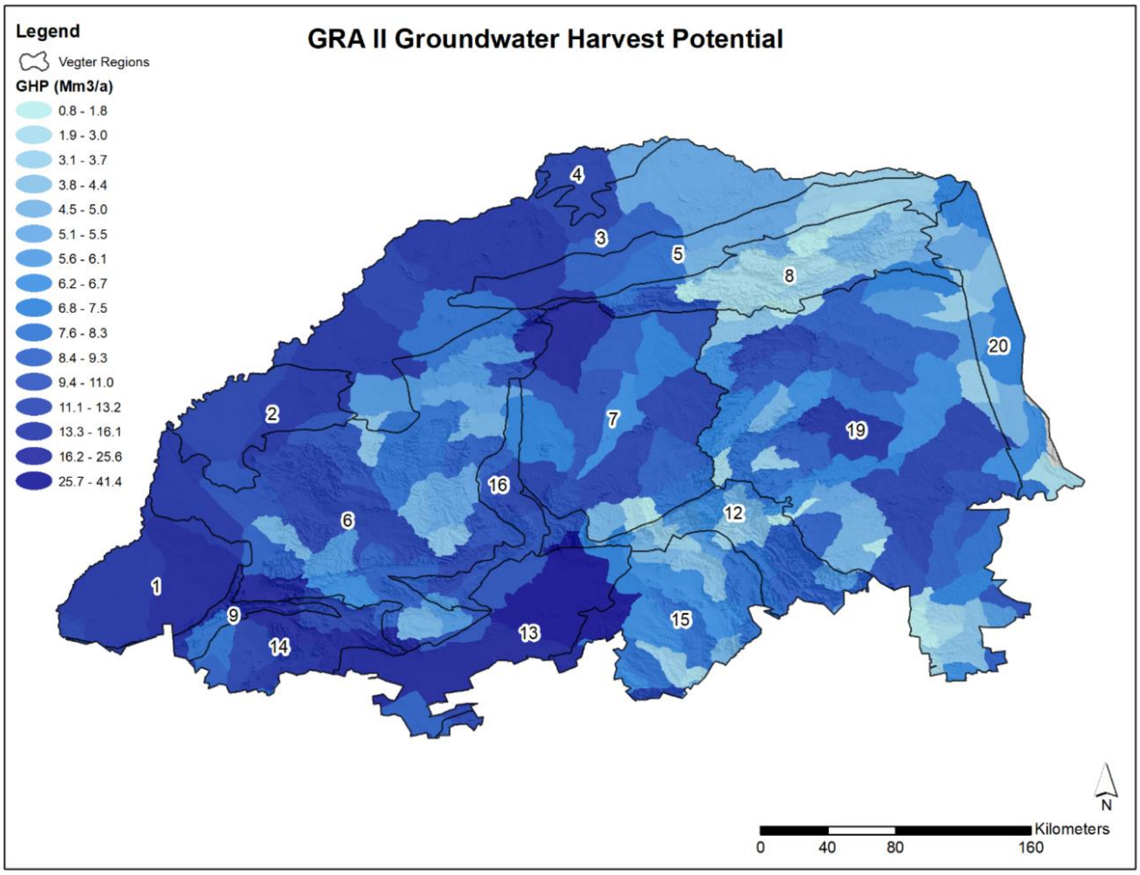


FIGURE 19 – GROUNDWATER HARVEST POTENTIAL MAP FOR THE LIMPOPO PROVINCE

2.2.5 Aquifer classification map of South Africa

The Aquifer classification map indicates the aquifer classification system of South Africa. Development was based on the British Geological Survey aquifer vulnerability classification (Parsons and Conrad, 1999). It classifies aquifers into three categories: Poor, Minor and Major respectively.

The yield of an aquifer and its water quality was used to determine its category. A major aquifer region would be a high yielding region with good water quality and a poor region is described as a low yielding aquifer of moderate to poor water quality. The map (Figure 20) was first produced by the CSIR in 1999 and this was recompiled by the Hydrological Services Directorate (HSD) in 2012 (DWS, 2016).

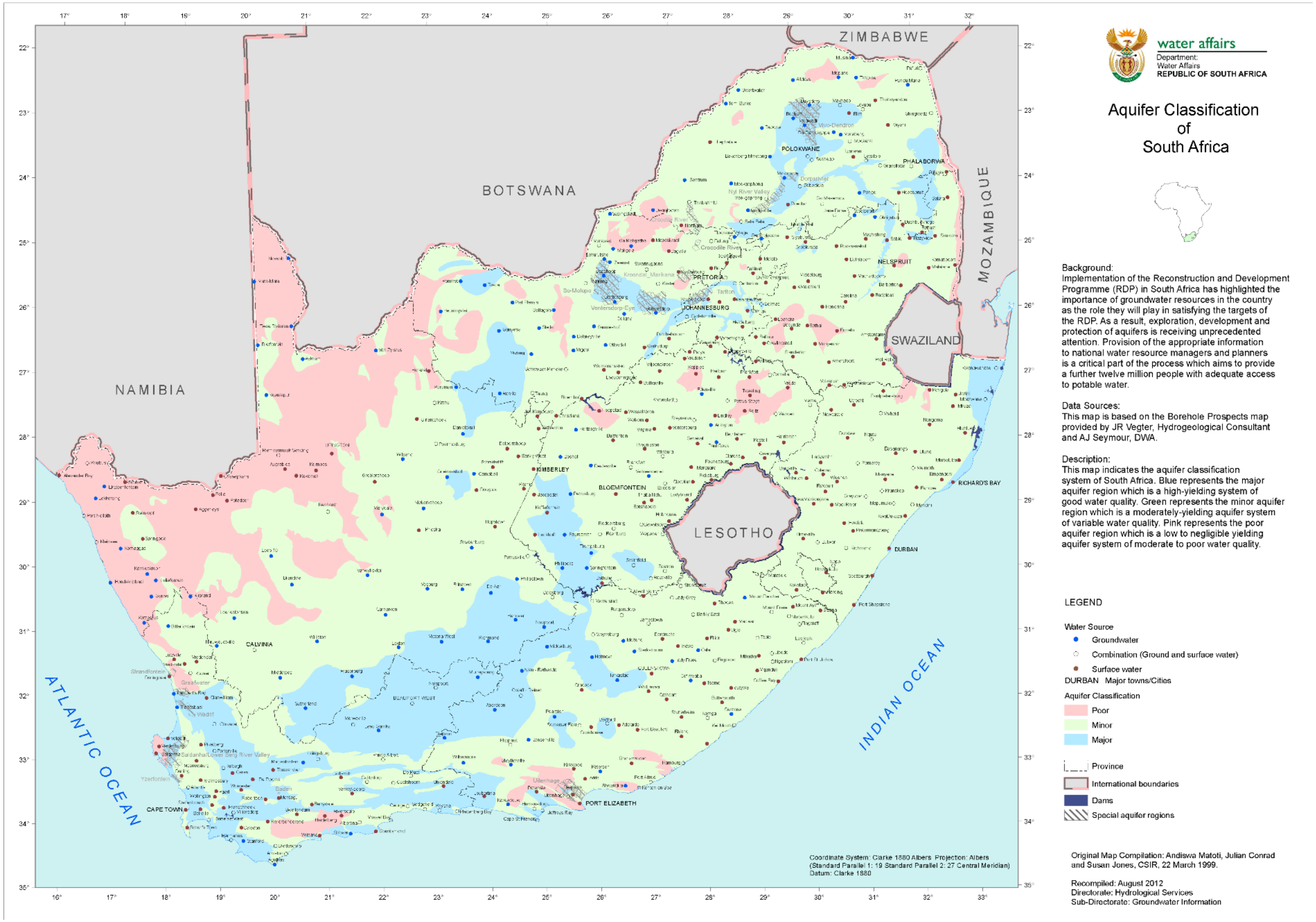


FIGURE 20 – AQUIFER CLASSIFICATION MAP OF SOUTH AFRICA

2.2.6 Aquifer vulnerability map

The aquifer vulnerability map indirectly provides insight into groundwater occurrence and therefore is included as part of the review of existing datasets. Aquifer vulnerability refers to the tendency or likelihood for contamination to reach a specified position in the groundwater system after introduction at some location above the uppermost aquifer. The DRASTIC aquifer vulnerability method generally makes use of seven (7) factors to calculate the vulnerability index value (Aller et al., 1987). The groundwater vulnerability of South Africa was also considered in terms of the DRASTIC method of assessment and produced by Parsons and Conrad (1999). The method considers the following factors which control the vulnerability of an aquifer to contamination from surface:

TABLE 2 - INPUTS FOR DRASTIC ASSESSMENT

Parameter	***Input dataset***
Depth to water table (D)	126 263 groundwater levels from the NGDB (for 4 280 of these, the mean groundwater level was calculated from time-series data) were interpolated to a groundwater level grid.
Recharge (R)	Recharge calculated as part of GRAII-3 project.
Aquifer material (A)	1:1 million Geology from CGS
Soils (S)	WR90 soils data set
Topography and slope (T)	DWAF 20m DTM resampled to 1X1km
Impact of the vadose (unsaturated) zone (I)	1:1 million Geology from CGS
Hydraulic conductivity (C)	1:1 million Geology from CGS

The overall DRASTIC equation is presented below, where r denotes the rating and w the weight for each of the inputs:

$$Index = D_r D_w + R_r R_w + A_r A_w + S_r S_w + T_r T_w + I_r I_w + C_r C_w \quad (1)$$

The ratings and weights used in the calculation are defined by Parsons and Conrad (1998). This vulnerability index is used to determine the aquifer's vulnerability to pollution and the index range from 1 to 200, where 200 represents the theoretical maximum aquifer vulnerability. The aquifer vulnerability map of the Limpopo Province is shown in Figure 21. The Vegter regions are once again shown for reference purposes.

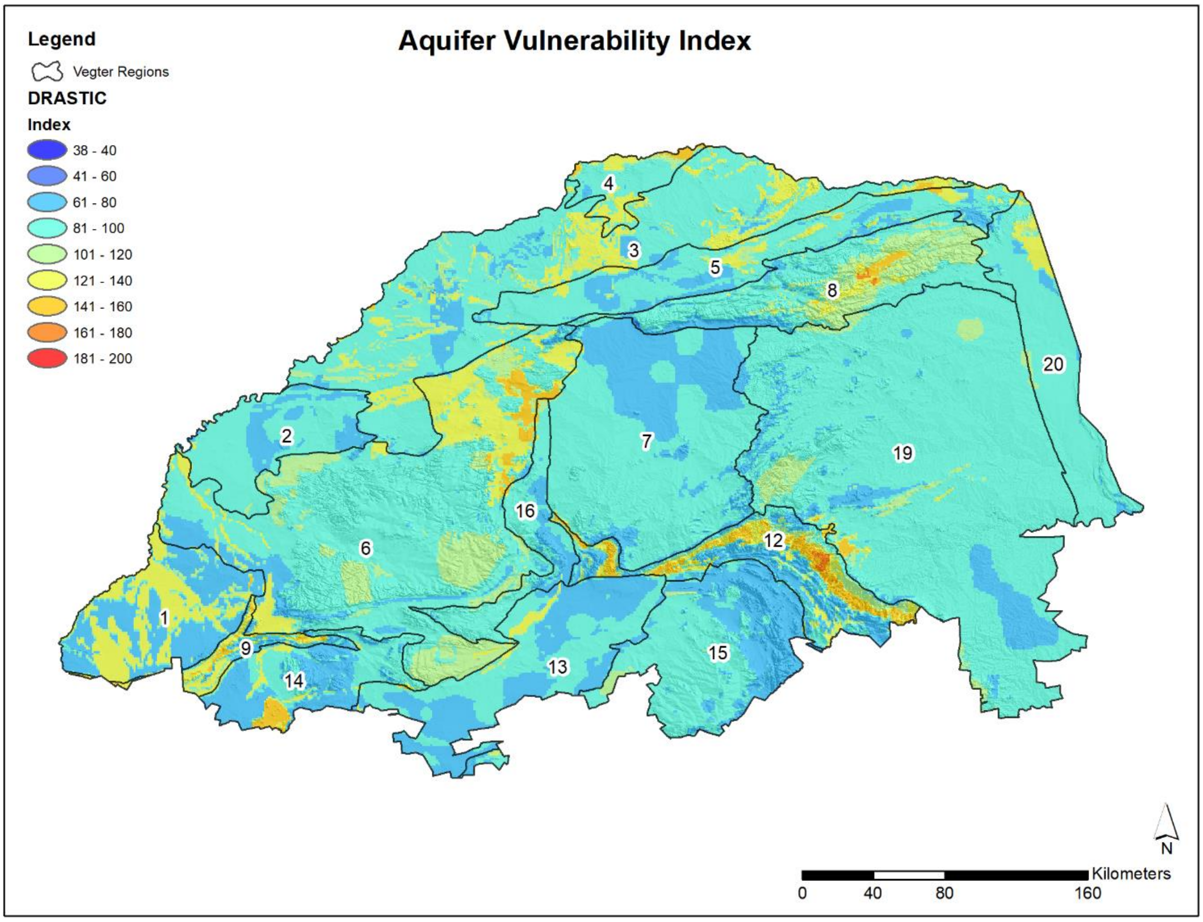


FIGURE 21 – AQUIFER VULNERABILITY MAP OF THE LIMPOPO PROVINCE

2.2.7 Inferred aquifer parameters

Some inferred data sets also exist, where a parameter is inferred from another. An example of this is where Dennis and Dennis (2011) estimated transmissivity (Figure 22) and aquifer storage coefficient (Figure 23) from the aquifer types obtained from the geohydrological maps. Aquifer parameters govern the yield that can be delivered by an aquifer and are therefore an important dataset if available.

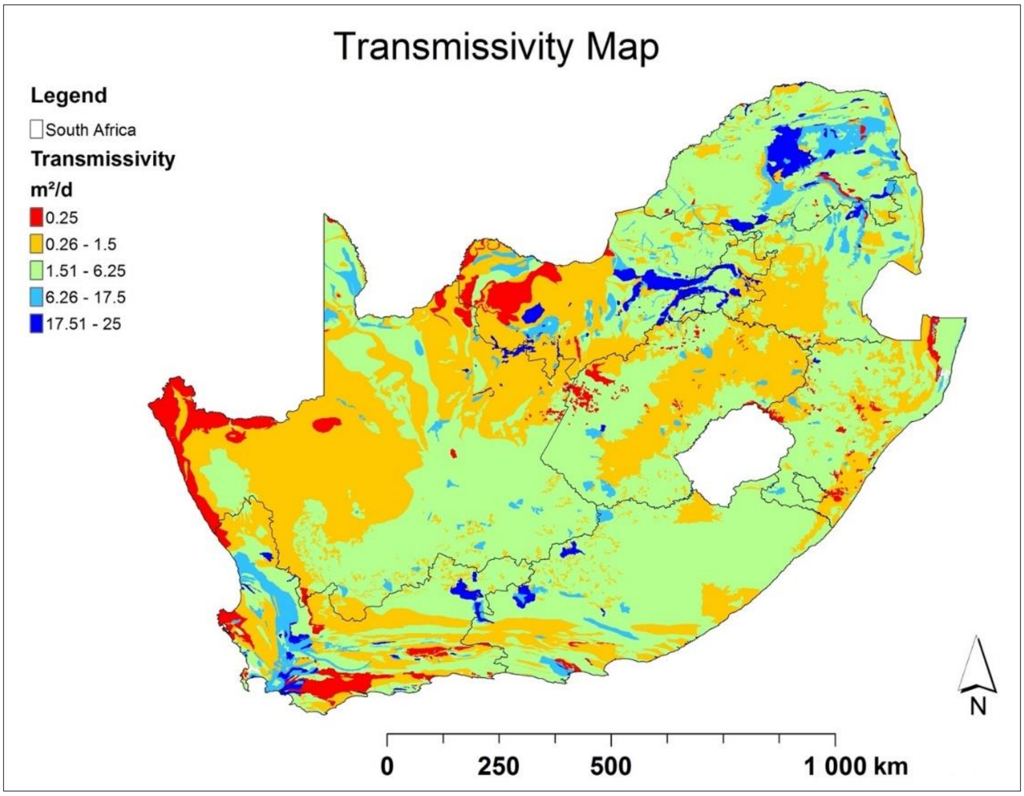


FIGURE 22 – INFERRED TRANSMISSIVITY MAP OF SOUTH AFRICA (DENNIS AND DENNIS, 2011)

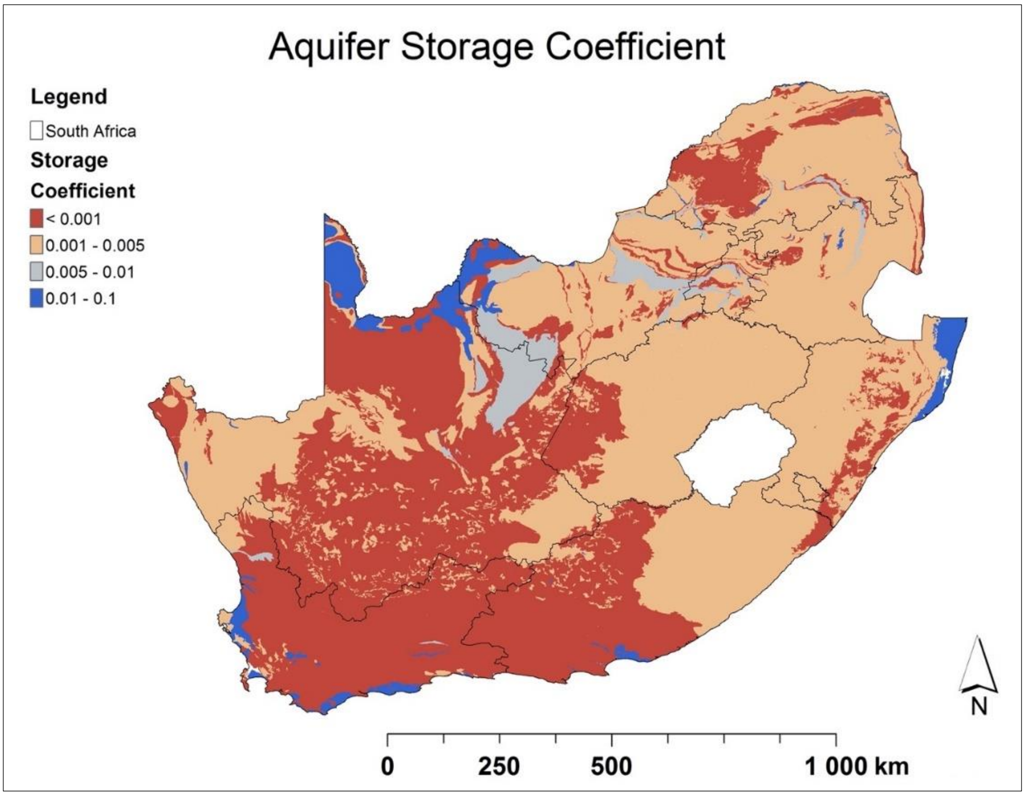


FIGURE 23 – INFERRED STORAGE COEFFICIENT MAP OF SOUTH AFRICA (DENNIS & DENNIS, 2011)

2.2.8 Limpopo Groundwater Atlas

The Limpopo Groundwater Atlas (Dennis & De Klerk, 2012) portrays the following parameters for the whole of the Limpopo Province, but has been divided into grid cells to offer a higher resolution view of the maps:

- Aquifer firm yield (based on the AFYM which is discussed in a later section)
- Surface geology with lineaments and boreholes (1:50,000 geological maps and GRIP data)
- Groundwater levels with recharge (GRIP data and recharge based on chloride values)

The purpose of the atlas was to update the water level, recharge and aquifer yield map based on the latest GRIP data and the AFYM. Example pages of the atlas are presented in Figure 24 to Figure 26.

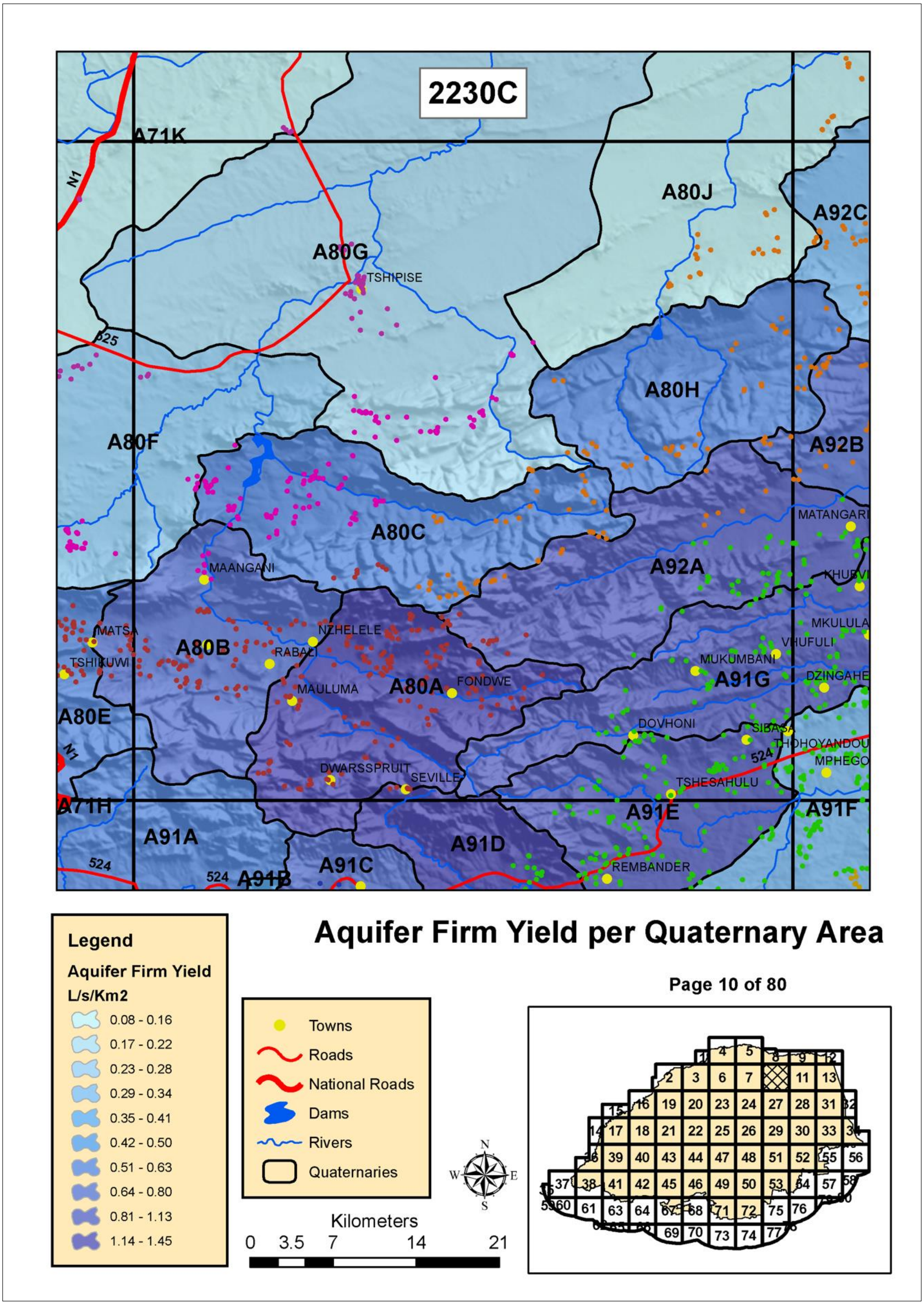


FIGURE 24 – LIMPOPO ATLAS 2230C AQUIFER FIRM YIELD PER QUATERNARY (DENNIS & DE KLERK, 2012)

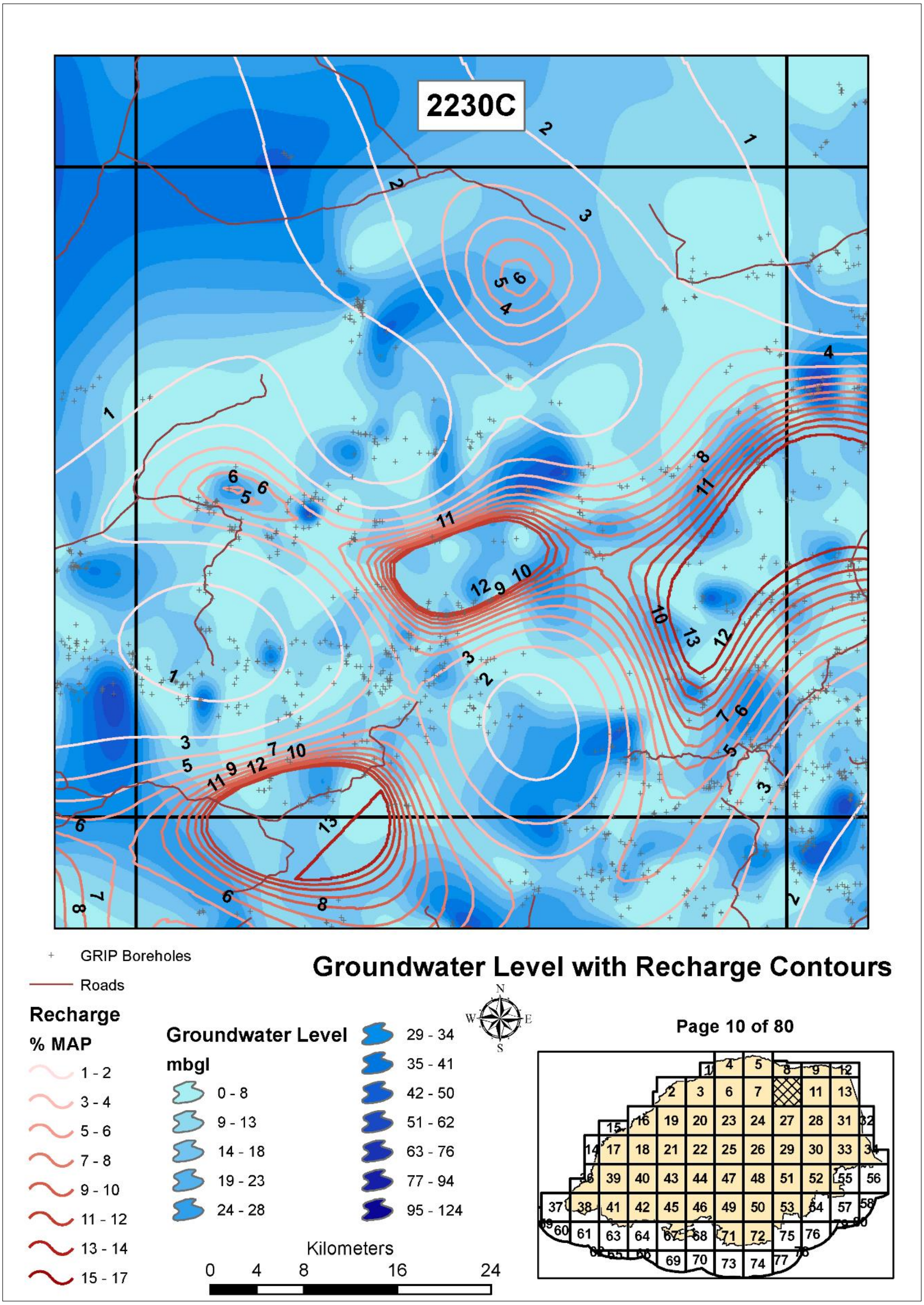


FIGURE 25 – LIMPOPO ATLAS 2230C GROUNDWATER LEVEL WITH RECHARGE (DENNIS & DE KLERK, 2012)

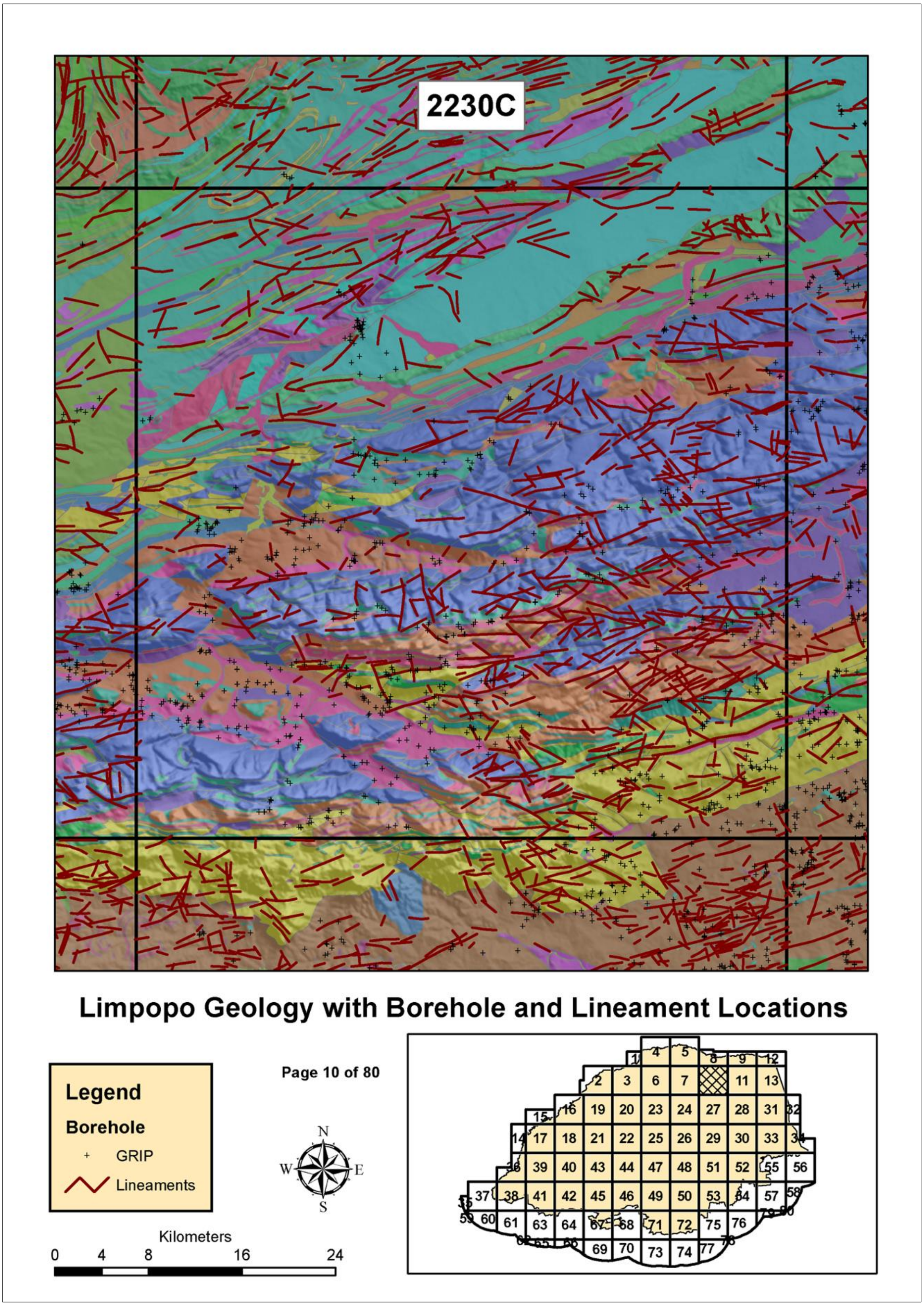


FIGURE 26 – LIMPOPO ATLAS 2230C GEOLOGY, BOREHOLES AND LINEAMENTS (DENNIS & DE KLERK, 2012)

2.3 Aquifer models

The harvest potential (Baron *et al*., 1998) and the GRA-II (DWAF, 2005) are commonly used by hydrogeologists to determine the average sustainable yield of an aquifer. Since GRA-II cover the whole of South Africa, it is almost impossible to verify all the data as it would simply cost too much. The GRA-II dataset is based on algorithms and calculations applied to the available data at the time and the resulting dataset is considered static since users cannot readily update the results as new information becomes available.

Groundwater modelling is seen as a specialist field and to set up a detailed groundwater model for the whole of South Africa is not plausible. Various detailed groundwater models do exist, but these only represent a fraction of the total area of South Africa. A simplified model is required that can be executed by non-modellers for any area in South Africa. Two such models are discussed in the sections that follow.

2.3.1 Groundwater Resource Directed Measures Water Balance

The initial Groundwater Resource Directed Measures (GRDM) groundwater model was represented by a simplistic water balance equation (Parsons & Wentzel, 2005) as shown in Equation 2, where Q represents volume per annum. By implication, all recharge excluding the Basic Human Need (BHN) and ecological reserve, which is protected by the water law (South Africa, 1998), can be extracted, which cannot happen in practice, therefore this model will over-predict the allocable volume that can be extracted from the aquifer.

$$Q_{Allocable} = Q_{Recharge} - Q_{BHN} - Q_{Ecological} \quad (2)$$

A graphical representation of the annual water balance model is shown in Figure 27.

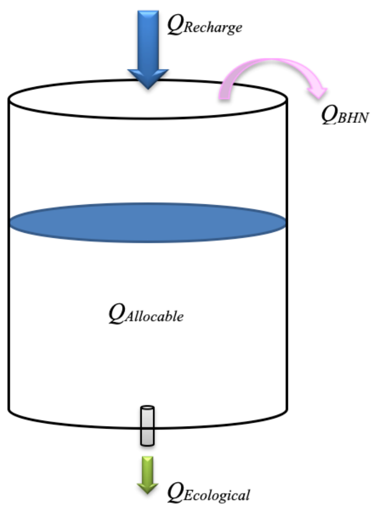


FIGURE 27 – GRDM WATER BALANCE MODEL

2.3.2 Aquifer Firm Yield Model

Murray *et al*. (2012) developed the concept of aquifer firm yields for an aquifer. This newly-developed Aquifer Firm Yield Model (AFYM) relies on the GRA-II dataset for model input and by implication can only model at the quaternary level. The authors have overcome this problem by creating a representative unit comprising fractions of single or multiple quaternaries based on the study area delineated by the user. The premise of the model is the Saturated Volume Fluctuation (SVF) equation that is valid for the saturated zone. Note this model applies to unconfined aquifers and therefore only applies to the shallow aquifer system. The formulation of SVF model translates a change in volume to a change in head as shown in Equation 3.

$$h_i = h_{i-1} - \frac{(Q_{re} - Q_e - Q_b - Q_{res} - Q_p)}{S_y A} \tag{3}$$

where

h_i	Head at month *i* (m)
h_{i-1}	Head at previous month
Q_{re}	Effective recharge rate in month *i* (m^3/month)
Q_e	Evapotranspiration rate in month *i* (m^3/month)
Q_b	Baseflow rate in month *i* (m^3/month))
Q_{res}	Reserve in month *i* (m^3/month))
Q_p	Abstraction rate in month *i* (m^3/month)
A	Area of aquifer (m^2)
S_y	Specific yield

The AFYM differs from the GRDM water balance model in the following respects:

- The AFYM accounts for the temporal nature of rainfall by making use of long-term monthly rainfall starting in 1920 available from the GRA-II project. The model is therefore run in monthly time step as opposed to the GRDM annual average model. This allow the model results to consider periods of drought.
- The AFYM model includes a temporal component for groundwater contribution to baseflow for the ecological reserve component, which is modelled making use of the Herold baseflow separation technique (Xu & Beekman, 2003).
- Evapotranspiration of the riparian zone is included in the AFYM, which is based on the FAO Penman-Monteith method (1992, cited in Schultze, 1997).
- Lastly the AFYM does not allow for the total abstraction of recharge out of the model. This is controlled by the following two mechanisms:

1. The concept of a “dead storage level” is introduced in analogy to the dead storage of a dam; the level at which no more water can be abstracted. Once this level is reached no more abstraction is allowed.
2. When abstraction takes place and groundwater contribution to baseflow is being used, no more abstraction is allowed.

The AFYM model remains a lumped box model as shown in Figure 28, but with the addition of temporal data and the implementation of the concept of dead storage for an aquifer.

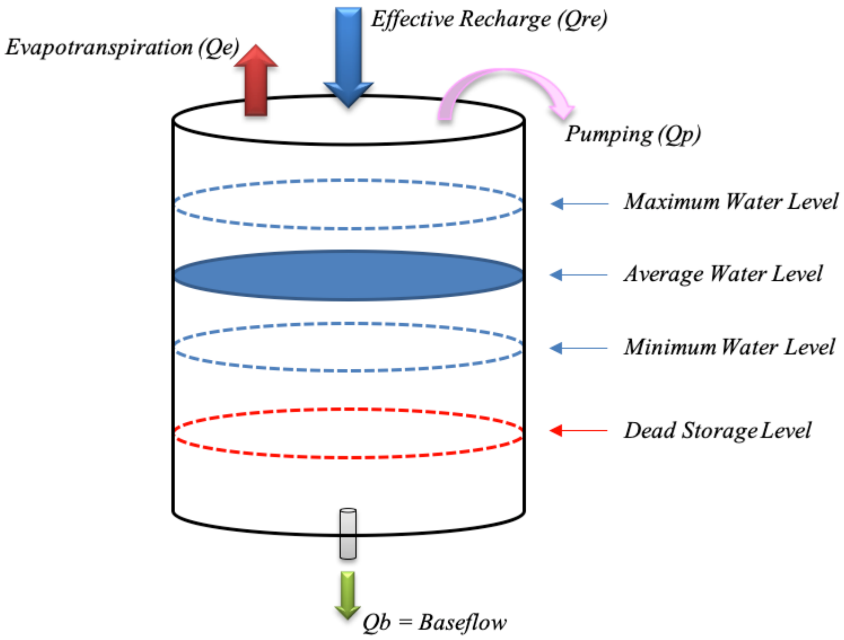


FIGURE 28 – AQUIFER FIRM YIELD MODEL (ADAPTED FROM MURRY ET AL., 2012)

The AFYM allows users to update model inputs for the specific area of interest to estimate the firm yield. Figure 29 shows the comparison of the AFYM with GRA-II and the Harvest Potential if the default GRA-II values are used for input to the AFYM in the Karoo. The benefit of the AFYM model being able to be updated by users is clear versus the static datasets of the GRA-II and Harvest Potential.

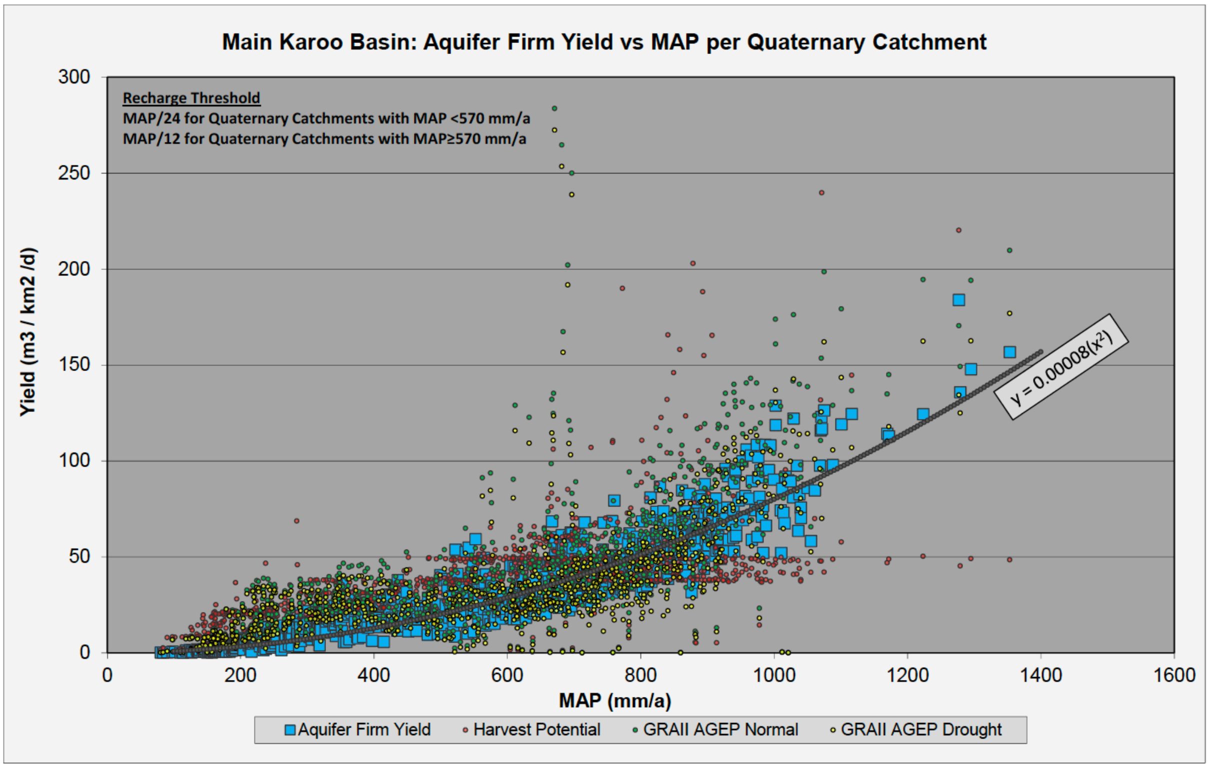


FIGURE 29 – COMPARISON OF THE AFYM WITH GRA-II AND HARVEST POTENTIAL (MURRY ET AL., 2012)

2.4 Interpolation techniques

2.4.1 Preamble

Delineation algorithms are based on the interpolation techniques applied to a specific parameter. It is a well-known fact that different interpolation techniques can produce different results from the same dataset as shown Figure 30. Data density plays a major role when interpolating and few interpolation algorithms are suited to extrapolate to areas where data distribution is sparse. The aforementioned can lead to large errors in the estimations.

It is important to understand that different interpolation methods have their strengths and weaknesses when using different datasets. It is not correct to generalise that a given interpolation method is better than the other without considering, the type and nature of the dataset and phenomenon involved. Therefore, it is important to select an appropriate interpolation technique when delineating.

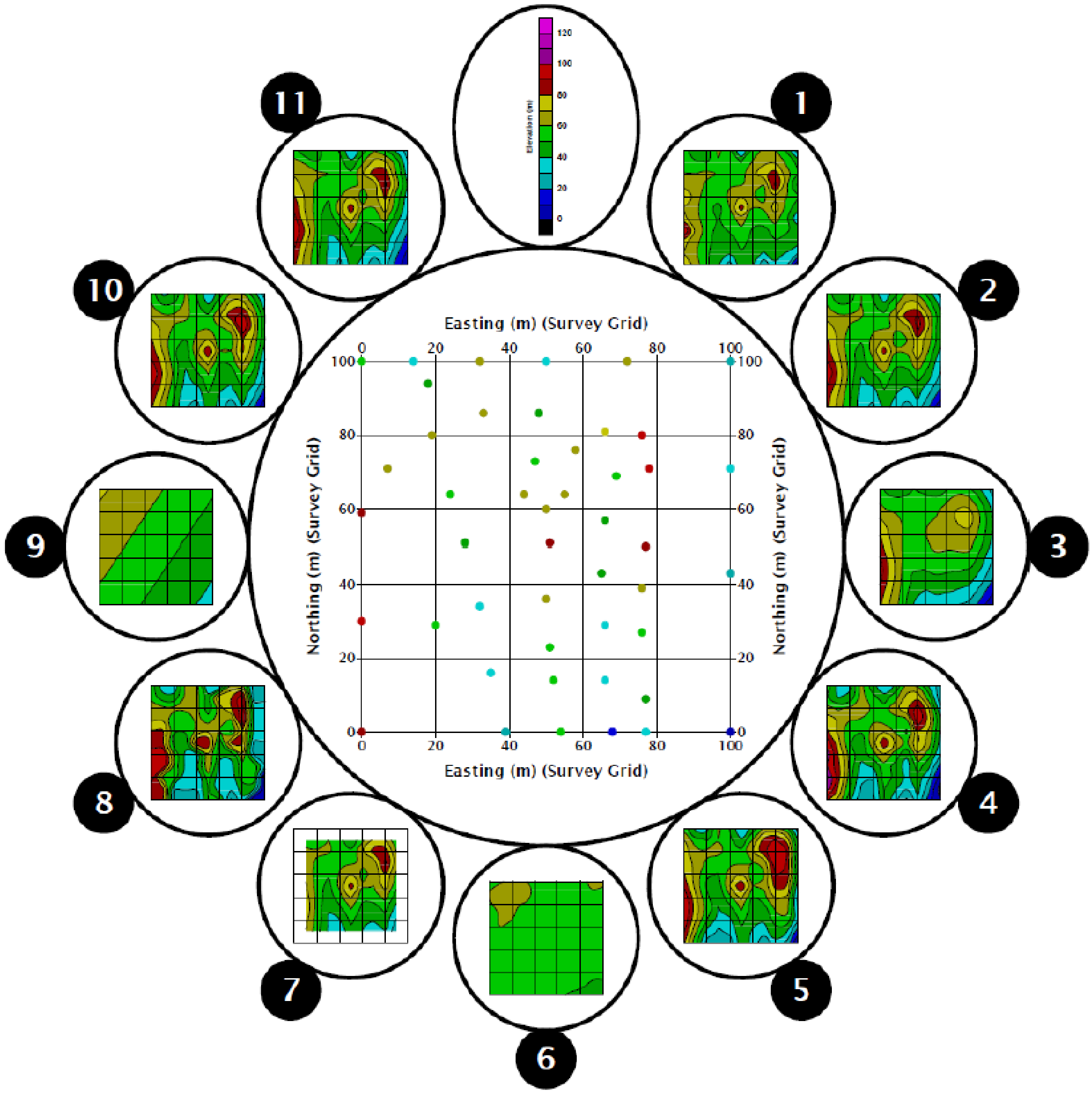


Figure 30 – Interpolation results from the same dataset (taken from Golden Software Surfer Documentation)

Different spatial interpolation methods have been developed in different domains for different applications. Tobler's first Law of Geography states that everything is related to everything else, but near things are more related than distant things. This forms the general principle of many interpolation methods. Spatial interpolation methods are classified into two major groups (Ikechukwu *et al*., 2017):

Type 1 - Mechanical/deterministic/non-geostatistical

Type 2 - Linear/stochastic/geostatistical

Three interpolation methods are considered for the purpose of delineation and are discussed in the following sections.

2.4.2 Inverse Distance Weighting (Type 1)

The Inverse Distance Weighting (IDW) function should be used when the set of points is dense enough to capture the extent of local surface variation needed for analysis. IDW determines cell values using a linear-weighted combination set of sample points. The weight assigned is a function of the distance of an input point from the output cell location. The greater the distance, the less influence the cell has on the output value. The mathematical detail for this technique is presented in Appendix A.

2.4.3 Splines (Type 1) (After Ikechukwu *et al*., 2017)

Splines belong to a group of interpolators called Radial Basis Functions (RBF). Methods in this group include Thin-Plate Spline, Regularized Spline with Tension, and Inverse Multi-Quadratic Spline. These models use mathematical functions to connect the sampled data points. They produce continuous surfaces while limiting the bending of the surface produced to a minimum. RBF models are best employed in smooth surfaces for which the available sample data size is large as their performance is less than optimum for surfaces with appreciable variations spanning short ranges. RBF does not force estimates to maintain the range of the sampled data in these models.

Spline functions are the mathematical equivalents of the flexible ruler cartographers used, called splines, to fit smooth curves through several fixed points. It is a piecewise polynomial consisting of several sections, each of which is fitted to a small number of points in such a way that each of the sections joins up at points referred to as break points.

Splines are normally fitted using low order polynomials (*i.e.* second or third order) constrained to join up. The smoothing spline function also assumes the presence of a measurement error in the data that needs to be smoothed locally. Among the many versions and modifications of spline interpolators, the most widely used technique is the thin-plate splines as well as the regularized spline with tension and smoothing. The mathematical detail for this technique is presented in Appendix B.

2.4.4 Kriging (Type 2)

Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial field. Kriging also generates estimates of the uncertainty surrounding each interpolated value. In a general sense, the kriging weights are calculated such that points near the location of interest are given more weight than those farther away. Clustering of points is also considered, so that clusters of points are weighted less heavily (in effect, they contain less information than single points) and this helps to reduce bias in the predictions.

Kriging will in general not be more effective than simpler methods of interpolation if there is little spatial autocorrelation among the sampled data points (that is, if the values do not co-vary in space). If there is at least moderate spatial autocorrelation, however, kriging can be a helpful method to preserve spatial variability that would be lost using a simpler method. Kriging can be understood as a two-step process: first, the spatial covariance structure of the sampled points is determined by fitting a variogram; and second, weights derived from this covariance structure are used to interpolate values for unsampled points or blocks across the spatial field. The mathematical detail for this technique is presented in Appendix C.

2.4.4.1 The variogram

A variogram (sometimes called a "semi-variogram") is a visual depiction of the covariance exhibited between each pair of points in the sampled data. For each pair of points in the sampled data, the gamma-value or "semi-variance" (a measure of the half mean-squared difference between their values) is plotted against the distance, or "lag", between them. The "experimental" variogram is the plot of observed values, while the "theoretical" or "model" variogram is the distributional model that best fits the data. Variogram models are drawn from a limited number of "authorized" functions, including spherical, exponential, and Gaussian models as shown in Figure 31.

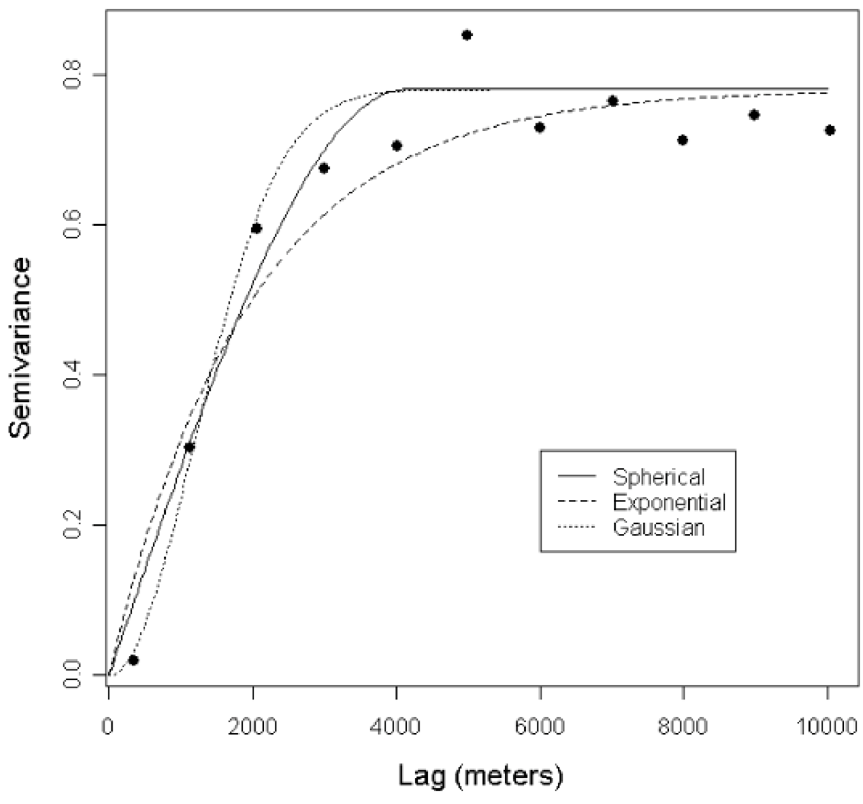


FIGURE 31 – EXAMPLES OF VARIOGRAM MODELS (BOHLING, 2005)

2.5 Groundwater occurrence delineation methodologies

2.5.1 South African Context

A delineation methodology was proposed for the sub-delineation of Vegter Region 65 (Northern Zululand Coastal Plain) by Dennis and Dennis (2014). The same methodology was applied by Holtzhausen (2018) in creating groundwater resource units for setting resource quality objectives for the Crocodile West Catchment.

The purpose of the delineation was to allow users to sub-delineate the Vegter regions based on available borehole parameters. The resulting delineation should be areas that are similar in terms of the geohydrological character. One of the major challenges still remain to get the delineation result continuous across boundaries, since delineation depends on contouring algorithms that in turn produce a result based on a given dataset.

The delineation methodology applied by Dennis and Dennis (2014) was based on the frequency analysis of available borehole parameters. This frequency analysis was conducted on the scale of the study area and the final borehole parameters chosen based on availability is presented in Table 3.

TABLE 3 – AVAILABLE BOREHOLE PARAMETERS FOR DELINEATION (DENNIS AND DENNIS, 2014)

Parameter	***Comment***
Water level	Average borehole water level measured from surface to water level position.
Blow yield	During the drilling of a borehole, water strikes are encountered and the “blow yield” measured gives an indication of the sustainable yield of the borehole.
Water strike depth	During the process of drilling, water strikes are encountered at specific depths. For the purposes of delineation, the deepest water strike for each borehole was used.
Water quality (EC)	The EC (Electrical Conductivity) of a borehole is an indication of the salinity of the water. Very saline water is associated with deep stagnant water, but can also be due to the mineralogy of the host rock.

A frequency analysis of the aforementioned parameters (Table 3) was done in such way to ensure that the histogram bin selection resulted in the majority of the dataset (> 90%) residing in the first four bins. This was done to combine the aforementioned parameters in a linear combination to calculate a delineation class. A summary of this frequency analysis of the example study area is shown in Table 4.

TABLE 4 – SUMMARY OF FREQUENCY ANALYSIS FOR VEGTER REGION 65 (DENNIS AND DENNIS, 2014)

Bin No	***Water Level***		***Blow Yield***		***Water Strike***		***EC***	
	Bin	*Count*	*Bin*	*Count*	*Bin*	*Count*	*Bin*	*Count*
1	5	1498	8	307	5	194	60	165
2	10	250	16	240	10	1216	120	49
3	15	54	24	474	15	85	180	17
4	20	28	32	656	20	26	240	14

A delineation class was formulated based on the water level, blow yield, water strike and EC parameter frequency analysis. Since the bin selection was done to ensure the bulk of the dataset is contained within the first four bins, only the bin number and associated range was of importance for the delineation process. The delineation class calculation applied is presented in Equation 4.

$$Delineation\ Class = \frac{Level + (5 - Yield) + Depth + EC}{4} \tag{4}$$

Each of the parameters in Equation 4 correspond to the physical bin number ranging from 1 to 4 which implies the minimum delineation class is 1 and the maximum delineation class is 4. Note that the yield bin was reversed to group high yields with shallow water levels and shallow water strikes, as this was the trend observed in the dataset for the example study area.

The visualisation of the delineation class is required to transform point data to representative polygon regions. The principle of Voronoi polygons involves generating a Voronoi polygon around each borehole and assigning the calculated delineation class to the generated polygon. Once this is done adjacent polygons with the same delineation class values are merged and this represents the simplification part of this operation. This process is shown for the example study area in Figure 32 and it is evident that a high borehole density is present in the northern part of the study area, which results in more detailed areas being generated.

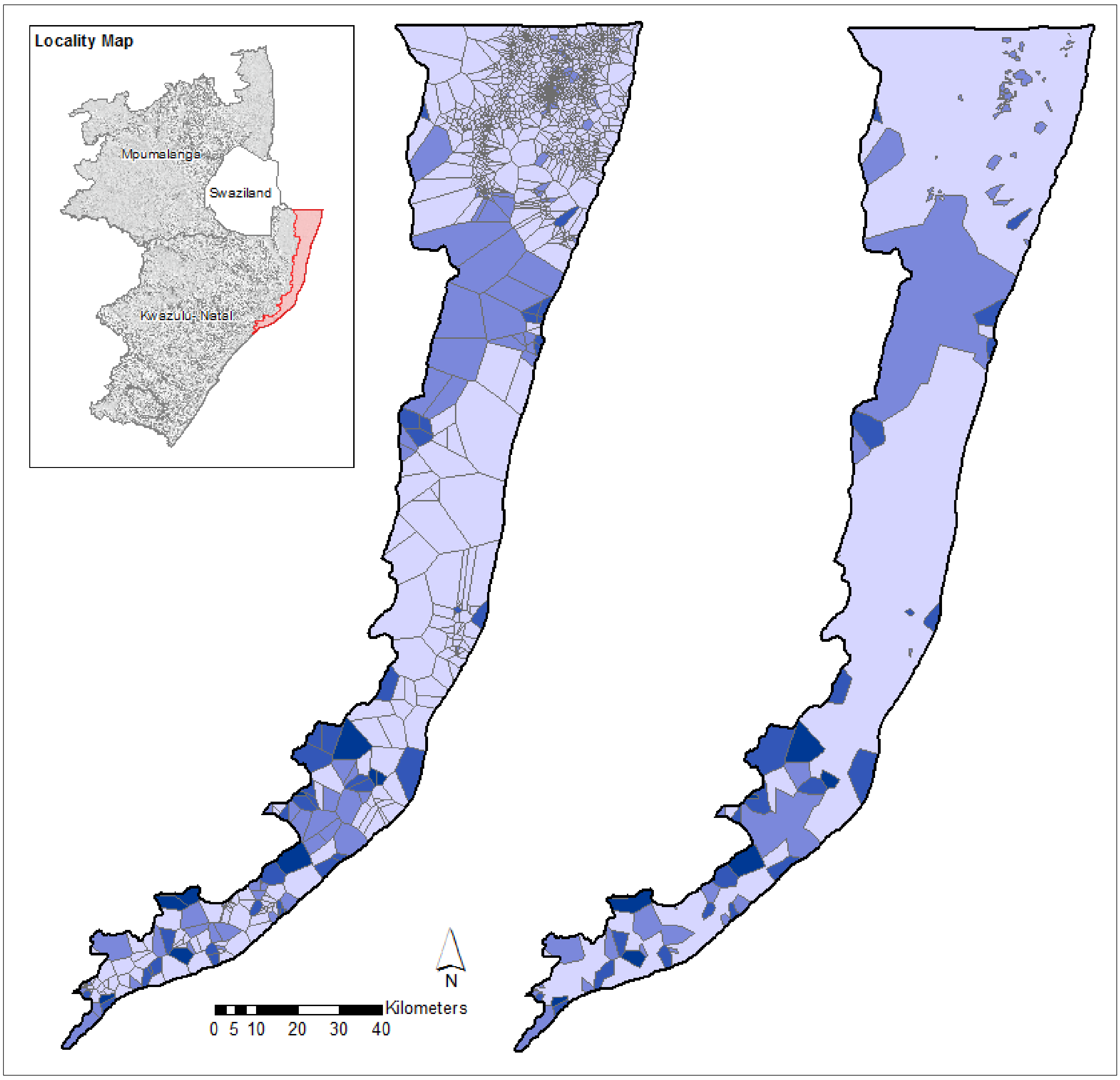


FIGURE 32 – VORONOI POLYGON SIMPLIFICATION EXAMPLE (DENNIS & DENNIS, 2014)

The final delineation map of the example study area is shown in Figure 33 together with the water level distribution. The importance of good data distribution is evident in this map. The final result is a delineation represented by four classes where each class is representative of the following four parameters and typical ranges as identified in the frequency analysis:

- Water level
- Blow yield
- Water strike
- EC

Although the delineation methodology seems to work for the example study area presented, it has the following flaws when applied over large areas:

- It is not always the case that the deeper water levels are associated with the higher EC values or that the shallow water levels are associated with higher blow yields. Therefore, it becomes increasingly difficult to formulate four unique classes in terms of the parameter combinations.
- The delineation class method considers the whole range of values for each parameter and transforms them to a delineation class based on the results of the frequency analysis. Therefore, when applied to large areas with a wide range of parameter values, delineation detail is lost in certain areas and will only be represented by a single class.
- Continuity is lost in delineation when considering adjacent areas which exhibit a contrast in parameter values, since the delineation classes for each area is formulated in terms of the frequency analysis of the total range of values.

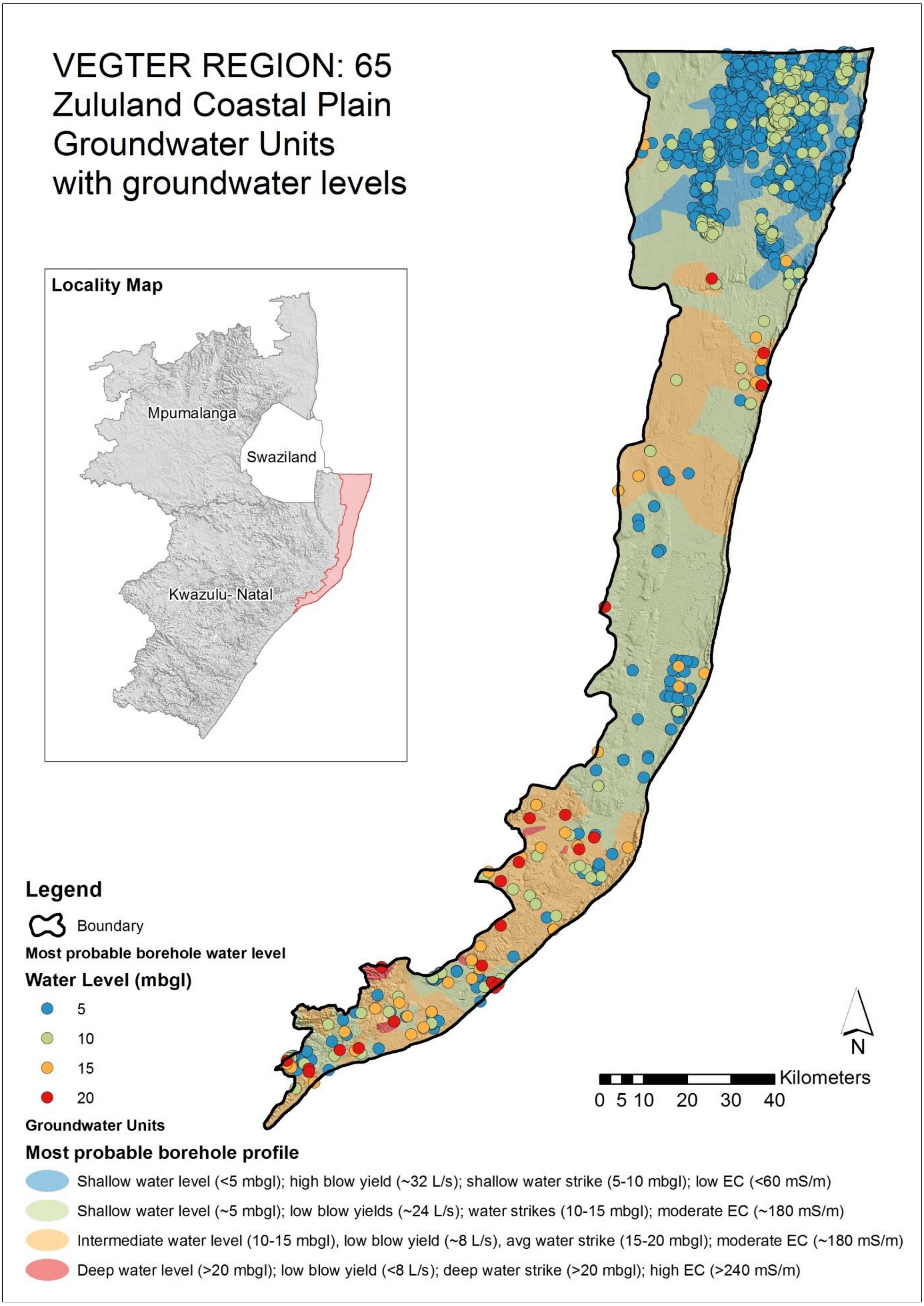


FIGURE 33 – VEGTER REGION 65 GROUNDWATER OCCURRENCE (DENNIS & DENNIS, 2014)

2.5.2 International literature

Internationally a vast number of studies have been done and published where mostly geographical information systems and spatial methods are utilised to delineate groundwater potential zones and to model groundwater occurrence. These indirect methods are used to model or estimate the processes and characteristics that affect the storage and transport of underground water (Javed & Wani, 2009).

The rapid analysis techniques are effective as a means of time and cost saving to narrow a large regional study down to smaller clusters when groundwater occurrence is studied. This approach is followed for various studies on developing countries mainly due to the lack of budget and resources. Silwal and Pathak (2018) mentioned similar studies completed for Iran, Iraq, Egypt, India and Nepal. The majority of these studies use factors such as lithology, rainfall, lineaments, slope, drainage etc. with thematic layers where a weight is assigned according to its relation to groundwater occurrence.

Integrated multi-criteria decision analysis (MCDA), remote sensing and GIS techniques are used for the purpose of the delineation of groundwater potential zones. The results normally lead to the characterization of different groundwater zones divided into three to six brackets according to their predicted yield (for example: low, medium and high) depending on the discretion of the authors. Most studies correlate its results with existing borehole data and in various cases more than 70% correlation is achieved (Fashae *et al*., 2013).

These correlations might be questionable especially when the existing borehole data is sparse and it is not clear if dry drilled boreholes were taken into consideration. The sample size of the existing borehole data could be a challenge and causes a small number of existing borehole information to be used when data is validated especially in areas where less amounts of exploration data is gathered. Therefore, it can be seen as an effective tool but not a solution to the real challenge of accurately predicting groundwater occurrence on a more localised scale. The importance of enough existing groundwater data with good distribution as mentioned in the previous section is again highlighted.

The conclusions reached from most studies showed that results or information gathered from additional projects should be used to update existing studies or datasets after the new information becomes available, but even more so when the results derived from the conclusions and recommendations were implemented or practically tested.

2.6 Statistical analysis of geohydrological data in the Limpopo Province

Holland (2011) investigated the variability of basement aquifer in the Limpopo Province making use of the Limpopo GRIP database. Holland (2011) found it necessary to split the Province into two main areas and referred to these areas as Limpopo plateau (Western portion) and the Letaba Lowveld (Eastern portion). These areas were delineated based on topography, surface drainages and geological domains.

Various statistics were used during this study including frequency histograms that illustrated the influence of dykes on transmissivity, cumulative frequency plots of borehole yield and transmissivity to characterise regional occurrence of groundwater and lineament frequency rose diagrams showing structural trends within an area. Holland (2011) identified potential factors that might have an influence on groundwater productivity and can be summarized as follows:

1. Geological setting which comprised the following analysis:
   - Lithological information obtained from borehole logs was analysed and it was found that Pegmatites and Amphibolite were clearly more productive than other geologies.
   - Grouping of boreholes according to geological or lithological setting and then using the corresponding mean values to compare them to each other. The highest borehole productivity was found to be within the primary alluvial aquifers and along the contact zones of granitic batholiths (inselbergs).
2. Influence of weathered layers and erosion on borehole productivity. No correlation was found between either transmissivity and weathering or between borehole yield and weathering. In addition, no concrete influence of erosion on groundwater could be established.
3. A third factor that was investigated was the topographic setting. As expected, the study found that the boreholes within the alluvial aquifers (along rivers) had above average transmissivities and yields in comparison with boreholes in mountainous areas. It was also concluded that topography and proximity to rivers and streams had a significant influence on the productivity of boreholes. Streams tend to follow zones of structural weakness like lineaments therefore more fractured, jointed and weathered geology can be expected in these areas.
4. Borehole productivity in relation to dykes (dolerite and diabase) was also investigated. It was found that the data from borehole logs indicated lower productivity where dykes were encountered than those that had no dykes present. Data from 1000 boreholes on the Limpopo plateau was analysed and it was found that drilling into a dyke would be a poorer target although drilling alongside (within 25m) from a dyke could still be a promising drill target. The opposite was found for another 1500 boreholes analysed in the Letaba Lowveld area where dykes in the proximity of boreholes had no influence on the productivity of these boreholes. Ultimately this indicated the fact that dykes should only be used as drilling targets after considering the area where they are located. The azimuth (dyke strike or bearing) was also analysed and found to be evidently higher yielding targets when striking North-East (30˚ to 45˚) or East-North-East (75˚ to 90˚).
5. The analysis of the relationship of borehole yields to lineaments indicated that lineaments have a positive influence on the productivity of boreholes and it should definitely be considered in production borehole drilling programmes.

2.7 Groundwater drill targets

A discussion of possible drilling targets is warranted here. There exists a general impression that dykes and lineaments are good targets for high yielding boreholes and therefore they are targeted in the field.

2.7.1 Preamble

The permeability of underground rock is an important factor for the ability of the aquifer to transmit water. The occurrence of fractures or fault zones is necessary to transmit the underground stored water in a sufficient amount to boreholes or other abstracting systems (Riemann & Blake, 2010). Geophysics is used to determine the location of underground structures or to better understand the underground characteristics and compilation of the unknown formations below the surface. When such information is available, it can be used to determine more accurate locations of underground structures such as faults, dykes, fractures, depth of weathering etc. These fractures are normally associated with underground water which means that more accurate drill sites can be located by geohydrologists in order to save time and money on the drilling/prospecting for higher yielding boreholes.

2.7.2 Lineaments

DWS incorporated lineament analysis into the GRIP project and it is seen as very important for groundwater explorations. Geological remote sensing consultants use ASTER (Advanced Spaceborne Thermal Emission and Reflectance) satellite imagery or Landsat satellite imagery to interpret and map geological lineaments. These lineaments reflect features such as faults, fractures, joints, dykes, contact zones and linear branches of drainage systems (Holland, 2011). These lineaments are often targeted by geohydrologists by means of geophysical equipment in order to determine exact locations, widths, dips and direction of the geological structures.

2.7.3 Dykes

A common geological structure to target would be Dykes (Dolerite also known as Diabase) as mentioned by Holland (2011). Dykes are intrusive, vertical or inclined dark grey, very hard rock features that cut through the host rock, causing a “baked” zone in the sedimentary host rocks (Cairncross, 2004). Dykes are known to cause cracks in the “baked zone” of the host rocks which act as conduits of water, which then accumulates next to the hard dyke feature. These zones are typical good drill targets. In instances where the dykes are narrower (<10m in width) it can be seen as a good aquifer, storing the groundwater through joints and cracks that developed in them (Bromley *et al*., 1994).

2.7.4 Data analysis

Although lineaments and dykes may be good targets, various studies have shown that this is not the rule. The results of some of the example studies are summarised as follows:

- Vegter (2001) found that in the Mopane-Pontdrif area (Region No. 3) drilling on dyke contacts had the best success rate.
- Vegter (2003a) indicated that the success ratio as well as the percentage higher-yielding boreholes was greater in the absence of dykes. This was found in two regions e.g. Region No. 7 (Pietersburg plateau, named after the Limpopo capital now known as Polokwane) and in Region No. 19 (Lowveld) that was investigated.
- Holland (2011) investigated borehole productivity in relation to dykes and found that the data from borehole logs indicated lower productivity where dykes were encountered than those that had no dykes present in certain areas. Data on the Limpopo plateau was analysed and it was found that drilling into a dyke would be a poorer target although drilling alongside (within 25m) from a dyke could still be a promising drill target. The opposite was found in the Letaba Lowveld area where dykes in the proximity of boreholes had no influence on the productivity of these boreholes.
- Dennis and Dennis (2014) also showed that boreholes intersecting dykes in the Northern Zululand Coastal Plain were associated with lower yields than those not intersecting dykes.

From these examples it can be concluded that the general perception that dykes and their immediate vicinity are always the most favourable drilling targets are questionable. Ultimately these studies indicate that dykes should only be used as drilling targets after considering the area where they are located.

2.8 Conclusion

These various datasets and models provide insight into groundwater occurrence on a regional scale, but it is not practical to use on local scale for water siting. Most of the datasets presented are interrelated and use the same data source as base data - even the aquifer models presented. The base data for the Vegter and GRA datasets is actual borehole data, but this borehole dataset is considered “old” in today’s terms. Therefore, it is evident that a “live” computational approach would be beneficial to continually update the groundwater occurrence as more borehole data becomes available.

3 METHODOLOGY

3.1 Introduction

From the literature review presented in Chapter 2, the following guidelines followed for successful mapping of groundwater occurrence in the Limpopo Province:

- Ensure using the most up to date and comprehensive borehole database
- Use lessons learnt from previous data projects and don’t reinvent the wheel
- The final delineation / identification methodology should be scale sensitive, i.e. it should be able to indicate groundwater occurrence on a local scale

3.2 Data sources

3.2.1 Background

Public domain borehole information in South Africa is generally stored in the National Groundwater Archive (NGA) and the Groundwater Resources Information Project (GRIP) database of which both are centralized databases. Over and above the two mentioned public domain borehole databases, various users and consultancies maintain their own local borehole databases which are not available to the public, but which contain valuable information from a perspective of refining the existing public domain databases.

3.2.2 National Groundwater Archive

A central web-enabled, fully accessible database had to be developed by the Department of Water and Sanitation due to their legal obligation to ensure the protection, development, conservation and management of the groundwater resources. At first a National Groundwater Database (NGDB) was established and later replaced by the NGA in 2009 which was web-based. This web-enabled database allows capturing, viewing, modifying and extraction of groundwater related data by registered users (DWAF, 2009a). The NGA database structure is modelled on the Standard Descriptors for Geosites (DWAF, 2004) and therefore makes provision for various borehole related parameters to be stored.

This database is maintained by Department of Water and Sanitation (DWS) and users obtain the required information by submitting a geo-request. The NGA database relies on users to upload captured borehole information to the database, but this is not happening for various reasons, e.g. data collection costs money and consultants keep borehole information to themselves as they see it as a competitive advantage and users often complain about the accessibility of data from the NGA through the web interface and therefore do not see the need to upload any of their own borehole information.

Various DWS systems also interface to the NGA e.g. CHART (DWS, 2015) provides read-only access to data from NGA, which is an online solution to analyse and interpret groundwater, surface water, meteorological data and present information to effectively monitor groundwater and assist in decision-making.

The historic NGA borehole distribution for the country is shown in Figure 34 (DWS, 2017) and it is clear that Limpopo has a higher borehole distribution when compared to the rest of the country, mainly due to the Limpopo GRIP project discussed in the next section.

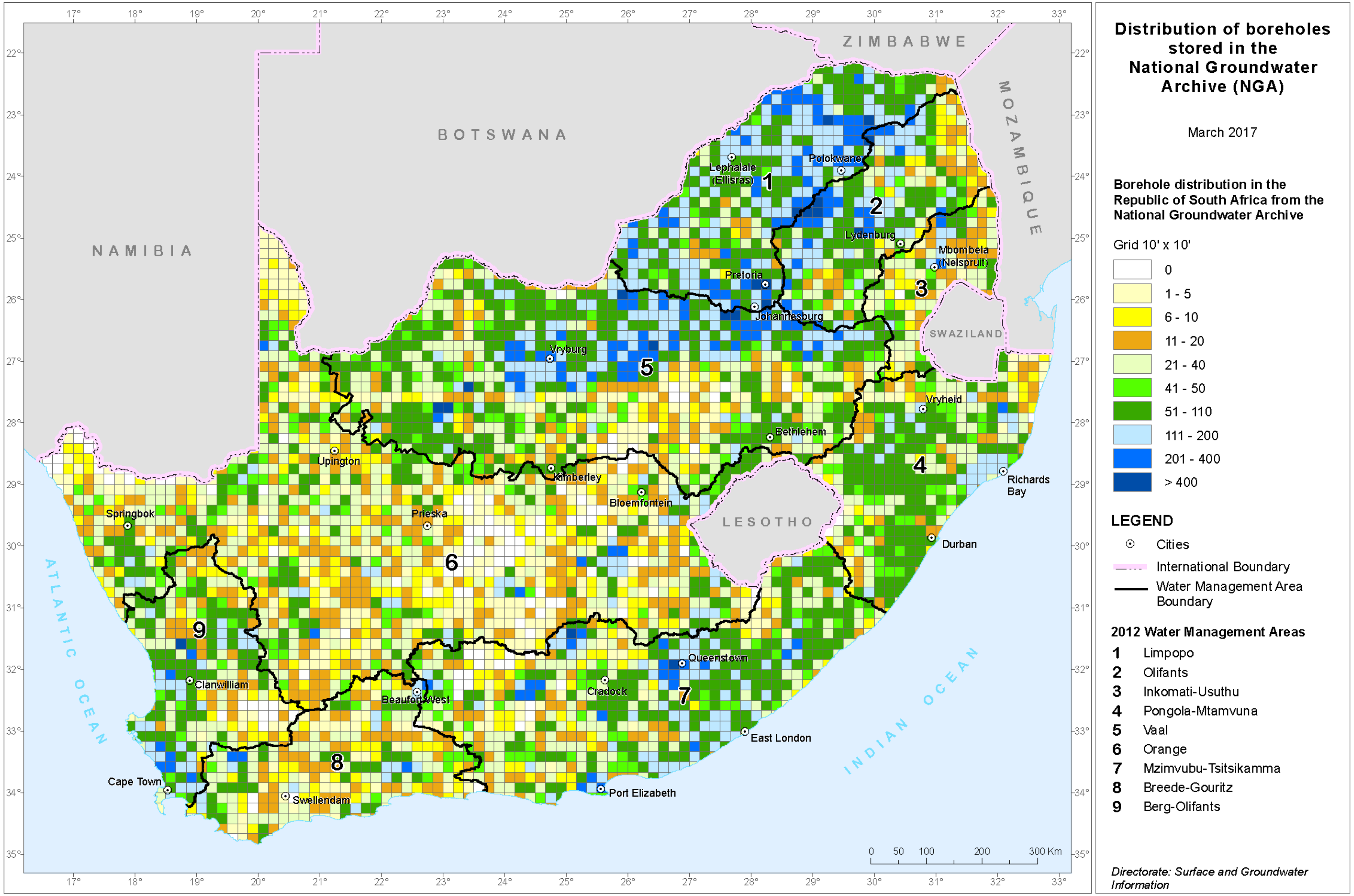


FIGURE 34 – NGA BOREHOLE DISTRIBUTION (DWS, 2017)

3.2.3 Groundwater Resources Information Project

The GRIP database was developed to capture borehole point data and is used for the management of groundwater with emphasis on water availability and aquifer characteristics. The GRIP was originally implemented in the Eastern Cape and Limpopo Province and would later be extended to KwaZulu-Natal. To date the GRIP has only been implemented within the Limpopo Province due to lack of funds and available human resources in other Provinces. For a national GRIP project to succeed it is paramount to receive support from government in the form of resources and funding (DWAF, 2009a). In addition, the project should be driven by people who understand the need, urgency and importance of verified data. The initial methods were developed by Dr Fanie Botha in his PhD thesis (2005) titled "A proposed method to implement a groundwater resource information project (GRIP) in rural communities, South Africa".

In general, the GRIP database is regularly updated through term-contracts set up by DWS and ultimately this data should also be uploaded to the NGA. The project was managed by appointing consultants within certain regions to perform the following tasks:

1. Ensure constant updates of the information
2. Assess available data and gather new data
3. Conduct yield tests and drilling of new boreholes within priority areas
4. Deliver a provincial groundwater planning report (DWAF, 2009b)

This data would then be uploaded to the central database and could be accessed by the consultants and the public through the DWS website. Some of the GRIP data already exists within the NGA, but to date not all GRIP data has been uploaded to the NGA.

As part of the GRIP project, a planning tool was developed to enable planners and managers to access data regarding boreholes within a certain project area or municipality. This also includes a geographical environment where data can be viewed and maps can be constructed by adding shape files, maps or even developing contour maps. In addition, a costing tool was also incorporated, for estimating the costs of developing a groundwater source in a certain area by using existing data from previous projects (Botha, 2005).

Since the GRIP data is considered as having a high confidence level and is uploaded to the NGA, only the GRIP database will be considered for the purposes of this study. A summary of the basic statistics for the GRIP database is presented in Table 5.

TABLE 5 – GRIP DATABASE BASIC STATISTICS

Database Feature	***Description***	***Count***
All boreholes	Total borehole count in the database.	26,912
Borehole water level (Figure 35)	The number of boreholes that have one or more water levels associated with it.	10,333
All water levels	The total number of water levels available in the database.	11,022
Borehole EC (Figure 36)	The number of boreholes that have one or more EC measurements	7,884
All EC readings	The total number of EC measurements available in the database.	12,057
Borehole chemistry (Figure 37)	The number of boreholes that have one or more chemistry analyses associated with it.	7,050
All chemistry	The total number of chemistry analyses available in the database.	8,701
Borehole yield (Figure 38)	The number of boreholes that have one or more yield values associated with water strikes.	10,002
All yield values	The total number of water strikes with yield available in the database.	10,425
Water strikes (Figure 39)	The number of boreholes that have one or more water strike associated with it.	2,386
All strikes	The total number of water strikes available in the database.	2,680
Borehole logs (Figure 40)	Total number of boreholes that have a borehole log.	3,403

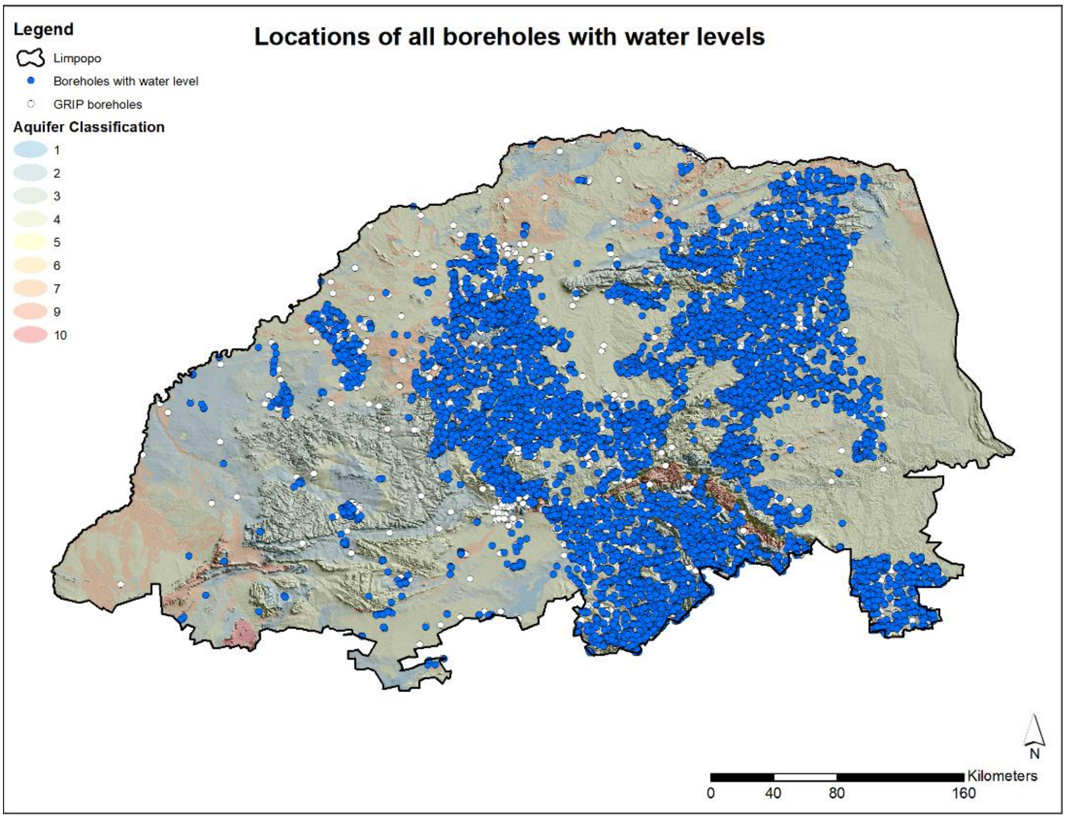


FIGURE 35 – WATER LEVEL DISTRIBUTION OF THE LIMPOPO GRIP DATABASE

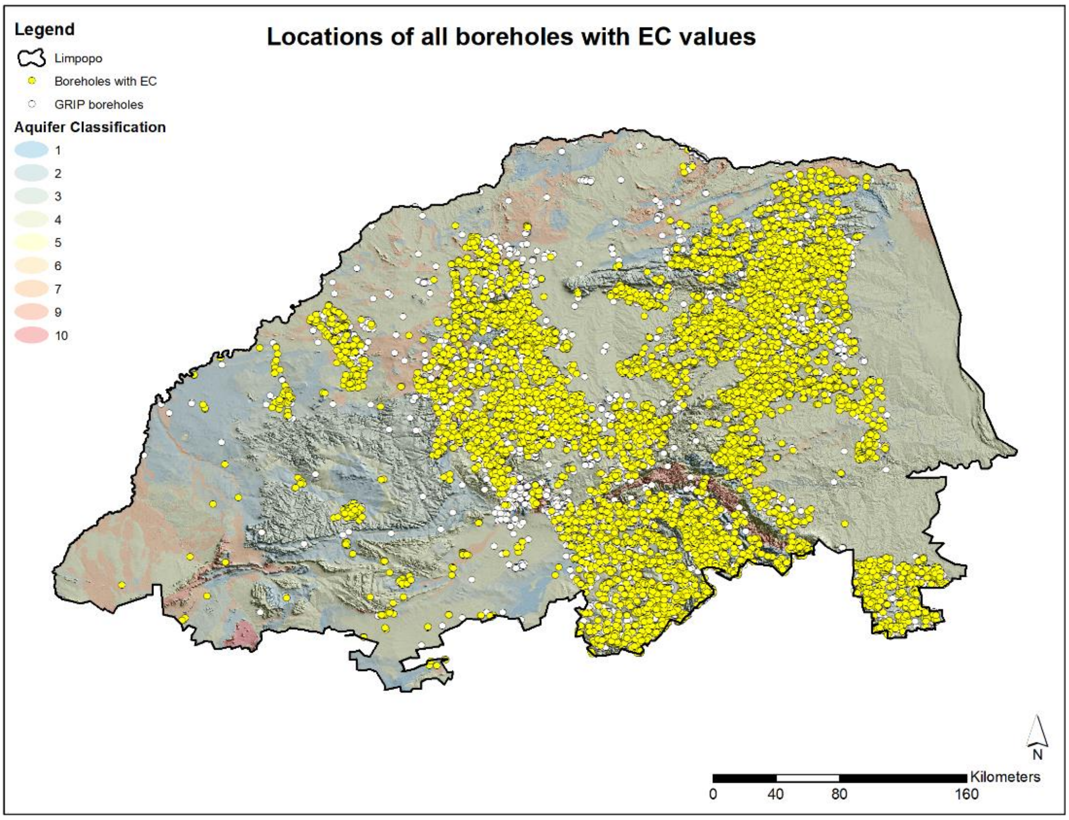


FIGURE 36 – EC DISTRIBUTION OF THE LIMPOPO GRIP DATABASE

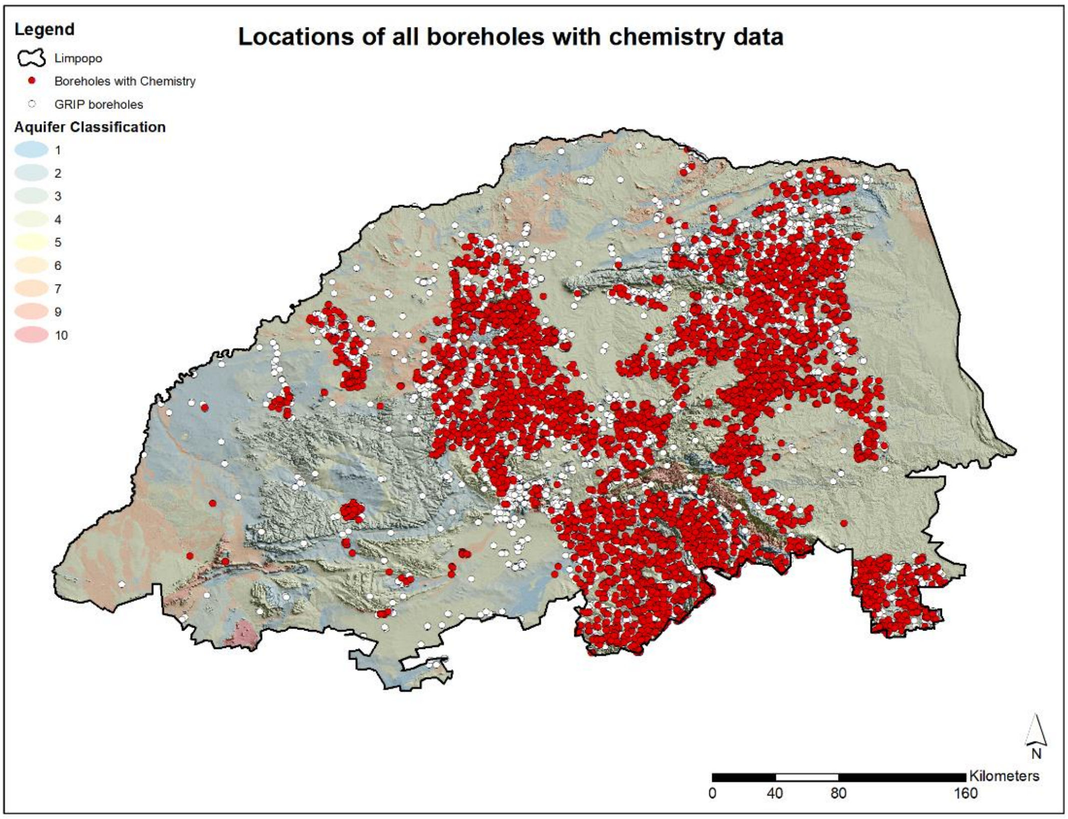


FIGURE 37 – CHEMISTRY DISTRIBUTION OF THE LIMPOPO GRIP DATABASE

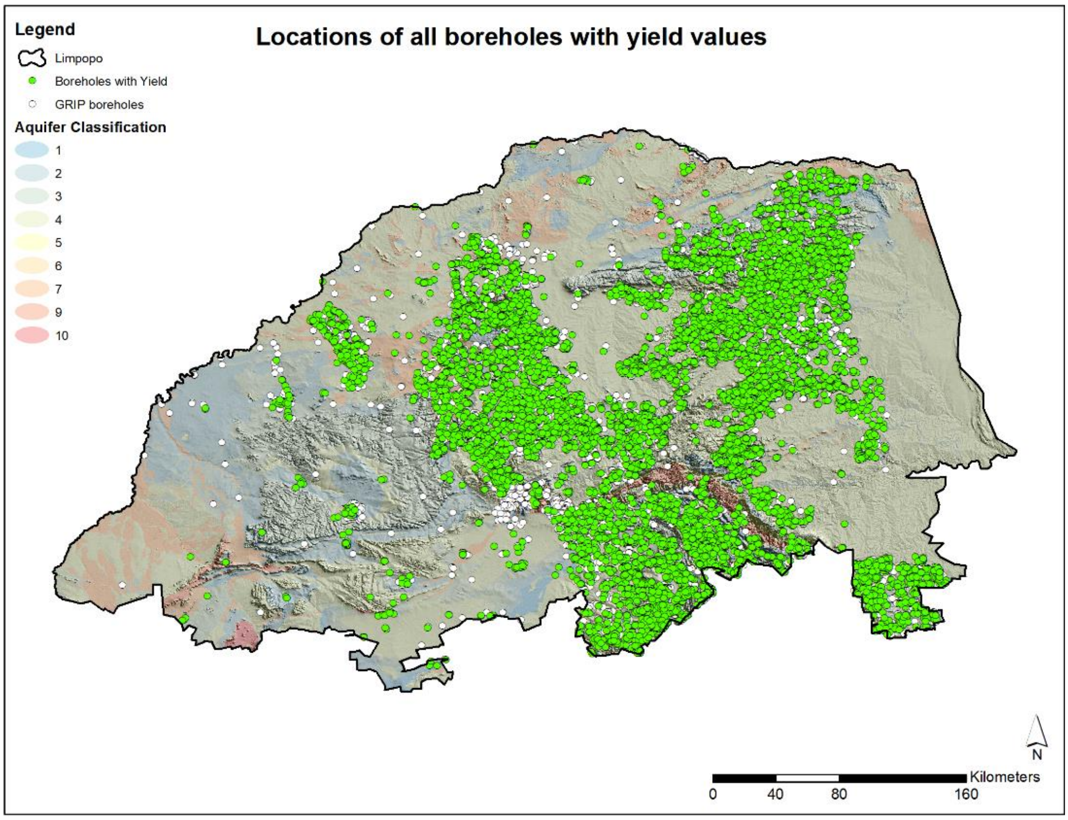


FIGURE 38 – YIELD DISTRIBUTION OF THE LIMPOPO GRIP DATABASE

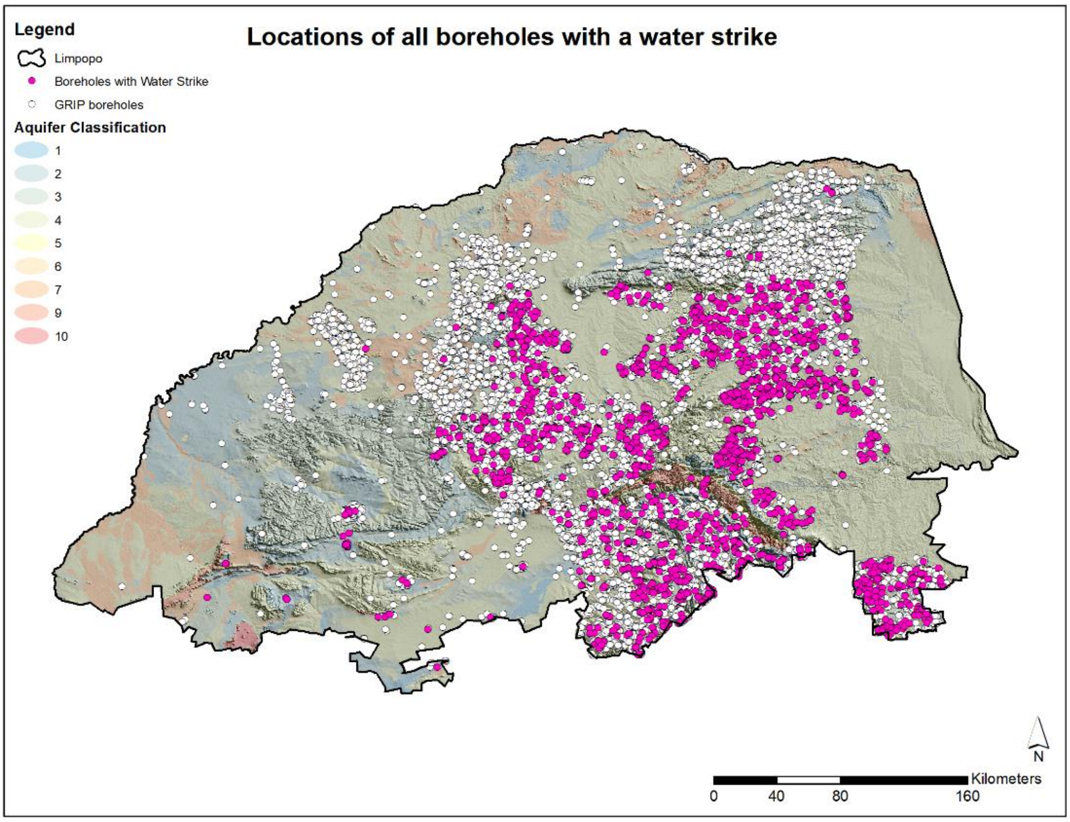


FIGURE 39 – WATER STRIKE DISTRIBUTION OF THE LIMPOPO GRIP DATABASE

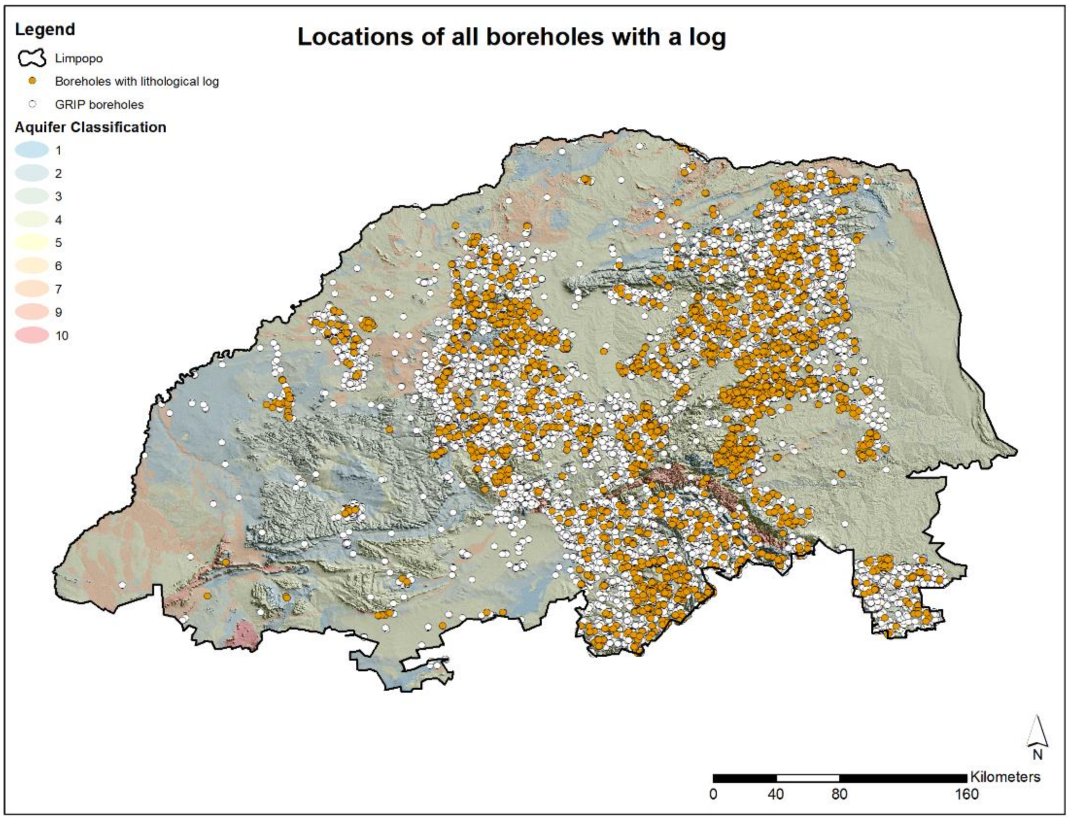


FIGURE 40 – BOREHOLE LOG DISTRIBUTION OF THE LIMPOPO GRIP DATABASE

3.2.4 Spatial datasets

This section presents a selection of the available spatial datasets that were considered in the analysis. The advantage of these spatial datasets is that a value of each of these sets can be assigned to an existing or new borehole entered into the database. By assigning each of the layers discussed in the sections that follow, the specific parameter can be used as a contextual parameter, which means statistical analysis can be expressed in the context of the specified parameter. As an example, the Vegter methodology uses simplified geology as the contextual parameter.

3.2.4.1 Surface Elevation (SRTM90)

The SRTM90 dataset is used to refine the borehole elevations based on the 20m elevation contour intervals used historically. The SRTM90 digital elevation model for the Limpopo is presented in Figure 41 with the Vegter regions as reference.

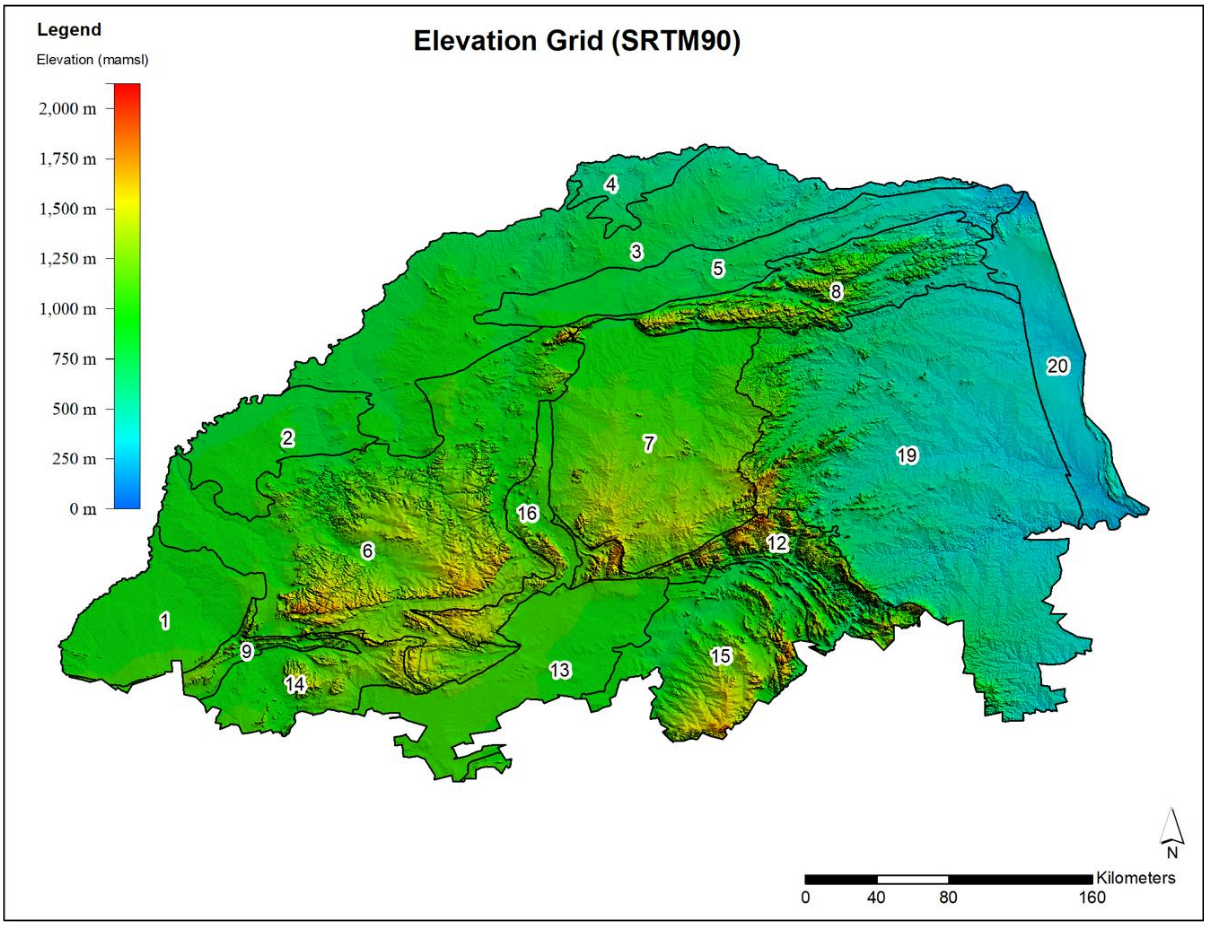


FIGURE 41 – DIGITAL ELEVATION MODEL OF THE LIMPOPO PROVINCE

3.2.4.2 Simplified surface geology and geological and structural lineaments

A simplified South African map series of surface geology data (GCS, 1974-2000) is used in the Vegter analysis. It should be noted that the simplified geology only represents the surface geology and that detailed lithological information is contained in the borehole logs where data is available. This dataset also contains the aquifer rating (1-10) as it relates to the geology which is used as aquifer as shown in Figure 42. The geological and structural lineaments are shown in Figure 43 which was generated from the 1:50,000 geological maps.

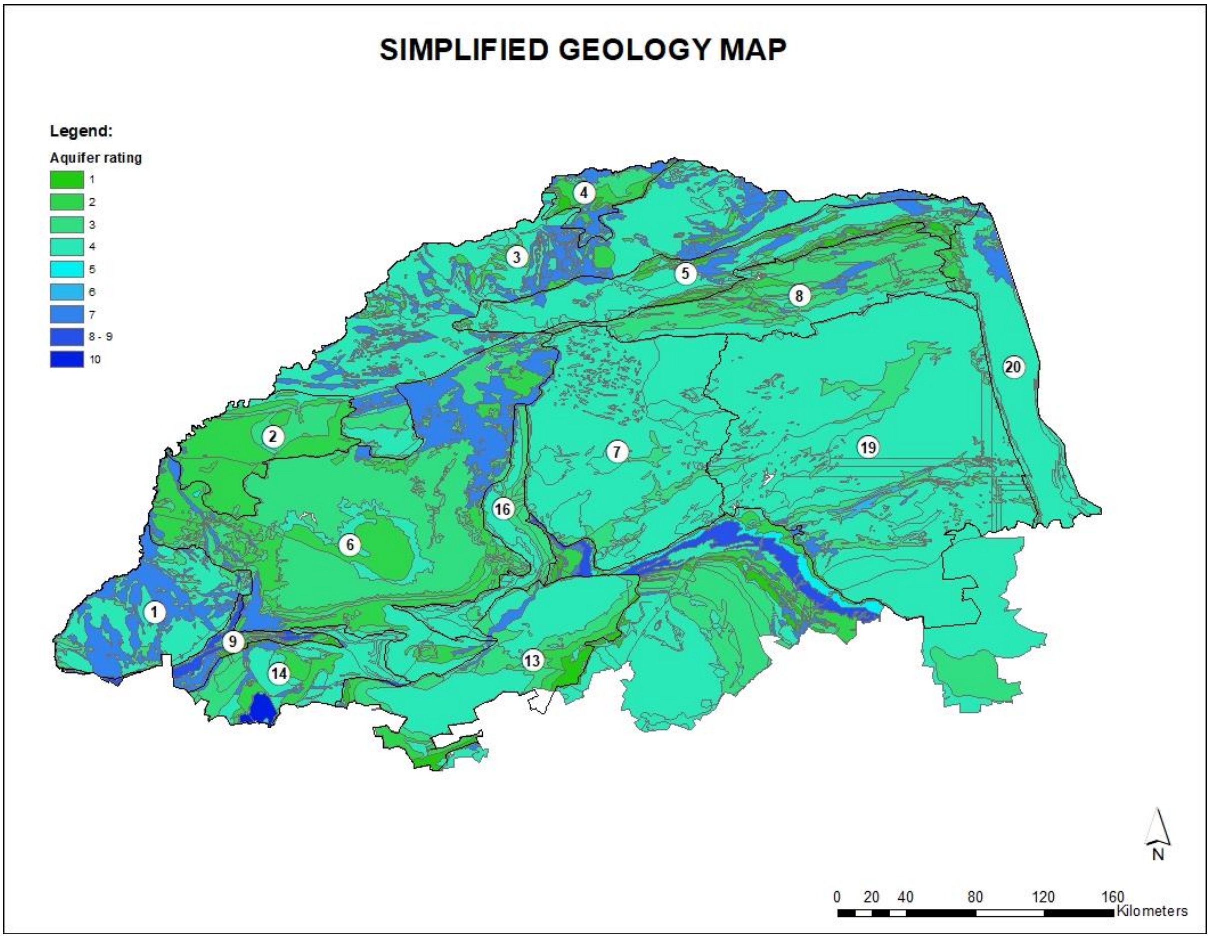


FIGURE 42 – SIMPLIFIED GEOLOGY MAP (AQUIFER CLASSIFICATION)

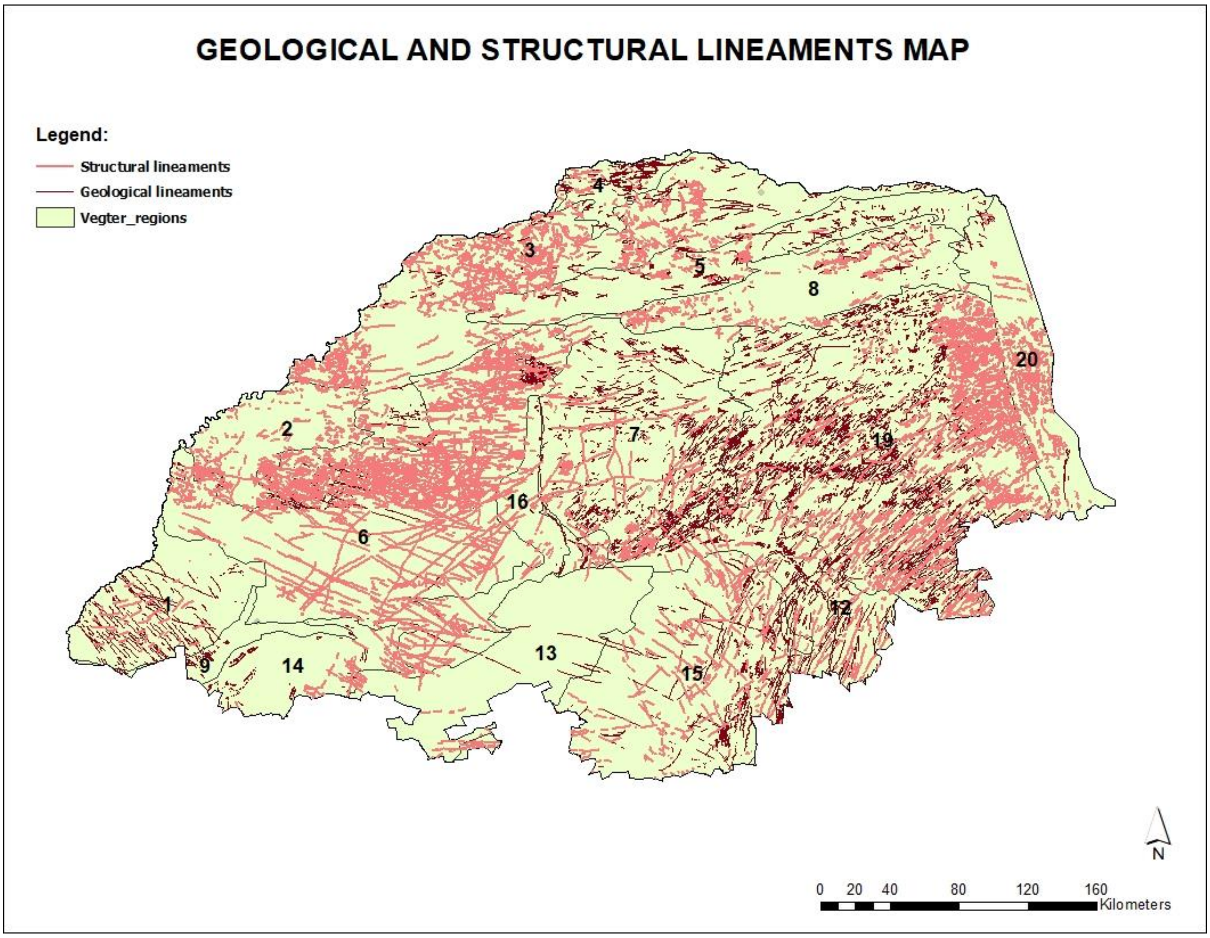


FIGURE 43 – GEOLOGICAL AND STRUCTURAL LINEAMENTS

3.2.4.3 Groundwater occurrence

The groundwater occurrence map based on the geohydrological map of South Africa is presented in Figure 14. The basis for this map is also the simplified geological map of South Africa.

3.2.4.4 Aquifer vulnerability

Aquifer vulnerability refers to the tendency or likelihood for contamination to reach a specified position in the groundwater system after introduction at some location above the uppermost aquifer. This vulnerability index (Figure 21) is used to determine the aquifer's vulnerability to pollution and the index ranges from 1 to 200, where 200 represents the theoretical maximum aquifer vulnerability.

3.3 Parameter Selection

The aim is to minimise the complexity through using data parameters that can be gathered from boreholes by general data users whilst keeping to the requirements for the Vegter analyses.

The parameters were selected and used based on the availability but mainly based on the distribution of sufficient data points. The borehole distribution is mainly focused around areas with higher rural populations and density as this is the areas where groundwater was rapidly developed as primary water source and during the GRIP project surveys; these areas was also the main focus points.

3.4 Database design

3.4.1 Preamble

To perform an analysis on the selected data, the selected data must be extracted or queried from the GRIP database and placed in a temporary database for analysis purposes. The choice of database software to use for this purpose was SQLite which is an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine. The code for SQLite is in the public domain and is therefore free for use for any purpose, commercial or private.

3.4.2 Rational database

Since the GRIP database is inherently a relational database, it makes sense to also create a relational database for the purpose of analysing the data. The newly-created database will be termed the Geo-Statistical Analysis (GSA) database for the purpose of this document.

The entity relationship diagram of the analysis database is presented in Figure 44. The design consists of five tables based on the parameter selection, where **SiteID** is the primary key to the *Site* table and the foreign key to the remaining tables. The *Site* table contains the borehole name, position and all common layer's attributes and has a 1-to-many relationship with the other tables as shown in Figure 44. The sections that follow provide a description of fields in the tables. Note that the data types are not repeated in the table and are shown in the entity relationship diagram.

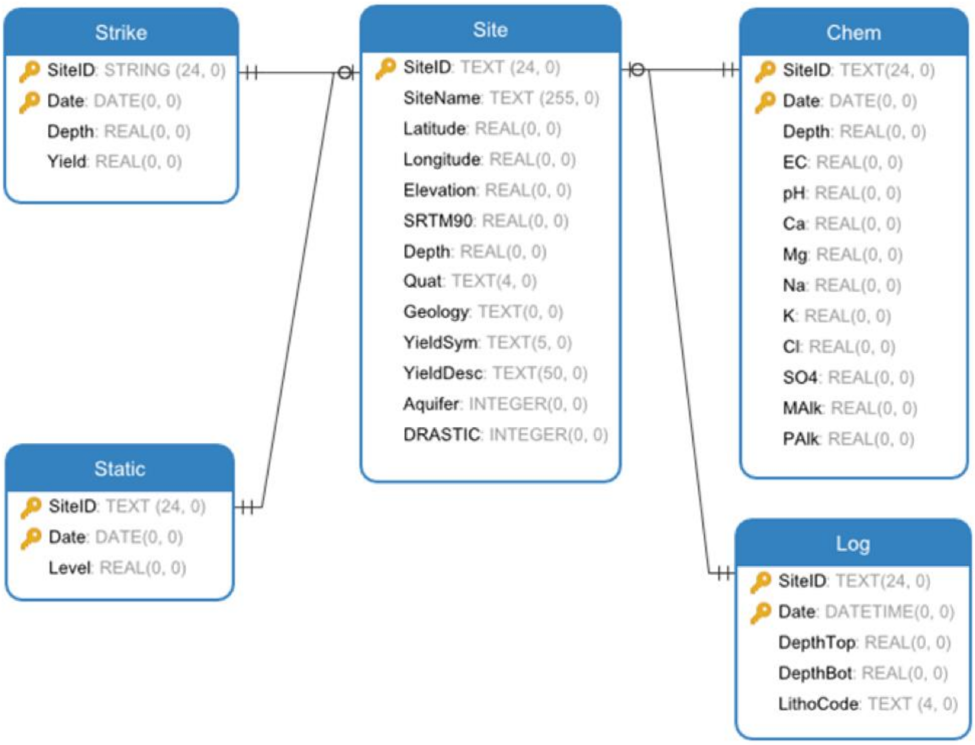


FIGURE 44 - ENTITY RELATIONSHIP DIAGRAM OF ANALYSIS DATABASE

3.4.2.1 Site Table

The field descriptions of the site (*Site*) table are presented in Table 6. The **SrcID** field is used to distinguish between the various data sources to allow the Vegter analysis to be done with only selected data sources.

TABLE 6 - SITE TABLE FIELD DESCRIPTIONS

Field Name	***Description***
SiteID	Unique Site ID
Site Name	Descriptive site name
Longitude	Longitude in decimal degrees
Latitude	Latitude in decimal degrees
Elevation	Elevation (mamsl)
SRTM90	DEM elevation (mamsl)
Depth	Depth of borehole (m)
Quat	Quaternary catchment
Geology	Simplified geology code
YieldSym	Yield map symbol
YieldDesc	Yield map description (occurrence)
Aquifer	Aquifer rating (1-10)
DRASTIC	Aquifer vulnerability (0-200)

3.4.2.2 Static level Table

The field descriptions of the static water level (*Static*) table are presented in Table 7. It is important to note that only static water levels are considered in the analysis and therefore pumped water levels should not be considered, as the natural state of the system is desired. Time series water level data is accommodated through the use of the **Date** field.

TABLE 7 - STATIC WATER LEVEL TABLE FIELD DESCRIPTIONS

Field Name	***Description***
SiteID	Unique Site ID
Date	The date of measurement
Level	Water level (mbgl)

3.4.2.3 Water strike Table

The field descriptions of the water strike (*Strike*) table are presented in Table 8. Generally, the blow yield was determined when the borehole was drilled and therefore is not considered as time series data. However, boreholes can be drilled deeper or redeveloped and new blow yield values can be obtained, therefore the **Date** field is part of the table design to accommodate any new blow yield values.

TABLE 8 – WATER STRIKE TABLE FIELD DESCRIPTIONS

Field Name	***Description***
SiteID	Unique Site ID
Date	The date of measurement
Depth	Depth of water strike
Yield	Blow yield (l/s)

3.4.2.4 Chemistry Table

The field descriptions for the chemistry (*Chem)* table are presented in Table 9. Time series water chemistry data is accommodated by using the **Date** field.

TABLE 9 - CHEMISTRY TABLE FIELD DESCRIPTIONS

Field Name	***Description***
SiteID	Unique Site ID
Date	The date of measurement
Depth	Depth of sampling (mbgl)
EC	Electrical conductivity (mS/m)
pH	pH
Ca	Calcium (mg/L)
Mg	Magnesium (mg/L)

Na	Sodium (mg/L)
K	Potassium (mg/L)
Cl	Chloride (mg/L)
SO4	Sulphate (mg/L)
MAlk	Methyl Orange Alkalinity (mg/L)
PAlk	Phenolphthalein Alkalinity (mg/L)

3.4.2.5 Borehole Log Table

The field descriptions of the borehole log (*Log*) table are presented in Table 10. Generally, the borehole log is determined when the borehole is drilled and therefore is not considered as time series data. However, boreholes can be drilled deeper and new borehole log values can be obtained, therefore the **Date** field is part of the table design to accommodate any new borehole values.

TABLE 10 – BOREHOLE LOG TABLE FIELD DESCRIPTIONS

Field Name	***Description***
SiteID	Unique Site ID
Date	The date of measurement
DepthTop	Start of the specified lithology measured from surface (m)
DepthBot	End of the specified lithology measured from surface (m)
LithoCode	Lithological code used for lookup of lithology details

3.5 Data analysis

3.5.1 Preamble

Vegter introduced a set of statistical analysis performed on boreholes in an area for the purpose of delineation and these statistics were based on frequency analyses of the available data. The aim of the Vegter analysis was the delineation of groundwater regions to provide guidelines for successful and cost-effective siting of boreholes (Vegter, 1995). One of the objectives of this study is to formulate a delineation methodology to express groundwater incidence based on results of the statistical analysis, which is in line with what Vegter has also done, therefore the Vegter statistical assessment is considered.

The statistical methods associated with Vegter's analysis method are summarised in Table 11. The spatial dependency arises from the fact that the physical study area boundary selection dictates the borehole selection, therefore the results will change as the boundary selection changes.

TABLE 11 – SUMMARY OF VEGTER'S STATISTICS

Description of statistics	***Spatial dependency***

The drilling success rate	X
Borehole yields in the various geologies	X
Distribution of boreholes per depth	X
Distribution of weathering and fracturing per depth	X
Strike frequency and cumulative strike frequency	X
Yield versus strike analysis	X
Yield versus dyke intersection	X

Considering the GRIP database and typical borehole information available, certain statistics are not readily available e.g. the true drilling success rate as generally only high-yielding boreholes will be logged and captured in the database. The statistics considered in this study is presented in the sections that follow.

3.5.2 Frequency analysis

The parameters considered for frequency analysis with respect to depth in this study are outlined as follows:

- Water level
- Water strike
- Borehole yield
- Borehole yield vs. dyke intersection
- Electrical conductivity

Frequency analysis is performed by constructing histograms for the selected parameters. Histograms are a form of bar chart, usually used to measure frequency distribution of data. The groups of data are called bins because each bin represents data points within a particular range.

A generic algorithm is presented in Figure 45 to generate a histogram based on a specific parameter ***p*** when a list of parameter values is available. The minimum value of the parameters is always considered to be zero. The standard deviation (Appendix D) is also displayed for each histogram.

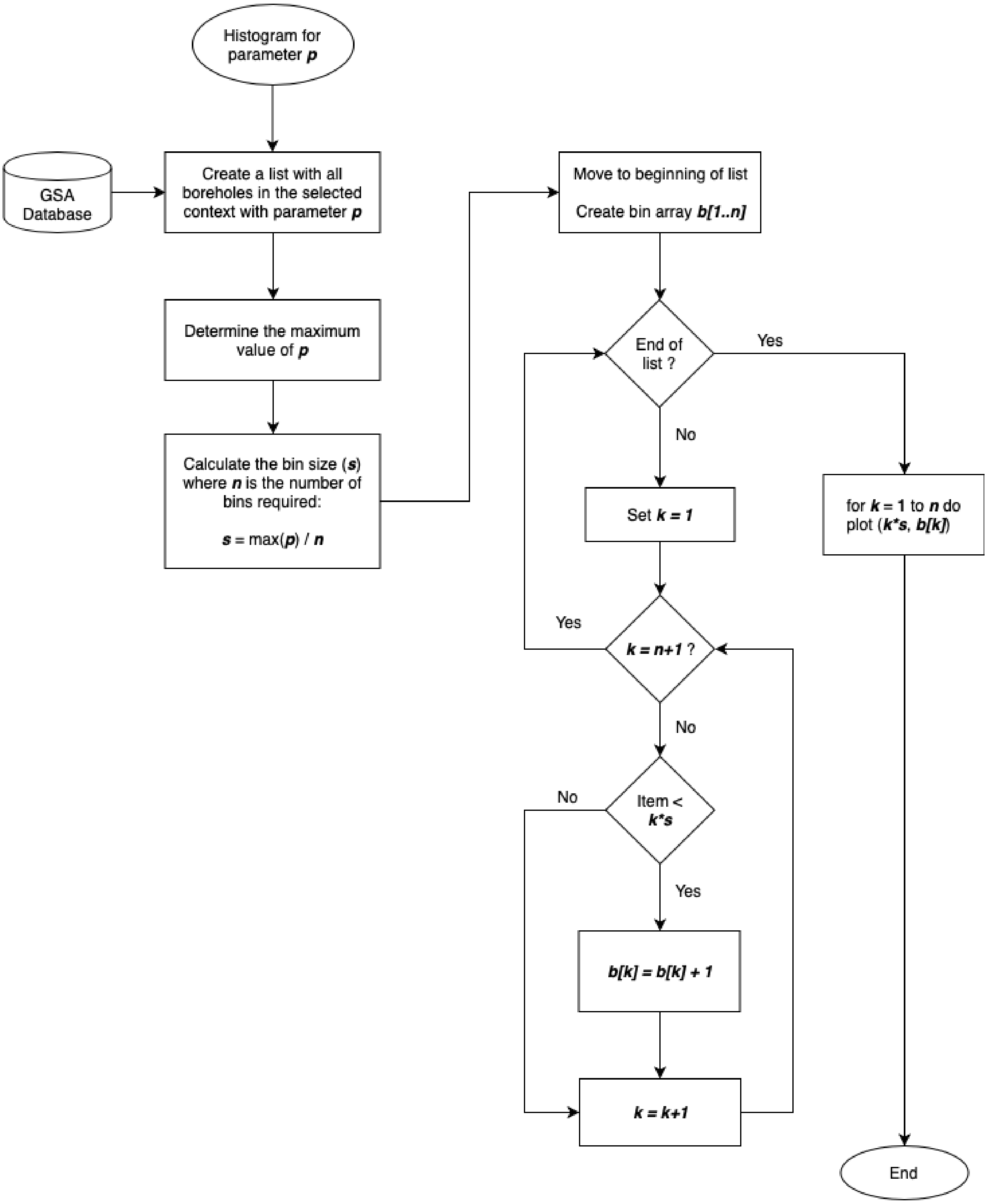


FIGURE 45 – GENERIC HISTOGRAM ALGORITHM

3.5.3 Water level correlation

The water level correlation algorithm uses the water level and associated elevation values for each borehole and plots these two parameters against each other in a scatter plot. Generally, a high correlation coefficient is obtained in unconfined aquifer systems as opposed to confined system which tends to show low correlation values. Having said that, most confined systems are punctured during drilling and depending on the casing depth, these systems also exhibit a high correlation figure. This aforementioned observation does not render this analysis inadequate, since it still provides a level of information on the general trend of specific areas.

The Pearson correlation coefficient (Appendix E) is calculated and used as the measure of correlation. The high-level flow chart of this algorithm is presented in Figure 46.

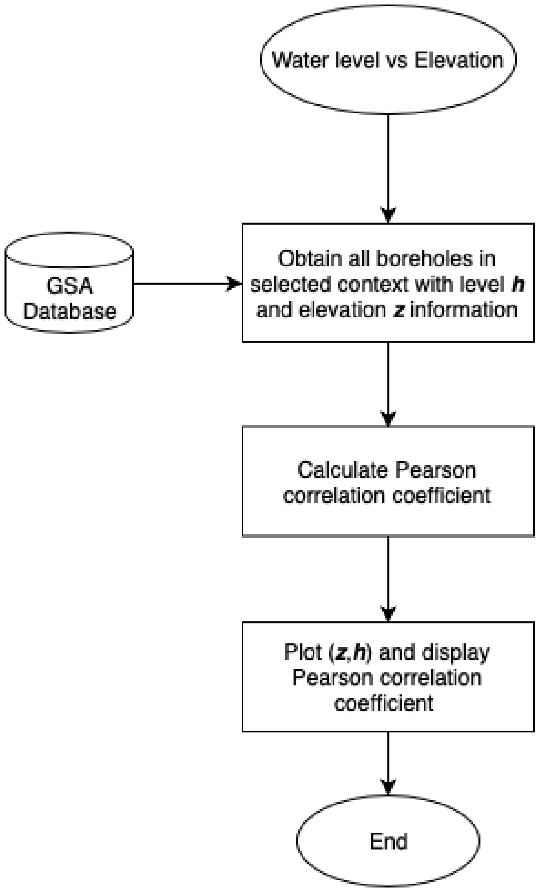


FIGURE 46 – WATER LEVEL CORRELATION ALGORITHM

3.5.4 Water character analysis

The water character is depicted by a Piper plot and the plotting procedure is outlined in Appendix F. The Piper diagram is a so-called multivariate display, simultaneously taking up to eight variables into consideration, projecting these variables to a single point on the diagram which related to the water character of the sample (Kovalevsky *et al.* 2004). The high-level flow chart for this algorithm is presented in Figure 47.

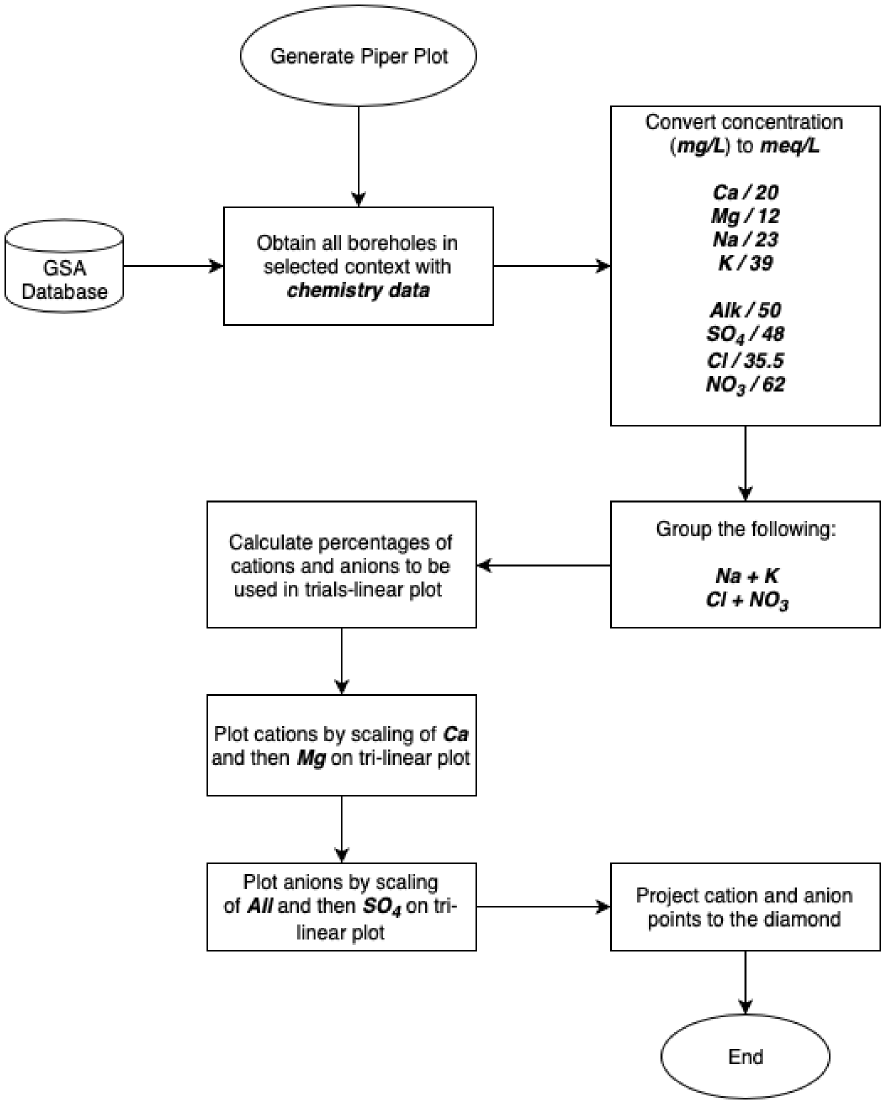


FIGURE 47 – WATER CHARACTER ALGORITHM

3.5.5 Contextual analysis

The frequency analysis described is calculated for the selection of boreholes considered within a study area. Each study area in turn comprises various parameters that vary spatially, e.g. surface geology. A selection of spatial parameters was assigned to each borehole position in the analysis database as discussed in a previous section and these spatial parameters can now be used as the context to view the traditional Vegter statistics. As an example, if geology is chosen the results will be grouped by geology and the calculated statistics will be available per geological unit. The Vegter methodology reported most statistics per geology and this is now extended to the specific context selection as summarised in Table 12.

Table 12 – Vegter's statistics in context

Description of statistic	***Temporal dependency***	***Spatial dependency***
Water levels in context	X	X
Water strike in context		X
Yield in context		X
EC in context	X	X
Borehole chemistry in context	X	X
Water level correlation	X	X

It is important to note that some of the borehole statistics can have both temporal and spatial dependencies. The temporal dependency arises from the fact that boreholes can have time series data e.g. water levels and chemistry that vary over time. Based on the time frame selected, different results can be obtained which will have an impact on the resultant spatial distribution of boreholes as boreholes will not be considered that lie outside the selected time window.

3.6 Delineation methodology

3.6.1 Preamble

This section discusses the delineation methodology developed during the course of the study. The purpose of the delineation is to allow users to delineate groundwater occurrence based on available borehole parameters. The resulting delineation should be areas that are similar in terms of the geohydrological character. The obtained delineation should also be continuous across boundaries and study areas. To address shortcomings of the delineation methodology applied by Dennis and Dennis (2012), the delineation class should be replaced with a calculated value that relates to a physical property.

3.6.2 Aquifer parameters

Since the groundwater occurrence is of interest, it makes sense to look at aquifer parameters as these parameters relate to possible borehole yields. When considering the available parameters, water level and abstraction or yield, the Cooper-Jacob equation comes to mind as this equation determines the drawdown in a borehole based on abstraction rate, duration of pumping and the associated aquifer parameters as expressed in Equation 5 (Kruseman & De Ridder, 1991).

$$s = \frac{2.3Q}{4\pi T} log\left(\frac{2.25Tt}{r^2S}\right) \quad (5)$$

where,

s = *Drawdown from static water level (m)*
Q = *Abstraction rate (m^3/d)*
T = *Transmissivity (m^2/d)*
S = *Storativity or Storage Coefficient*
t = *Time of pumping (days)*
r = *Borehole radius*

Storativity is not readily available from the GRIP database unless pump test data is available that can be analysed. The Cooper-Jacob equation (Equation 5) is less sensitive to uncertainty in the storativity due to the fact that the storativity only appears in the log-term of the equation. Some estimations of storativity already exist in map form (Figure 8, Figure 16 and Figure 23) and the fact that Equation 5 is not too sensitive in uncertainty in this parameter, certainly makes Equation 5 an attractive option to determine yield if transmissivity is known or vice-versa as the remaining parameter values can be provided.

A method of calculating the sustainable yield is described in unpublished groundwater course notes by the Canadian International Development Agency (CIDA), and has been used by the Geological Survey in Swaziland. Swaziland's Geological Survey carries out twenty-four-hour constant discharge tests on boreholes to be equipped with motorised pumps and eight-hour tests on boreholes to be equipped with hand-pumps or windmills. An approximate daily production yield is calculated using Equation 6.

$$Q = 0.068Ts \tag{6}$$

where,

s = *Available drawdown from static water level (m)*
Q = *Abstraction rate (m^3/d)*
T = *Transmissivity (m^2/d)*

Equation 6 can be rearranged to make transmissivity the subject of the equation and the resultant expression is presented in Equation 7.

$$T = \frac{14.7Q}{s} \tag{7}$$

When analysing a pump test making use of the Cooper-Jacob equation (Equation 5), the transmissivity can be calculated making use of the drawdown that occurs over 1 log-cycle of time (Kruseman & De Ridder, 1991), therefore resulting in Equation 8 where Δs represents the drawdown over one log-cycle.

$$T = \frac{2.3Q}{4\pi\Delta s} = \frac{0.183Q}{\Delta s} \tag{8}$$

Equation 7 and Equation 8 are similar in form, but differ by a factor of 80. This is attributed to the fact that Equation 8 typically uses a constant pump rate over 24 hours and in Equation 7 a sustainable pumping rate was used.

If Equation 5 is directly applied to calculate the transmissivity, by using an estimation for storativity, using a borehole radius of 0.08m and assuming the abstraction rate is over a period of a year (typical sustainability criteria), the following relationship (Figure 48) is obtained when compared to Equation 8 and assuming $\Delta s = s$:

$$F = -0.434 \ln(S) + 5.1 \tag{9}$$

where,

S = *Available drawdown from static water level (m)*
F = *Conversion factor between transmissivity calculated by Equation 4 and Equation 7*

$$T_{Eq5} = F\left(T_{Eq8}\right) \tag{10}$$

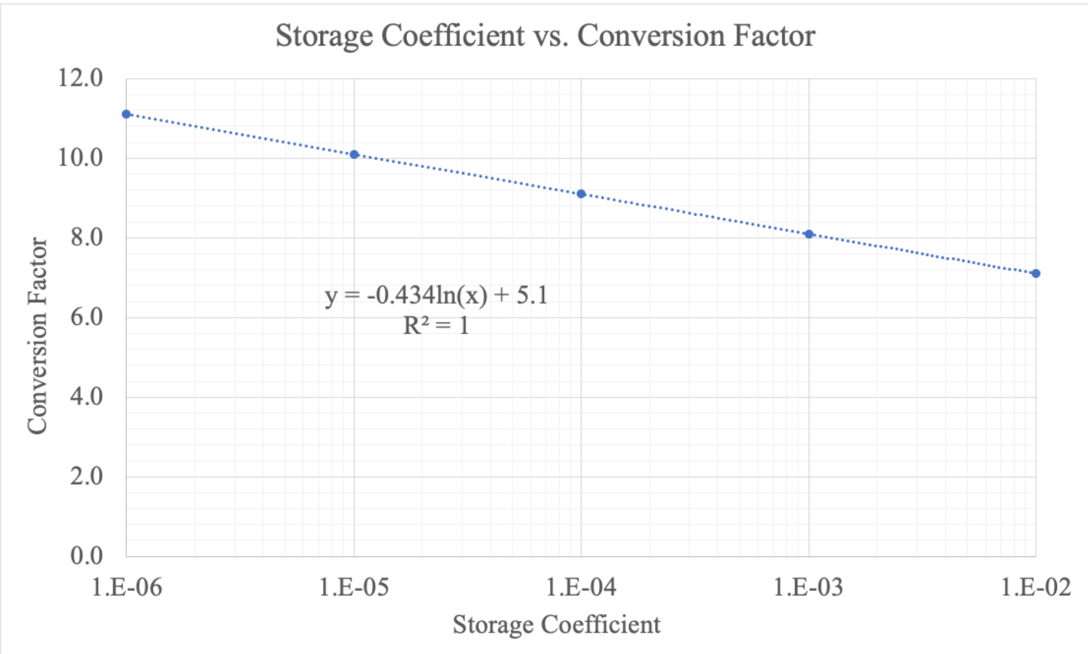


FIGURE 48 - STORAGE COEFFICIENT VS. CONVERSION FACTOR

The magnitude of the conversion factor is related to the abstraction duration (365 days), since Equation 8 is not explicitly a function of time (1 log-cycle of time assumed). The goal is to calculate a parameter that relates to a physical parameter to be used for delineation purposes. Table 13 lists the parameter assumptions made with regard to the available borehole parameters and the parameters of Equation 5:

TABLE 13 – PARAMETER ASSUMPTIONS FOR COOPER-JACOB EQUATION

Cooper-Jacob	***Assumptions***
Q	The blow yield is representative of Q
s	The drawdown can be calculated as $s = Strike - Level$
S	Figure 8, Figure 16 or Figure 23

The only Province wide database of transmissivity values is that of the GRA-II dataset (

FIGURE 15 – GRA-II TRANSMISSIVITY OF THE LIMPOPO PROVINCE

). According to DWAF (2006), as no regional aquifer transmissivity information are available, borehole yields may be used in a qualitative way as a proxy for transmissivity and the following relationship is applied, where *T* is in m^2/d and *Q* expressed in *L/s*:

$$T = 10Q \tag{11}$$

Since the Limpopo GRIP database has recorded transmissivity values of pump tests conducted, this database was used to select boreholes for which transmissivity values were recorded. In addition to the recorded transmissivity values, the parameters as described in Table 13 were obtained for the same dataset, which resulted in 502 boreholes being used in the analysis.

Transmissivity values based on Equation 5 and Equation 8 were computed to compare to that of the transmissivity values obtained from pumping tests and the GRA-II dataset (

FIGURE 15 – GRA-II TRANSMISSIVITY OF THE LIMPOPO PROVINCE

). The probability distribution of the results is shown in Figure 49.

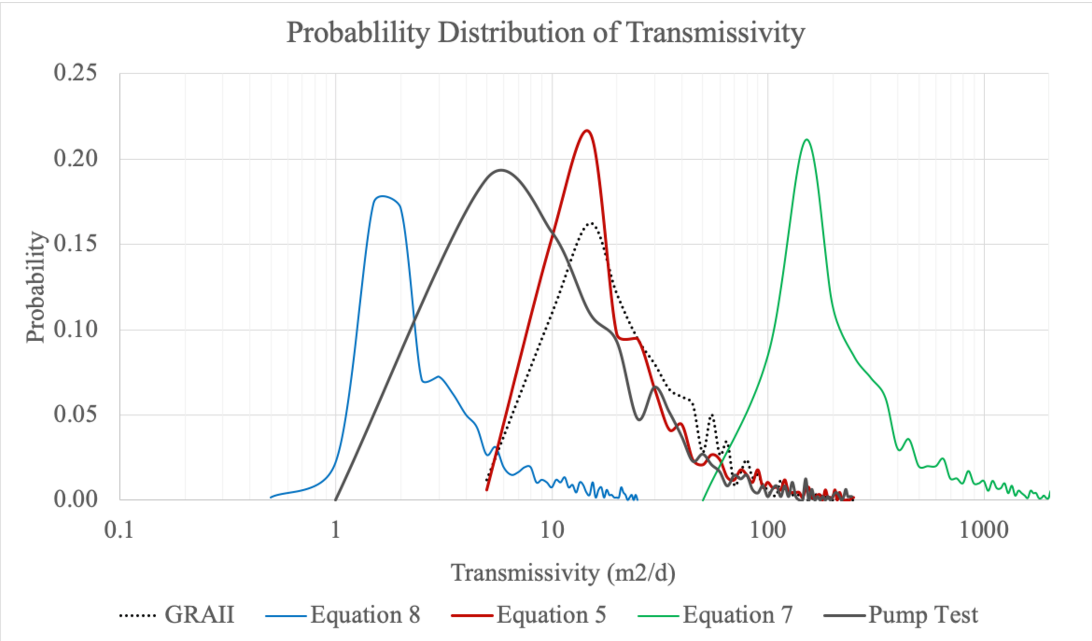


FIGURE 49 - PROBABILITY DISTRIBUTIONS OF CALCULATED TRANSMISSIVITY

It is clear that the distribution of transmissivity values obtained from Equation 5 and that of the GRAII dataset compares well. The distribution of the higher transmissivity values from the pump test data also compares well with that of Equation 5 and the GRAII dataset, although the mean value from the pump test data is almost an order of magnitude lower. It is clear that Equation 7 and Equation 8 represent transmissivity distributions that are much higher and lower respectively when compared to the rest.

Due to the difference in mean values between the pump test data and Equation 5, it is not expected that a high correlation will exist between these two datasets. The correlation between the transmissivity obtained from the pump test data and that of Equation 5, Equation 7, Equation 8 and the GRAII dataset is presented in Figure 50. The poor correlation is evident between the datasets, but as expected, Equation 5 and the GRAII data shows the closest values to that of the pump test data analysis.

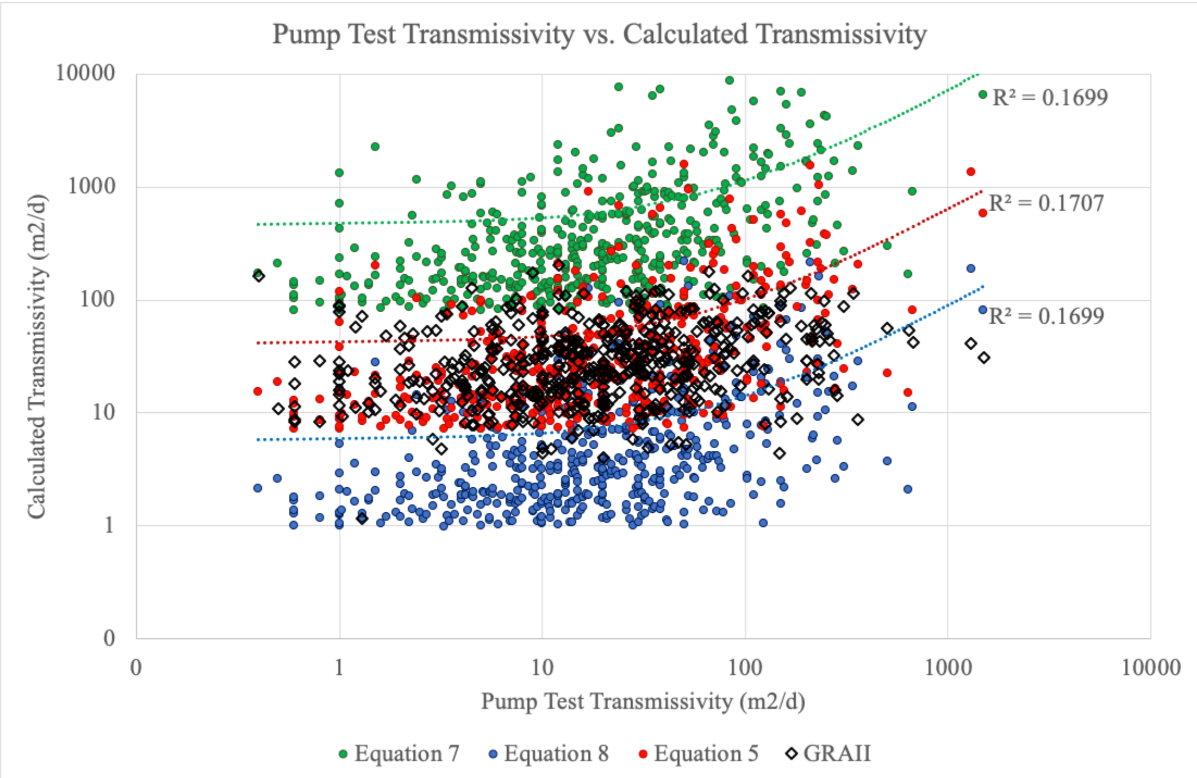


FIGURE 50 - PROBABILITY DISTRIBUTIONS OF CALCULATED TRANSMISSIVITY

Since transmissivity values are not readily available across the country, Equation 5 will be used to calculate an inferred transmissivity for delineation purposes of groundwater units that exhibit a similar response. The term inferred is used due to the fact that the transmissivity calculated using Equation 5 does not reflect the actual transmissivity that is obtained through analysing a pumping test.

3.6.3 Classification

In the previous section it was shown that it is possible to calculate an inferred transmissivity by means of an analytical equation (Equation 5) based on parameter values available in the database. The transmissivity was inferred using representative parameters and the said equation, based on some assumptions.

In general, if it is difficult to derive parameters that would describe the problem of estimating a certain parameter, due to the fact that the level of knowledge regarding this specific field is not yet advanced enough to compute analytically, classification algorithms can be used to compute results.

Since there is some knowledge to compute a transmissivity result, classification is used here to verify the existing analytical relationship proposed (Equation 5 with parameters form Table 13).

Classification algorithms are able to compute many probable results across many parameters, if there are sufficient examples with a known outcome - this is called inference. It is not needed to know what the actual analytical relation between the parameters and the outcome is, as classification algorithms will internally detect this relation, although they will not be able to express it. If an algorithm is able to express the detected relation, it is called a machine learning algorithm. Because not all relations can always be expressed with signs and words, the classification algorithms have higher classification accuracy than machine learning algorithms.

Considering various parameter combinations available from the database, the two governing parameters identified by various classification schemes was *yield* and *available drawdown* which is in line with two of the parameters presented in Table 13. Using a trial and error approach the following relationship was determined for transmissivity classes considered in the classification algorithm:

$$TC_n = 5^n \tag{12}$$

where,

n = *Enumerator:* $0 \leq n \leq 3$
TC = *Transmissivity Class*

Figure 51 shows the first classification tree considering both *yield* and *available drawdown (AD)* to predict the transmissivity. The four transmissivity classes specified by Equation 12 and the associated classification probabilities are also presented in Figure 51.

On closer inspection it is clear that the *available drawdown* is only used to classify two different instances of a transmissivity of 25 m^2/d, which is one class. Therefore, *available drawdown* can be omitted as a classification criterion, resulting in the last classification tree presented in Figure 52.

It should be noted that the yield values in the classification trees refer to the unit m^3/d.

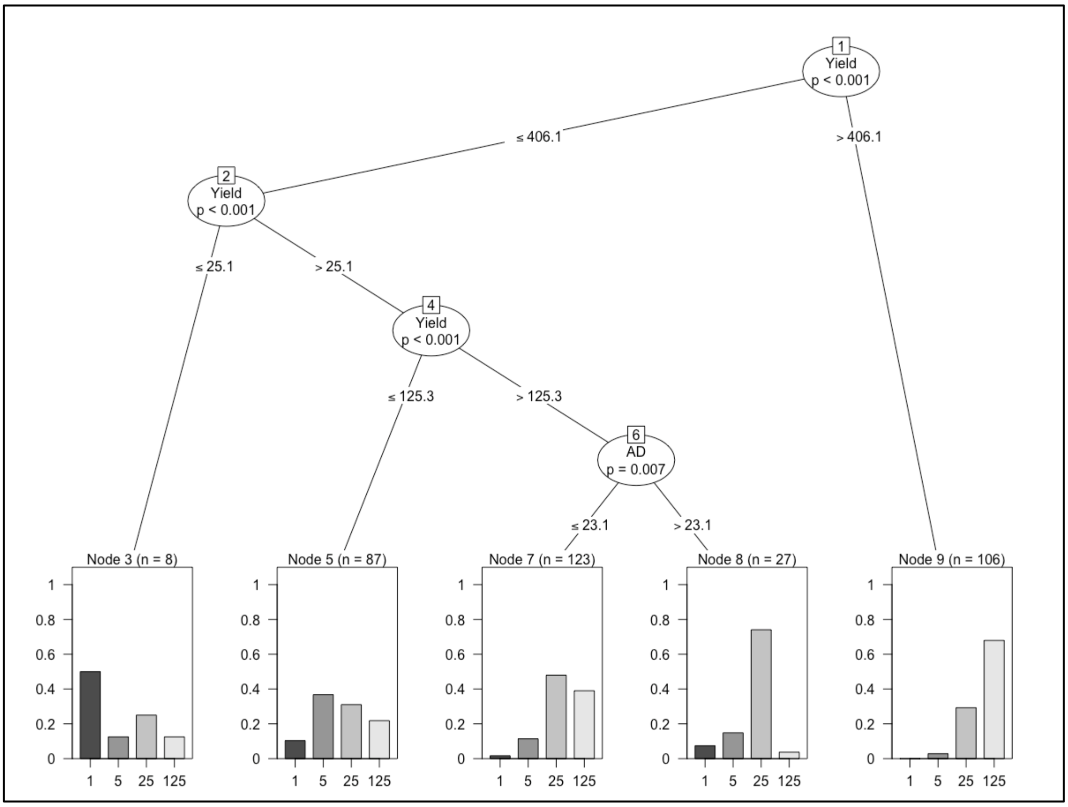


FIGURE 51 – CLASSIFICATION TREE CONSIDERING BOTH YIELD AND AVAILABLE DRAWDOWN

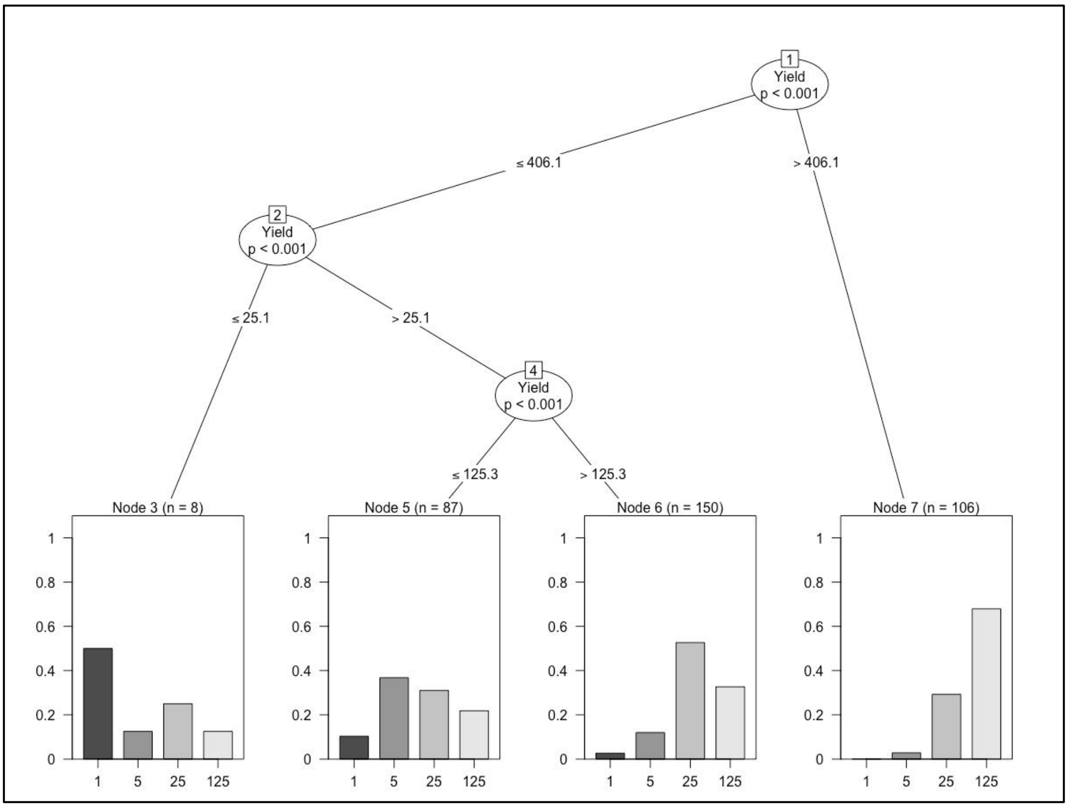


FIGURE 52 – CLASSIFICATION TREE ONLY CONSIDERING YIELD

The indicated probabilities refer to the probability of predicting the correct transmissivity class based on the transmissivities obtained from the pumping test data and are summarised in Table 14. Note that the last transmissivity class represents all transmissivities greater than 25 even though the maximum is indicated as 125.

TABLE 14 – CLASSIFICATION TREE PROBABILITIES

Transmissivity class *(m^2/d)*	***Prediction probability*** *(%)*
0 - 1	50%
1 – 5	39%
5 - 25	52%
> 25 (25 - 125)	70%

When classing all the datasets according to Equation 11 the following correlations were obtained with respect to the actual classed transmissivities:

TABLE 15 – CLASSED DATASETS CORRELATION WITH CLASSED TRANSMISSIVITIES

Classed dataset	***Pearson Correlation***
Classification tree	0.42
Equation 5	0.38
GRAII	0.24

Considering Table 15 it seems that the classification tree outperforms Equation 5 and based on this result, the classification tree (Figure 52) is recommended for delineation purposes. As more data becomes available in terms of actual transmissivities across the Limpopo Province, the existing classification tree should be updated as necessary.

3.6.4 Delineation algorithms

Finally, applying the appropriate interpolation technique on the inferred transmissivity class will then represent the delineated groundwater units. This delineation is purely dependent on yield values although it has been shown that similar results are obtained when applying Equation 5 which requires additional parameters.

The recommended interpolation technique to apply would be the one which results in the smallest RMSE (Appendix G) between observed and simulated values. The interpolation technique may vary from one study area to another depending on the yield data distribution.

Since water quality can be highly variable due to various factors, it is not explicitly incorporated into the inferred transmissivity classification delineation, but it is recommended to rather use as secondary delineation criteria based on specific user requirements.

The following delineation algorithms are presented here:

- Delineation classification tree (proposed method)
- Cooper-Jacob delineation (based on Equation 5)

Note the selected interpolation techniques were discussed in Chapter 2 and mathematical detail surrounding these methods are available in Appendix A - Appendix C.

3.6.4.1 Delineation Classification Tree

The delineation algorithm is represented in a flow chart as shown in Figure 53 and is based on the inferred transmissivity classification tree (Figure 52) discussed earlier.

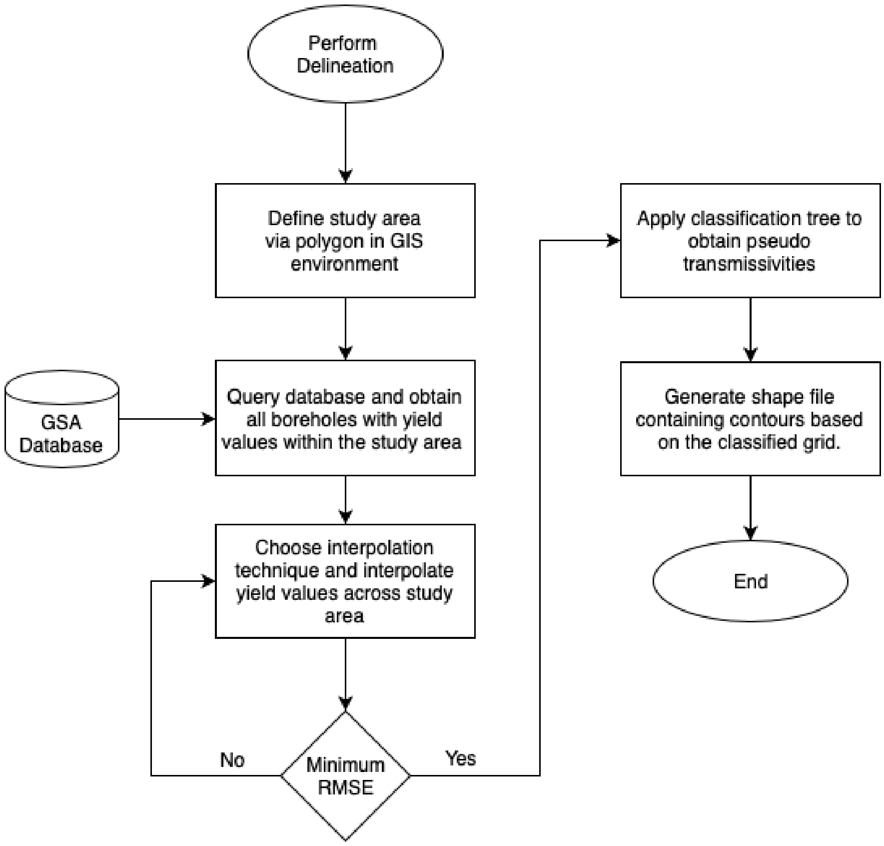


FIGURE 53 - INFERRED TRANSMISSIVITY DELINEATION ALGORITHM

3.6.4.2 Cooper-Jacob Delineation

As an alternative method to calculate the inferred transmissivity, the Cooper-Jacob equation (Equation 5) is used. Interpolated grids are calculated for water level, strike position and yield values. These grids are then used as input parameters to Equation 5 together with a $r = 0.08\ m$ and $t = 365\ days$. The inferred transmissivity is then calculated according to Equation 5. The algorithm flow chart is shown in Figure 54. Note that checks are performed for when the water strike is above the water level and when blow yield is zero.

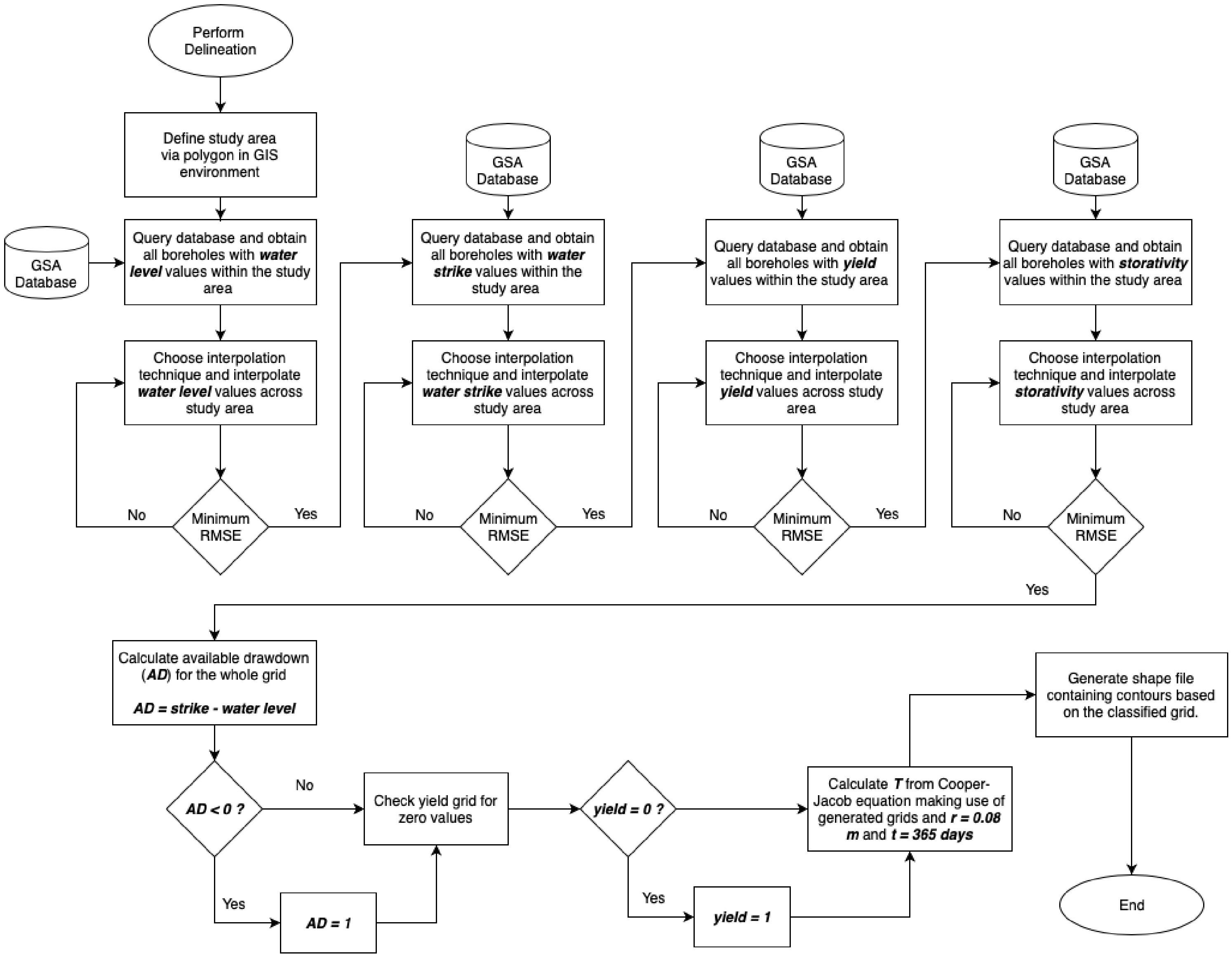


FIGURE 54 - INFERRED TRANSMISSIVITY DELINEATION ALGORITHM BASED ON COOPER-JACOB

4 RESULTS AND DISCUSSION

4.1 Preamble

This section presents a case study based on the proposed delineation methodology developed as well as the statistical analysis of the groundwater occurrence. Both the delineation based on the classification tree and the Cooper-Jacob method are presented here for comparison purposes.

4.2 Study area

It is evident from the literature review that different geology types are known to be water bearing and others not so much. For this reason, it is expected that different geology types may have a different influence on groundwater occurrence (depth, yields and water levels) and it would be deemed necessary to study the characteristics of the different geologies to provide the reader with the necessary facts, characteristics and correlations between lithostratigraphic units. It is known that Vegter already divided the Limpopo Province into different groundwater regions. This was done mainly based on rock type, geological age and lithostratigraphy (Holland, 2012).

The study area as indicated in Figure 55 considered here is Vegter Region 7 due to the high availability of transmissivity data to test the delineation methodology.

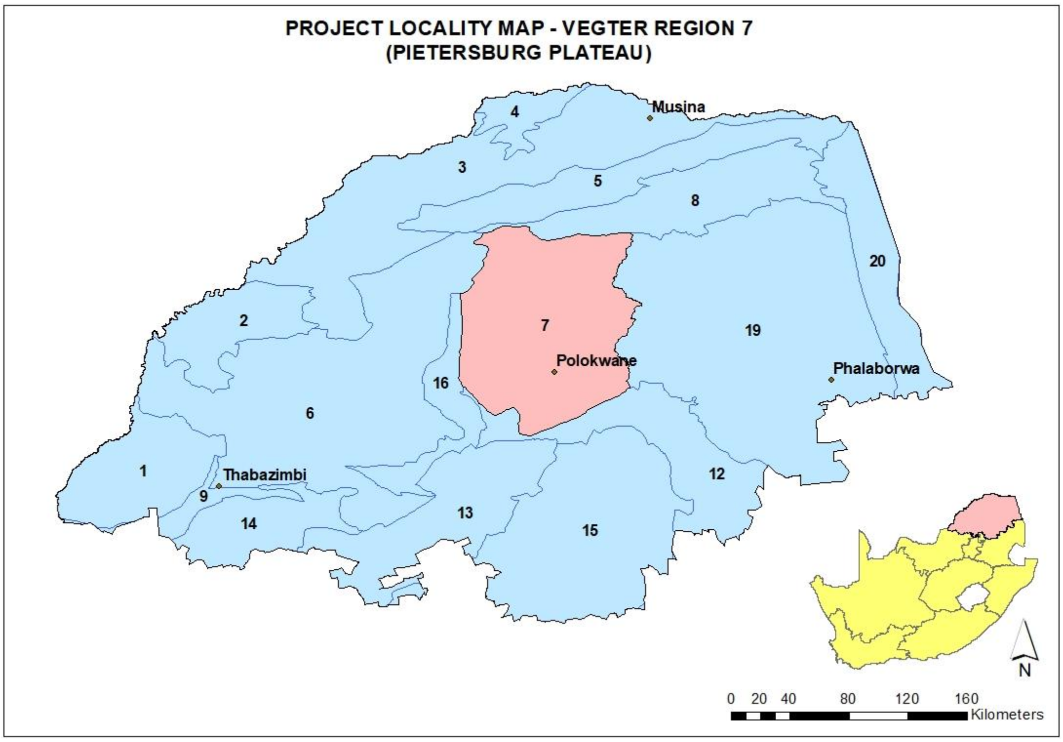


FIGURE 55-LOCALITY MAP OF THE PROJECT AREA

The simplified geological map of the study area is shown in Figure 56 and the groundwater occurrence map is shown in Figure 57.

Vegter mentioned the importance of groundwater for this region due to its large number of rural communities and agricultural land that exist in the area. Rivers only flow after seasonal rains and no major storage dam exists within this area. The alluvial deposits along the major rivers are important sources of groundwater. According to the 1:250 000 geological map series, more than 20 types of surface geology are intersected by this area although mainly consisting of gneisses and granites from the Swazian to the Randian ages. Notable features include two narrow belts of supracrustal rocks in the vicinity of Polokwane and a swarm of diabase dykes that all trend in the north-easterly direction.

Rainfall occurs during the summer months averaging between 400-650mm per annum and typical savannah biome vegetation is found here. A dry semi-arid climate with warm to hot overall temperatures occurs in this region with some occasional winter temperatures dipping below 0° Celsius.

Even though Vegter delineated Region 7 as an individual Vegter region, the diverse surface geology and groundwater occurrence is evident from the aforementioned maps. This just highlights the challenge in delineating groundwater regions with exhibit a similar character.

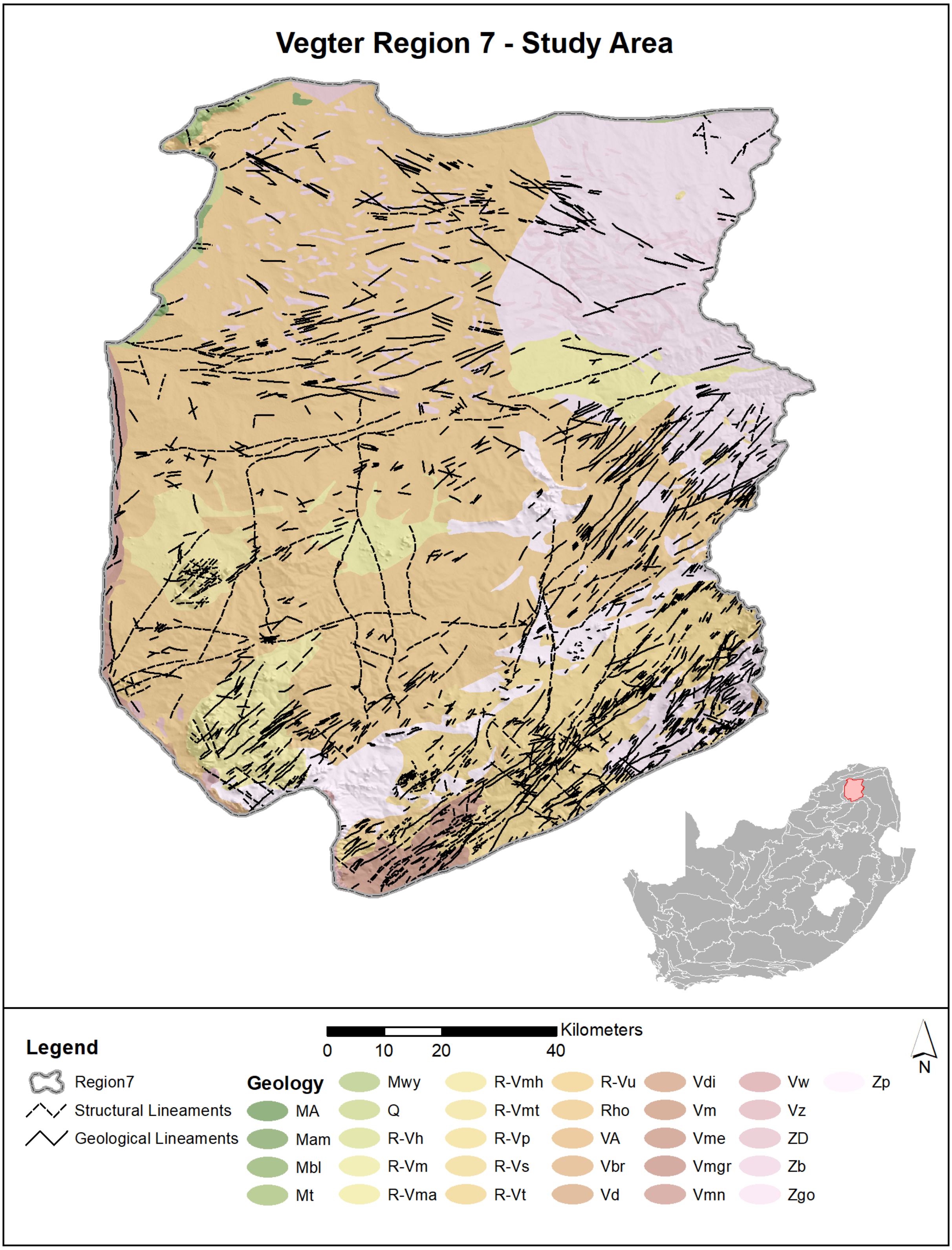


FIGURE 56 – VEGTER REGION 7 SIMPLIFIED GEOLOGICAL MAP

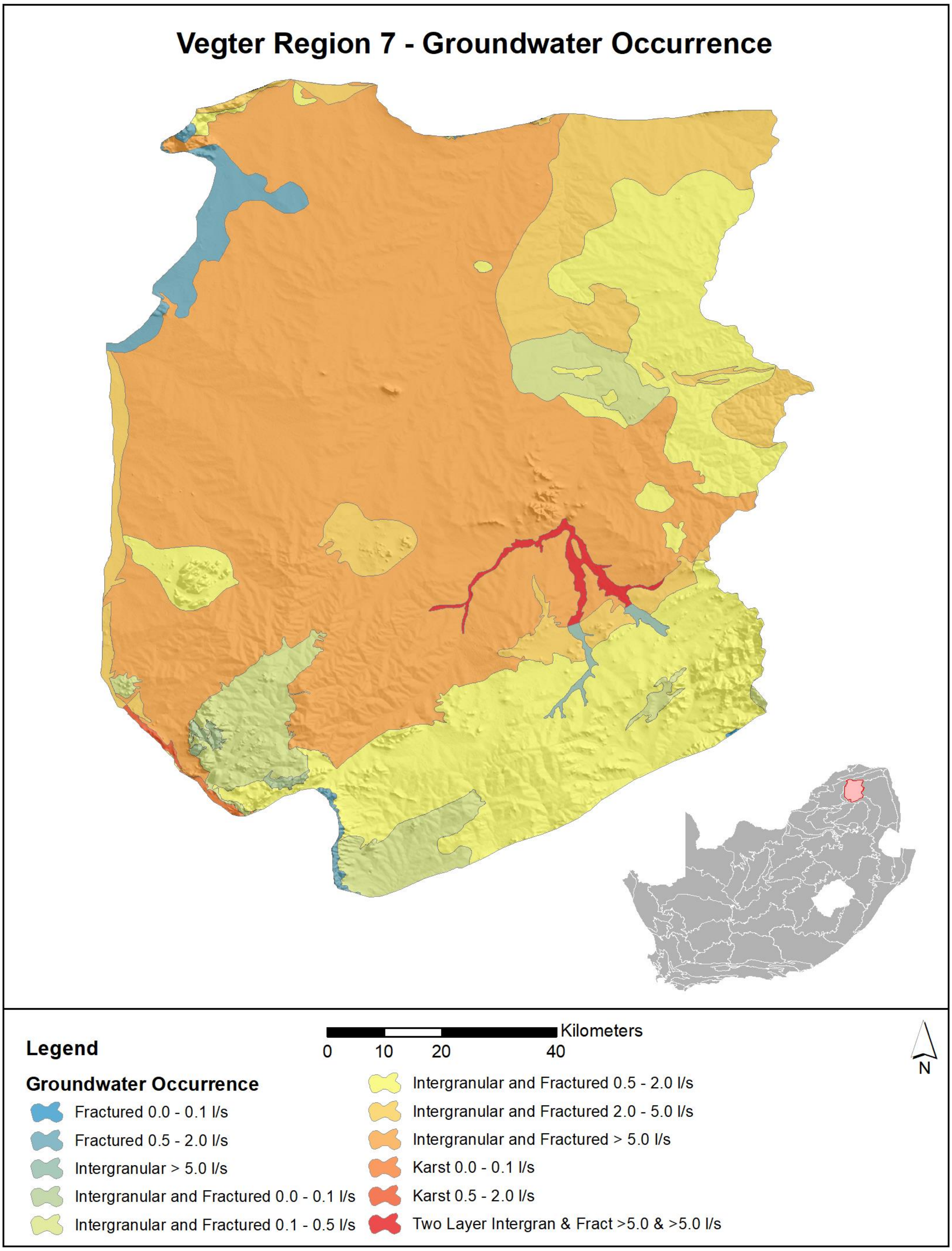


FIGURE 57 – VEGTER REGION 7 GROUNDWATER OCCURRENCE MAP

4.3 Groundwater occurrence delineation

The transmissivity map of the area based on the GRAII dataset is presented in Figure 58. It is clear from the comparison of the aforementioned maps that a correlation exits between the simplified geology and groundwater occurrence which is expected as the simplified geology was used as basis for the geohydrological map which presents the groundwater occurrence. There are however no distinct correlations between the GRA-II transmissivity map and the other maps. It is evident from the GRA-II transmissivity map that the gridded data originates from boreholes.

It should be noted that geology is a 3D feature and that the geological maps only represent surface geology. Drilled boreholes can intersect multiple lithologies and are therefore not necessarily representative of the surface geology in which they were drilled. The transmissivities for the GRA-II are calculated making use of Equation 11 (DWAF, 2006).

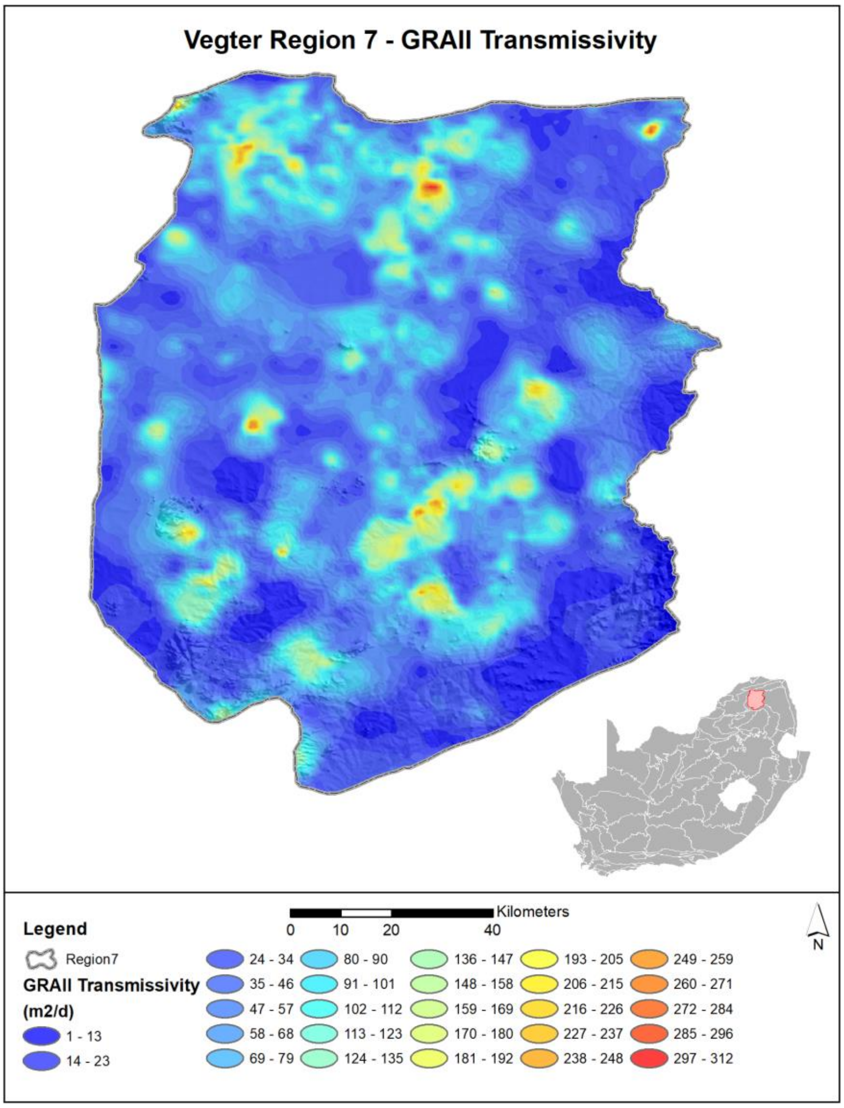


FIGURE 58 – VEGTER REGION 7 GRAII TRANSMISSIVITY MAP

4.3.1 Classification tree delineation

The borehole distribution within a 25 km buffer zone around Vegter Region 7 is shown in Figure 59. The buffer zone is chosen to allow for proper interpolation across the study area boundary, an approach used to ensure continuous delineation across boundaries. The boreholes with actual transmissivity values compare well with the boreholes that have yield values. This is important for comparison purposes to see how well the inferred transmissivity compare to that of the actual transmissivity.

Spline interpolation with minimum curvature is applied to the actual transmissivity values as well as the inferred transmissivities produced by the classification tree. Both result sets together with the GRA-II transmissivity dataset were classed according to Equation 12 and the results are shown in Figure 60. The boreholes are shown to see how the results relate to borehole distribution. The correlation of the classification tree to that of the actual transmissivity values is presented in Table 16.

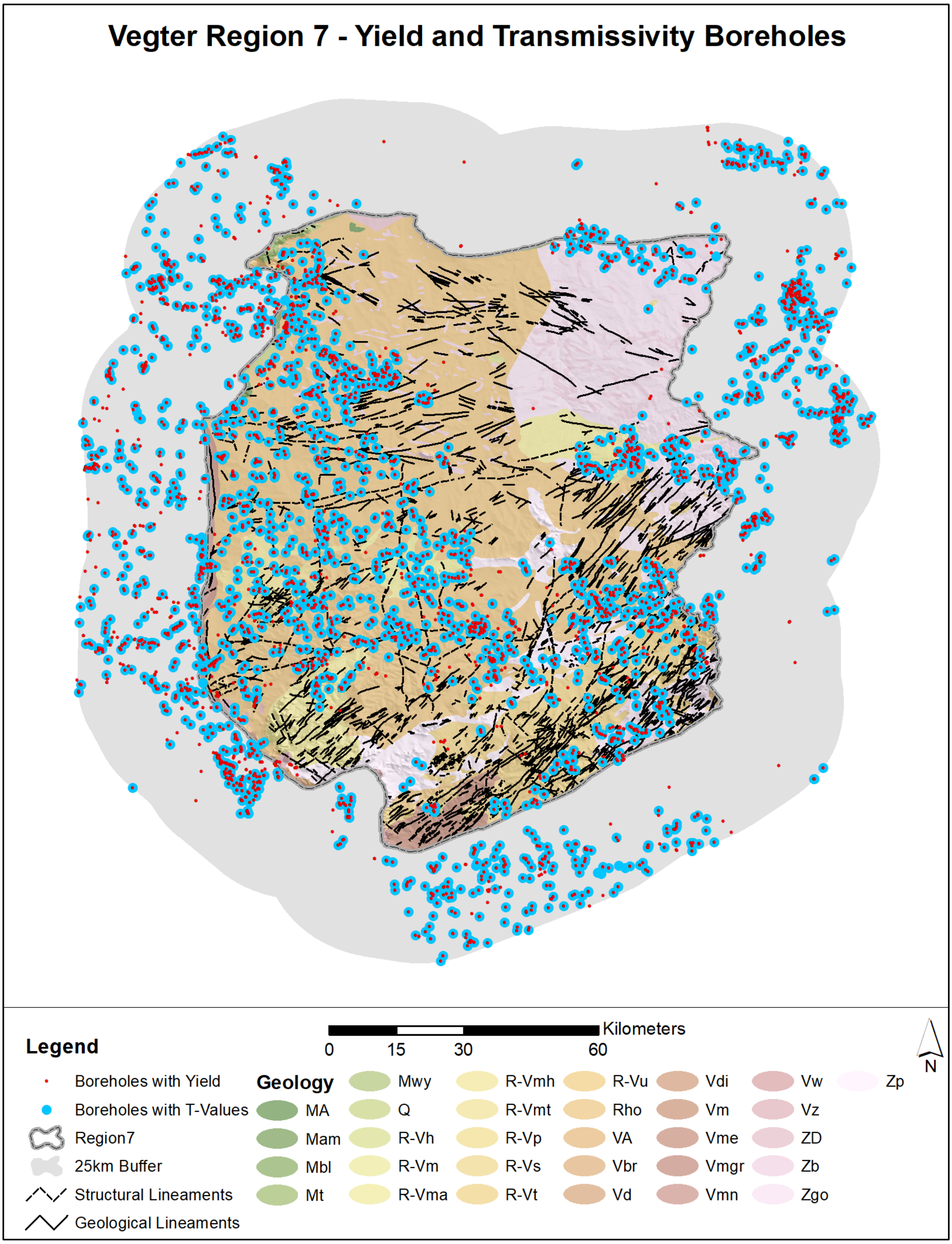


FIGURE 59 – VEGTER REGION 7 BOREHOLES WITH YIELD AND TRANSMISSIVITY VALUES

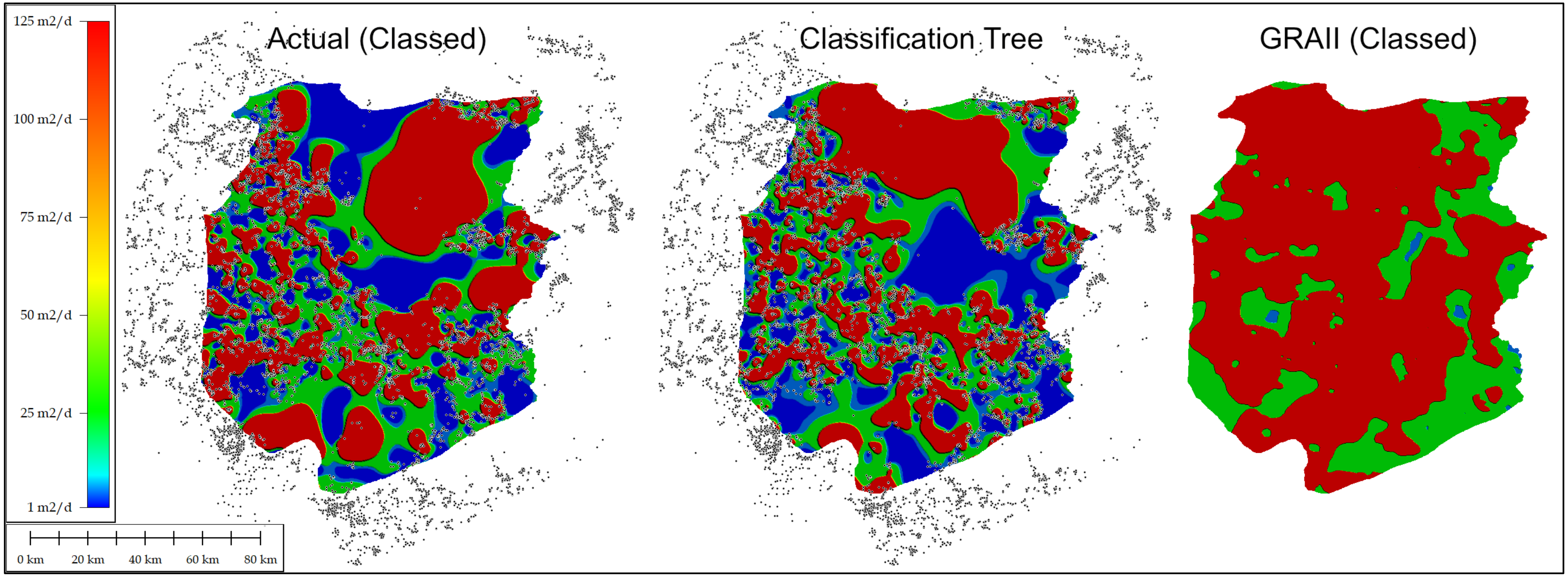


FIGURE 60 – VEGTER REGION 7 CLASSIFICATION TREE RESULTS

TABLE 16 – REGION 7 CLASSED TRANSMISSIVITY CORRELATION WITH CLASSIFICATION TREE

Classed dataset	***Pearson Correlation***
Classification tree	0.32
GRAII (Classed)	0.12

4.3.2 Cooper-Jacob delineation

The borehole distribution within a 25km buffer zone around Vegter Region 7 is shown in Figure 61. The buffer zone is chosen to allow for proper interpolation across the study area boundary. The boreholes with actual transmissivity values compare well with the boreholes presented in Figure 61. This is important for comparison purposes to see how well the inferred transmissivity compares to that of the actual transmissivity.

Spline interpolation with minimum curvature was applied to the actual transmissivity values as well as the inferred transmissivities produced by the Cooper-Jacob method (Equation 5). Both result sets together with the GRA-II transmissivity dataset are classed according to Equation 12 and the results are presented in Figure 62. The correlation of the Cooper-Jacob delineation (Equation 5) to that of the actual transmissivity values are presented in Table 17.

The Cooper-Jacob method exhibits a slight decrease in correlation when compared to the classification tree method which is also the observation in Table 15. Furthermore, the Cooper-Jacob delineation process is computationally more expensive due to the multiple interpolation grids required (Figure 54).

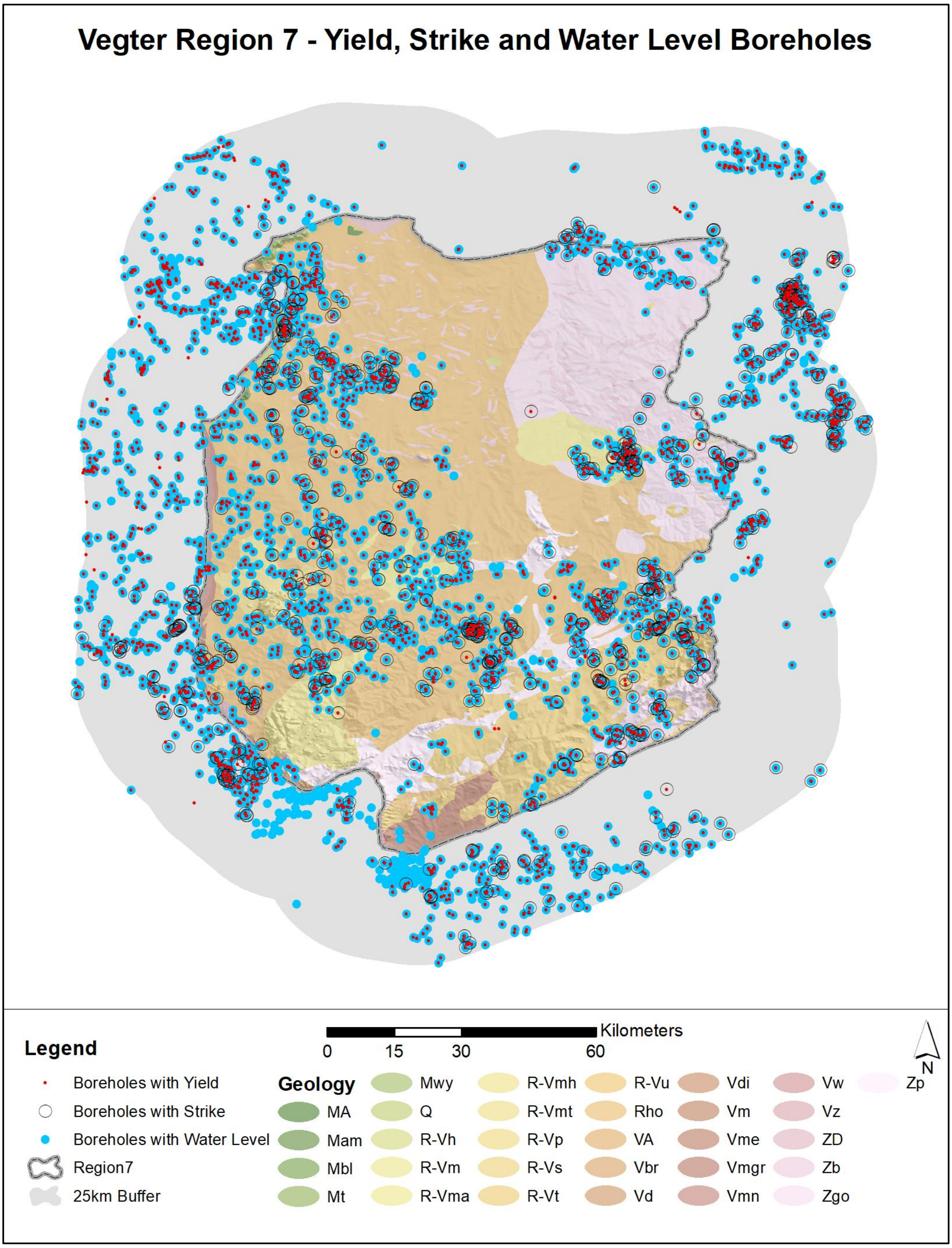


FIGURE 61 – VEGTER REGION 7 BOREHOLES WITH YIELD, STRIKE AND WATER LEVEL VALUES

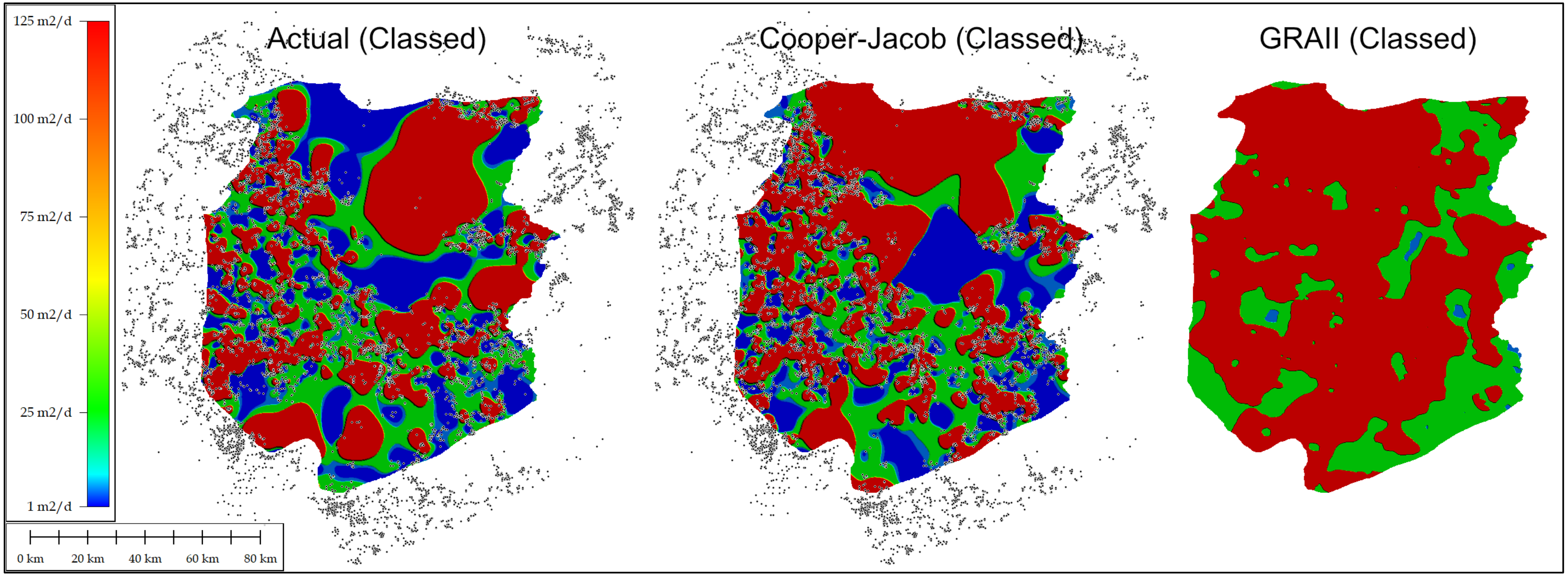


FIGURE 62 – VEGTER REGION 7 COOPER-JACOB RESULTS

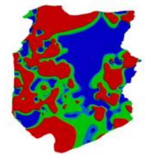
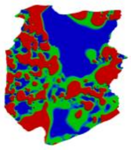
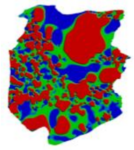
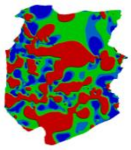
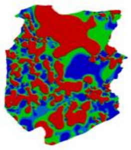
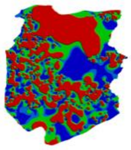
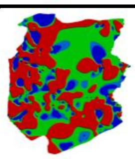
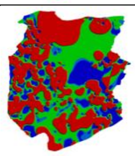
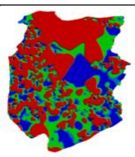
TABLE 17 – REGION 7 CLASSED TRANSMISSIVITY CORRELATION WITH CLASSIFICATION TREE

Classed Dataset	***Pearson Correlation***
Cooper-Jacob (Classed)	0.28
GRAII (Classed)	0.12

4.3.3 Data density

Since the delineation methods make use of spatially distributed data, it is accepted that data density will affect the delineation result. The higher the data density the more refined the delineation results will be. The data sets in the preceding sections were reduced twice randomly by 50% to inspect the effect of lowering the data density. The correlation among the various datasets is presented in Table 18 where *T* denotes actual transmissivity and *CT* denotes the inferred transmissivity obtained from the classification tree. *CJ* denotes the inferred transmissivity obtained from the Cooper-Jacob method.

TABLE 18 – EFFECT OF DATA DENSITY ON CORRELATION

				
		$25\%_{T}$	$50\%_{T}$	$100\%_{T}$
	$25\%_{CT}$	0.19	0.02	0.13
	$50\%_{CT}$	0.13	0.22	0.25
	$100\%_{CT}$	0.13	0.05	0.32
	$25\%_{CJ}$	0.20	0.12	0.04
	$50\%_{CJ}$	0.28	0.15	0.11
	$100\%_{CJ}$	0.20	0.11	0.28

It is clear that data distribution will play a major role in the delineation results. The best correlation obtained was 0.32 when using 100% of the available dataset and making use of the classification tree.

The correlation between the inferred transmissivities for the classification tree and the Cooper-Jacob method (Equation 5) is presented in Table 19.

TABLE 19 – REGION 7 CORRELATION BETWEEN CLASSIFICATION TREE AND COOPER-JACOB

Dataset	***Pearson Correlation***
25%	0.63
50%	0.66
100%	0.72

4.3.4 Comparative results

The report on the Vegter analysis of Region 65 (Project K5/2251/1) was presented in selected sections of this document for illustration purposes and therefore it is fitting to compare the newly-developed delineation methodology results with that of the delineation done for Vegter Region 65.

The comparison is shown in Figure 63 and as expected, distinct differences exist between the delineation results as the previous methodology used four parameters to determine a delineation class. The new methodology only relies on borehole yield values to perform delineation based on a classification tree.

What is of interest in these results is that the majority of the boreholes in the northern part of the region have all been classified into a single inferred transmissivity class due to the high yields as opposed to the detailed delineation regions observed from the previous methodology mainly driven by the borehole density in that region (Figure 33). As mentioned in an earlier section, the advantage of the new delineation methodology is that the results are continuous across the boundaries.

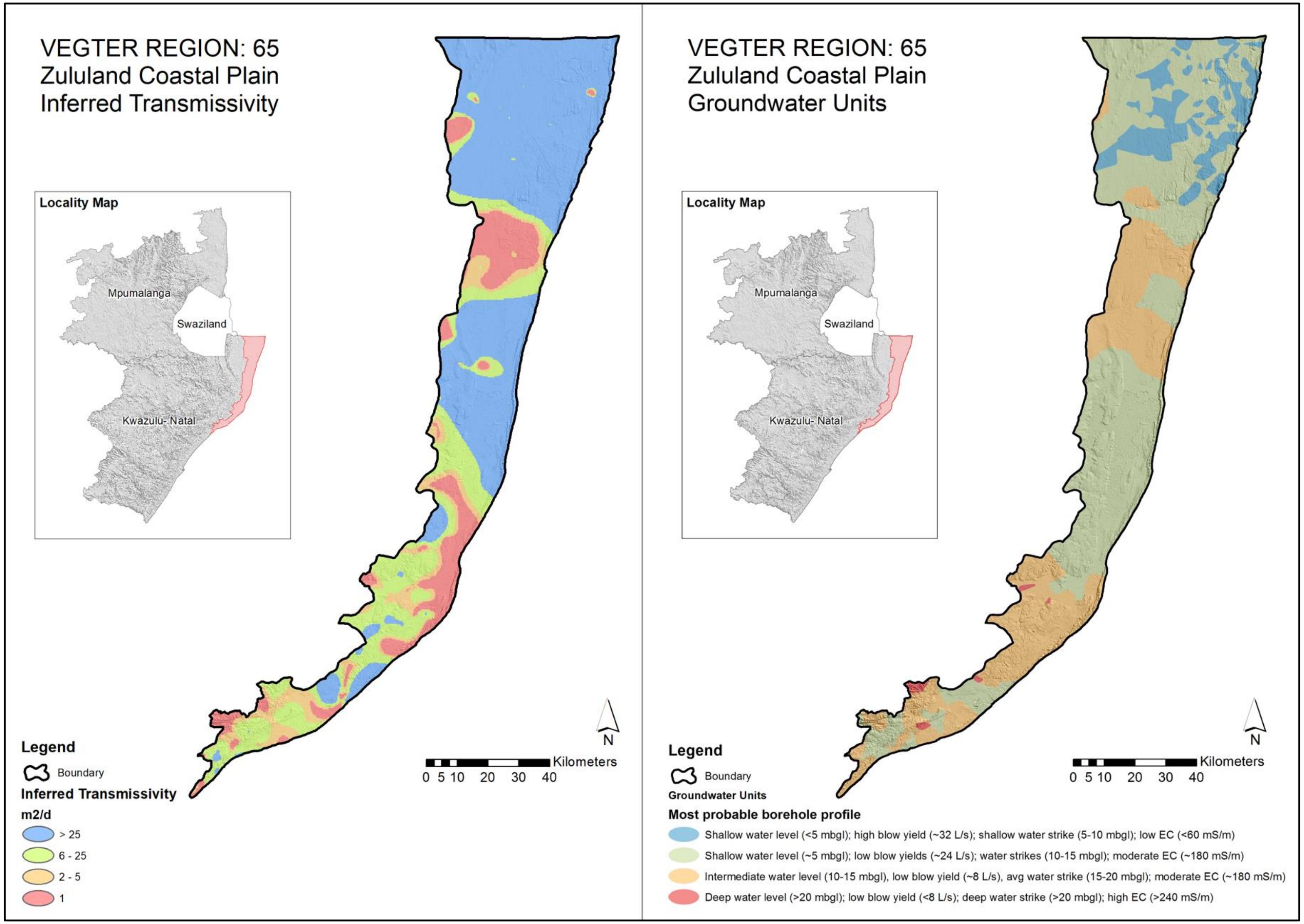


Figure 63 – Vegter Region 65 delineation, showing the differences In the results found from a previously used delineation methodology and the newly developed methodology

4.4 Statistical analysis

In this section the frequency analysis of identified parameters, the water level correlation analysis and the water character are discussed.

4.4.1 Frequency analysis

Frequency analyses were performed on the *water level*, *water strike*, *yield* and *EC* values of the study area. The bin sizes for the histogram generation were selected as shown in Table 20.

TABLE 20 - SELECTED BIN SIZES FOR FREQUENCY ANALYSIS

Parameter	***Bin Size***
Water level	5 mbgl
Water strike	5 mbgl
Borehole yield	1 l/s
Electrical conductivity	25 mS/m

The results of the frequency analysis are presented in Figure 64 where SD refers to standard deviation in each of the graphs presented. Each of the histograms shown has to present the borehole count associated with each bin on the y-axis and the specific parameter with associated units on the x-axis. The histograms have specifically not been normalised as to also present the borehole count.

Based on the results of Figure 64, the most probable value for each of the parameters presented in Table 20 is given in Table 21. Note these must be considered in the context of the specified bin size. By decreasing the associated bin size, a higher resolution result can be obtained.

TABLE 21 – MOST PROBABLE VALUE FOR PARAMETERS ANALYSED

Parameter	***Value***
Water level	10 mbgl
Water strike	25 mbgl
Borehole yield	1 l/s
Electrical conductivity	125 mS/m

Most of the distribution functions exhibit a broad response which indicates high variability over the area. The study area chosen is quite large in size (12028 km^2) and comprises multiple geologies that explain this high variability in the parameters.

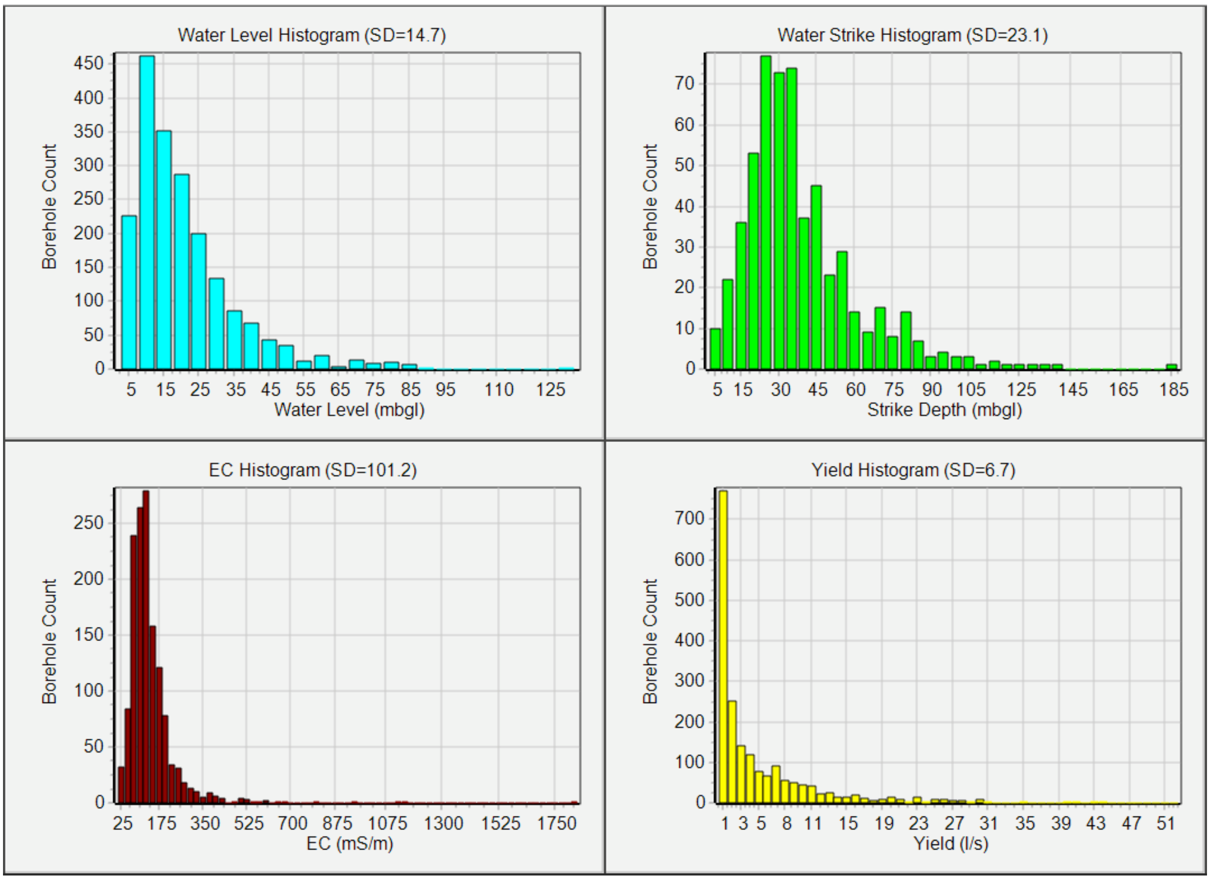


FIGURE 64 - RESULTS OF FREQUENCY ANALYSIS OF VEGTER REGION 7

The frequency analysis of the borehole yield of boreholes intersecting a dyke and those where no dyke intersection is present is shown in Figure 65.

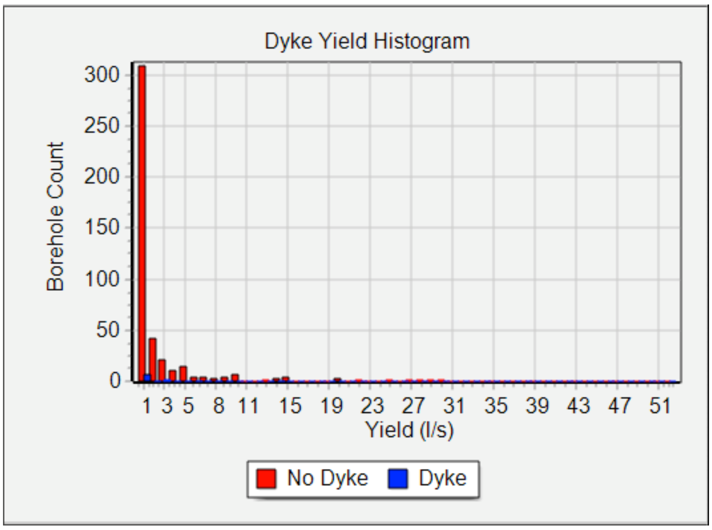


FIGURE 65 – FREQUENCY ANALYSIS OF BOREHOLES INTERSECTING AND NOT INTERSECTING DYKES

The zoomed view of Figure 65 is presented in Figure 66 to get a better view of the results of the boreholes intersecting the dykes. From the results it is clear that minimal yielding boreholes were intersecting dykes and the high yielding boreholes are not associated with any dyke intersection.

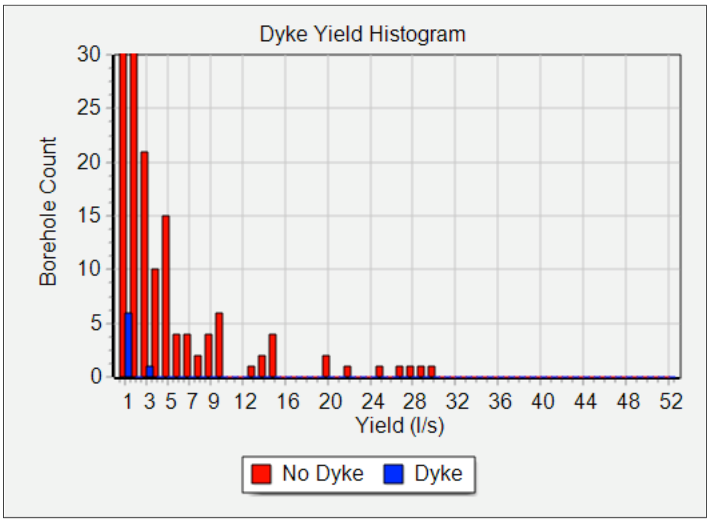


FIGURE 66 – ZOOMED VIEW OF BOREHOLES INTERSECTING AND NOT INTERSECTING DYKES

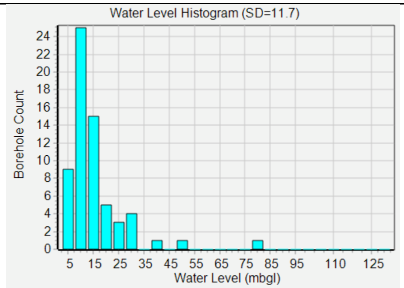
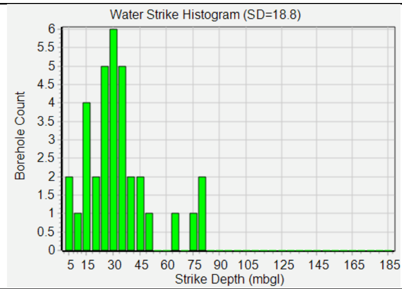
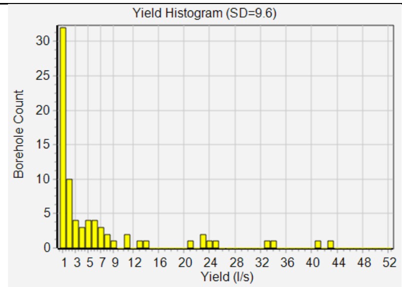
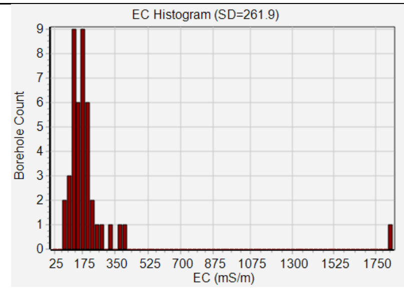
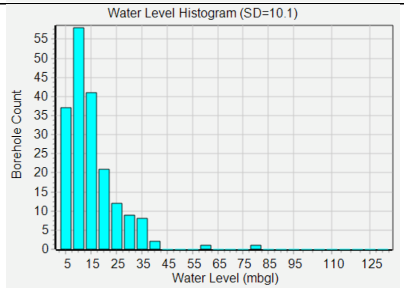
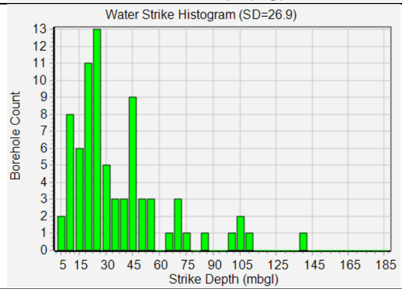
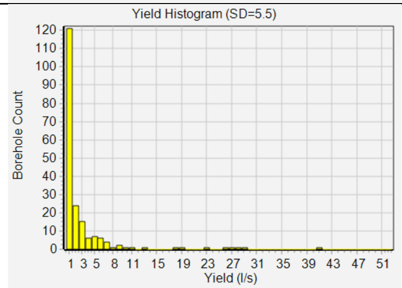
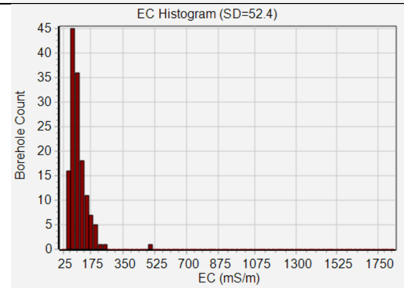
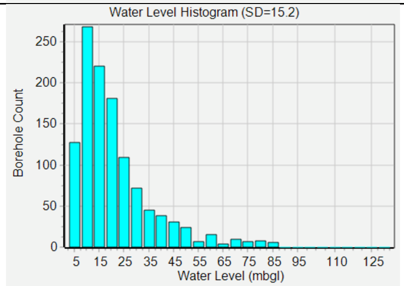
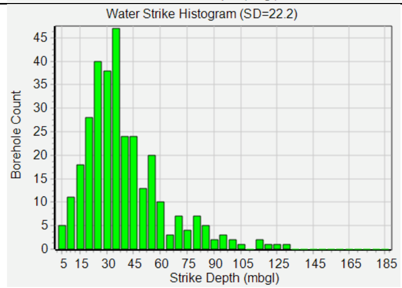
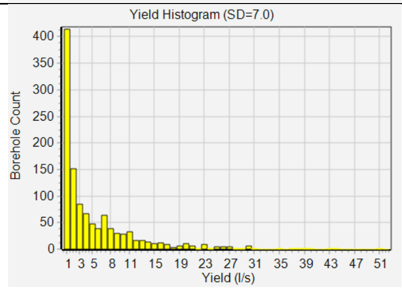
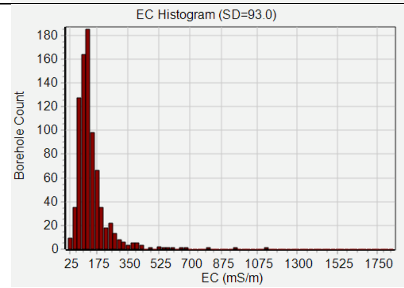
The study area was further analysed in the context of the simplified geology and the results obtained are summarised in Table 23. Note a total number of 20 surface geology types are intersected in the study area, but only the results of six are presented due to the number of boreholes present in each geology type. Frequency analysis requires multiple data points to produce usable results. An information summary of these surface geologies is provided in Table 22- Geological information summary.

Since the yield relationship between boreholes that intersect dykes and those that do not were very clear for the total study area, and this analysis is not repeated here in the context of the geology as the yield analysis already provides the required information.

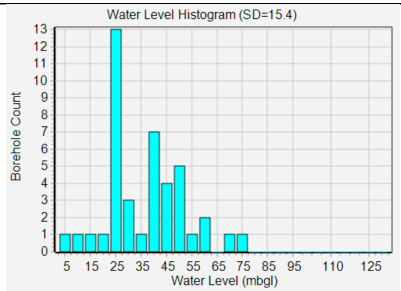
TABLE 22- GEOLOGICAL INFORMATION SUMMARY

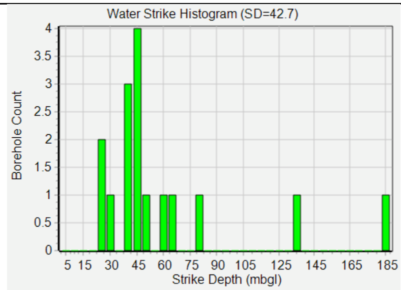
Geology	Strat name	Era	Description
R-Vt	Turfloop granite	Randian	Medium to coarse-grained grey and pink biotite granite.
R-Vma	Matok granite	Randian	Coarse-grained porphyritic, pink and grey biotite granite, in places hornblende granite
Rho	Houtrivier gneiss	Randian	Leucocratic migmatite and gneiss, grey and pink hornblende-biotite gneiss, grey biotite gneiss; minor muscovite bearing granite, pegmatite and gneiss.
Zb	Bandoliers complex	Swazian	Gneiss, Marble, calc-silicate rocks, Metapelite, amphibolite, maphic granulite, magnetite quartzite.
Zp	Pietersburg group	Swazian	Lava, quartzite, conglomerates, schist, part of the Bandolierskop complex.
Zgo	Goudplaats gneiss	Swazian	Dark-grey, in places light-grey biotite-hornblende gneiss with anatectic mobilisates; banded gneiss.

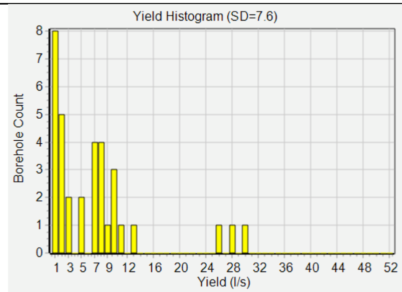
TABLE 23 - SUMMARY OF FREQUENCY ANALYSIS IN THE CONTEXT OF GEOLOGY

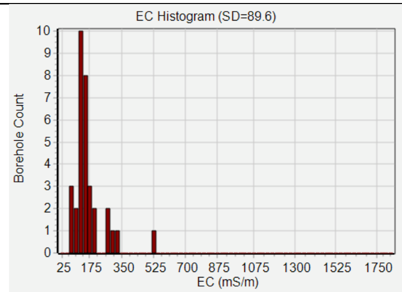
<table>
<tr><th>Geology</th><th>Water level</th><th>Water strike</th><th>Borehole yield</th><th>Electrical conductivity</th></tr>
<tr><td>R-Vma</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>R-Vt</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>Rho</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
</table>

Zb

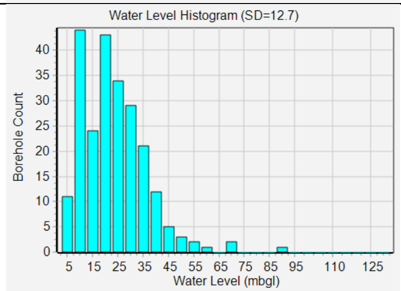

Water Level Histogram (SD=15.4)
Borehole Count
Water Level (mbgl)

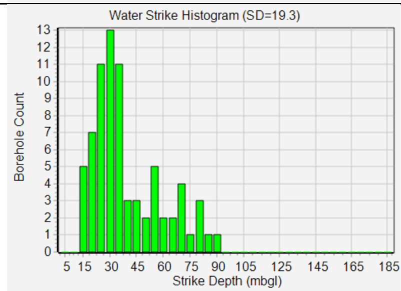

Water Strike Histogram (SD=42.7)
Borehole Count
Strike Depth (mbgl)

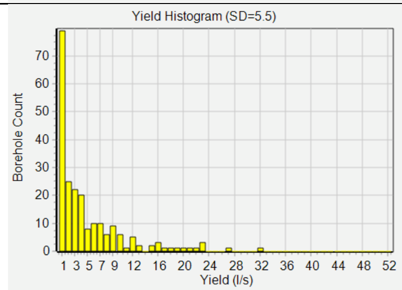

Yield Histogram (SD=7.6)
Borehole Count
Yield (l/s)

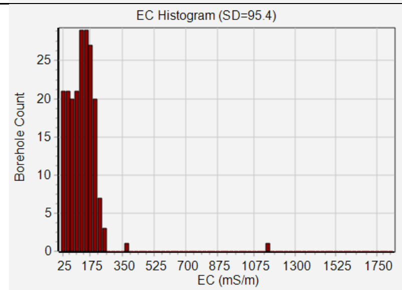

EC Histogram (SD=89.6)
Borehole Count
EC (mS/m)

Zgo

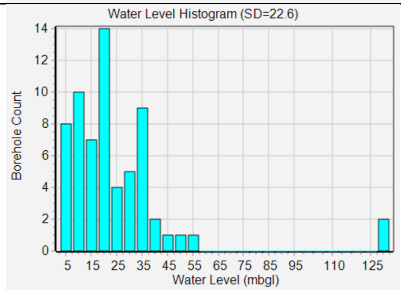

Water Level Histogram (SD=12.7)
Borehole Count
Water Level (mbgl)

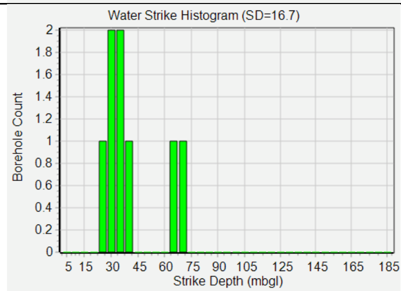

Water Strike Histogram (SD=19.3)
Borehole Count
Strike Depth (mbgl)

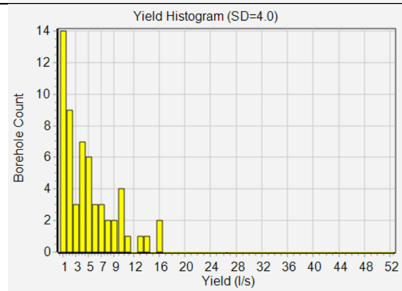

Yield Histogram (SD=5.5)
Borehole Count
Yield (l/s)

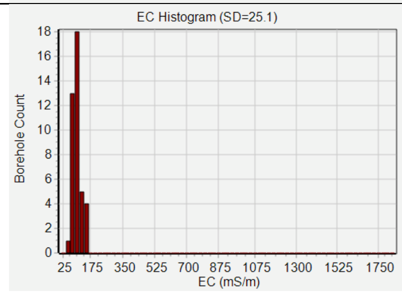

EC Histogram (SD=95.4)
Borehole Count
EC (mS/m)

Zp


Water Level Histogram (SD=22.6)
Borehole Count
Water Level (mbgl)


Water Strike Histogram (SD=16.7)
Borehole Count
Strike Depth (mbgl)


Yield Histogram (SD=4.0)
Borehole Count
Yield (l/s)


EC Histogram (SD=25.1)
Borehole Count
EC (mS/m)

4.4.2 Water-level analysis

The water level correlation analysis for the study area is shown in Figure 67. The borehole elevations for the study area are estimated to vary between 800 mamsl and 1700 mamsl according to Figure 67. For the most part a strong correlation exists between water level and surface topography, but selected boreholes do deviate from this correlation. This aforementioned deviation could be due to the aquifer type or abstractions that took place at the time of water level measurement.

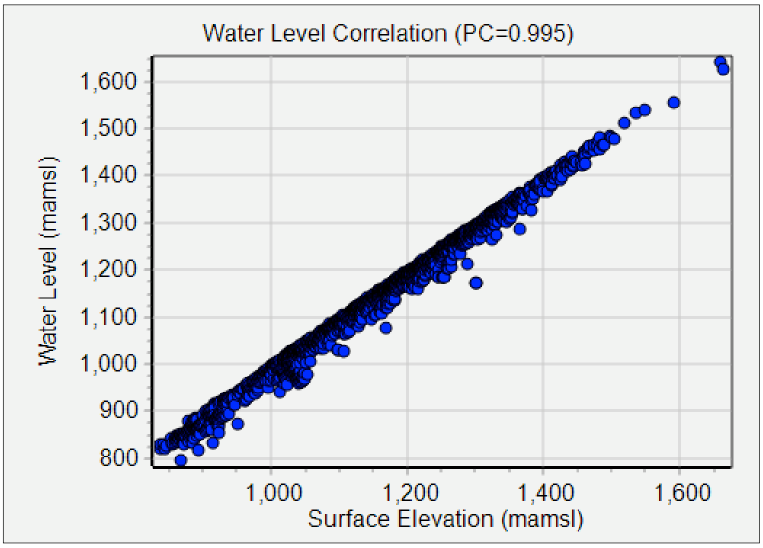
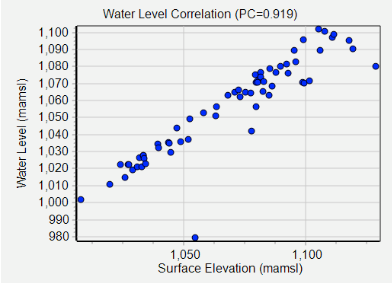
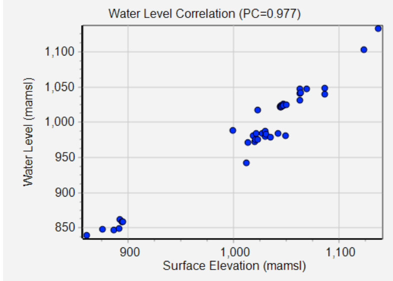
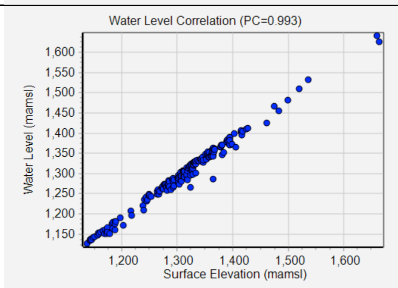
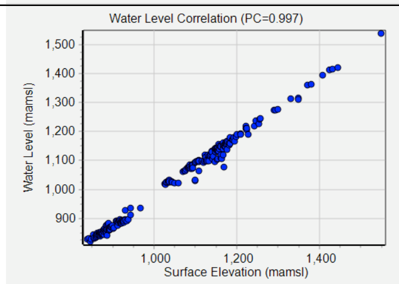
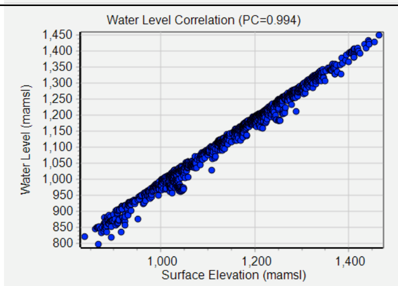
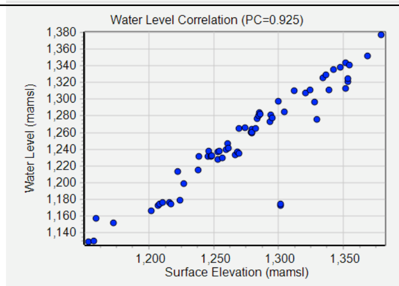


FIGURE 67 - VEGTER REGION 7 WATER LEVEL CORRELATION ANALYSIS

The summary of the water-level correlation in the context of surface geology is presented in

Table 24. It is noted that in some surface geologies the water levels correlate more with the surface elevation than in others but still the difficulty exists to establish concrete answers or even to try and predict certain properties to be expected solely by looking at the surface geology data. The reason is that the surface geology might only represent a small percentage of the total geological profile as intersected when drilling. When the geology in borehole logs is compared to the mapped surface geologies one will find that it differs drastically from the geology encountered when drilling occurred. Where more geological borehole log data is available, more accurate assumptions can be made in this regard.

TABLE 24 - SUMMARY OF FREQUENCY ANALYSIS IN THE CONTEXT OF GEOLOGY

<table>
<tr><th>Geology</th><th>Water level correlation</th><th>Geology</th><th>Water level correlation</th></tr>
<tr><td>R-Vma</td><td>
</td><td>Zb</td><td>
</td></tr>
<tr><td>R-Vt</td><td>
</td><td>Zgo</td><td>
</td></tr>
<tr><td>Rho</td><td>
</td><td>Zp</td><td>
</td></tr>
</table>

4.4.3 Water character analysis

The water character of the study area was analysed in the context of geology. The results for the different geologies are presented in Figure 68 to Figure 71. The results are presented over multiple Piper diagrams so as to not clutter the display.

It was expected that the water character would be dependent on the geology, but as can be seen from the results, this is not the case and a high variability of water character is observed within each geological formation. This could possibly be explained by the fact that the geology used is the surface geology and that each of the boreholes traverse various lithologies which all contribute to the final water character. Another factor to consider is the sampling method used for each borehole and the depth at which the sample was taken as all these factors could influence the measured water character.

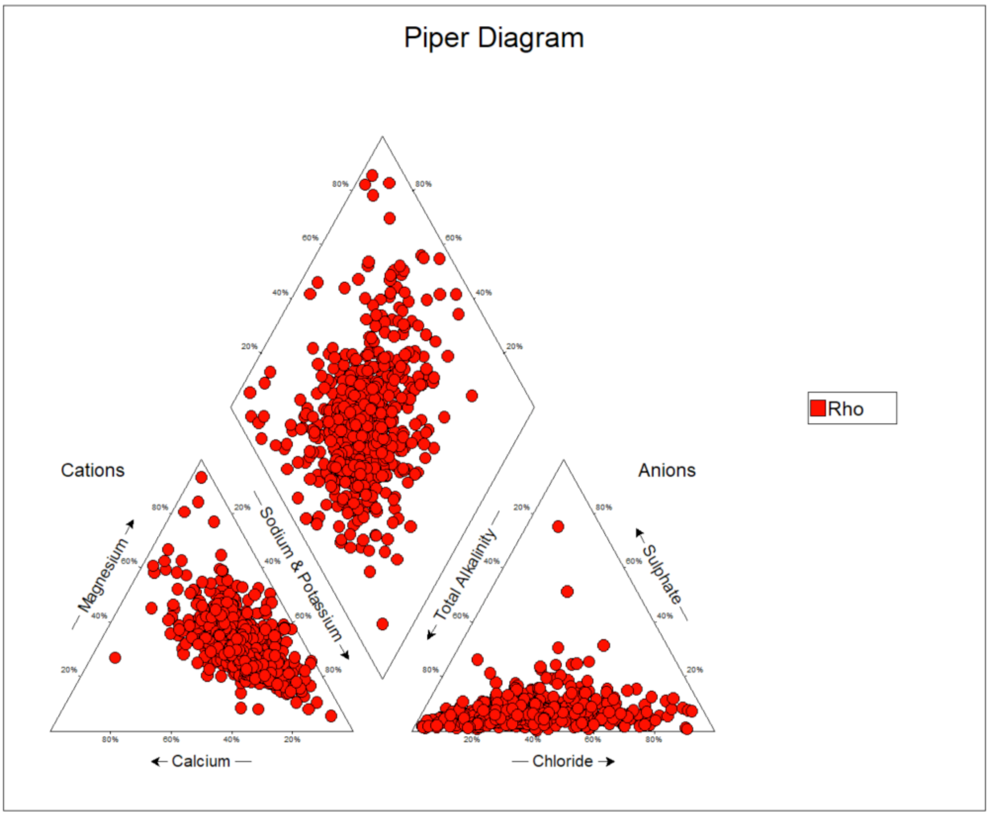


FIGURE 68 – PIPER DIAGRAM FOR THE RHO GROUP

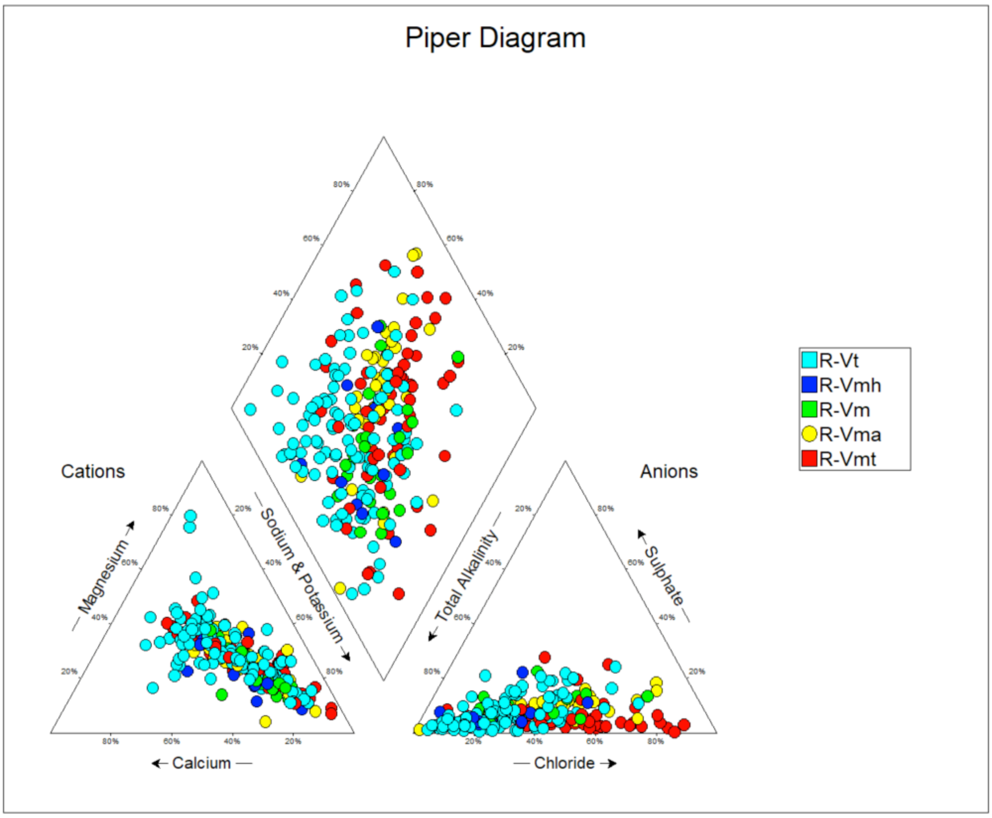


FIGURE 69 – PIPER DIAGRAM FOR THE R-VT, R-VMH, R-VM, R-VMA AND R-VMT GROUPS

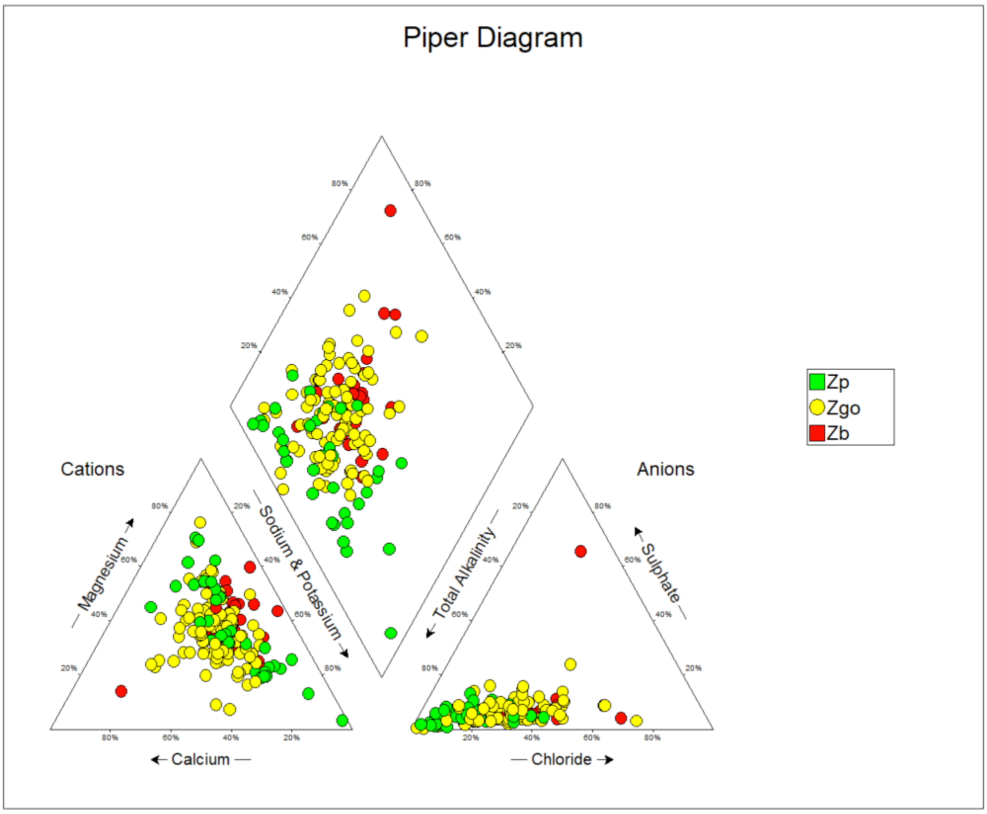


Figure 70 – Piper diagram for the Zp, Zgo and Zb groups

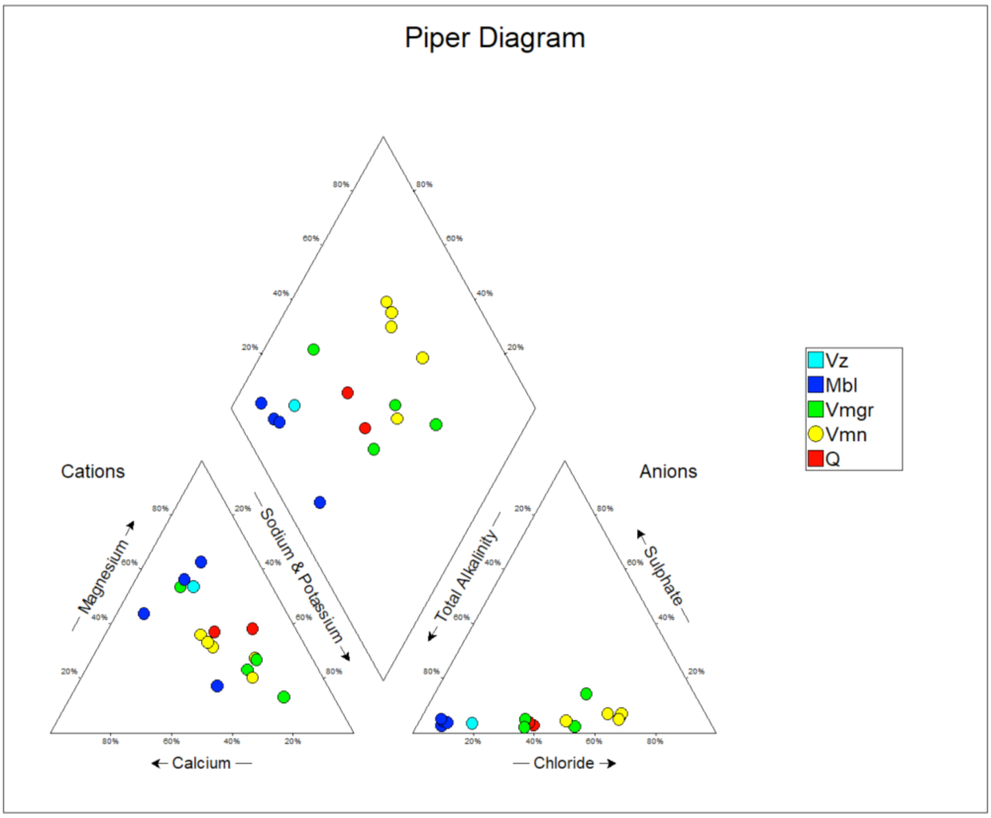


Figure 71 – Piper diagram for the Vz, Mbl, Vmgr, Vmn and Q groups

4.4.4 Conclusions

From the illustrated results in this section no defined conclusion could be reached to confirm opinions of the known industry that poorer water quality can be expected from deeper water strikes. In some cases, correlations between surface elevations and the depth to the water table were not as prominent as would normally be expected. One must admit that more concrete results had been expected here, but the results ultimately proved that underground water on its own and the relationship between geology and the occurrence of groundwater is more complex than just looking for answers in regional data such as surface geologies.

The overall conclusion reached is that areas should be investigated on a smaller scale to determine specific properties that can be expected for a certain project area and then this knowledge should be used to compile databases that can be used for future references. Studies on regional scale are of great assistance for the planning of groundwater projects, but geophysical investigations and especially expertise gathered in a specific region will ultimately help to make informed decisions on what properties to look for when searching for feasible amounts of groundwater. The changes of underground geological formations logged sometimes show different geology within a few metres from each other and may be the difference between a good yielding and a dry borehole; therefore, surface geology must be used as a starting point in an investigation rather than a groundwater target.

The Vegter statistical analyses do, however, provide a picture of the general groundwater occurrence as it relates to exiting boreholes in the area and remains a useful tool for the geohydrologist, that needs to describe the geohydrology of an area.

5 CONCLUSIONS

The aim of this study was to make use of existing borehole data and associated datasets to formulate a delineation methodology to delineate areas based on groundwater occurrence within the Limpopo Province.

A relational database (GSA) was successfully created by extracting the relevant borehole information from the Limpopo GRIP database as this database provided the most comprehensive dataset for the Province. The purpose of this database was twofold (i) executing statistical analysis of selected borehole parameters and (ii) performing groundwater occurrence delineation making use of the current borehole data.

The relevant borehole parameters were identified by looking at data availability and following the Vegter approach using frequency analysis of *water level*, *water strike*, *borehole yield*, *borehole yield vs. dyke intersection* and *EC*. Over and above the frequency analysis water level correlation and water character were also considered as part of the groundwater occurrence analysis. The water level analysis contributed to an understanding of the groundwater occurrence, but the water character results were too variable across an area to contribute to the understanding of the groundwater occurrence.

In addition to borehole specific information the following spatial layers were used to add a contextual filter field to the database, which included *simplified geology*, *quaternary catchment*, *groundwater occurrence*, *aquifer type*, *borehole depth*, *borehole elevation* and *aquifer vulnerability*. The aforementioned contextual field allowed for the statistical analysis proposed by Vegter (1995) to be carried out in the context of the selected parameter. Traditionally the Vegter analysis was only conducted in the context of the *simplified geology*. It is concluded that the selected parameters used in the frequency analysis together with the contextual field provides the level of detail required to characterise the groundwater occurrence in an area.

The existing delineation methodology proposed by Dennis and Dennis (2014) and also implemented by Holtzhausen (2018) was assessed and the following shortcomings were identified when applied over large areas:

- It is not always the case that the deeper water levels are associated with the higher EC values or that the shallow water levels are associated with higher blow yields. Therefore, it becomes increasingly difficult to formulate four unique classes in terms of the parameter combinations.
- The delineation class method considers the whole range of values for each parameter and transforms them to a delineation class based on the results of the frequency analysis. Therefore, when applied to large areas with a wide range of parameter values, delineation detail is lost in certain areas and will only be represented by a single class.

- Continuity is lost in delineation when considering adjacent areas which exhibit a contrast in parameter values, since the delineation classes for each area are formulated in terms of the frequency analysis of the total range of values.

Based on the aforementioned shortcomings it was proposed that delineation should be based on a physical property and not a calculated class. This led to the investigation of using aquifer parameters (storativity and transmissivity) to express the groundwater occurrence. The Cooper-Jacob equation and an equation used by the Canadian International Development were identified that relate aquifer parameters and borehole yield. It was shown that by using estimations for the storativity in the Cooper-Jacob equation together with recorded borehole yields an inferred transmissivity could be obtained. To validate the relationship between transmissivity and borehole yield, a classification algorithm based on borehole yield values was developed that also computed an inferred transmissivity of the area.

The correlation between the actual and inferred transmissivity resulted in roughly 42%. The inferred transmissivity was also calculated making use of the Cooper-Jacob equation under certain assumptions, which resulted in a correlation of 38% compared to the actual transmissivity. The reason for these low correlations is attributed to the fact that the mean value of actual transmissivities obtained from pump tests is an order of magnitude smaller than the inferred transmissivities, although the probability distribution of higher values seems to be quite similar. It is concluded that the inferred and actual transmissivities might never exhibit a high correlation, but the inferred transmissivity sufficiently represents a physical parameter that can be used for delineation purposes.

The inferred transmissivity was selected to represent this physical parameter as required by one of the objectives of the study to allow for continuity over adjacent areas and retain delineation detail in areas with highly contrasting parameter values. The delineation is then performed selecting an appropriate interpolation by considering the RMSE of the selected method to be applied to the inferred transmissivity.

Investigation into the data density confirms that it plays a major role in the final result as expected and as would be the case for any other methodology relying on point source data.

A case study showed a higher correlation between the classification tree results and actual results versus that of the Cooper-Jacob approach. Both of the aforementioned methods roughly had a 30% correlation with the actual transmissivity value, but as stated earlier the purpose of the inferred transmissivity is to base delineation on a physical parameter and it is not likely that the inferred transmissivity will ever show high correlation with the actual transmissivity, but it serves the delineation purpose.

Finally, it is concluded that the newly-developed delineation methodology succeeds in producing continuous delineation across boundaries and accommodates areas with highly contrasting parameter values as the delineation is related to a physical parameter value and not a calculated class based on the available data range. The delineation methodology was also applied to Vegter Region 65 for comparison purposes against the existing delineation. It was concluded that the newly-developed delineation methodology performs better than the existing one and addressed the identified issues discussed earlier.

6 RECOMMENDATIONS

It recommended that a buffer be placed around the study area under consideration to ensure proper interpolation across the study area boundary, as interpolation methods perform poorly when extrapolating. This simple measure increases continuity across adjacent study areas as additional boreholes outside the study area are also taken into consideration.

Although both the Cooper-Jacob and classification tree methods were considered, it is recommended to only use the classification tree, as it slightly outperformed the Cooper-Jacob method and is not nearly as computationally expensive as the Cooper-Jacob method which requires multiple grid interpolations.

The current classification tree is based on observed behaviour of borehole yields and actual transmissivities in the Limpopo Province. It is, however, recommended to revise the classification algorithm as more data becomes available across the Province as this might improve the overall result when more data is available.

Finally, it is recommended that the methodology be implemented within a software tool due to the computational effort involved in both the statistical and delineation calculations. This will allow for quick assessments of any area and the methodology could possibly be further expanded to include the whole of South Africa.

7 REFERENCES

Aller, L., Bennett, T., Lehr, J., Petty, R. and Hackett, G. (1987) DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. National Water Well Association, Dublin Ohio / EPA Ada, Oklahoma. EPA-600/2-87-035.

Baron, J., Seward, P. and Seymour, A. (1998) The Groundwater Harvest Potential Map of the Republic of South Africa. Technical Report GH 3917. Directorate of Geohydrology, Department of Water Affairs and Forestry, Pretoria, South Africa.

Bohling G. 2005. Introduction to Geostatistics and Variogram analysis. Kansas Geological Survey. USA.

Botha, F.S. (2005). A proposed method to implement a groundwater resource information project (GRIP) in rural communities, South Africa. PhD Thesis, University of the Free State, South Africa.

Botha, F.S., Du Toit, W.H., Holland, M. and Weidemann, R.V. (2011) Can groundwater be successfully implemented as a bulk water resource within rural Limpopo Province? Analysis based on GRIP datasets. Conference proceedings: Our source of security in an uncertain future, Pretoria, 19-21 September 2011.

Bromley, J., Manstrom, B., Nisca, D. and Jamtlid, A.A. (1994) Airborne geophysics: Application to groundwater study in Botswana. *Ground Water,* 32 79-90.

Cairncross, B. (2004) Field guide to rocks and minerals of Southern Africa. Struik New Holland Publishing, Cape Town, South Africa.

CGS (1974) Geological series 2426 Thabazimbi (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (1978) Geological series 2428 Nylstroom (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (1981) Geological series 2230, Messina (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (1985) Geological series 2230, Tzaneen (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (1985) Geological series 2328, Pietersburg (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (1986) Geological series 2430 Pilgrim's rest (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (1993) Geological series 2326, Ellisras (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (2000) Geological series 2228, Alldays (1:250 000). Council for Geoscience, Pretoria, South Africa.

CGS (2003) Simplified Geological Map of South Africa. http://www.syfnx.co.za/cgs-site/images/Maps/rsageology.gif. Accessed: 14 Apr 2016.

De Cock, N., D'Haese, M., Vink, N., Van Rooyen, C.J., Staelens, L., Schonfeldt, H.C., D'Haese, L. (2013) Food security in rural areas of Limpopo Province, South Africa.

Dennis, I. and Dennis, S.R. (2011). Climate Change: Impacts on South African groundwater resources. In proceedings of 2011 Geological Society of South Africa Groundwater Conference 2007, Pretoria, South Africa.

Dennis, I. and Dennis, S.R. (2014) Hydrogeology of groundwater region 65: Northern Zululand Coastal Plain. WRC Report No. K5/2251/1. Water Research Commission, Pretoria, South Africa.

Dennis, S.R. and De Klerk, T. (2012) Limpopo Groundwater Atlas. Department of Water Affairs, Pretoria, South Africa.

Du Toit, W.H., Mullin, H. and Jonck, F. (1998) Hydrogeological map series of South Africa, Phalaborwa 2330 sheet (1:500000). Government Printer, Pretoria, South Africa.

DWAF (1996) South African Water Quality Guidelines, Volume 1 Domestic Use. Department of Water Affairs and Forestry, Pretoria, South Africa. ISBN 0-7988-5339-5.

DWAF (2004) Standard Descriptors for Geosites. Department of Water Affairs and Forestry, Pretoria, South Africa.

DWAF (2006) Groundwater Resource Assessment II – Task 2C Groundwater Planning Potential. Department of Water Affairs and Forestry, Pretoria, South Africa.

DWAF (2009a) Review of GRA1, GRA2 and international assessment methodologies. Contract No. WP 9390. Department of Water Affairs and Forestry. Pretoria, South Africa.

DWAF (2009b) Strategy and Guideline Development for National Groundwater Planning Requirements, A proposed GRA3 Methodology. DWAF, Pretoria, South Africa.

DWAF (2010) Groundwater Strategy 2010. Department of Water Affairs and Forestry. Pretoria, South Africa.

DWAF (2013) National Water Resources Strategy. Department of Water Affairs and Forestry. Pretoria, South Africa.

DWS (2015) CHART. https://www.dwa.gov.za/groundwater/chart.aspx. Date of access: 26 Sep. 2016.

DWS (2016) Aquifer classification of South Africa. https://www.dwa.gov.za/groundwater/ACSA.aspx Date of access: 12 Oct. 2016.

DWS (2017) Distribution of boreholes stored in the National Groundwater Archive (NGA). Available from www.dwa.gov.za/Groundwater/Documents.aspx. Data of access: 21 Oct. 2017.

Fashae, O.A., Tijani, M.N., Talabi, A.O., Adedeji, O.I. (2013) Delineation of groundwater potential zones in the crystalline basement terrain of SW-Nigeria: an integrated GIS and remote sensing approach. *Applied Water Science* (2014), 4:19–38.

Haupt, C.J. (2001) Water Resources Situation Assessment. Directorate of Water Resource Planning. Department of Water Affairs and Forestry. Groundwater Resources of South Africa. Compiled by WSM Engineers.

Holland, M. (2011) Hydrogeological characterisation of crystalline basement aquifers within the Limpopo Province. South Africa. PhD Thesis, University of Pretoria.

Holland, M. (2012) Evaluation of factors influencing transmissivity in fractured hard-rock aquifers of the Limpopo Province. *Water SA*, 38, 3.

Holtzhausen, P.H. (2018) Groundwater Resource Quality Objectives for the Crocodile West Catchment. MSc Dissertation, North-West University, South Africa.

Ikechukwu, M.N., Ebinne, E., Idorenyin, U. and Raphael, N.I. (2017) Accuracy assessment and comparative analysis of IDW, Spline and Kriging in spatial interpolation of landform (topography): An experimental study. *Journal of Geographic Information System*. ISSN Online: 2151-1969.

Javed, A. and Wani, M.H., 2009. Delineation of groundwater potential zones in Kakund watershed, eastern Rajasthan, using remote sensing and GIS techniques. *Journal Geological Society of India*, 73(2), 229-236.

Kovalevsky, V.S., Kruseman, G.P. and Rushton, K.R. (2004) Groundwater studies: An international guide for hydrological investigations. IHP-VI, Series on Groundwater No. 3. UNESCO. ISBN 92-9220-005-4.

Kruseman, G.P. and De Ridder, N.A. (1991) Analysis and evaluation of pumping test data (2nd Edition). International Institute for Land Reclamation and Improvements. The Netherlands.

Mondal S.M., Pandey, A.C. and Garg, R.D., 2008. Groundwater prospects evaluation based on hydrogeomorphological mapping using high resolution satellite images: A case study in Uttarakhand. *Journal of Indian Society of Remote Sensing*, 36 (1), 69-76

Murray, R., Baker, K., Ravenscroft, P., Musekiwa, C., Dennis, R. 2012. A Groundwater planning-toolkit for the main Karoo basin: Identifying and quantifying groundwater-development options incorporating the concept of wellfield yields and aquifer firm yields. 38(3):407-416.

Parsons, R. and Conrad, J. (1999) Explanatory notes for the Aquifer classification Map of South Africa. WRC Report No. KV 116/98. Water Research Commission, Pretoria, South Africa.

Parsons, R. and Wentzel, J. (2005) Groundwater Resource Directed Measures. WRC Project K5/1427, Water Research Commission, Pretoria, South Africa.

Riemann, K., Blake, D. (2010) Groundwater Reserve Determination for current and potential Wellfield Development of TMG Aquifers. WRC Report No KV 236/10. Water Research Commission, Pretoria, South Africa.

Rosewarne, P., Woodford, A. and Girman, J. (2006) How much groundwater does South Africa have? www.srk.co.uk/groundwater/PDFs/1_A_Woodford.pdf. Date of access: 07 Jul. 2014.

Schulze, R.E. (1997) South African of Agrohydrology and Climatology, Water Research Commission, WRC Report TT82/96, Water Research Commission, Pretoria, South Africa.

Seymour, A. (1994) Groundwater Resources of the Republic of South Africa. Geographic Information Systems, ARCIMS Mapservices, www.dwaf.gov.za.

Sikakwe, G.U. (2018) GIS-based model of groundwater occurrence using geological and hydrogeological data in Precambrian Oban Massif South eastern Nigeria. *Applied Water Science* (2018), 8:79.

South Africa (1998). South African National Water Act, Act No. 36 of 1998. Pretoria, Government Printer, 1998.

StatsSA (2014) General household survey 2013. Statistics South Africa, Pretoria, South Africa. http://beta2.statssa.gov.za/publications/P0318/P03182013.pdf. Date of access: 07 Jul. 2014.

Van Tonder, G.J., Van Sandwyk. L. and Buys, J. (1996) Program TRIPOL, version 1.0, Institute for Groundwater studies, and Water Research Commission.

Vegter J.R. (2001b) Hydrogeology of groundwater: Region 3 - Limpopo Granulite-Gneiss Belt. WRC Report No TT 136/00, Water Research Commission, Pretoria, South Africa.

Vegter J.R. (2003b) Hydrogeology of groundwater: Region 19 - Lowveld. WRC Report No TT 208/03, Water Research Commission, Pretoria, South Africa.

Vegter J.R. (2006) Hydrogeology of groundwater: Region 26 - Bushmanland. WRC Report No TT 285/06, Water Research Commission, Pretoria, South Africa.

Vegter, J.R. (1995) An explanation of a set of national groundwater maps. WRC Report No. TT 74/95, Water Research Commission, Pretoria, South Africa.

Vegter, J.R. (2000) Groundwater Development in South Africa and an introduction to the Hydrogeology of Groundwater Regions. WRC Report No. TT 134/00, Water Research Commission, Pretoria, South Africa.

Vegter, J.R. (2001a) Hydrogeology of groundwater: Region 1 - Makoppa Dome. WRC Report No TT 135/00, Water Research Commission, Pretoria, South Africa.

Vegter, J.R. (2003a) Hydrogeology of groundwater regions, Region 7, Polokwane/Pietersburg Plateau. WRC Report No. TT 209/03. Water Research Commission, Pretoria, South Africa.

Xu, Y. and Beekman, H.E. (2003) Groundwater recharge estimation in Southern Africa. UNESCO IHP Series No. 64, UNESCO Paris. ISBN 92-9220-000-3.

APPENDIX A – INVERSE DISTANCE WEIGHTING

The following is taken directly from Ikechukwu *et al*. 2017

This method assumes that the value at an unknown location can be approximated as a weighted average of values at points within a certain cut-off distance, or from a given number of the closest points (typically 10 to 30). Weights are usually inversely proportional to a power of distance which, at unsampled locations, leads to an estimator as contained in equation below:

$$F(s) = \sum_{i=1}^{n} w_i z(s_i) = \frac{\frac{\sum_{i=1}^{m} z(s_i)}{|s - s_i|^p}}{\sum_{j=1}^{m} \frac{1}{|s - s_i|^p}} \tag{1A}$$

where p is a parameter (typically = 2). IDW is a method that is easy to use and readily available; it frequently does not produce the local shape implied by data and produces local extrema at the data points. Some modifications have given rise to a class of multivariate blended IDW surfaces and volumes.

The assumption for IDW is that measured points closer to the unknown point are more like it than those that are further away in their values. The weight is given as:

$$\lambda = \frac{\frac{1}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \tag{2A}$$

Where d_i is the distance between x_0 and x_i, p is a power parameter, and n is the number of measured points used for the estimation. The main factor affecting the accuracy of IDW is the value of the power parameter. Weights diminish as the distance increases, especially when the value of the power parameter increases, so nearby samples have a heavier weight and have more influence on the estimation, and the resultant spatial interpolation is local.

The choice of power parameter and neighbourhood size is arbitrary. The most popular choice of p is 2 and the resulting method is often called inverse square distance or inverse distance squared (IDS). IDW is referred to as "moving average" when p is zero, "linear interpolation" when p is 1 and "weighted moving average" when p is not equal to 1.

APPENDIX B – SPLINE INTERPOLATION

The following is taken directly from Ikechukwu *et al*., 2017

Regularized spline with tension

For regularized spline with tension and smoothing, the prediction is given by:

$$z(s_0) = a_1 + \sum_{i=1}^{n} w_i R(v_i) \quad (1B)$$

where a_1 is a constant and $R(v_i)$ is the radial basis function given by:

$$R(v_i) = -[E_1(v_i) + In(v_i) + C_e] \quad (2B)$$

and

$$v_i = \left[{}^{\phi h_0}/_{2}\right]^2 \quad (3B)$$

where $E_1\ (v_i)$ is the exponential integral function, C_E = 0.577215 is the Euler constant, ϕ is the generalized tension parameter and h_0 is the separation between the new and interpolation point. The coefficients a_1 and w_i are obtained by solving the system,

$$\sum_{i=1}^{n} w_i = 0 \quad (4B)$$

$$a_1 + \sum_{i=1}^{n} w_i \left[R(v_i) + \delta_{ij} \frac{\varpi_0}{\varpi_i}\right] = z(s_i); j = 1, \dots, n \quad (5B)$$

where $\varpi_0/\varpi i$ are positive weighting factors for a smoothing parameter at each location. The tension parameter ϕ determines the distance over which the given points influence the resulting surface, while the smoothing parameter controls the vertical deviation of the surface from the sample locations. The use of an appropriate combination of tension and smoothing produces a surface that correctly fits the empirical knowledge about the expected variation.

Thin plate spline

Thin plate splines (TPS), previously called Laplacian smoothing splines, for modelling climatic data. A basic solution to the bi-harmonic equation has the form

$$Z(r) = r^2 \log{(r)} \quad (6B)$$

where r is the distance between sample points and unsampled locations. The relation below approximates the surface with minimum bend

$$f(s) = a_1 + a_2x + a_3y + \sum_{i=1}^{n} w_i z(|s_i - s_0|) \quad (7B)$$

where the terms a_1, a_2x, a_3y model the linear portion of the surface defining a flat plain that best fits all control points using least squares, the last term models the bending forces due to m sampled points, w_i are control points coefficients and $|s_i - s_0|$ is the separation of sampled point s_i and location s_0. The unknowns a_1, a_2x, a_3y and w_i are evaluated using the relation:

$$L^{-1} = (w|a_1 a_2 a_3)^T \quad (8B)$$

where

$$L = \begin{bmatrix} K & P \\ P^T & 0 \end{bmatrix} \quad (9B)$$

and V is a vector of point heights. K is a matrix of the distance between sampled points and P is a matrix of the sampled points coordinates. L^{-1} is obtained by calculating the inverse of L. Once the unknowns are evaluated, one can compute $f(s)$ to determine the heights of unknown points. TPS computes a smoothing factor by limiting the Generalized Cross Validation function, GVC, making for a comparatively sturdy model as limiting the GVC improves the accuracy of estimations and is less reliant on the accuracy of the model itself. TPS gives a determination of spatial accuracy.

Inverse multi-quadratic spline

The relation below gives the inverse multi-quadratic spline function

$$f_i(s) = \frac{1}{1 + \|s - s_i\|^2} \quad (10B)$$

where $|s - s_i|$ is the Euclidean distance between control points s_i and the unknown points. The surface is modelled by the function

$$z(s) = \sum_{j=1}^{n} a_j f_j(s) \quad (11B)$$

where the weights a_i are selected to ensure exact estimations at each data point such that

$$z_i = z(s_i) = \sum_{j=1}^{n} a_j f_j(s_i), i = 1, \dots, n \quad (12B)$$

and is computed by the relation

$$z = Fa \quad (13B)$$

where z is replaced by a vector of sampled data values, F is a square function matrix given by,

$$L = \begin{bmatrix} K & P \\ P^T & 0 \end{bmatrix} \quad (14B)$$

The estimation function generated with these weights is smooth and exact at sampled data points.

Splines have been widely seen as highly suitable for estimation of densely sampled heights and climatic variables. Among its disadvantages, the inability to integrate larger amounts of auxiliary maps in modelling the deterministic part of change as well as the arbitrary selection of the smoothing and tension parameters has been widely criticized. Predictions obtained from splines therefore are largely dependent on decisions like the order of polynomial used and the number of break points taken by the user. Splines may also be modelled not to be exact to avoid the generation of excessively high or low values common with some exact splines. Unlike the IDW methods, the values predicted by RBFs are not constrained to the range of measured values, i.e., predicted values can be above the maximum or below the minimum measured value.

APPENDIX C – KRIGING INTERPOLATION

This section is adapted from Van Tonder *et al*., 1996.

Notations

Regionalized variables are variables distributed in space (and/or time). Mathematically, one can state that a regionalized variable is simply a function that describes the value of a characteristic quantity *z* at point x = (x,y) in space. We denote This quantity *z* is denoted as *z*(x).

A *random variable* is, by definition, a variable that can attain different numerical values, subject to a certain probability distribution. The random function associated with a random variable, z(x), is conventionally denoted by Z(x).

The basic estimation problem can now be defined as obtaining some estimate for the function Z(x) at a site x_0, where no observations on Z(x) are available.

The *estimator* for a function Z(x) at a site x_0 will be denoted as $Z^*(x_0)$.

The *observations* available for a given set of *n* regionalized variables or data points are denoted as $Z(x_i)$, i=1,...,n.

Semi-variogram computation

Because of the nature of Kriging, a semi-variogram, computed from the regionalized variables or data points, is needed to estimate the manner in which the mean values of the phenomena behave over the region, often referred to as the *drift* or *trend* of the regionalized variable. A mathematical function is then fitted to the semi-variogram values, to obtain certain parameters that are needed for interpolation by Kriging and Bayesian Kriging.

A semi-variogram describes the connection between two points at distance h from each other. It can be estimated by the function γ(h) in the following equation:

$$\gamma(h) = \frac{1}{2n(h)} \sum^{n(h)} \left[z(\mathbf{x}_i) - z(\mathbf{x}_j) \right]^2 \quad \text{for all } i,j \leq n$$

such that

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} = h \ ,$$

where d_{ij} is the distance between two points x_i and x_j, $\gamma(h)$ is the semi-variogram value of lag *h*, *n* the number of observations and *n(h)* is the number of pairs (x_i,x_j), such that $d_{ij} = h$.

In practice, the approximation for the semi-variogram γ(h) is computed for fixed values of h, given by some *basic lag distance*, a, say, thus for γ(a), γ(2a), γ(3a), and so forth, where γ(a) is computed for all pairs (x_i,x_j) such that $a-1 < d_{ij} \leq a$. Further, h is approximated by the mean distance between all pairs (x_i,x_j) used in the computation of γ(a). Thus γ(d)≡γ(a) for

$$d = \frac{1}{n(h)}\sum^{n(h)} d_{ij}$$

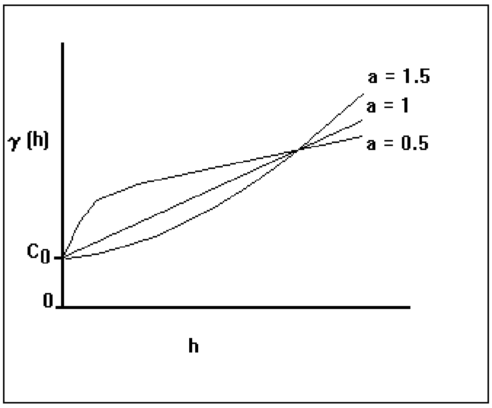
Experience indicates that γ(h) generally tends to increase with h, until it reaches a maximum value, called the *sill* at some lag a. This distance is customarily referred to as the *range* of γ(h) The semi-variogram is then approximated with either one of the six model equations, given in the next section.

Fitting

Interpolation by Kriging requires a theoretical semi-variogram and functional terms of an assumed trend in a given neighbourhood of the input data. The neighbourhood is determined by the number of nearest points that are used for interpolation. The most commonly used theoretical semi-variograms are shown in the following figures. The quantity, C_0, which corresponds to γ(*0*), is usually referred to as the *nugget effect*.

$$\gamma(h) = 0 \qquad h = 0$$

$$\gamma(h) = Ch^a + C_0 h > 0, \quad 0 < a < 2$$

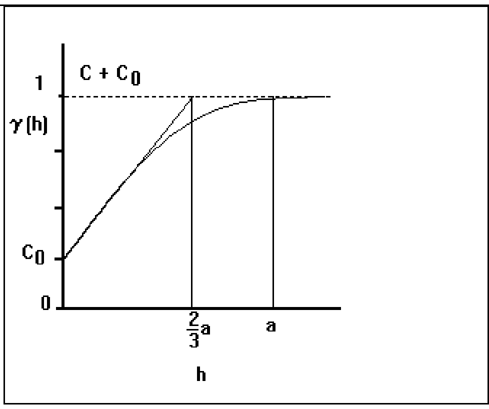


TYPE 1, POLYNOMIAL SEMI-VARIOGRAM

$$\gamma(h) = 0 \quad h = 0$$

$$\gamma(h) = C\left[1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right] + C_0 0 < h < a$$

$$\gamma(h) = C + C_0 \quad h > a$$

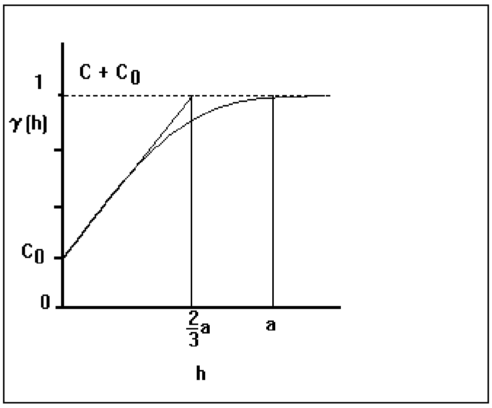


TYPE 2, SPHERICAL SEMI-VARIOGRAM

$$\gamma(h) = 0 \quad h = 0$$

$$\gamma(h) = C\left[1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right] + C_0 0 < h < a$$

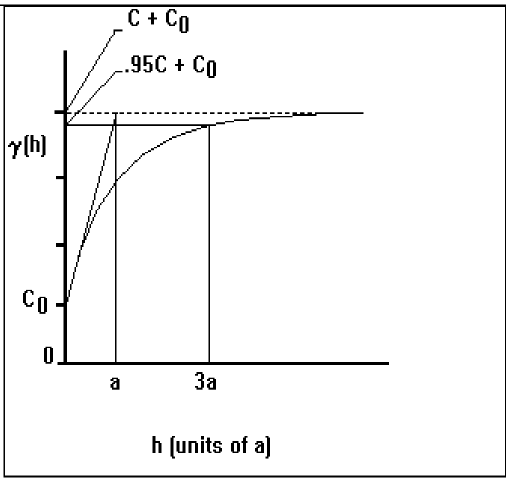
$$\gamma(h) = C + C_0 \quad h > a$$



TYPE 2, SPHERICAL SEMI-VARIOGRAM

$$\gamma(h) = 0 \qquad h = 0$$

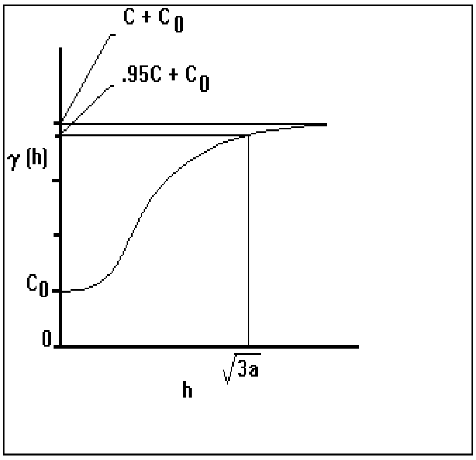
$$\gamma(h) = C\left[1 - e^{-\left(\frac{h}{a}\right)}\right] + C_0 \qquad h > 0$$



TYPE 3, EXPONENTIAL SEMI-VARIOGRAM

$$\gamma(h) = 0 \qquad h = 0$$

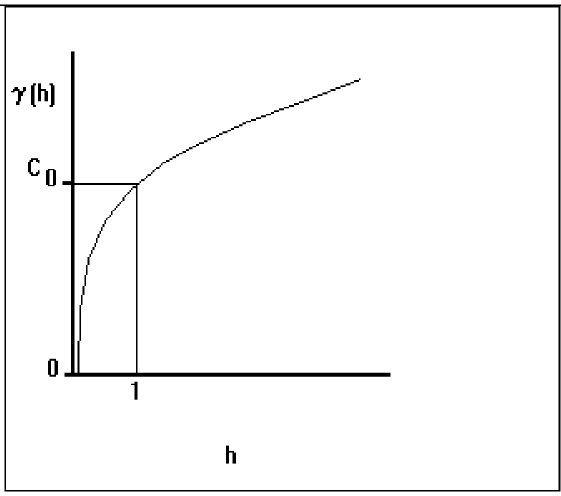
$$\gamma(h) = C\left[1 - e^{-\left(\frac{h}{a}\right)^2}\right] + C_0 \qquad h > 0$$



TYPE 4, GAUSSIAN SEMI-VARIOGRAM

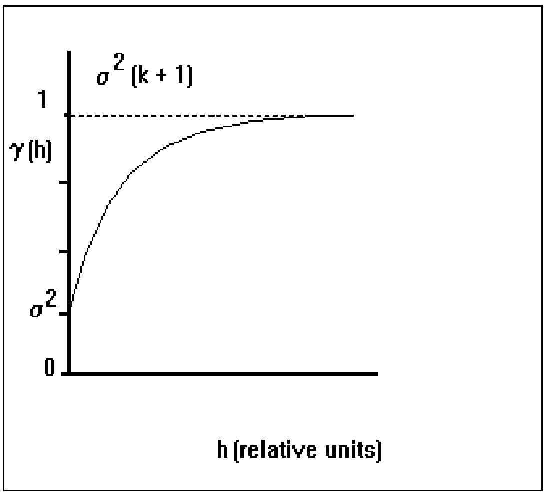
$$\gamma(h) = 0 \qquad h = 0$$

$$\gamma(h) = a\ln(h) + C_0 \qquad h > 0$$



TYPE 5, DE WIJSIAN SEMI-VARIOGRAM

$$\gamma(h) = \sigma^2 \qquad h = 0$$
$$\gamma(h) = \sigma^2\left[1 + k\left(1 - \rho^h\right)\right] \qquad h > 0$$



TYPE 6, KRHO OF DE WAAL SEMI-VARIOGRAM

Formulation

The most appropriate way to describe the spatial variability of environmental variables, is to present them with random functions. This approach has the advantage that it allows one to describe an environmental variable in statistical terms, through the Theory of Regional Variables. The best-known estimation method, based on this approach, is Ordinary Kriging or *Kriging*, as it is conventionally known.

Interpolation done with Kriging, where the mean value of Z(**x**) is unknown, is given by:

$$Z^*(\mathbf{x}_o) = \sum_{i=1}^{n} w_i Z_i$$

where the weight function w_i is calculated by solving the system of linear equations

$$\sum_{j=1} w_j \gamma_{ij} = \gamma_{io}, \text{ with } \sum_{i=1} w_i = 1 \text{ where } \gamma_{ij} = \gamma(d_{ij})$$

The semi-variogram function, $\gamma(h)$, as a function of *h*, must be known for all values of h. This condition requires the approximations of the semi-variogram with any of the models in the previous section.

Since Kriging is a linear procedure, difficulties are experienced if the variable to be estimated contains a non-linear trend, or drift as it is called in geostatistical literature. To solve this, *Universal Kriging* was developed, but it is numerically unstable and often singular.

APPENDIX D – MEAN AND STANDARD DEVIATION

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1A}$$

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} \tag{2A}$$

where,

N	=	Total number of boreholes
x	=	Parameter value of interest
i	=	Index of parameter list
$\overline{x}$	=	Mean value of *x*
s	=	Standard deviation

APPENDIX E – PEARSON CORRELATION

The Pearson's correlation coefficient determines the degree to which two variables are linearly related. The results range from -1 (negative linear relationship) to 1 (positive linear relationship), and 0 indicates no relationship present between the predicted data and the observed data. The Pearson correlation coefficient is calculated as follows:

$$PCC = \frac{\sum(Oi - \bar{O})(Pi - \bar{P})}{\sqrt{\sum(Oi - \bar{O})^2 \sum(Pi - \bar{P})^2}} \qquad (1B)$$

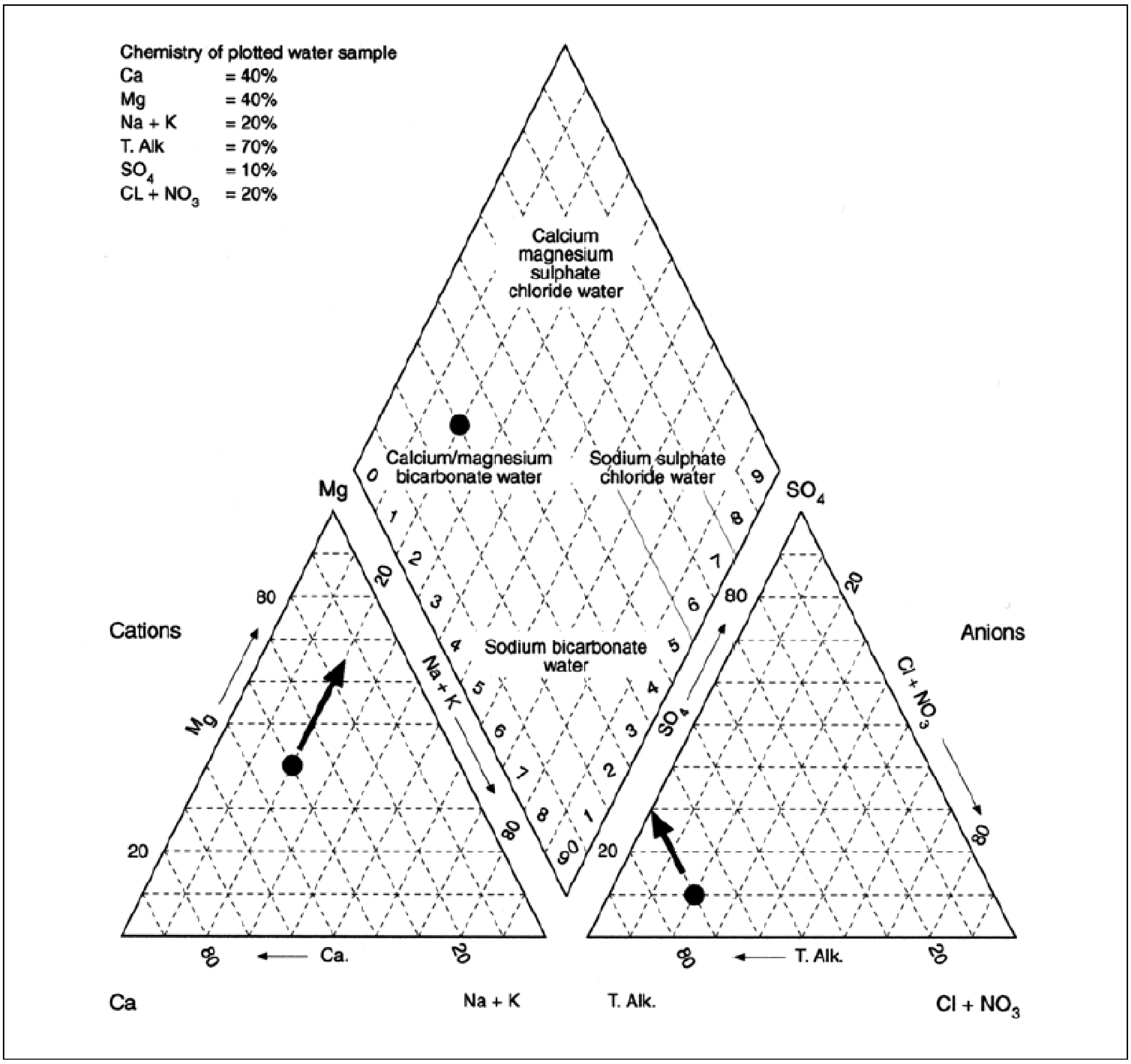
where,

PCC	=	*Pearson Correlation Coefficient*
i	=	*Total number of observation points*
P_i	=	*predicted value*
O_i	=	*the observed values for the n observations*
$\bar{O}$	=	*the mean of the observed values*

APPENDIX F – PIPER DIAGRAM PLOTTING PROCEDURE

Piper Diagram Plotting Procedure

Plotting procedures for the Piper diagram: Convert mg/l to meq/l by division ($Ca/20$, $Mg/12$, $Na/23$, $K/39$, T.Alk./50, $SO_4/48$, $Cl/35.5$, $NO_3/62$. Add Na +K and Cl + NO_3. Calculate percentage of cations and anions. Plot cations by scaling off Ca, then Mg. Plot anions by scaling off T.Alk. then SO_4. Project cation and anion points to triangle



APPENDIX G – ROOT MEAN SQUARE ERROR

RMSE provides a measure of the error size, but is sensitive to outliers as it places a lot of weight on large errors and is calculated as follows:

$$RSME = \left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]^{1/2} \quad (1G)$$

where,

n	=	Number of observations
i	=	Index
p	=	Predicted value
o	=	Observed value