An assessment of the physical drivers for farm dam distribution in Midlands KwaZulu Natal, using GIS and Remote Sensing

Master of Science Research Report

Report prepared by:

Jonathan Tsoka (1778306)

Supervisors:

Professor Jasper Knight

Dr. Elhadi Adam

This research proposal is done in partial fulfillment of the requirements of the Master of Science degree in GIS and Remote Sensing at Witwatersrand University, Johannesburg, South Africa

March 2018

DECLARATION

I, Jonathan Tsoka, declare that this research report is my own unaided work. It is being submitted to the Degree of Master of Science in Geographical Information Systems and Remote Sensing to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature of candidate

23rd day of March 2017 in Johannesburg

ABSTRACT

The interest in farm dams emanates mainly from their use for livestock watering, irrigation and fisheries enhancement on a sustainable basis. While management information on large dams in South Africa is largely available, it is lacking for farm dams which cumulatively store large volumes of water. As a result they are barely considered as part of the water resources of a river basin. Data acquisition methods for obtaining information about farm dams are costly, time consuming and labour intensive. This study was an attempt to map farm dams and establish the factors driving their spatial distribution pattern in the Midlands, KwaZulu-Natal, South Africa, using cost effective, time saving and less laborious GIS and remote sensing techniques.

A classified April 2017 Landsat 8 satellite image was used to identify all water bodies in the Umgeni River basin U2 quaternary catchment (U2) while Google Earth was subsequently employed for differentiating farm dams from other water bodies. There were approximately 2000 water bodies that were identified by the classification. These included large national dams, pools in golf courses, ponds and disused mine dumps. A total of 864 farm dams in the U2 region quaternary catchment was observed.

Six physical factors, namely slope, aspect, elevation, land use, soil type and geology were assessed to establish to what extent they influenced the siting of farm dams. The results indicated the importance of soils and land use as farm dams were mainly found in clusters in areas where agricultural farm land is also found since water is required for crop and livestock production.

The influence of other factors such as slope, geology and elevation were observed in the spatial distribution maps. They all gave significant p-values in their univariate analysis. Of the six variables only aspect gave non-significant results while the rest were significant. A binary multivariate logistic regression was created for forecasting future farm dam sites and to establish which sites are poorly sited. The other four factors were fitted to the model except aspect and geology type which had no significant p-values in the model. The model had an Akaike information criterion (AIC) score of 293.42 and had the best combination of variables relative to other models. It was validated using 500 farm dam sites and it predicted 86% correctly.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1. INTRODUCTION

1.1 Background

South Africa is a water scarce country affected by the vagaries of climate (De Winnaar et al., 2007). It is located in the semi-arid tropics, and according to Hughes and Mantel (2010), has a mean annual precipitation of 450 mm which is significantly lower than the world average of approximately 860 mm. In terms of distribution, western regions receive relatively less rainfall than the east that receives approximately 1000 mm/year. The potential evaporation rates range from 1100 to 3000 mm/year and the resultant average yearly rainfall to annual runoff ratio is 8.6%, which is generally low (Hughes and Mantel, 2010).

According to Biswas et al. (2009), the role of water cannot be understated as it is critical in the production of raw materials, ensuring food security, sustenance of bionetworks and socio-development of a country. There are global water scarcity issues however, emanating from the increasing demand and pressure on the already limited water resource, resulting from climate change, population growth and subsequent increase in the need for raw materials, and food production (Bressers et al., 2010), and South Africa is not spared. It is imperative that action is taken to address these challenges in order to address an impending water crisis (Hughes and Mantel, 2010) in the world at large and in South Africa in particular.

Globally, agriculture accounts for 70% withdrawal of all fresh water, while with irrigation consumes 66% of the withdrawals (Tingey, 2014). In South Africa the agriculture sector consumes 60% of all available water resources (Baleta and Pegram, 2014), and according to the DWAF (2004) irrigation accounts for 50% of the total water use. The variability of stream flows within and between rainy seasons has necessitated the need for storage (DWAF, 2004). In order to guarantee reliable supplies of water all year round for irrigation and livestock watering, and countering midseason droughts and dry spells, farming communities heavily depend on farm dams (Boardman, 2009; Stephens, 2010). Hughes and Mantel (2010) report that the number of farm dams has significantly increased in the past years in South Africa. Maaren and Moolman (1986) reported a nine-fold increase in farm dam numbers between 1944 and 1984 in certain selected catchments in South Africa.

In light of the accelerated developments of farm dams over time, it is important to understand their purpose and distribution (Shao et al., 2012). Knowledge of their spatial distribution and capacities at all basin levels is of fundamental importance for water resources management as they provide a surface for evaporation loses and affect downstream flows (Neal et al., 2002; Schreider et al., 2002; Callow and Smettem, 2009). Although they are important in catchment water resource systems, farm dams are seldom included in catchment management plans due to lack of information on their size, locations and purpose (Sawunyama et al., 2006).

Successful documentation of spatial distributions and estimation of capacities of farm dams has been done by Liebe et al. (2005), and Annor et al. (2009) in Upper East Ghana, Sawunyama et al. (2006) in Zimbabwe, Rodrigues et al. (2011) in Preto, Brazil, and Roohi and Webb (2012) in Australia. Their approaches were generally similar as they employed medium resolution Landsat satellite imagery which is relatively cheap, time-efficient, and reliable. Rodrigues et al. (2011) adopted the methodology used by Liebe et al. (2005). They mapped farm dams and developed power relationships between surface areas and capacity. On the other hand, Meigh (1995) employed an indirect method, of estimating farm dam capacity in Botswana by using 1: 50,000 topographical maps and satellite imagery.

This research seeks to use GIS and Remote Sensing techniques in assessing the spatial distribution pattern of farm dams and development of a predictive model for suitable locations of farm dams in the Midlands, KwaZulu-Natal (KZN). In KZN there are a few large dams that have been built for mainly urban water supply (DWA, 2010). A large number of farm dams have been constructed to retain water for agriculture, however, there are no known numbers of farm dams in the area as there are many unregistered ones (Letty et al., 2012). This presents a challenge of knowing the exact quantities of the farm dams and this lack of information has negative impacts on water resources planning in the catchment. Given the important role played by farm dams in agricultural production in KZN and in South Africa at large, there is need to understand their distribution for effective water resources planning and management.

1.2 Study area

Figure 1-1 shows the location of the study area, which is in the U2 sub-catchment. The U20 tertiary catchment is found in the Umgeni River basin in the South East of South Africa. The study area is the U2 sub-catchment in the Mgeni river system which has a total area of 4442.20 km^2.

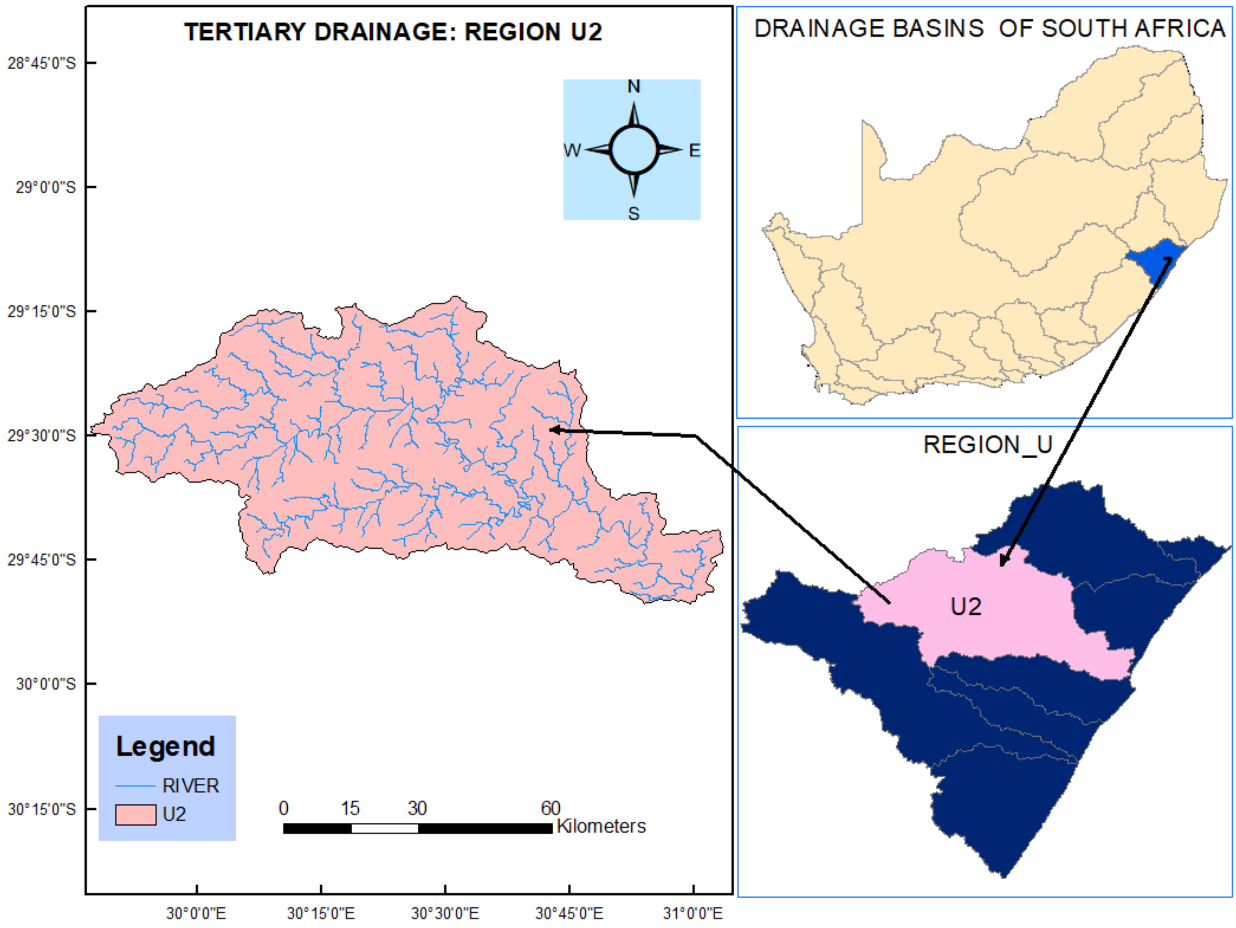

Figure 1-1. Map of the study area U2 sub-catchment (NGI, 2013)

1.2.1 Geology

Figure 1-2 illustrates the geology of the study area. It is comprised of different rock sub-types that formed over different time periods. The Dwyka formation of the Karoo supergroup, exists in central U2 sub-catchment, where a continuous belt spans from north to south. It is also found in the eastern parts of U2 sub-catchment and generally covers 302.5 km^2.

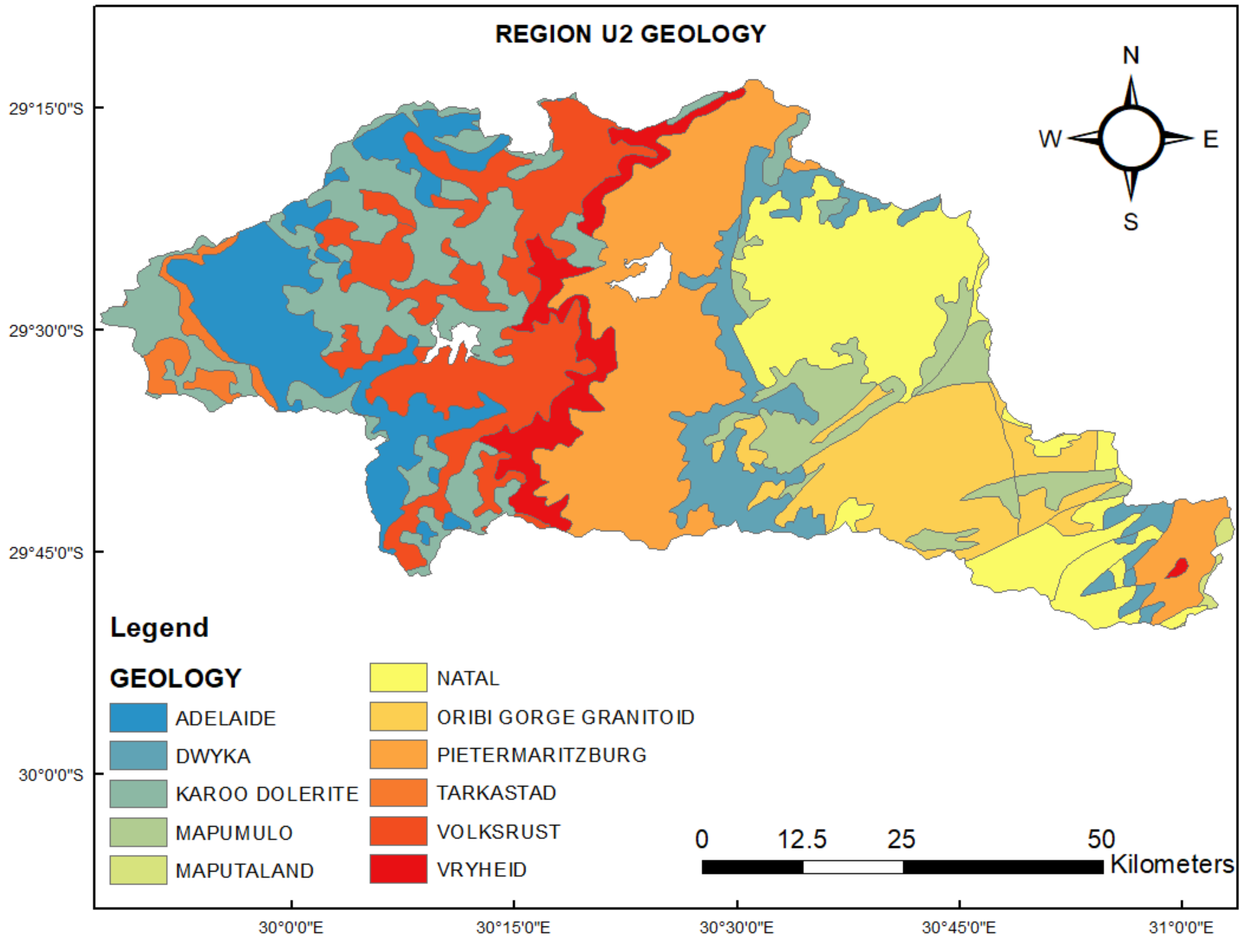

Figure 1-2. Geology map of the study area U2 sub-catchment (NGI, 2013)

Karoo Dolerite exists mainly in the western areas of the study area and covers an area of 606.6 km^2. It is a network of dolerite sills, sheets and dykes, mainly intrusive into the Karoo supergroup. The Adelaide and Tarkastad subgroups are present predominantly in the west, covering 446.5 km^2 and 69.4 km^2 respectively. The Adelaide subgroup is a mud-rock, and is a subordinate of sandstone, while the Tarkastad subgroup is a red and greenish-grey mudstone, which is fine- to medium-grained sandstone. The Pietermaritzburg subgroup is shale with thin siltstones and sandstones in the uppermost part. It is mainly found in the central area of the study area and a small part exists in the far east. It covers an area of 801.4 km^2 in the study area. Other subgroups that have the group Ecca as their parent rock type are Vryheid and Volksrust. The Vryheid sub group is fine to coarse-grained sandstone, shale, coal seams, covering 199.3 km^2 and exists in the central areas of the study area, close to the Pietermaritzburg formation. On the other hand, the Volksrust subgroup is basically mud-rock and is found in the central areas and spreads westwards of the study area. It exists close to the Pietermaritzburg and Vryheid subgroups covering an area of 583.5 km^2. The Natal Group exists in the eastern part of the U2 sub-catchment, covering an area of 674.7 km^2. It is generally reddish, feldspathic and micaceous sandstone with subordinate quartz arenite, mud-rock, granulestone and conglomerate. Also spreading from the central areas of U2 sub-catchment towards the east is the Mapumulo rock group, which is made up of heterogenous layered paragneisses and migmatites with a wide compositional range. This rock group covers an area of 262.3 km^2. The Maputaland group exists in the east, covering the least area (10.4 km^2). It is a calcarenite, clayey sand, red and grey dune sand, limestone, conglomerate. The Oribi gorge granitoid group is found in areas in central U2 sub-catchment and spreads towards the east covering an area of 447.5 km^2.

1.2.2 Soils

According to Bainbridge (2004), the soils exhibit similar characteristics to those of the Maloti Drakensburg Mountains, as they are old and give an impression of the age and high rainfall in the area. There are generally shallow and highly weathered soils on the slopes of the Drinkkop Mountain and the dolerite ridges. The soils in the area generally fall into three categories; Cambisols, Ferralsols, and Luvisols. According to Fey (2010), Cambisols have subsurface horizons that are sandy loam or finer, with at least 8 percent clay by mass and a thickness of 15 cm (6 inches) or more. Their infiltration rates range from 20 – 30mm/hr. Bainbridge et al. (1986) defined Ferralsols as red and yellow weathered soils whose colours result from an accumulation of metal oxides, particularly iron and aluminum (from which the name of the soil group is derived). They are technically defined by a fine-textured subsurface layer of low silt-to-clay ratio, high contents of kaolinitic clay and iron and aluminum oxides, and low amounts of available calcium or magnesium ions. Generally, their infiltration rate is between 1–5mm/hr. According to MacVicar (1977), Luvisols are characterized by mixed mineralogy, high nutrient content, and good drainage which makes them suitable for a wide range of agriculture, from grains to orchards to vineyards. Luvisols form on flat or gently sloping landscapes under climatic regimes that range from cool to warm temperate regions. They have a loam texture with surface accumulation of humus overlying an extensively leached layer that is nearly devoid of clay and iron-bearing minerals. Generally, their infiltration rates range between 10 – 20 mm/hr.

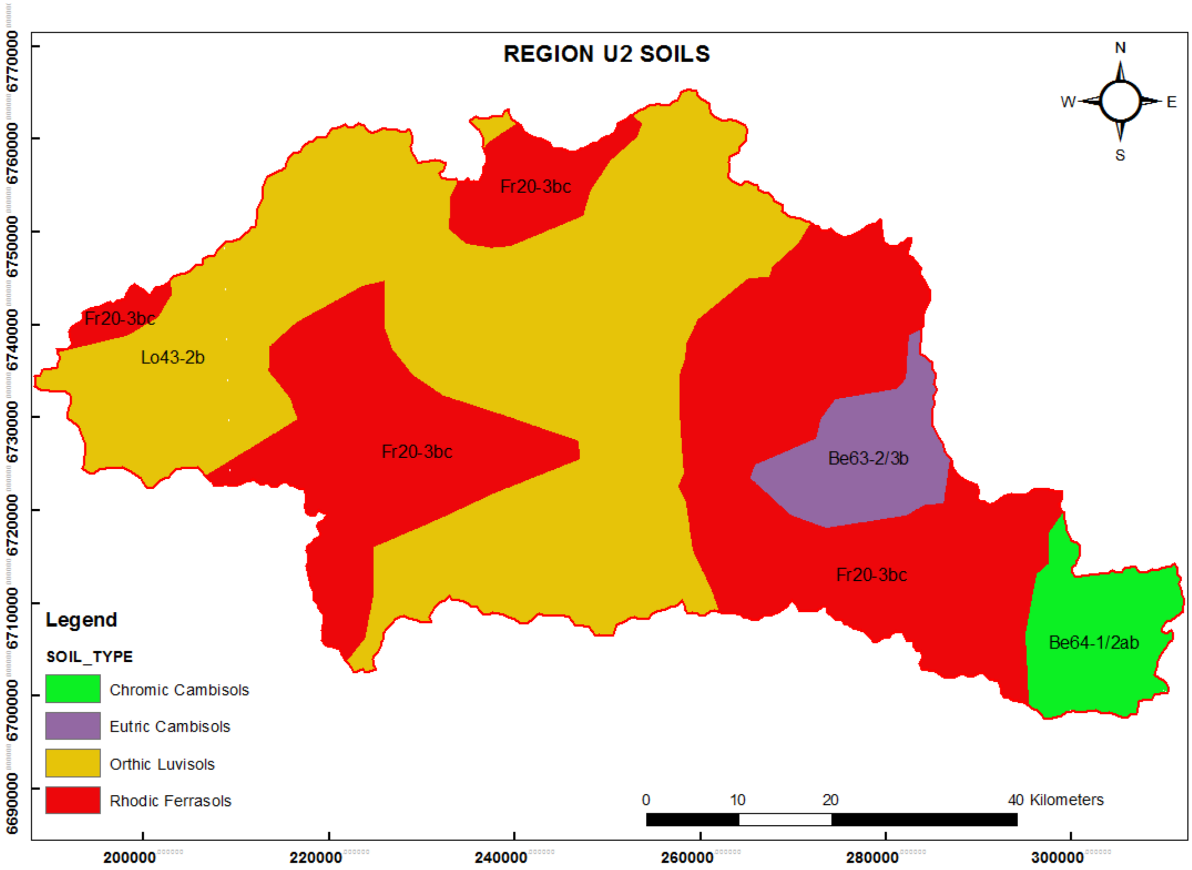

Figure 1-3. Soil map of the study area U2 sub-catchment (ESDC, 2017) https://esdac.jrc.ec.europa.eu/content/european-soil-portal

1.2.3 Topography

The topography of the U2 sub-catchment is shown in Figure 1-4. The highest elevation is 2071 m asl while the lowest is -13 m asl. The areas in the west of the U2 sub-catchment are the highest as they are part of the Drakensburg mountain range. To the far east of the U2 sub-catchment are the lowest areas where the catchment drains. In the study area, rivers will naturally flow from the west to the east, draining into the Indian Ocean. The terrain in the area is generally rugged and this is evident from the illustration and also the high elevation range ($Elevation_{max}$ - $Elevation_{min}$). The average elevation is 923 m asl.

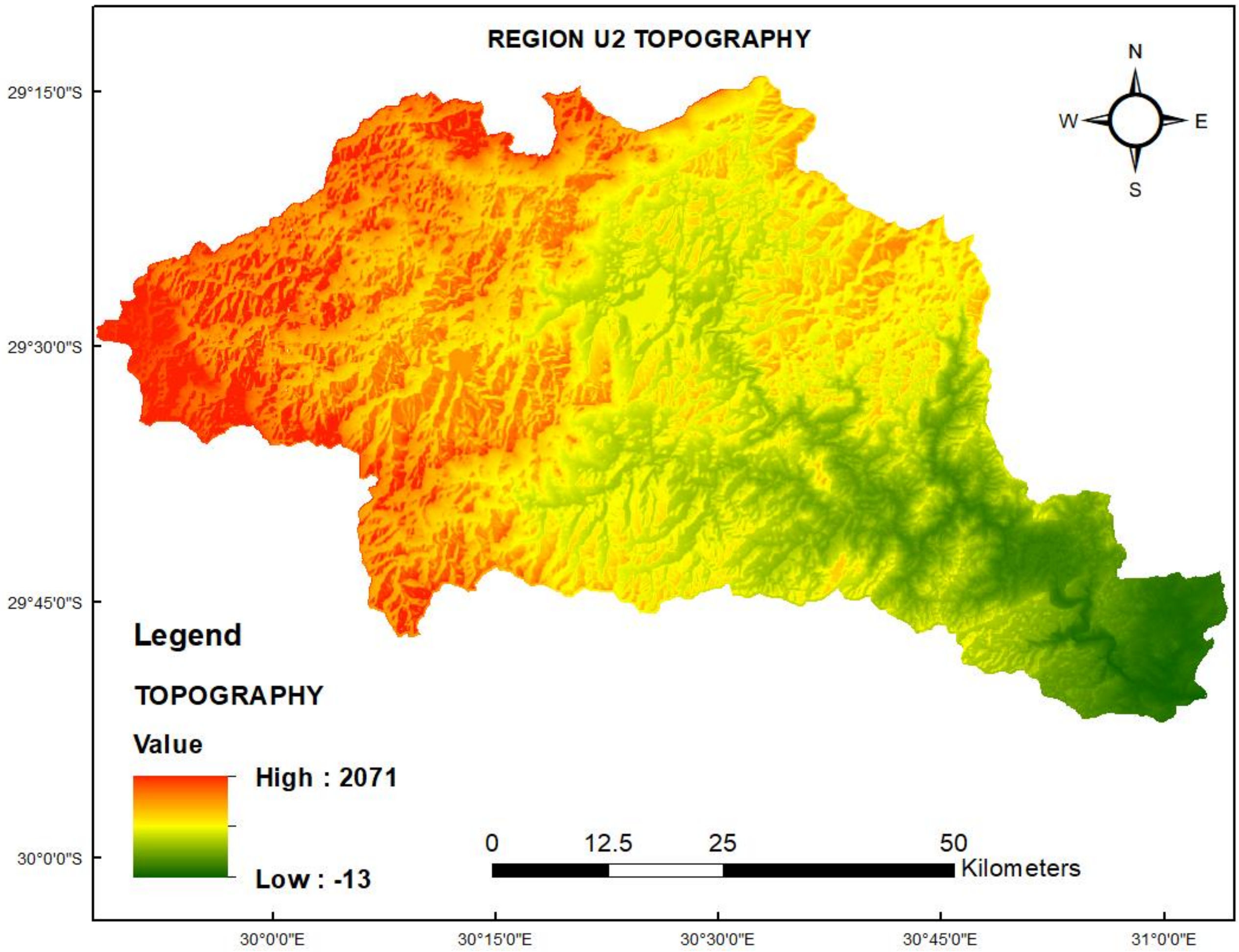

Figure 1-4. Region U2 topography (USGS, 2017)

1.2.4 River patterns

Figures 1-5, 1-6 and 1-7 present the river patterns and general flow pattern of rivers and streams in the U2 region. The region's topography has an influence on the river patterns of the U2 sub-catchment. The general direction of flow is from the west to the east. Figure 1-5 shows the different stream orders while Figure 1-6 focuses on the main rivers supplying the major dams in the Region. Figure 1-7 depicts the catchment areas of the water bodies in the region. There are five major dams in the U2 region, namely Midmar, Albert Falls, Nagle, Henley and Inanda. They supply water to Pietermaritzburg, the towns in the Midlands area and Durban.

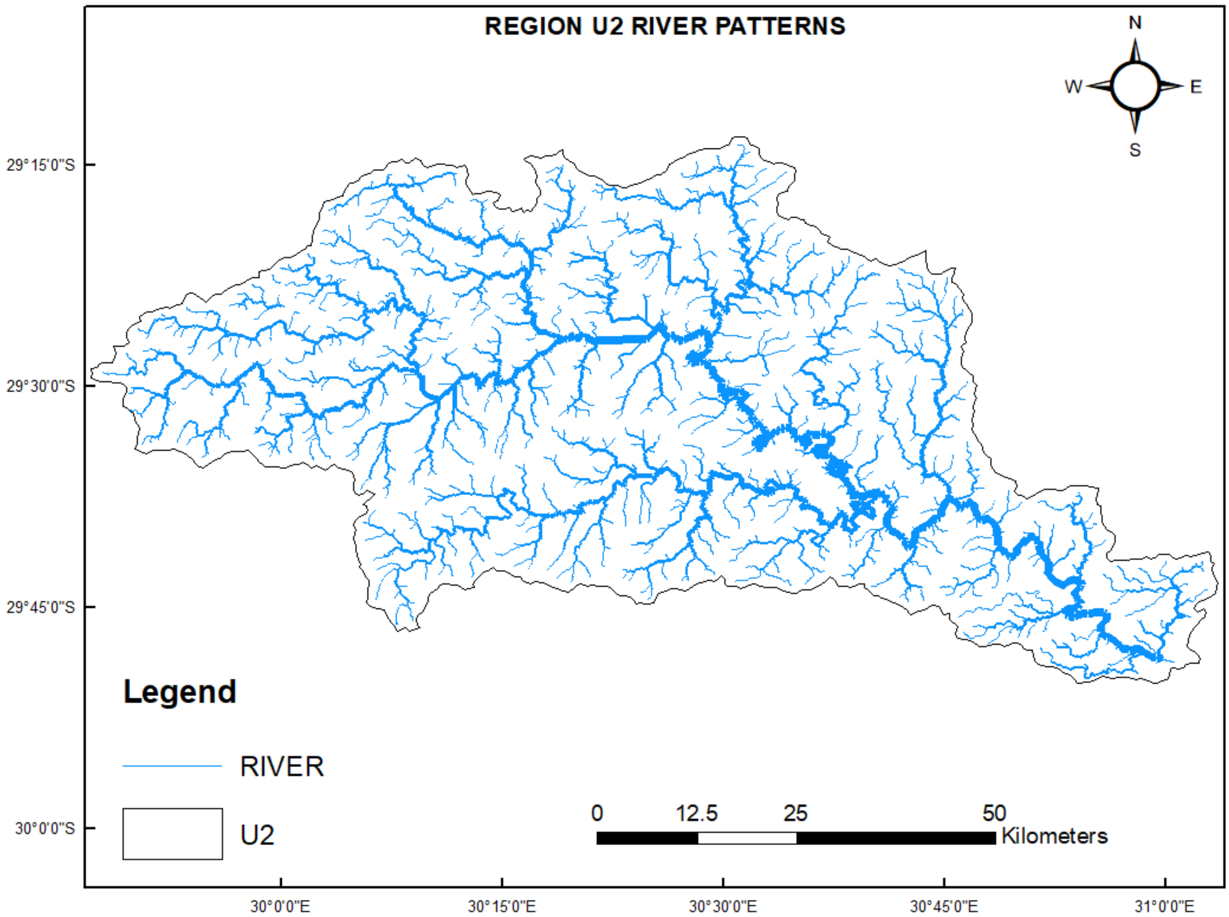

Figure 1-5. Region U2 river patterns (NGI, 2013)

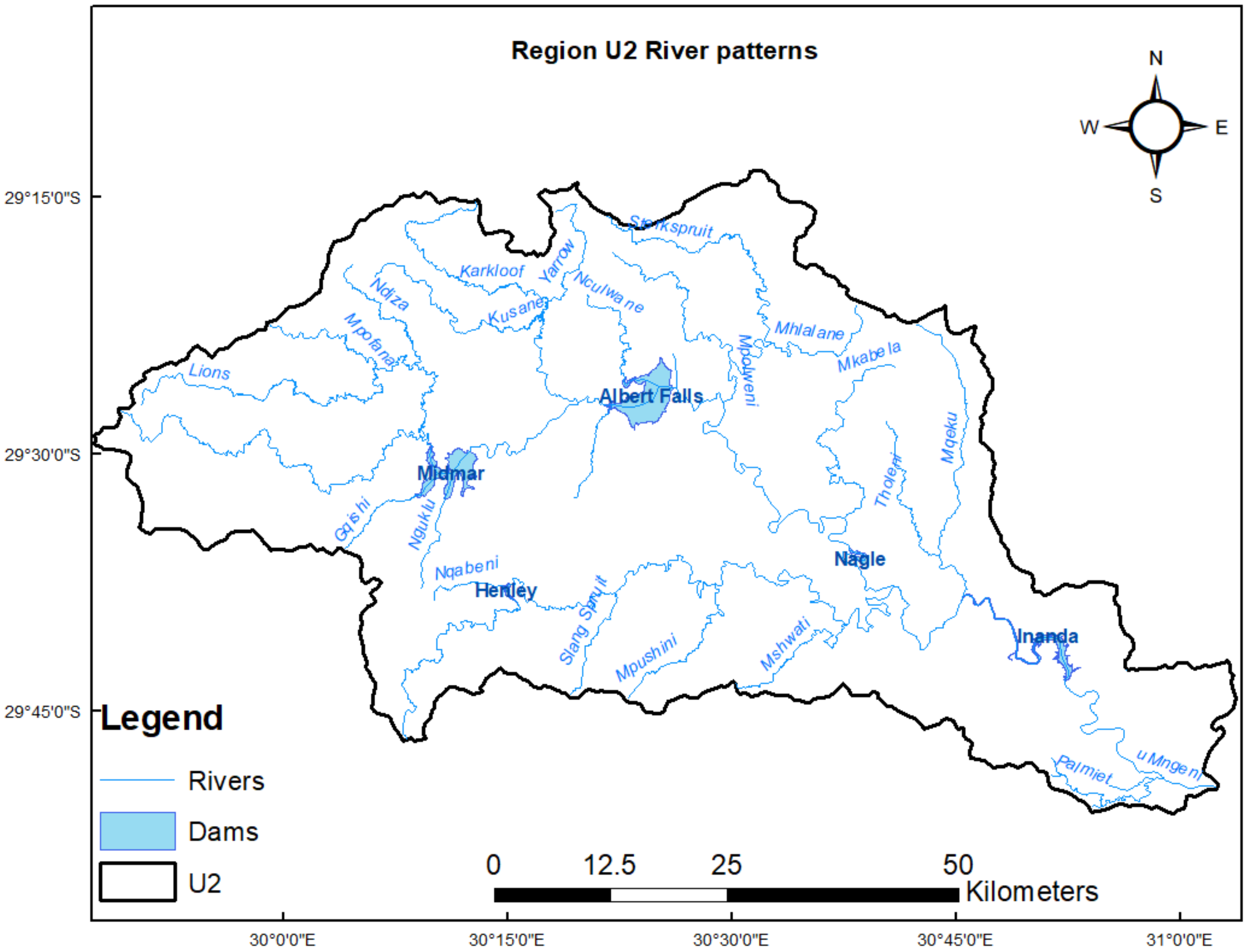

Figure 1-6. Rivers and major dams in Region U2 (NGI, 2013)

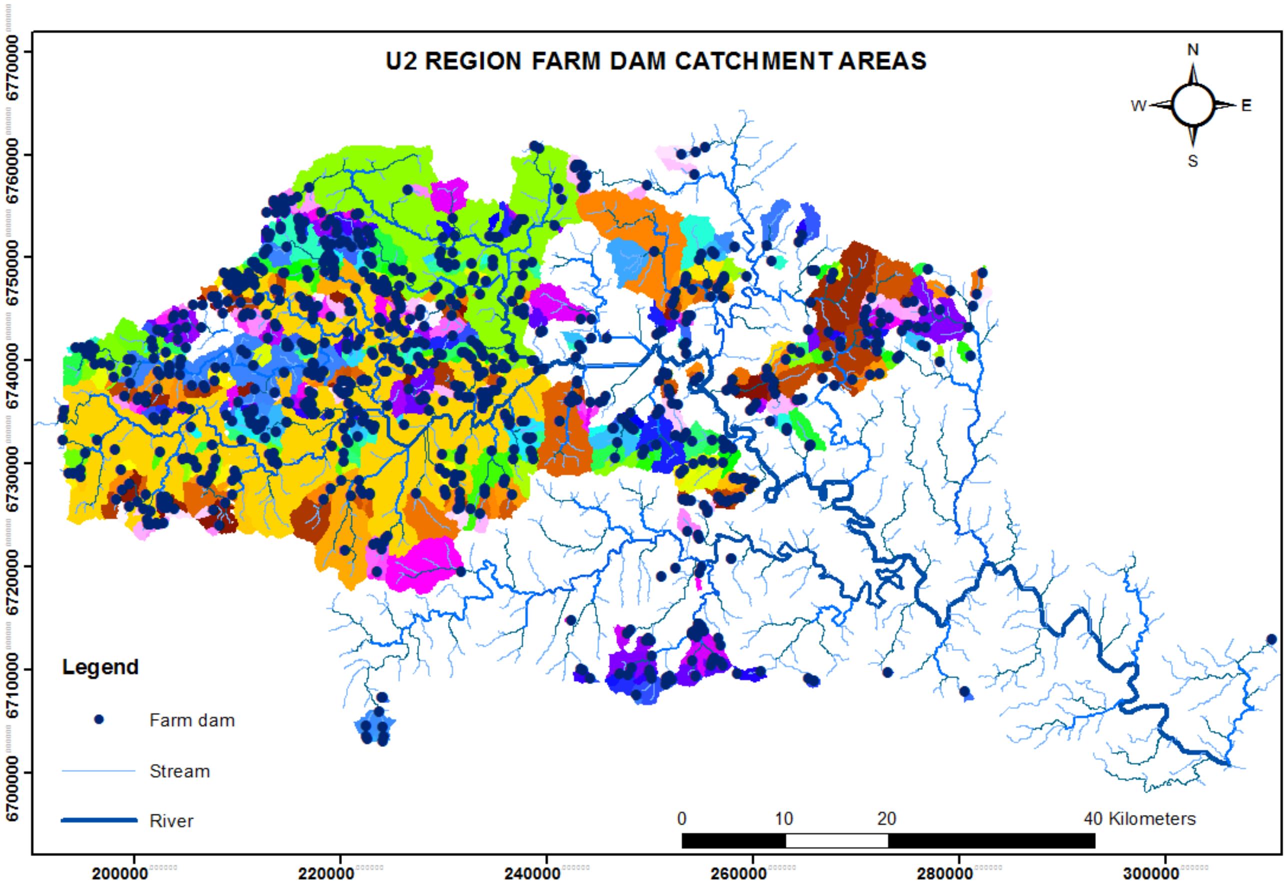

Figure 1-7. Farm dam catchment areas (NGI, 2013)

1.2.5 Climate

Figure 1-8 shows the mean monthly rainfall and temperature for Howick for 1979 – 2014. It is in a climatic zone called the Köppen classification Cfb which has a temperate summer rain climate. Rainfall is mostly received in the summer months, thus October to March; however, there are occasional winter showers. It receives an annual average rainfall of approximately 1100 mm per year. The peak rainfall months are December to February (DWAF, 2004).

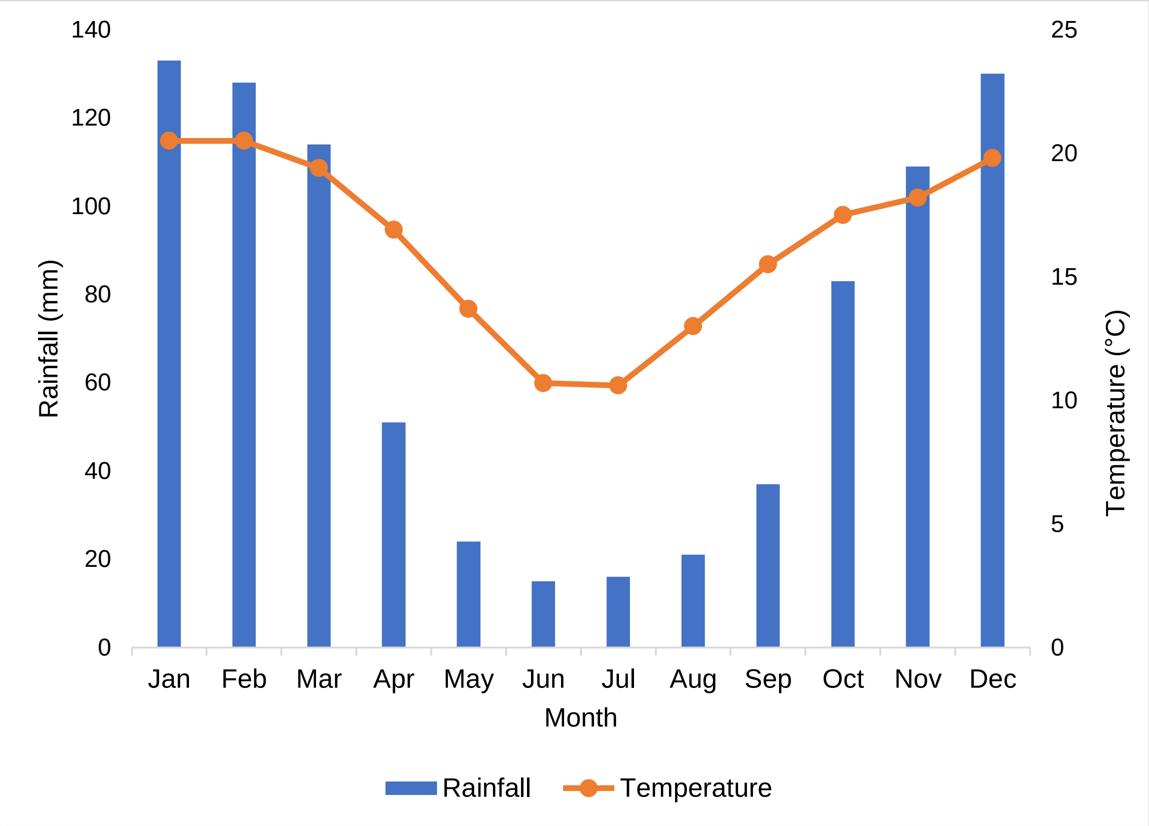

Figure 1-8. Mean monthly rainfall and temperature for Howick from 1979 – 2014 (Global Weather, 2017)

The spatial distributions patterns of evaporation and rainfall are similar where relative high humidity is experienced in summer. The daily mean peak humidity gets up to 68% in the month of February and a mean low of approximately 60% in July. The evaporation is 1214 mm and the natural runoff is 152 mm (DWAF, 2004).

1.2.6 Farming activities

According to Troskie (2013), agriculture is a major economic and social force in the KZN province. The province has good rainfall, fertile soils and is classified into three regions on the basis of the varied topographic and climatic conditions. The three different geographic areas are the lowland region found along the Indian Ocean, plains (grasslands) in the central region, and two mountainous areas which are the Drakensberg and the Lebombo mountains. This varied topography and climate translates to a range of agricultural activities. Of the 6.5million hectares of farming land, 18% is good for crops and 82% suitable for livestock activities (Meissner et al., 2013).

The province is home to a significant percentage of South Africa's small-scale farmers. Sugarcane is the trademark agricultural commodity in this region, with soybeans (9%), maize (7%), dry beans (7%) and wheat being another field crop (Hendriks, 2014). There is a variety of subtropical fruits and horticultural crops grown in the province. There are large forestry plantations in areas around Vryheid, Eshowe, Richmond and Harding. Larger than the province's maize sector and with an annual trade worth around R62 million, indigenous medicinal plants is a commodity flying below the radar. The province does well in livestock production, having 19% of the country's cattle, 13% of its goats and 10% of its pigs (Meissner et al., 2013). It has 10.4% of the country's layers (eggs) and 12.1% of its broilers (SAPA, 2015). There are a number of dairy farms in the Midlands and 15% of the country's milk producers live in this province (Meissner et al., 2013).

1.3 Problem statement

According to Sawunyama (2006), water resources planning and management present serious challenges in South Africa for various domestic and agricultural uses. The water resource is increasingly scarce (Oberholster and Ashton, 2008) as the population and subsequent demands continue to increase in South Africa. There has also been expansion of industry and agricultural crop production. Agrawala (1998) predicted that future growth rates of population and expected socio-economic development would not be sustained by the water resources available in South Africa if the trends of water use did not change. In 2005, almost all of the country's freshwater resources had been fully allocated (Oberholster and Ashton, 2008).

South Africa lies in the semi-arid region where rainfall is highly variable and droughts are often experienced (Rodrigues et al., 2011). Farm dams play an important role in ensuring climate-change-proof crop and livestock production by harvesting surface runoff during the rainy season. The harvested water is used to counter midseason dry spells and droughts. Studies have shown a significant increase in farm dam numbers in many parts of South

Africa due to drought events (Nathan and Lowe, 2012). This is evidence of efforts to build resilience against variability of rainfall in agricultural crop production.

Farm dam presence, distribution and capacity must be known in order to have efficient water management planning (Sawunyama, 2006). While management information on large dams in South Africa is largely available, it is lacking for farm dams (Liebe et al., 2009), which cumulatively store large volumes of water. Despite being this important, farm dams are barely regarded as important contributors of the water resources of a river basin.

Efficient management and sound planning of water resources are hindered by inadequate farm dam data, their storage capacities and spatial distribution (Sawunyama, 2006; Annor et al., 2008). Collection of planning data for dams is generally expensive and time consuming. However, GIS and remote sensing techniques offer means for acquiring the data in a shorter time and at a relatively cheaper cost. This study seeks to use GIS and remote sensing techniques to establish the spatial distribution of farm dams and to also develop a predictive model for extrapolation of sites where these dams can be sited in the future and in other catchments.

1.4 Aims and objectives

1.4.1 Aim

The aim of this research is to assess the factors that influence the spatial distribution of farm dams in the Midlands, KwaZulu-Natal, South Africa.

1.4.2 Objectives

The objectives of the research are:

- To map the distribution of farm dams and determine their spatial distribution pattern using R- Studio;
- To determine by Principal Component Analysis the factors influencing the spatial distribution of farm dams using SAS software suite; and
- To develop a spatial predictive model for farm dam sites using R-Studio.

1.5 Research report structure

The first chapter gives a background of the study and also presents its aims. The second chapter presents the theoretical background and literature review of GIS and remote sensing methods that have been used in the past for similar studies, and the different types of farm dams. The third chapter gives a description of the data collection process and analysis.

Chapter four presents the results and discussion. In the fifth chapter, a discussion of the research findings is done. Chapter six gives the conclusions and recommendations.

2. LITERATURE REVIEW

2.1 Overview of farm dams

According to the Department of Sustainability and Environment (2007), a farm dam is a water harvesting structure, usually excavated, having a bank and a barrier to check stream flow on a water way, and other works located on a farm. Nathan and Lowe (2012) further define farm dams as private structures that are used to intercept catchment runoff that would otherwise have contributed to streamflow (or recharge to aquifers) with the basic purpose of capturing runoff when it is available and storing it for later use. Based on the classification by the Department of Sustainability and Environment (2007), there are two common types of farm dams and they are differentiated by where they are constructed and how they harvest water. One type is constructed in gullies or waterways, making use of natural formations to direct water to the dam, and the other is built on sloping ground to capture surface run-off (often directed to the dam from contour drains). The larger ones are relatively fewer in number and are used for commercial irrigation and the small ones which are more, have a storage area of between 5 and 10 hectares (Roohi and Webb, 2012).

In the recent past, there has been a lot of research on farm dams in Australia and other regions but not in South Africa. Matteau et al. (2009) indicate that the impacts of small farm dams on the hydrology and environment of a catchment have received relatively less attention when compared to assessments done on large farm dams. There have been a number of studies, however, on the impacts of farm dams on hydrological connectivity and run-off estimation in agricultural landscapes (Callow and Smetten, 2009); hydrological impacts (Nathan and Lowe, 2012); impacts of stream flows and sustainability in a context of climate change (Habets et al., 2013); and sediment delivery (Verstraeten and Prosser, 2008).

This study focuses on assessing spatial distribution of farm dams using GIS and Remote Sensing techniques. There are a number of contributions that have been made especially in Australia, India and Ghana on the use of aerial photography and remote sensing technology to estimate reservoir storage volumes (Nathan and Lowe, 2012; Rodriguez et al., 2012); counting farm dams using Landsat imagery (Dare et al., 2002; Roohi and Webb, 2012; Shao et al., 2012; Bhagat and Sonawane, 2011).

2.2 Mapping farm dams

2.2.1 Use of Landsat imagery and maps

There is general agreement among authors (e.g. Sawunyama et al., 2006; Nathan and Lowe, 2012; Roohi and Webb, 2012) that aerial photographs, satellite imagery and topographic maps can be used in the determination of the distribution and numbers of farm dams. The approach of using these resources is cost-effective, working best and efficiently over small areas, although it tends to get expensive and labour intensive over larger areas. In studies where large areas were focused on using Landsat data for dam inventories, image enhancement was employed to improve their accuracy. There has also been use of Landsat data which has been proven to have a number of advantages with regards to costs, availability and a long archiving period which dates back to 1972, captured repeatedly with a 6-band spectral resolution (Roohi and Webb, 2012). Shao et al. (2012) highlight that high-resolution imagery such as Spot 5 is expensive and labour intensive when used in trying to identify farm dams.

Roohi and Webb (2012) and Shao et al. (2012) acknowledge the importance of dam usage in determination of spatial distribution. In their study on in Western Victoria, Australia, they used a combination of Landsat imagery and Google Earth. Their approach was cost-saving, affordable, proficient and consistent and it can be replicated in various types of environments. They used imagery from 1973–2004 and they managed to identify 99% of the farm dams in the area. They discovered that in the period 1973–1993, farm dams developed increased by 279% as a result of increased livestock production. After 1993 however, the number of dams declined possibly because of unavailability of suitable sites, and a change in land use from dryland pastures to dryland cropping which does not require watering points.

2.2.2 Statistical analysis of farm dam distribution

Shao et al. (2012) established statistical relationships to estimate the number of farm dams based on climatic and geophysical factors in Australia by using regression analysis. They worked mainly with medium to small dams as their model had limitations in predicting farm dams. The results of their research indicated that more dams will be found in areas where there is high rainfall, and also found out that farm dams with a relatively smaller volume were found in flatter areas, but size increased with an increase in gradient up to 3°, thereafter decreasing. Although there is large uncertainty in farm dam development, and their variations in use and size (Dare et al., 2002; Shao et al., 2012), it is of importance to quantify farm dam spatial distribution for effective water resources management.

Savadamuthu (2002) in a study on the impact of farm dams on streamflow in the Upper Marne catchment in Australia also indicated an increase in number of farm dams in the area, and attributed this to the intense growth of the agricultural sector, particularly vineyards. The dams were developed mainly for irrigation purposes. Studies conducted in the same area by Good (1992) and Nathan (1999) were concerned with computations of the water balance and employed stochastically selected farm dams from the known spatial distribution. Whilst these approaches were useful in giving the estimates for overall impacts of dams for the entire catchment, their weakness was they were unable do the same for individual sub-catchments. Since rainfall, land use and farm dam concentration are not uniform, modeling the catchment using the actual spatial distribution of dams provided a more accurate impression of the impacts of farm dams on stream flows. Development of these farm dams was also done in high rainfall areas.

2.2.3 Using remote sensing for dam counting and delineation

Dare et al. (2001) detected, counted and classified dams in Upper Wimmer catchment in Australia. They applied density slicing and spatial filtering techniques to coloured aerial photographs at a resolution of 1: 40 000.

Through density slicing and spatial filtering of the colored aerial photographs at a resolution of 1:40 000, Dare et al. (2001) detected, counted and classified dams in Upper Wimmer catchment in Australia. This exercise was done in order to enable quantification of storage dams in the area and to establish the impacts of farm dams on stream flows in the catchment. The study was complemented by high resolution satellite imagery (5 m resolution) for an 8 km by 8 km area of the catchment which was also acquired, processed and used to extract farm dams in automated way. They concluded that automatic interpretation of the high resolution remotely sensed imagery had great potential in mapping, monitoring and measuring of dams in remote parts of Australia.

Bhagat and Sonawane (2011) focused on determining the number and extent of the small farm reservoirs by using satellite data. In their work, they used Landsat 7 for identifying water bodies in their study area in India. Surface Wetness Index (SWI), Normalised Difference Vegetation Index and a Slope map were used for identification and delineation of water bodies. Surface Wetness Indices are important indicators of surface wetness, thus they show the presence and distribution of surface water bodies (Huang et al., 2002). In their study, the SWI based classification produced maps showing water and non-water surfaces. An overall accuracy of 91.74% was achieved for water bodies and they accounted for 3.8% of the total study area. Identification and delineation of vegetated and non-vegetated land can be done using the Normalised Difference Vegetation Index (NDVI) based on the ETM +

dataset (Kiage et al., 2007). The NDVI based classification produced maps showing the distribution of ‘vegetation’ and ‘non-vegetation’. They used the slope map for masking shadows included in the estimated water bodies.

Both Dare et al. (2001) and Bhagat and Sonawane (2011) used simple and time saving methods to map farm dams and also estimate catchment storage. However, it should be pointed out that high resolution satellite imagery used by Dare et al. (2001) can be very expensive whereas the medium resolution Landsat 7 used by Bhagat and Sonawane (2011) is freely available. Although high resolution satellite imagery provides a more detailed platform for identifying farm dams, they require advanced classification algorithms and skills set in order to meaningfully classify them.

Rodrigues et al. (2012) estimated reservoir storage volumes based on Landsat 7 satellite image of 30 m resolution in Preto River Basin in Brazil. They prepared an inventory of dams from remotely sensed Landsat data that they classified. They achieved 96% overall accuracy in their Maximum Likelihood classification, just missing one feature which was a lagoon. They also administered a questionnaire survey and conducted a semi-structured interview at farm households before they collected survey data for farm dams to validate the classification. They asked questions in line with attribute information about the reservoir which included checking for availability of designs, maps, area and depths of the reservoirs, management information which included maintenance, age and main use. They managed to identify a total of 252 dams and selected those with areas of between 1 and 50 ha for modeling. From inventory, 147 dams had a size of between 1 ha and 50 ha. They managed to visit 47 sites and did 42 interviews.

Their method was efficient as they saved time and costs by employing Landsat imagery for identifying and counting the farm dams. Their classification accuracy was very good as they achieved an accuracy of 97%. Their work is very relevant in the study area as it now provides water resources planning information, a basis for decision making and a guide in identifying the reasons why farm dams were constructed.

2.3 Predictive modeling of farm dam site location

Singh et al. (2017) used GIS-based multi-criteria decision analysis in evaluating potential sites for rain water harvesting sites in catchments in India. They prioritized rainwater harvesting sites and recharge structures that were critical in ensuring that rainwater harvesting interventions were implementable. For mapping rainwater harvesting potential, they used slope, runoff co-efficient, and the drainage density of catchments. Subsequently, they identified feasible sites by employing suitability criteria and GIS-based Boolean logic. They subdivided their study area into 4 rainwater harvesting potential zones, based on the potential from high to moderate. The Multi-Criteria Decision Analysis (MCDA) and integrated geospatial approaches they employed in their research are time saving, cost-effective, reliable and enhance efficient planning and management of rainwater at a larger scale.

Another GIS-based decision support system (RHADESS) for selection of and determination of suitability of Rain Water Harvesting sites was used in the Upper Orange and Thukela Water Management Regions in South Africa by Kahinda et al. (2009). Their work was similar to that carried out by Singh et al. (2017) in that they also carried out suitability analysis to identify the best and most suitable sites for rainwater harvesting sites. The RHADESS model they used was built in ARCVIEW3.3's model builder and assesses Rain Water Harvesting potential in the 19 Water Management Areas (WMA) in South Africa. The RHADESS user interface enables users to upload their data and is agile, giving options to alter parameters within the system. The RHADESS tool is a useful tool and may be useful in determining farm dam sites. In the event that it has short comings, a similar tool can be developed in ARCVIEW for assessment of suitability of dam sites for the study area initially, and can be replicated to other sites across South Africa.

The spatial distribution of *Muscari latifolium*, a prevalent bulbous plant species in Turkey, was modelled by Yilmaz et al. (2016). The modelling was done to determine the important factors influencing the spatial distribution of the plant species and to create a predictive surface for predicting sites with presence of *M. latifolium.* The Boosted Regression Trees (BRT) method was used for building models based on presence–absence data collected from the Gonen Dam basin in West Anatolia. Location data for the model had 105 points, thus 35 presence and 70 absence. The main variables in the model were monthly and seasonal climate, fine scale land surface, geological and biotic. Univariate analysis was performed on the variables to assess their importance in the multivariate model. Results of the analysis showed that there were 14 factors that influenced the spatial distribution of *M. latifolium.*

Although some of the predictive spatial modeling methods given in this section are not all for farm dams, the approaches used by Kahinda et al. (2009), Yilmaz et al. (2016) and Singh et al. (2017) can be adopted for farm dam analysis. The parameters they looked at in Kahinda et al. (2009) and Singh et al. (2017) are similar to the ones assessed in this this study while the methodology used by Yilmaz is closely related the one used for this study.

3. METHODOLOGY

3.1 Methods

This section presents the details of the procedures and tools that were employed to accomplish the objectives, together with a discussion of the data requirements, and the techniques to be implemented. A desk study which entailed reviewing of literature on spatial distribution of farm dams, and the use of remote sensing technology in delineation of farm dams was done. The review provided information on the nature of data required. The methodology was split into three phases. The first phase focused on primary and secondary data acquisition and pre-processing. The second phase looked at the data processing, which involved classification of satellite imagery, and DEMs while the third looked at mapping, modeling and interpretation of the data. Table 3-1 presents the general datasets that were used for the study. The specific datasets are given in detail for each activity.

Table 3-1: Summary of the general dataset requirements

Data type	Data set	Source of Data	URL
Farm Dams	•Dam location •Capacity	•Landsat 8 •Google Earth	URL: https://lta.cr.usgs.gov
Hydrological data	•River systems •Drainage systems	•Department of Water Affairs • NGI	
DEMs	•SRTM	•USGS	URL: https://lta.cr.usgs.gov
Land Use	•Cropped land •Grasslands •Bare soil •Settlements •Plantations •Water Bodies	•USGS •Google Earth	URL: https://lta.cr.usgs.gov
Soil data	•Soil types	•European Soil Data Centre	
Geology data	•Rock types	•City of Johannesburg	

3.2 Landsat data

Cloud free Landsat 8 satellite imagery for the study area Path 168 and Rows 80 and 81, with 1 arc second resolution (30 m), was acquired from the USGS website. The image was acquired on 14 April 2017 at 0840 hrs. A summary of the details of the acquired images is given in Table 3-2.

3.2.1 Landsat 8 satellite imagery

Table 3-2: Landsat 8 data specifications (USGS, 2017)

Year	Date of Acquisition	Sensor Identifier	Satellite	Output file	Path	Row	Resolution (m)
2017	2017–04-14	OLI_TIRS	Landsat 8	GeoTIFF	168	81	30
2017	2017–04-14	OLI_TIRS	Landsat 8	GeoTIFF	168	80	30
Multispectral_LC08_L1TP_168081_20170331_20170414_01_T1_MTL							
Multispectral_LC08_L1TP_168080_20170331_20170414_01_T1_MTL							

3.2.2 Atmospheric Correction

According to Matthew et al. (2002) atmospheric correction is a technique for conversion of data from spectral radiance to spectral reflectance performed to reduce the effects of molecular scatter and absorption of radiation caused by dust and water molecules. Atmospheric correction is a procedure of fundamental importance for classification of Landsat imagery data as it eliminates scatter and absorption effects from the atmosphere, and providing surface reflectance characteristics. Liang et al. (2001) state that light intensity is reduced by absorption with a haziness effect while Electromagnetic Energy (EME) is redirected in the atmosphere by scattering, causing an adjacency effect where neighboring cells are shared. These two phenomena compromise the quality of the resultant image.

The Landsat 8 satellite image was corrected for atmospheric and radiometric errors caused by atmospheric scatter and absorption of Electromagnetic energy. Radiometric calibration coefficients were used to convert images to Top-of-atmosphere radiance. The radiometric correction of satellite imagery can be done using a number of techniques some of which include FLAASH, dark subtraction, QUAC, and emissivity normalization. The FLAASH module in ENVI SAT 5.3 was used for radiance calibration in order to rectify surface reflectance and also eradicate unwanted materials and noise (Lu et al., 2004). According to Kaufman et al. (1997), consider the MODTRAN4 radiation transfer code in FLAASH suitable for carrying out atmospheric corrections.

The Rural for aerosol module was selected as most parts of the study area lie in rural areas or farming communities, initial visibility was set at 20-40km and the 2-Band (K-T) for aerosol retrieval was selected.

3.2.3 Image classification and validation

Image classification is defined by Mesev (2010) as a technique that entails identification and categorization of pixels and comparing their digital numbers and spatial orientation to produce meaningful geographical features. According to Mas (1999), the pixels are grouped into classes representing information of interest to land managers or researchers. In this study, the land use land cover classes (LULC) were produced from the April 2017 Landsat 8 image of the U2 sub-catchment.

Supervised classification was used for image classification. The spectral signatures of the pixels found in each training area are determined by the classifier, which in turn uses this information to compute the measures of central tendency (means, medians, and variance) of the classes in relation to the input bands. Pixels are assigned to classes based on the spectral signatures they closely match. It is important to ensure selection of training sights in the entire extent of the image in order to get an accurate classification of each land cover. There are a number of classifiers used for performing supervised classification and they include Support Vector Machines, the Gaussian Maximum Likelihood, Parallelepiped Classifier and Random Forest.

The Random Forest (RF) classifier was selected for use in this study. It was selected ahead of other supervised classifying techniques because of its relatively higher classification accuracy, faster processing speed and higher stability (Chan and Paelinckx, 2008). Breiman (2001) defines the RF classifier as an ensemble classifier that employs classification and regression trees to make a prediction. According to (Belgiu and Drăguţ, 2016) trees are created by drawing a subset of training samples through the bagging approach. The in bag samples (about two thirds of the sample) are used to train the trees while the out of bag (the remaining one third of the sample) are used for cross-validation, thus checking how well the model has performed in classifying. Resources such as topographic maps with dam points and Google Earth were used for ground truthing. A confusion matrix was constructed and used for computing the user and producer accuracies (Adam et al., 2014). The Kappa coefficient will be used to check the accuracy of the classification.

The user directed the RF algorithm to specify the land cover classes of interest and also defined the training sites, which are areas in the map known to be characteristic of a specific land cover type. In order to ensure development of an efficient classifier, representative training and validation samples for each Land Use Land Cover Class (LULC) were obtained from the Landsat 8 image. The LULC classes for which the samples were collected bare land (BS), grassland (GL), plantations (PL), cropped land (CL) built up areas (ST) and water bodies (WB). A total of 220 training and validation samples per class were collected. Table 3-3 presents the scheme used to classify the land cover and land use classes adopted from Anderson et al. (1976).

Table 3-3: Land use land cover classes attributes

LULC	**Characteristics**
Bare land	There are exposed rocks and non-vegetated ground. The land has little potential of supporting life and there is a vegetation cover of less than one third of the area
Built-up areas	Cities, towns, industrial, residential, and commercial structures
Water bodies	Streams, rivers, ponds and reservoirs
Agricultural land (Cropped, Pastures, Plantations)	The land is used primarily for the production of fibre and food and includes parcels with commercial and horticultural production

The Normalised Difference Vegetation Index (NDVI) is an important indicator used for identifying vegetated and non-vegetated land (Kiage et al., 2007). An NDVI based classification produced distribution maps of ‘vegetated’ and ‘non-vegetated’ as shown in Figure 3-1. It can be seen from its mathematical definition that the NDVI of an area containing a dense vegetation canopy will tend to have positive values (0.3 to 0.8) while clouds and snow fields will be characterized by negative values of this index. Other targets on Earth visible from space include water bodies and bare soils whose threshold values are very low. Water typically has an NDVI value less than 0, while bare soils between 0 and 0.1 and vegetation over 0.1. NDVI was applied in selection of training sites for the different land use land cover classes. In Figure 3-1 the water bodies and bare surfaces which have low reflectance values are seen as red parcels.

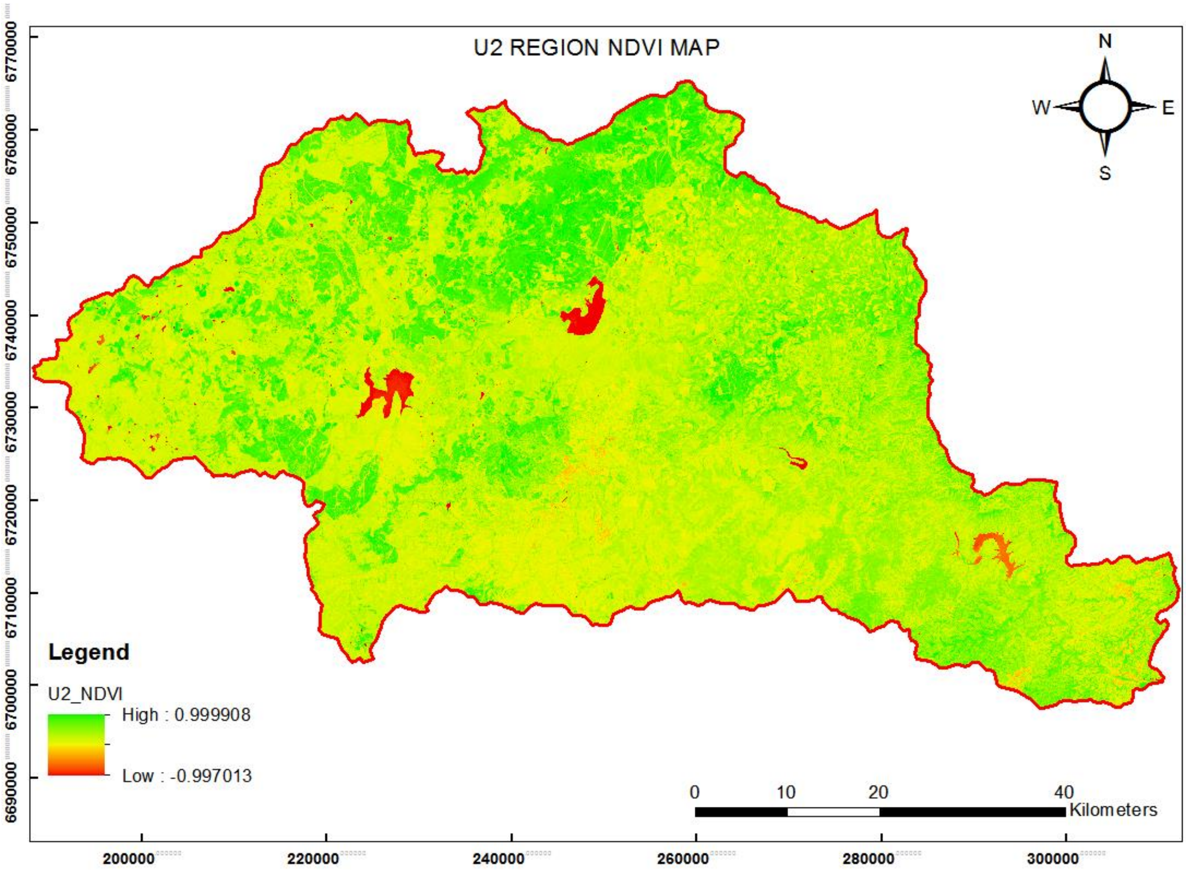

Figure 3-1. NDVI map for Region U2

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

where 0 < NDVI < 1

The spectral signatures defined in the training set formed the basis for classification of the samples. Google Earth was used for ground-truthing the LULC classification (Knorn et al., 2009).

Spectral confusion, noise and weaknesses of classification algorithms result in classification errors on classified images (Liu and Cai, 2012). Lu et al. (2004) state that a quality check of the of the classified Landsat image upon completion of classification is essential. The assessment was done in order to determine the accuracy of categorization of pixels in the area of study. According to Congalton and Green (2008), a confusion matrix is created by employing ground truth regions of interest (ROIs) for the purpose of calculating the kappa statistic, overall accuracy, producer accuracy and user accuracies. The overall accuracy is the proportion which represents the probability of a correct classification on the map for a randomly selected point (Richards, 2012). The formulae used for computing accuracy is:

$$Overall\ accuracy = \frac{\sum(Correctly\ \ classified\ classes\ along\ diagonal)}{\sum(Row\ total\ or\ Column\ total)} \qquad (2)$$

The probability of the labelled image pixel being classified correctly is referred to as producer's accuracy (Richards, 2012).

$$\textit{Producer's accuracy} = \frac{\textit{Number of the correctly classified classes in a column}}{\textit{Total number of verified items in that column}} \qquad (3)$$

According to Richards (2012) the user's accuracy is computed by dividing the number of correctly classified samples for a class by the total number of verified samples belonging to the class. The formulae is given as equation (4).

$$User's\ accuracy\ = \frac{Correctly\ classified\ number\ of\ items\ in\ a\ row}{Total\ number\ of\ verified\ items\ in\ that\ row} \qquad (4)$$

The Kappa coefficient is defined by Adam et al. (2014) as score of the relative difference between the actual agreement and the agreement expected by chance. It runs on a scale of 0 to 1 with 1 indicating perfect agreement between the classification and the ground truth points, and 0 indicating a lack of an agreement (Dorn et al., 2015). The kappa coefficient is computed by the formulae given in equation (5).

$$K\ = \frac{N\sum_{i=1}^{n} m_{i,i} - \sum_{i=1}^{n}(G_i\, C_i)}{N^2 - \sum_{i=1}^{n}(G_i\, C_i)} \qquad (5)$$

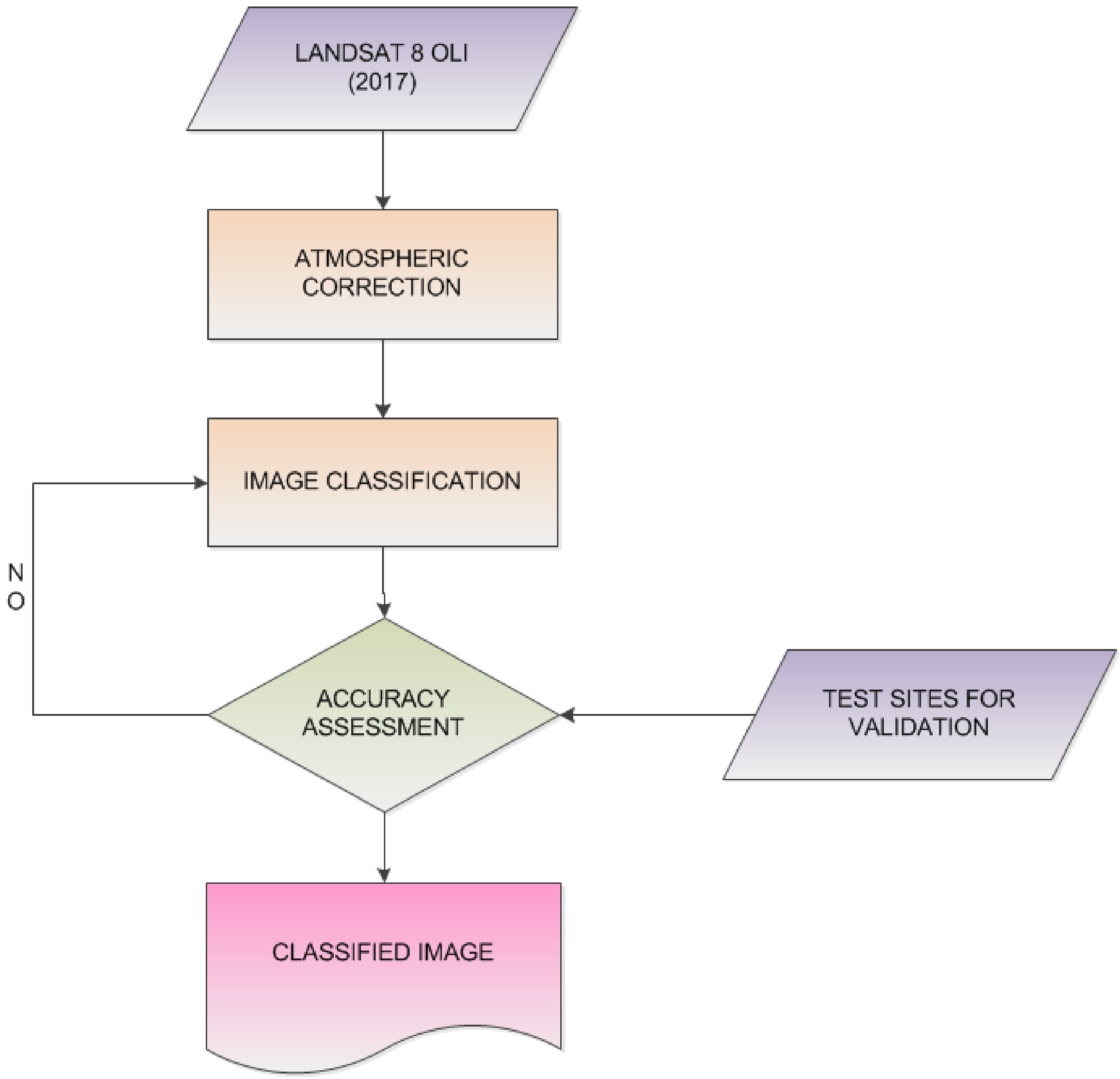

Figure 3-2. Image classification framework

Figure 3-2 gives a logical framework of the classification procedure that was followed in the study.

3.3 Digital Elevation Model data

3.3.1 Overview

According to Burrough and McDonnell (1998), a Digital Elevation Model (DEM) is a quantitative model of a part of the earth's surface in digital form, consisting of spatial distribution of elevations on a regular grid (Forkuor and Maathuis, 2012). DEMs provide one of the most useful digital datasets for a wide range of users. According to Nikolakopoulos and Chrysoulakis (2006), there are 2 post-processed elevation datasets that are widely used for a number of applications due to their near global coverage. These are Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), and Shuttle Radar Topography Mission (SRTM) Digital Elevation Models (DEMs). Both SRTM and ASTER data have been widely applied in geomorphology, hydrology, tectonic, and others since being made freely accessible to public (Ouerghi et al. 2015). The SRTM DEM was selected for use in the study because it is comparatively more accurate than the ASTER DEM (Ouerghi et al., 2015).

The SRTM DEM GEOTIFF files covering the study area were acquired from the earth explorer web page. The extent of the study area is covered by 3 tiles. As such, 3 GEOTIFF files of 1 arc second resolution were extracted and mosaicked in ARCMAP, using the data management tools in the following sequence:

Arc Toolbox > Data Management Tools > Raster > Raster Dataset > Mosaic to New Raster

3.3.2 Mosaicking, Clipping and Filling

The mosaicking technique produces DEMs of large areas from various input DEMs with a common coordinate system. Mosaicking of DEMs also includes compositing their data sets for the same area by averaging the available height values within overlap areas (Knöpfle et al., 1998). The 3 tiles have a common coordinate system and cell depth. They were merged and created a new raster surface, thus a new single DEM with a larger extent than the study area. Thereafter, a GIS shapefile for the U2 study area was used to clip the new DEM so that it could cover only the extent of the study area.

DEMs require pre-processing before use as they possess grids of no data (voids) and some erroneous data that needs correction (Forkuor and Maathuis, 2012). According to ESRI (2013), a DEM free of sinks, thus a depression-less DEM, is suitable as input data to the flow direction process. The presence of sinks may result in an erroneous flow-direction raster. In some cases, there may be legitimate sinks in the data. Sinks can be located using the Sink tool. This tool requires a direction raster that is created by the Flow Direction tool. The sink tool was used to identify sinks in the DEM. The Fill function was used to fill the identified sinks, creating a new filled Raster surface.

3.3.3 Hydrological network extraction

DEMs have been widely used as input data for defining the flow directions in distributed hydrological models for discharge simulation due to their high efficiency in representing the spatial variability of the Earth's surface (Beasely et al.1980; Beven and Kirkby 1979; Fortin et al. 2001). Zhang et al. (2017) highlight other applications of DEMs which include extraction of watershed characteristics such as channel networks, hillslope, flow length, sub catchments, soil erosion and flood simulation. Figure 3.3 illustrates the steps taken to extract the hydrological network from the DEM.

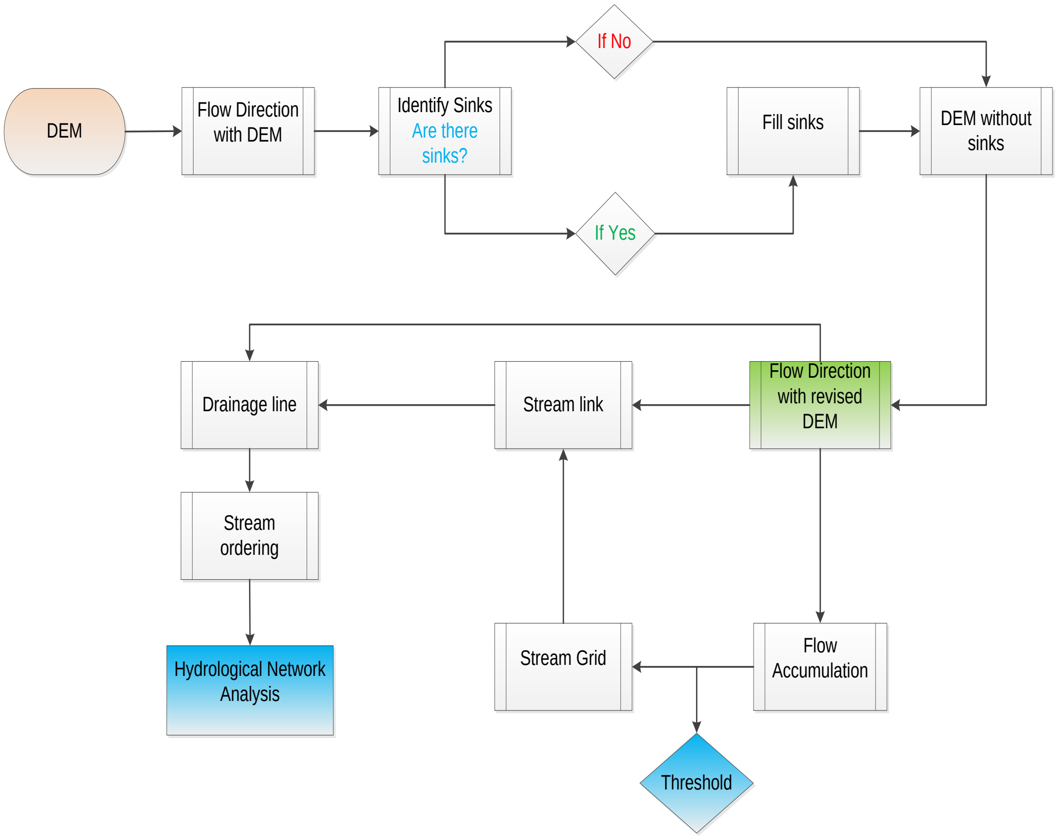

Figure 3-3. Conceptual workflow for hydrological network extraction procedure in ARCGIS Adapted from Hosseinzadeh (2011)

Before starting the flow direction computation process, it is important to assess the DEM and check for the existence of sinks, should the sinks exist, the steps mentioned in the previous section can be taken. The flow direction raster was created using the filled DEM. The flow direction raster indicates the direction of the downslope neighbor for each cell by colour coded direction. The maximum drop is given by the expression: Difference in elevation between cells / Distance between cell centers x 100.

The flow accumulation raster was created thereafter, using the flow direction raster surface as input. The flow accumulation raster shows the number of cells contributing to flow for each cell in the grid. Each cell carries a contributing weight of one. The resultant flow accumulation raster was input for the conditional flow raster grid. This raster grid performs a conditional if / else evaluation on each of the input cells of an input raster. The conditional flow grid was used as the input stream raster, and the flow direction raster as the input flow direction raster. The resulting raster was a Strahler (Stahler, 1952). The Stahler hierarchy is the default stream ordering pattern in ARCGIS. The Stahler number is a modification of the Horton number.

Having extracted the stream orders raster, it was used as input for stream network raster, and the flow direction raster was used as the input flow direction raster. The resulting raster was a stream network. It was converted to a polyline features and stream orders 5 – 11 were extracted.

3.3.4 Soil data and geology data acquisition

Soil data

Soil characteristics are very important factors in dam site assessment (Stephens, 2010). Soil porosity should be minimized in order to achieve a watertight seal for dams. Fine materials such as clay, help in blocking soil pores by compacting them. Compaction is required at the correct moisture content such that clay clods are crushed and air voids squeezed out.

GIS vector shapefile data with a scale of 1:250000 scale data for South Africa was acquired from the European Soil Data Centre (ESDAC). The data was projected to the UTM projection and was used in mapping the distribution of farm dams and converted to .tiff and asc. raster file formats and used as input for the multivariate logistic regression model.

Geology data

According to Stephens (2010) dam sites are not identical with respect to geology. Each dam construction project must be investigated individually. With regard to geology, some dam sites have relatively consistent and uniform geology, i.e. one rock type with a simple structure and a regular pattern of surface weathering. Geology sites are often complex and have various rock types with different physical characteristics such as strength, durability and susceptibility to weathering. A geology shapefile with a scale of 1: 250000 scale for

South Africa was acquired from the City of Johannesburg. The data was projected in UTM projection and was also used to map the distribution of farm dams.

3.4 Determination of spatial distribution of farm dams

3.4.1 Farm dam distribution maps

Distribution of farm dams refers to the statistical distribution of specific properties such as size, and to their distribution over space (Liebe et al., 2005). The datasets used were farm dam point data. Farm dam polygon feature shapefiles were converted to point data in ArcGIS and had their attribute data such as area attached to them. The dams were counted per grid and it was determined which grids had dam presence and absence and spatial pattern of dams. Distribution maps with respect to respect to topography, land use, soil type, rock type, aspect and slope were produced. Distribution maps with respect to surface area were also produced.

3.4.2 Farm dam spatial point pattern analysis

According to Diggle (2003), a point process is defined as a 'stochastic mechanism which generates a countable set of events. In this study, the events are dam sites. Locations of the dam sites generated by a point process in the area of study are called a point pattern (Parmigiani, 2009). Additional covariates such as area covered by dam and its use were recorded and attached to the locations of the observed events. A point pattern analysis of the dam sites was carried out for dam site location point data for the study area. A script written in R-Studio was used to perform the point pattern analysis. Analysis was done using the using, the quadrat method and nearest-neighbour methods respectively which test Complete Spatial Randomness (CSR) of spatial point pattern data.

Quadrat Method

The quadrat method is essentially a direct test of the Complete Spatial Randomness (CSR) hypothesis (Grabarnik, 2002). The study area was enclosed by a rectangular region closely fitting to its extent on all four sides, and is sub divided into congruent rectangular sub cells (quadrats). The CSR hypothesis asserts that the cell-count distribution for each quadrat must be the same. This approach was used to determine the farm dam point density in the study area.

Nearest-Neighbor Methods

According to Akasapu et al. (2011), the nearest-neighbour analysis compares the distance from each point to its closest neighbour with the nearest-neighbour distance expected for a random distribution of point within a specified location with known boundaries. Nearest neighbour was used to determine if there is a spatial pattern such as clustering and dispersion for the dam points. Nearest neighbour analysis examined the distances between each point and its closest neighbour. Nearest Neighbour functions such as the G, F and K were used to determine whether the spatial point pattern of dams was clustered or dispersed.

3.5 Evaluation of factors influencing spatial distribution of farm dams

A principal component analysis was carried out in order to determine factors of importance in farm dam distribution in the study area. A principal component analysis is a process defined by Wold et al. (1987) as a method that extracts the dominant patterns in a data matrix in terms of a complementary set of score and loading plots. The aim of a principal component analysis is to extract the important information from the Table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps (Abdi and Williams, 2010).

Sample site data was prepared in an excel sheet and loaded into the SAS 9.4 statistical package. The data was run and the results obtained showed six principle components. The components were analysed using the correlation matrix, Eigen vector Tables, scree plots and Eigenvalues of the Correlation Matrix. The variables that had the highest loadings were identified.

3.6 Predictive Model

This section presents the steps followed in the assessment of the locational properties of the distribution of farm dam sites using logistic regression, and the creation of a prediction surface using multivariate logistic regression. A script written in R-Studio was used to perform the logistic regression and multivariate logistic regression.

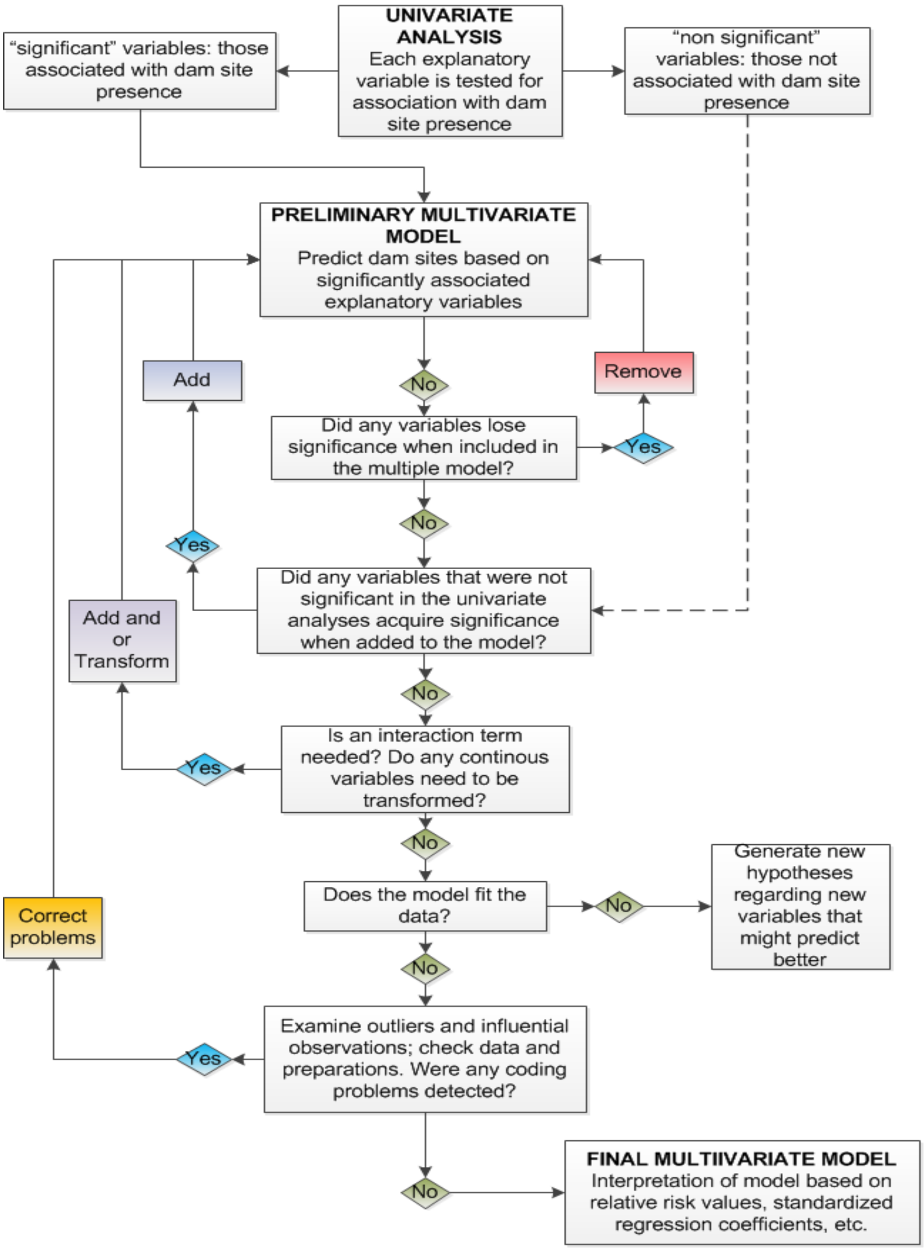

Figure 3-4. Multivariate logistic regression model for dam site location prediction

Figure 3-4 illustrates the fitting process of the logistic regression models, including the iterative checking and re-fitting procedure before arriving at the final multivariate logistic model that was followed in the study.

3.6.1 Covariate creation and data set up

The elevation (SRTM DEM), soil, geology, land cover class raster datasets and dam site and non-site point locations data were prepared and added to the model in R. The dam site and non-site point locations data was a vector dataset of point locations representing dam sites (coded 1), as well as a set of non-sites (coded 0). These sites and non-sites were a `binary' or `presence-absence' dataset that together formed the dependent variable in the regression. The raster datasets were converted from .tiff file ASCII file format using the conversion tools in ARCGIS. Table 3-4 shows the data requirements and the file formats for the model. Slope and aspect were generated from the elevation data. A preliminary investigation of the variables was done by plotting a map of each covariate next to a histogram of the cell frequencies on that map.

Table 3-4: Data requirements for the model

Covariate	**Data format**
Elevation	ASCII format (.asc)
Slope	ASCII format (.asc)
Aspect	ASCII format (.asc)
Soil	ASCII format (.asc)
Geology	ASCII format (.asc)
Land use	ASCII format (.asc)
Dam site and non-site point data	Shapefile format (.shp)

3.6.2 Univariate Analysis

An assessment of the statistical relationship between the dependent variable and each of the covariates ensured. Each of the covariates was individually tested for association with dam site presence. This gave a preliminary assessment of how useful each covariate may be as a predictor. In cases where the p-values were greater than 0.05, the variable was discarded and regarded as unlikely to be of much use for prediction.

As an alternative, the likelihood ratio test method of testing the significance of models was also used. It makes an assessment of the difference between the error of not knowing the independent covariates and the error when the independents are included in the model (Crema et al., 2010). This test follows a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters in the model with independents compared to the model with only the intercept.

3.6.3 Visual assessment

In order to further examine the relationship or a lack of, between the predictor and response variables, simple histograms of sites and non-sites against the covariates were plotted and visually assessed. The visual assessment of the histograms suggested the possibility that breaking down the variables into carefully chosen sub-groups could increase their predictive power. The process created some new, nominal-scale (categorical) variables. Table 3-5 shows the covariates after being divided into sub groups.

Table 3-5: Subgroups of covariates created to improve the model

Covariate	Sub-group	Sub-group range
Elevation	Low elevation model	0 – 700 m
	Medium elevation model	700 – 1 600 m
	High elevation model	1 600 – 2 100 m
Aspect	South East facing model	90° - 180°
	North East facing model	0° - 90°
Slope	Gentle model	0° – 5 °
	Steep model	5° – 15°
	Very steep model	15° – 88°
Soil	Fr20 model	
	Lo43 model	
Geology	Adelaide model	
	Karoo model	
	Pietermaritzburg model	
	Volksrust model	
Land use	Grassland model	
	Agricultural land (cropped) model	

Each of the subgroup's relationship with the dependent variable was tested to establish if the variables could be of use in the prediction of the dependent variable. Variables that had a significant p value at $\alpha = 0.05$ significance level were considered for the model, and those that were not significant were again discarded.

3.6.4 Multivariate Logistic Regression

Key variables were identified using the univariate analysis. A binary multivariate logistic regression model was created using important predictor variables. Despite univariate results, the multivariate outcomes do not show high levels significance for all covariates. Since this depends on how the covariates combine with one another, different combinations were tested (by dropping or adding covariates) in order to get the best model. The step AIC () function in the MASS package in R Studio was used to compute the AIC values for the different models. The function allows for assessment of the relative merits of different models (starting from the one provided by the user) (Crema et al., 2010). The best model was chosen on the basis of the Akaike Information Criteria (AIC). AIC provides a measure of the performance of a model in relative but not absolute terms, and is useful for rough comparisons between models that use different combinations of covariates (Crema et al., 2010). Typically, the smaller the AIC, the better the model performance.

3.6.5 Creating the Prediction Surface

The coefficients of the logistic regression and raster map algebra were used to generate a prediction surface. This computed a prediction for each cell using the equation:

$$logit\ (p)\ =\ \alpha +\ \ \beta_{1x1} + \beta_{2x2} + \cdots . + \beta_{nxn} \tag{6}$$

Where: α is the intercept, β are the coefficients and x are the covariates. This gives the natural logarithm of the odds of sites presence divided by the odds of sites absence. From this, the relative probability of a site existing at a particular location can be calculated using the following equation:

$$p_i = \frac{exp(V_i)}{1 +\ exp(V_{i)}} \tag{7}$$

After picking out the model with the best combination of variables based on AIC, a log odds expression was created. Thereafter a relative probability surface was generated.

4. RESULTS

4.1 Overview

The chapter presents the results of the study on the spatial distribution of farm dams. It gives the results of the image classification, mapping of the distribution of farm dams, spatial distribution of farm dams, principal component analysis of farm dam location variables and multivariate logistic regression.

4.2 Landsat Image classification

The land use land cover map of the study area (2017) in Figure 4-1 and the histogram in Figure 4-2 show that 54% of U2 region was covered by grasslands (GC). Figure 4-2 also illustrates that 20% of the land was covered by plantations (PL), 13% by bare soil (BS), 10% by cropped agricultural land (CL), 2% by water bodies (WB) and 0.7% by built up areas (ST).

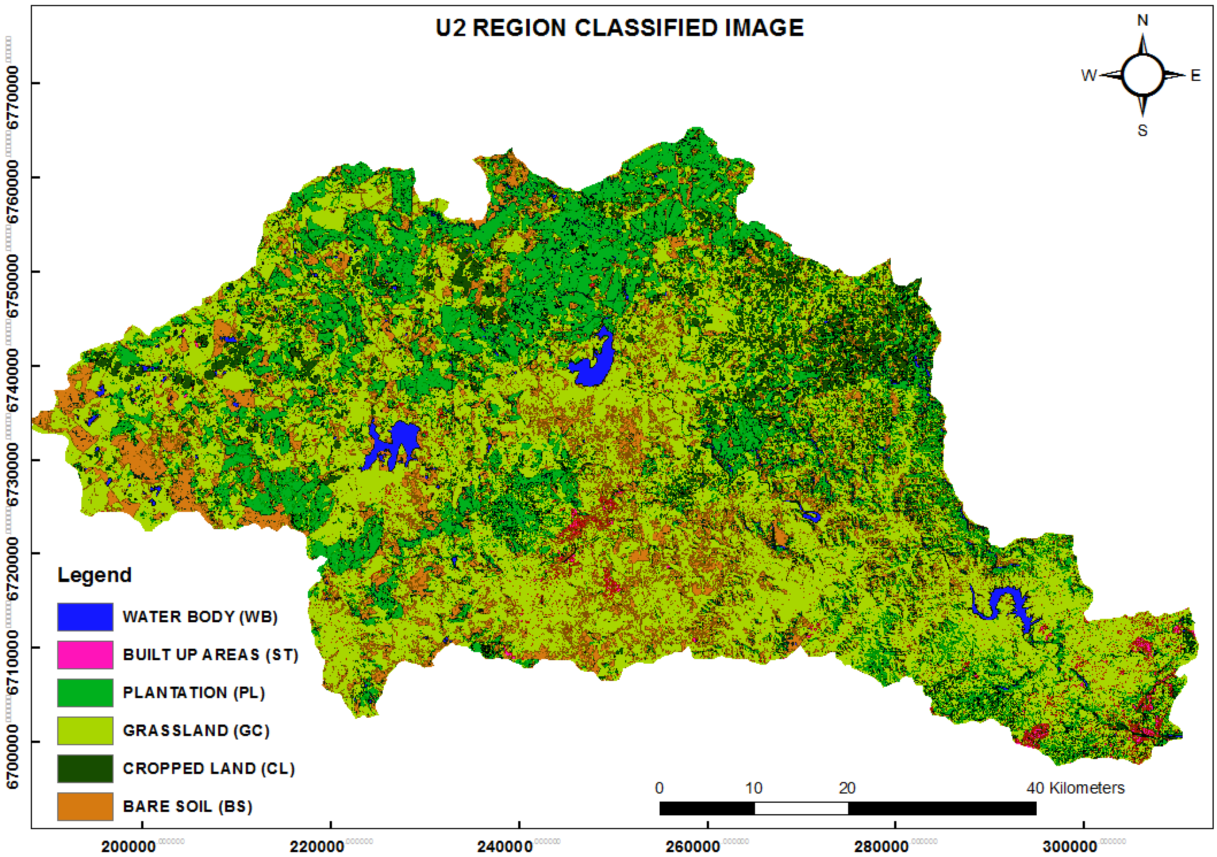

Figure 4-1. Region U2 classified Landsat 8 image

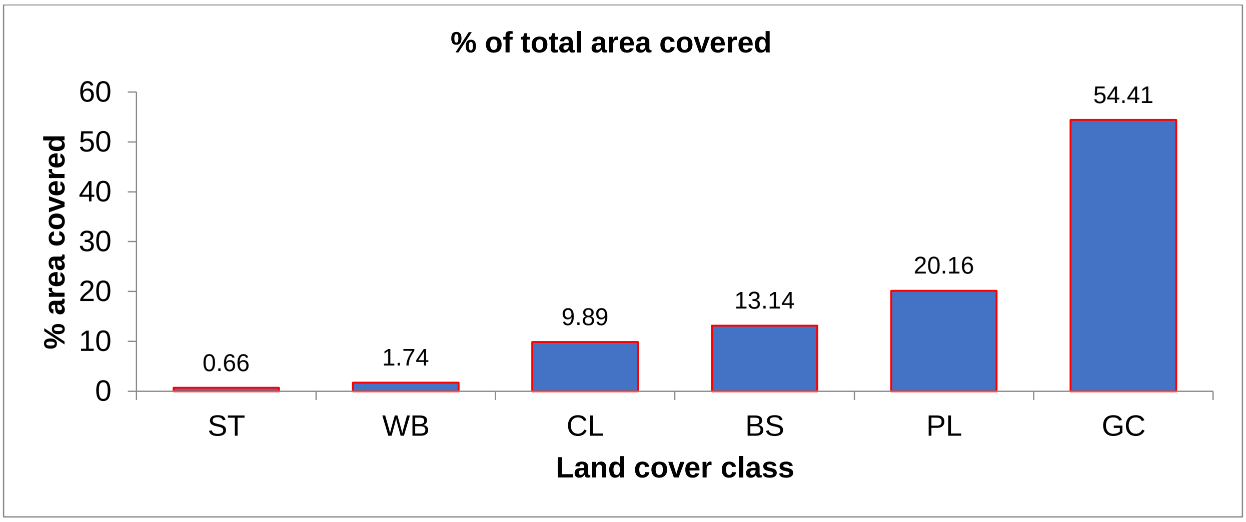

Figure 4-2. Proportions of area covered by land use land cover classes

Accuracy Assessment

Accuracy of the Landsat image that was classified using RF classifier was produced by using data from the confusion matrix produced from the R-Studio platform. Ground control points extracted from Google Earth that were not used in training, were used in the construction of a confusion matrix.

Table 4-1: Confusion matrix for validation

Land cover class	Producer Accuracy (%)	User Accuracy (%)
Agricultural land (cropped) (CL)	97.8	94.7
Plantations (PL)	98.7	93.8
Water Bodies (WB)	98.7	100
Built up areas (ST)	100	100
Grassland (GC)	95.4	91.3
Bare soil (BS)	94.3	95.7
Kappa Coefficient (K)		0.956
Overall Accuracy (%)		96.4

According to Anderson et al. (1976), the recommended threshold value for accuracy assessment is 85%. The technique used in this study produced a level of accuracy and reliability that meets the threshold. Table 4-1 shows that the overall accuracy for the classified Landsat image was 96%.

Water bodies and built up areas had higher user and producer accuracies compared to other land cover classes. The grass cover class had 91.3% accuracy and had the lowest user's accuracy. The producer's accuracy for the bare soil class was the lowest with a value of 94.3%. The reason why all classes could not attain 100% accuracy was because some pixels had mixed grass and bare soil and were problematic and resulted in misclassification leading to low producer and user accuracies.

4.3 Spatial distribution of farm dams

4.3.1 Spatial point pattern analysis of farm dam data

According to Parmigiani (2009), Complete Spatial Randomness (CSR) is often the standard against which datasets are tested for spatial distribution point pattern data. It assumes that points follow a homogeneous Poisson process over the study area. The distribution of farm dams in the study area was tested for complete spatial randomness using the quadrat and nearest neighbor methods.

Quadrat Analysis

Quadrat analysis is the most basic form of CSR, carried out using the quadrat count function in Spatstat (R-Studio).

The intensity and density of farm dams is not constant across the study area. Figures 4-3 and 4-4 indicate greater intensity and density in the north-western part of the study area where there are relatively more farm dams per unit area. The data resembles a bull's eye effect.

The following hypothesis was developed and tested using a χ^2 test for CSR using quadrat counts:

H_0: The farm dam distribution pattern exhibits complete spatial randomness (CSR)

H_A: The farm dam distribution pattern is spatially clustered or dispersed

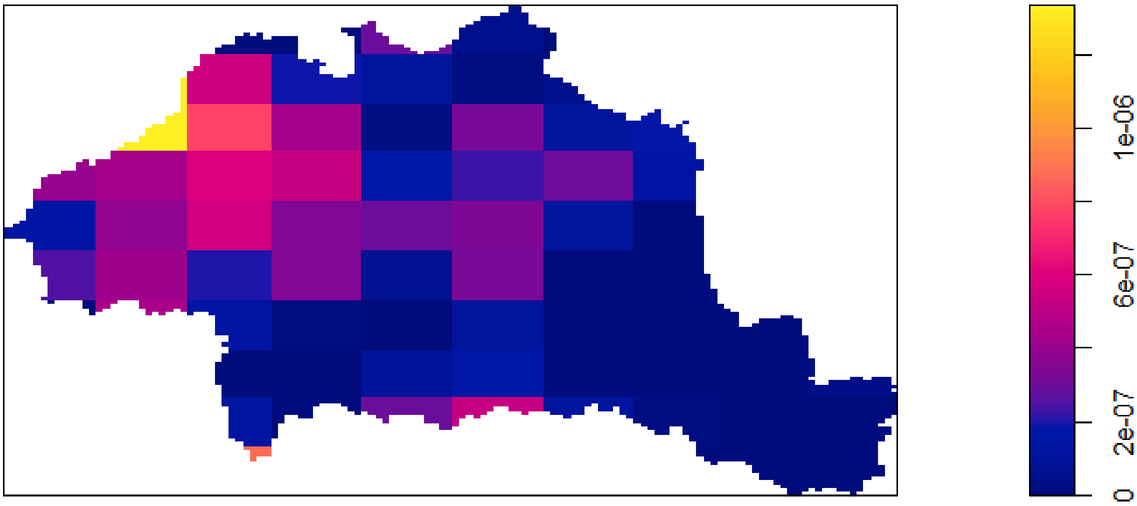

Figure 4-3. Intensity of farm dams

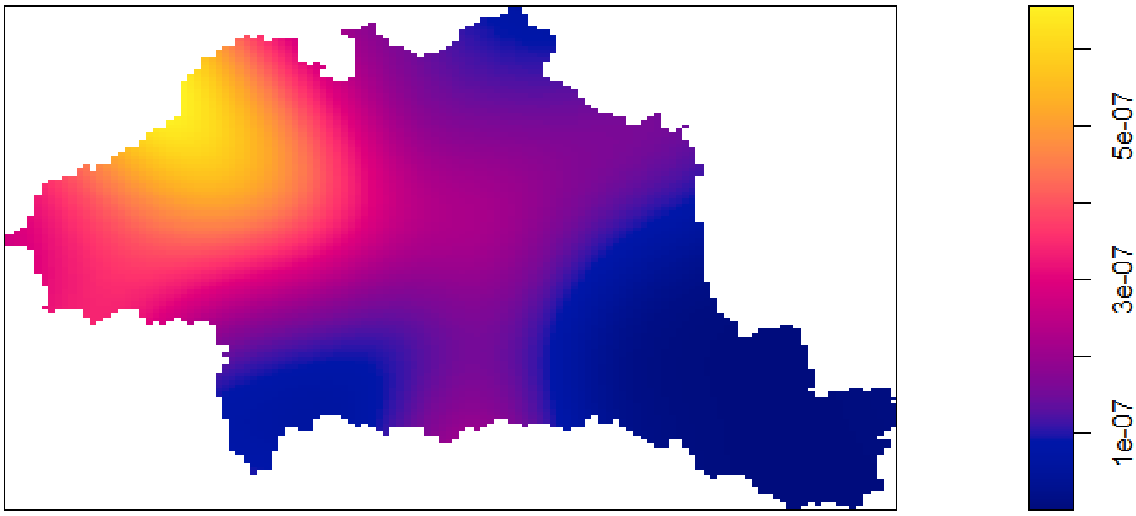

Figure 4-4. Density of farm dams

Table 4-2: Chi-square test p-values for quadrat analysis

Quadrat size	X^2	DF	P-value
10 X 10	1095.4	70	< 0.001
20 X 20	1507.2	257	< 0.001

Results from Table 4-2 indicate that for the 2 quadrat sizes that were tested, the p-values of the χ^2 tests are both significant (< α=0.05). Therefore the null hypothesis is rejected as there is sufficient evidence to conclude in favor of the alternative. The farm dam distribution pattern is spatially clustered or dispersed and there is no CSR. In order to determine whether the farms are clustered or dispersed, Nearest-Neighbour Methods were used and assessed.

Nearest-Neighbour Methods

Nearest-Neighbor methods are distance based measures that quantify interaction in spatial point patterns via G function (even-to-nearest-event distances), F-function (point-to nearest-event distances) and L-function (event-to-event distances).

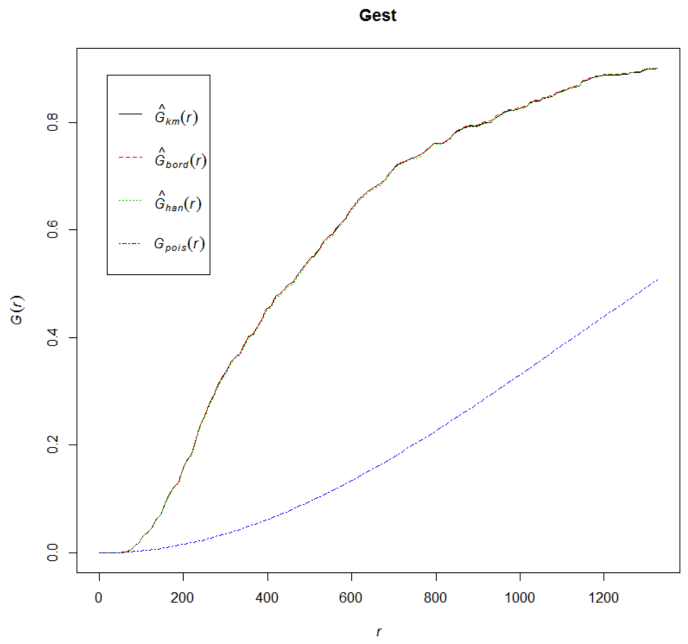

Figure 4-5. Gest function – event to nearest event distance function

The G-function also known as the event-to-nearest-event distances function examines the frequency distribution of the nearest neighbor distances to the point. The shape of the G function illustrates how the events are spaced in a point pattern. The graph in Figure 4-5 indicates a clustered distribution pattern for farm dams between 300 – 1200 m between each other.

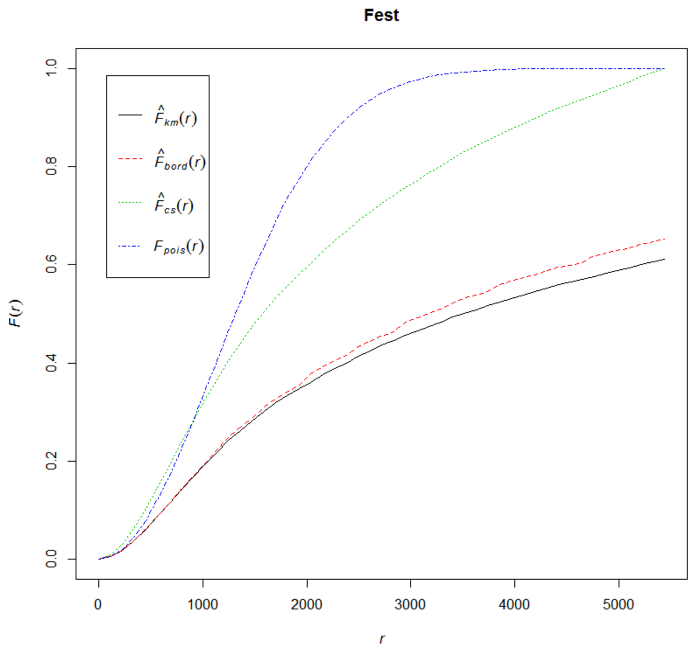

Figure 4-6. Fest - nearest neighbour analysis function

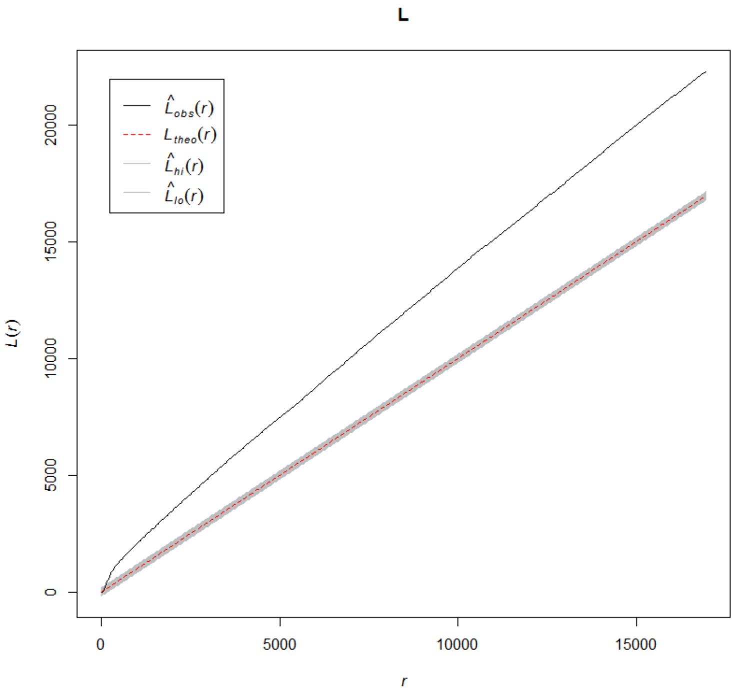

Figure 4-7. Lest function – point to nearest event

The L function in Figure 4.7 measures the distribution of all distances from a point to its nearest event. The L function is also known as empty space function because it measures the average space left between events. When the observed L value is larger than the expected L (theoretical) value for a particular distance, the distribution is more clustered than a random distribution at that distance. From the graph the L-function (L_{obs}) line presented by the solid black line is above the L_{theo}. This means the distribution of farm dams over a 15 km distance in the study area is clustered. The clustering is statistically significant since the L-function (L_{obs}) (observation line) above the confidence limits L_{hi} (95%) and L_{lo} (5%).

Based on evidence from the nearest-neighbor-methods, there was clustering of farm dams in the study area over short and long distances. The clustering of farm dams may be a result of the distribution of land under crop production and livestock farms. The farms are situated close to each other.

4.3.2 Mapping the distribution of farm dams with respect to size

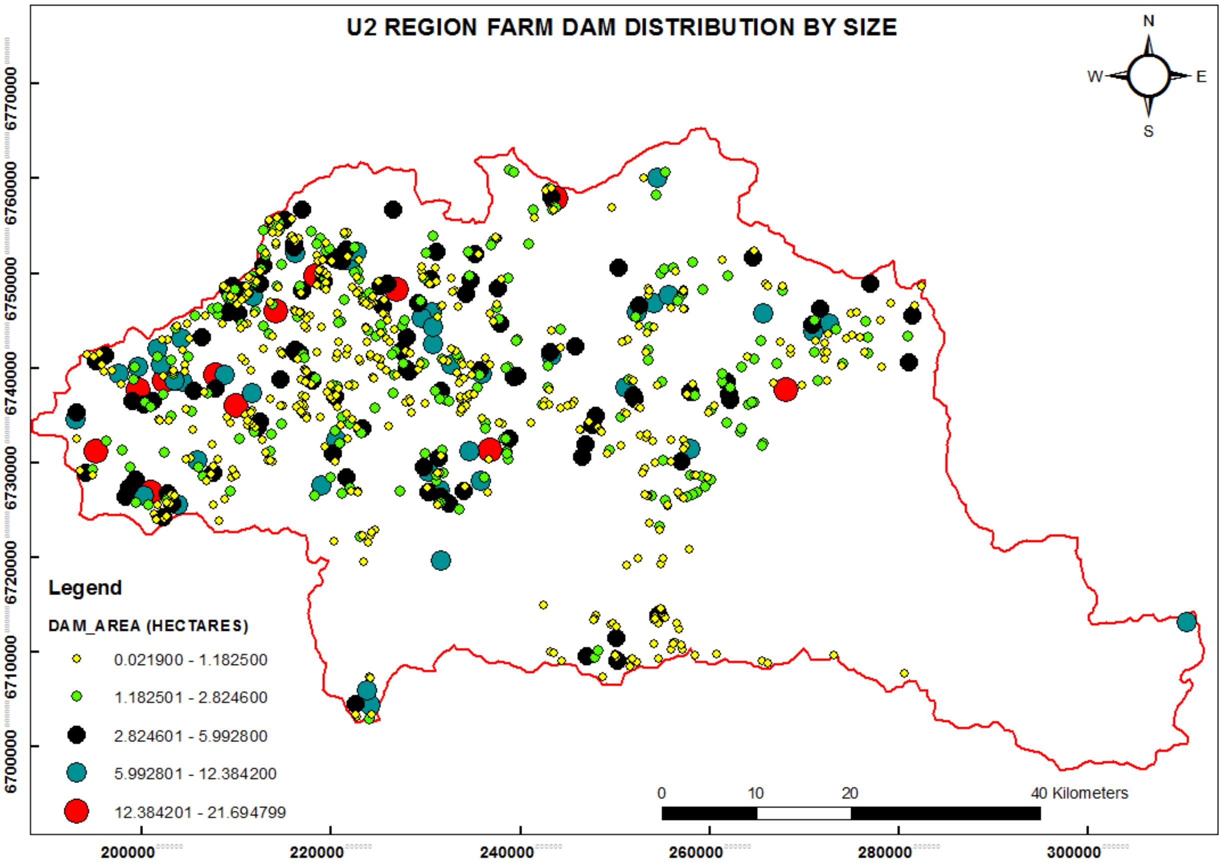

Figure 4-8. Map of the distribution of farm dams with respect to surface area (ha)

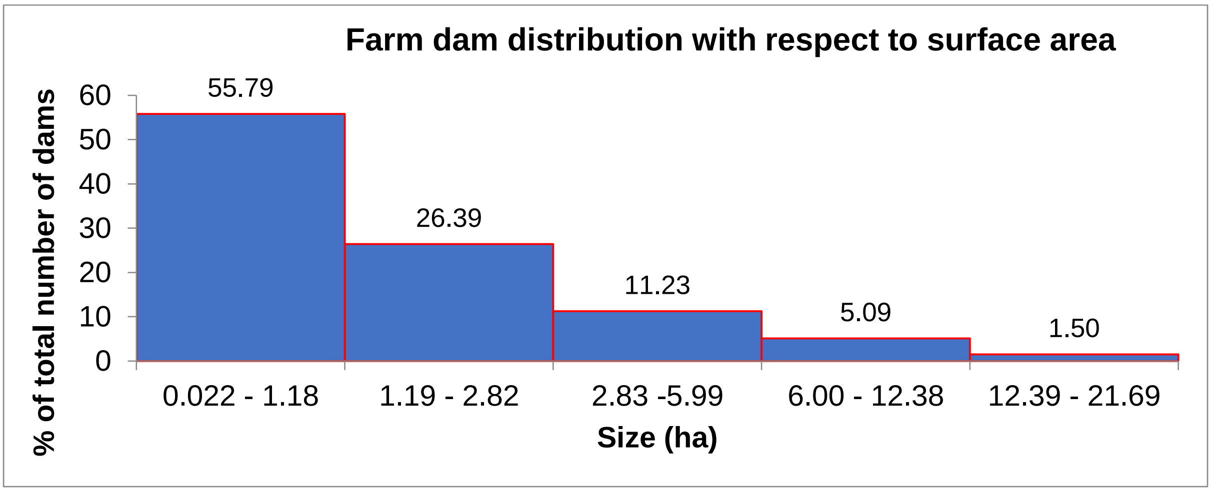

Figure 4-9. Distribution of farm dams with respect to surface area

The map for the study area in Figure 4-8 and the histogram in Figure 4-9 show the distribution of farm dams by surface area covered. With respect to farm dam size, the surface area covered ranges from 0.02 – 22 ha. About 90% of the largest farm dams are found in the north western region. From the histogram in Figure 4-9, it can be observed that dams between 0.022 – 6 ha surface area account for 90% of the total number of farm dams, while the rest account for 10 %.

4.3.3 Mapping the distribution of farm dams with respect to geology

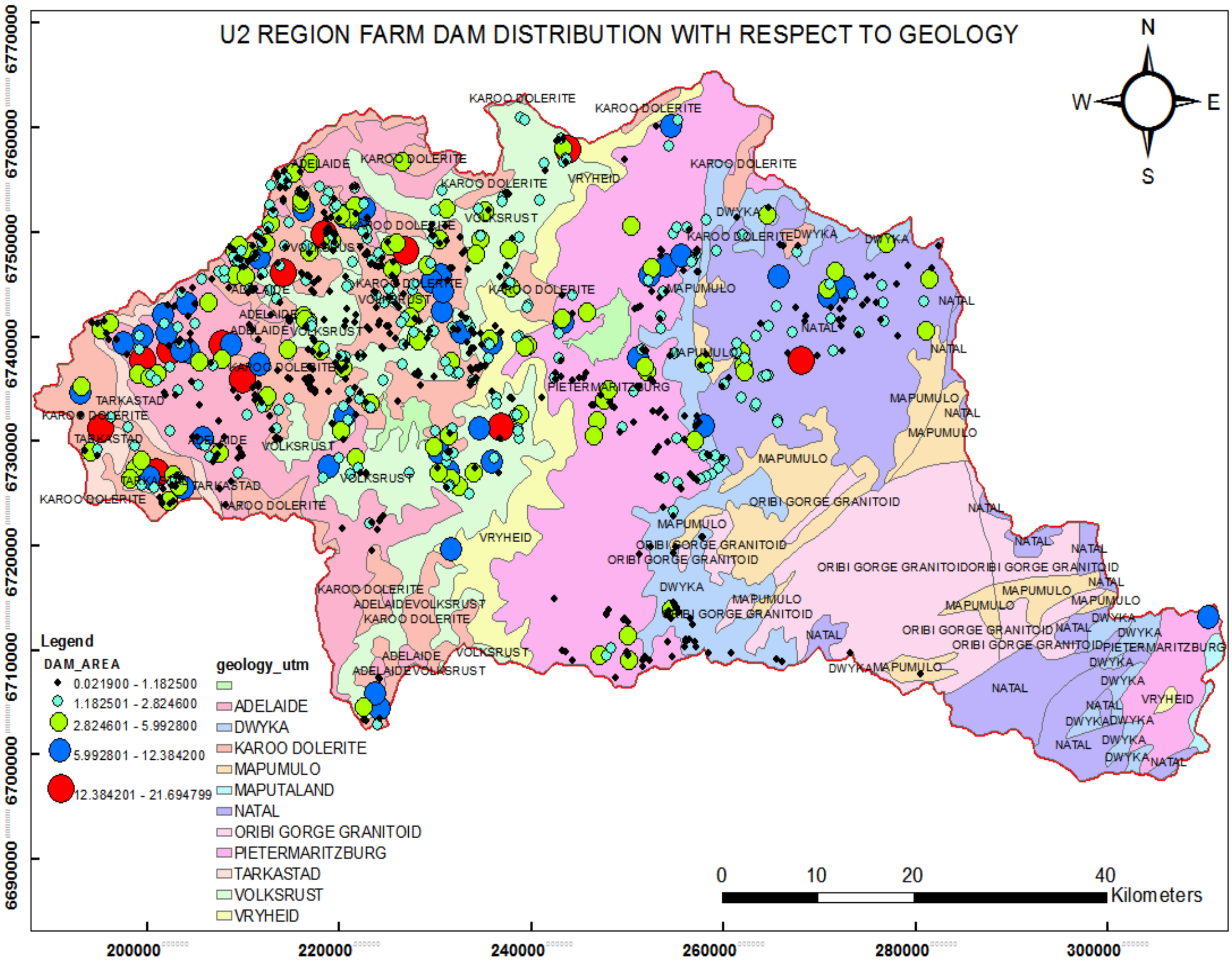

Figure 4-10. Map of the distribution of farm dams with respect to geology

Table 4-3: Distribution of farm dams with respect to rock formation

Geology	Number of Dams	% of Total number of dams	Area km^2)	Density (/100 km^2)	Rank
Dwyka	55	6.37	302.5	18	5
Karoo	233	26.97	606.6	38	1
Adelaide	139	16.09	446.5	31	3
Tarkastad	21	2.43	69.4	30	4
Pietermaritzburg	129	14.93	801.4	16	6
Vryheid	10	1.16	199.3	5	8
Volksrust	202	23.38	583.5	35	2
Natal	69	7.99	674.7	10	7
Mapumulo	5	0.58	262.3	2	9
Oribi gorge granitoid	1	0.12	447.5	0	10
Maputaland	0	0.00	10.4	0	11

The geology map for the U2 region in Figure 4-10 and Table 4-3 illustrate the distribution of farm dams constructed with respect to geological formation type. Approximately 27% of U2 region farm dams were constructed on the Karoo formation and has a density of 38 dams / 100 km^2 which is the highest. Adelaide and Volksrust rank second and third, accounting for 23% and 16% of the total number of dams in U2 region respectively. Adelaide has a density of 35 dams / 100 km^2 while Volksrust has 31 dams / 100 km^2. Pietermaritzburg rock formation covers the largest proportion of the study area; however it ranks sixth in the density of farm dams with 16 dams / 100 km^2. Vryheid and Mapumulo formations account for 1.2% and 0.6 % of U2 region farm dams each with a density of 5 and 2 dams / 100 km^2 respectively. Oribi gorge granitoid and Maputaland rock formations have farm dam densities of 0.

4.3.4 Mapping the distribution of farm dams with respect to elevation

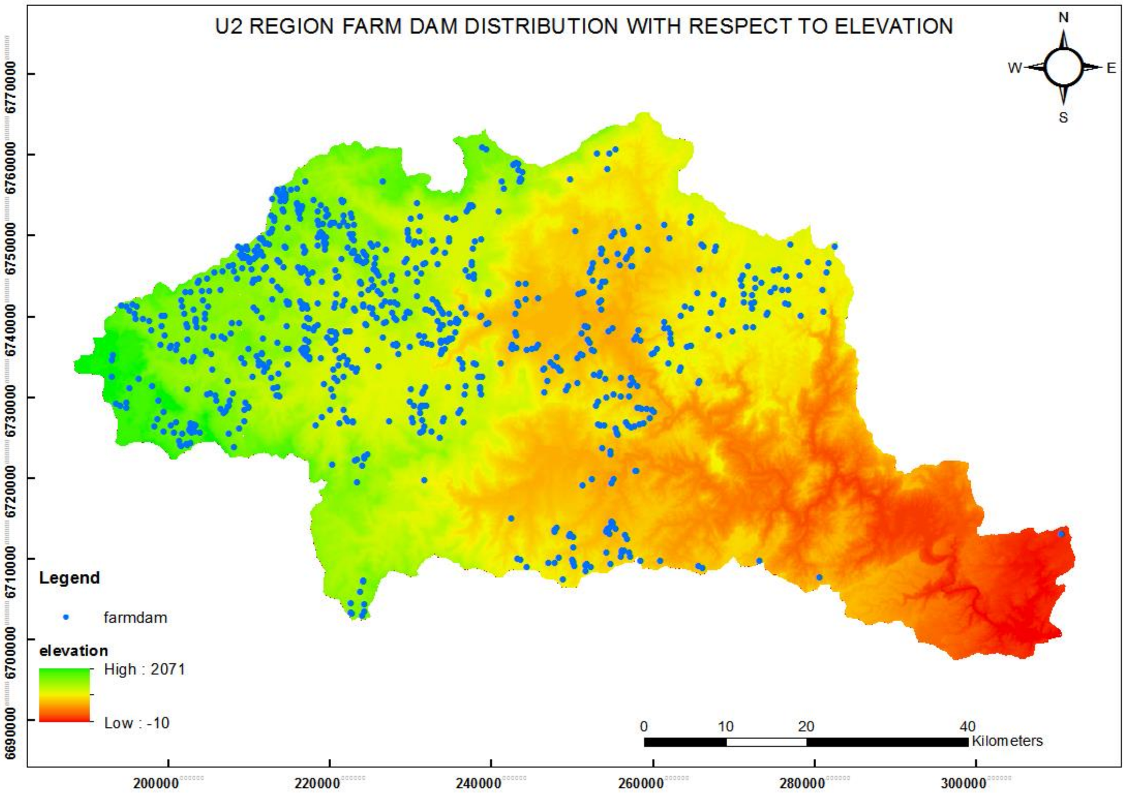

Figure 4-11. Map of the distribution of farm dams with respect to elevation

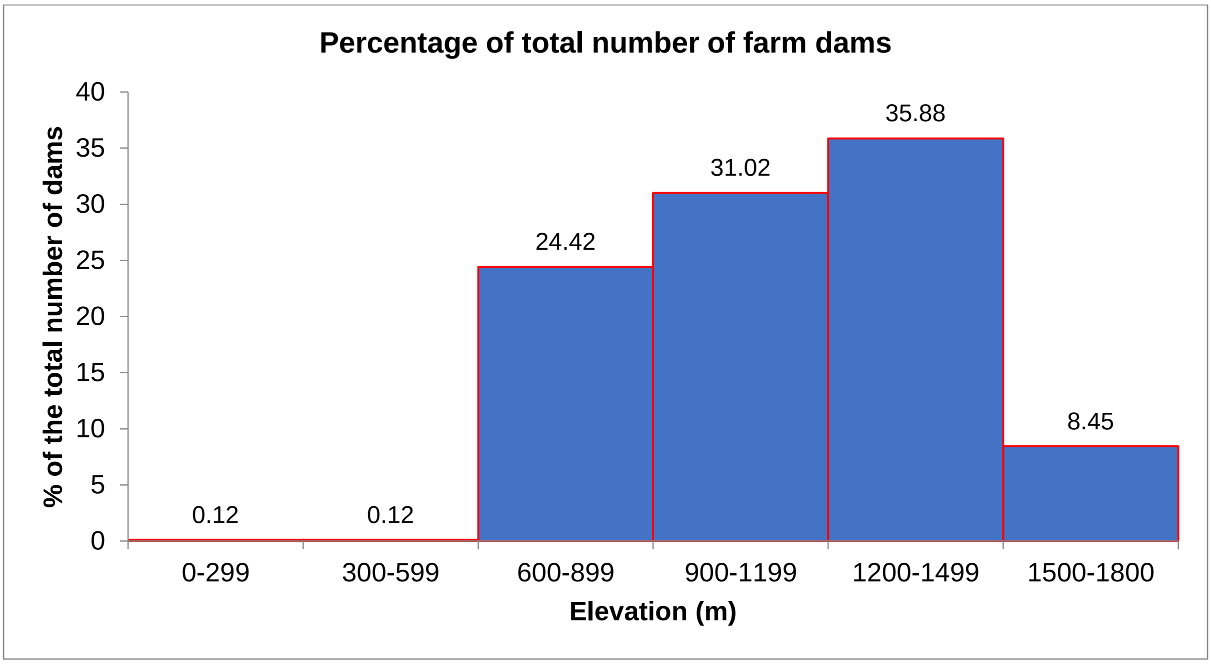

Figure 4-12. Distribution of farm dams with respect to elevation

The elevation map for the U2 region in Figure 4-11 and the histogram in Figure 4-12 show the distribution of farm dams constructed with respect to elevation. Approximately 36% of U2 region farm dams were constructed on altitudes of are between 1200–1500 m ASL. Areas that lie at elevations between 900-1200 m ASL account for 31% of the total number of farm dams in the U2 region while areas between 600-900 m ASL have 24%. Areas below 600 m have a combined total of 0.24%.

4.3.5 Mapping the distribution of farm dams with respect to slope

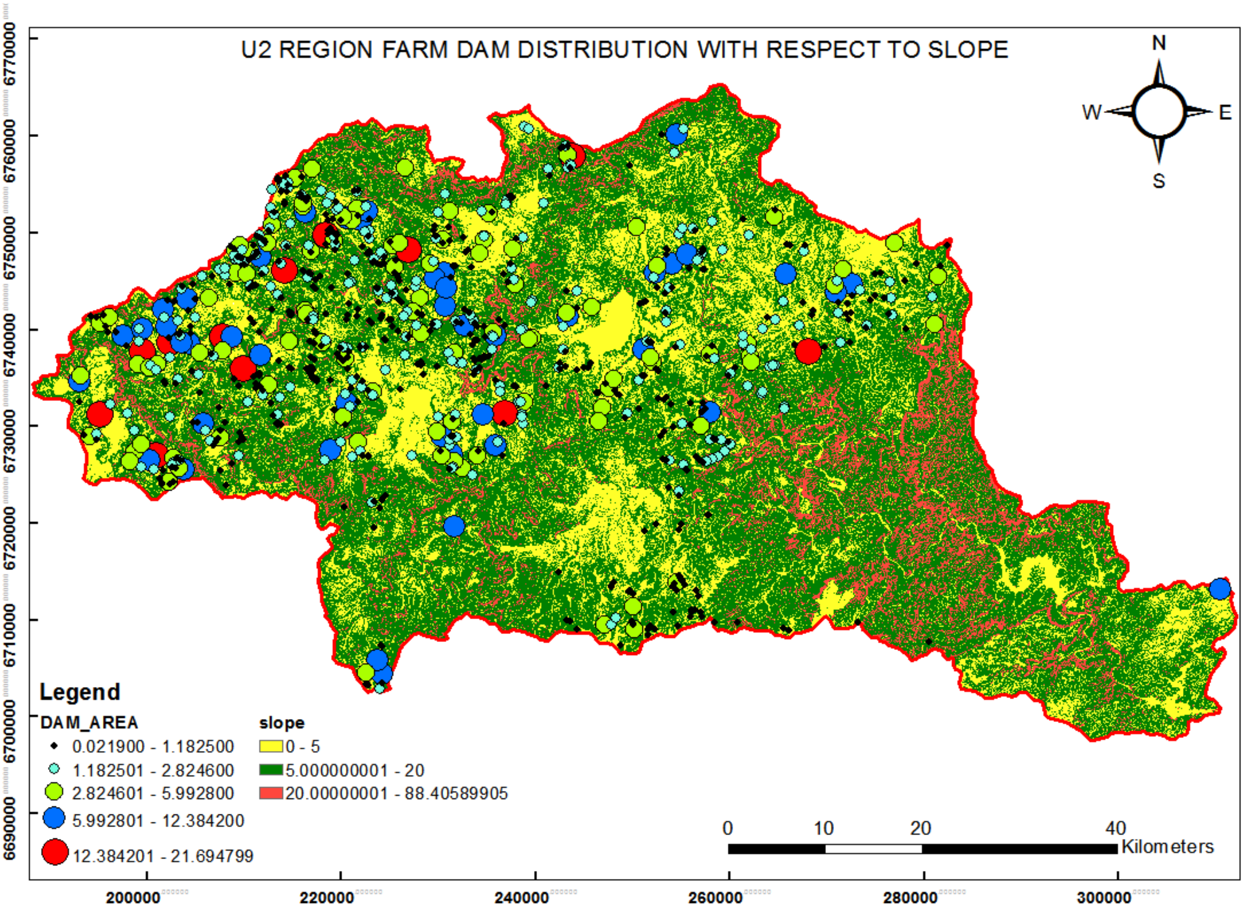

Figure 4-13. Map of the distribution of farm dams with respect to slope

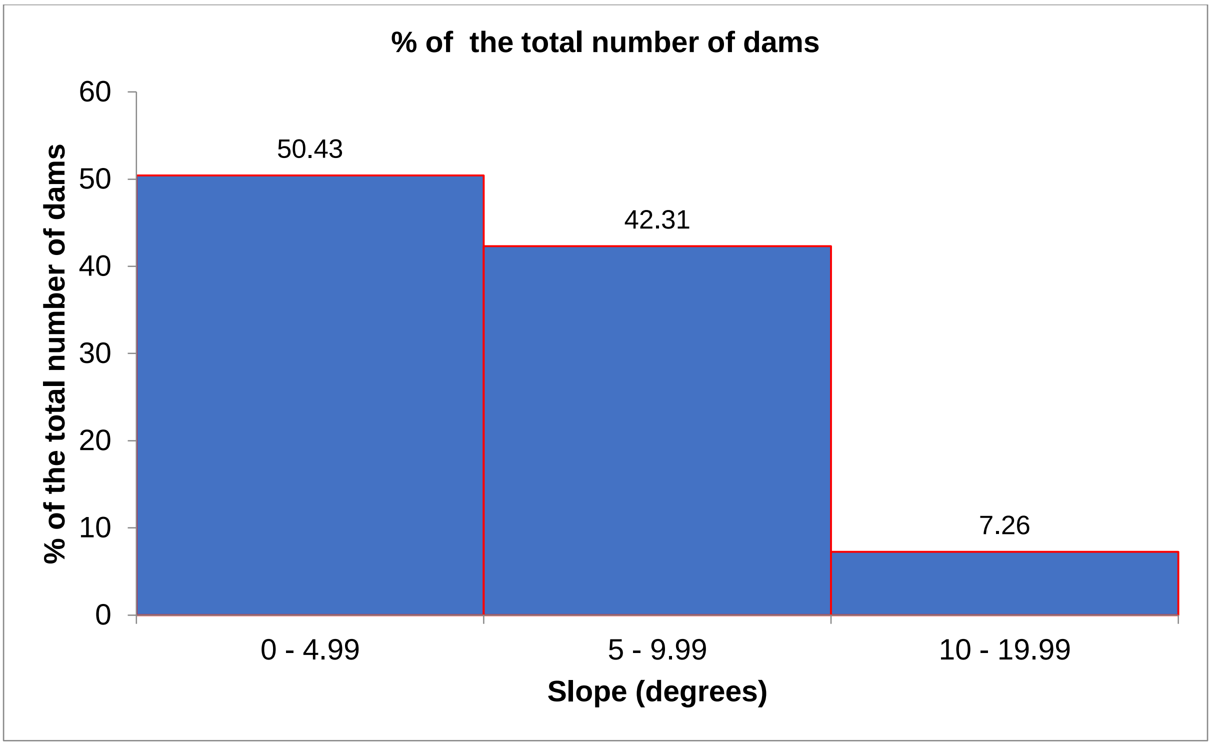

Figure 4-14. Distribution of farm dams with respect to slope

The slope map for the U2 region in Figure 4-13 and the histogram in Figure 4-14 illustrate the distribution of farm dams constructed with respect to slope. Approximately 50% of the total numbers of farm dams were built on gentle slope, thus areas in the slope range 0°– 5° slope while 42% of lie in areas with a moderately steep slope of range of 5°-10°. The slope range of 10°–20° accounts for the remaining 8%.

4.3.6 Mapping the distribution of farm dams with respect to soil

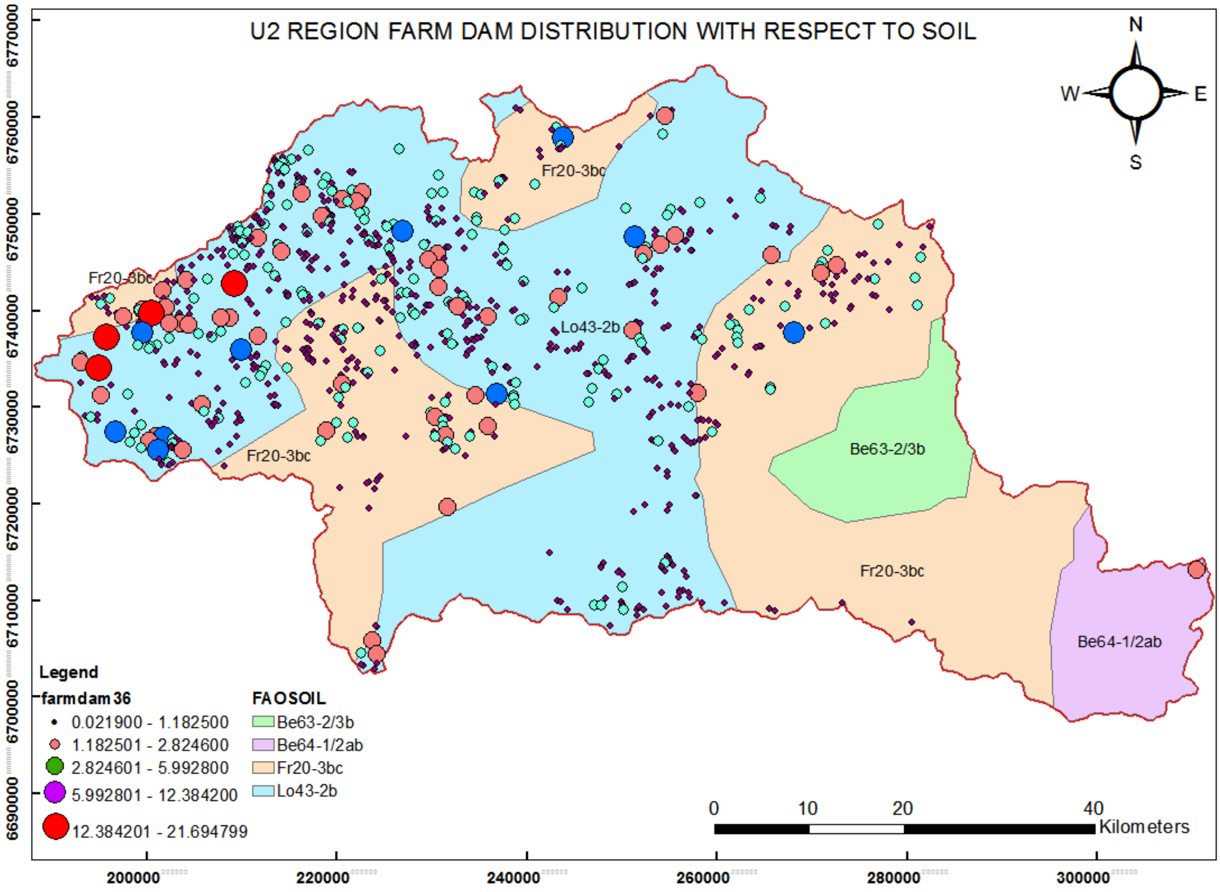

Figure 4-15. Map of the distribution of farm dams with respect to soil

Table 4-4: Distribution of farm dams with respect to soil formation

Soil type	Number of Dams	% of Total number of dams	Area (km^2)	Density (dams/100 km^2)
Be63-2/3b	1	0.1	231.4	0.4
Be64-1/2ab	1	0.1	241.1	0.4
Fr20-3bc	249	28.8	1810.08	13.8
Lo43-2b	613	70.9	2159.6	28.4

The soil map for the U2 region in Figure 4-15 and Table 4-4 illustrate the distribution of farm dams constructed with respect to soil formation type. Approximately 71% of U2 region farm dams were constructed on the Orthic Luvisols (Lo43-2b). Humic Ferralsols (Fr20-3b) have a dam density of 29 dams/100 km^2, accounting for 28.8%. The Cambisols, thus, Be63-2/3b and Be64-1/2ab, have a dam each in their areas and account for 0.1%. Their chemical and physical properties are good and may be used as sites for dams however, in this case they are found in areas occupied by the cities and towns and there is minimal agricultural activity.

4.3.7 Mapping farm dam distribution with respect to land use

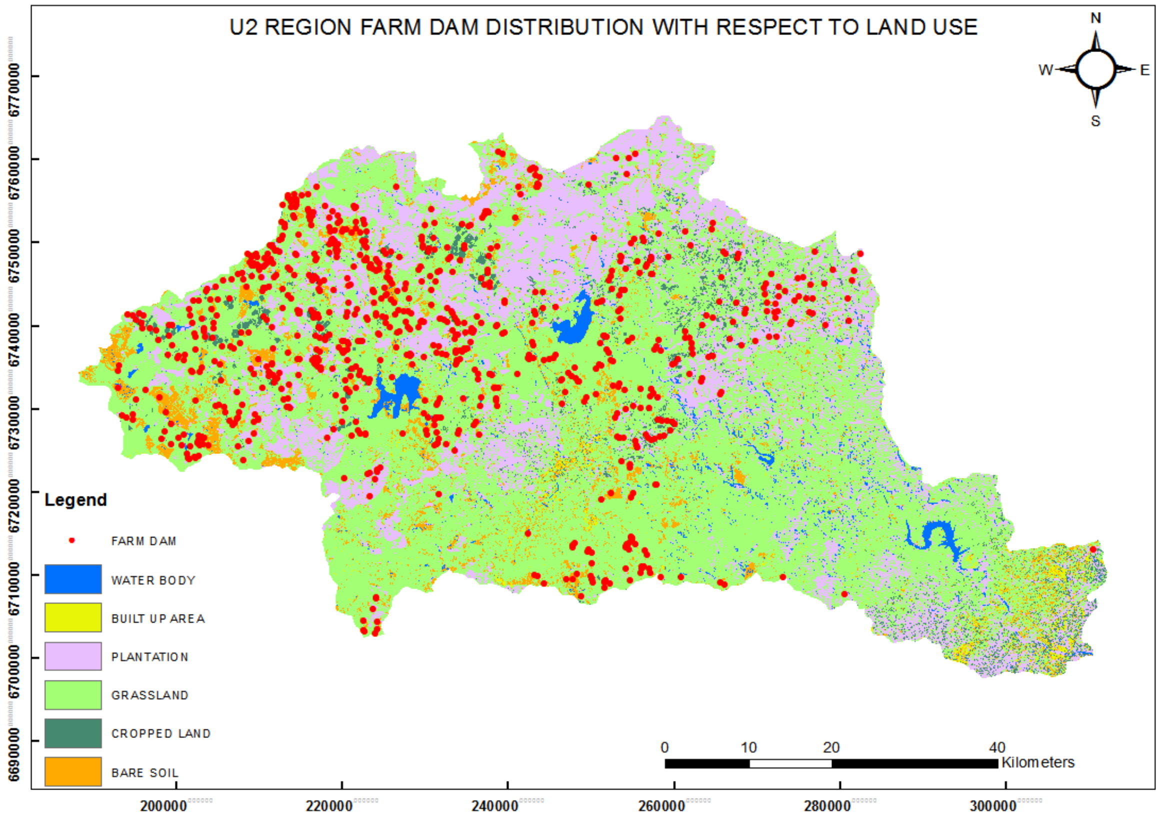

Figure 4-16. Map of the distribution of farm dams with respect to land use

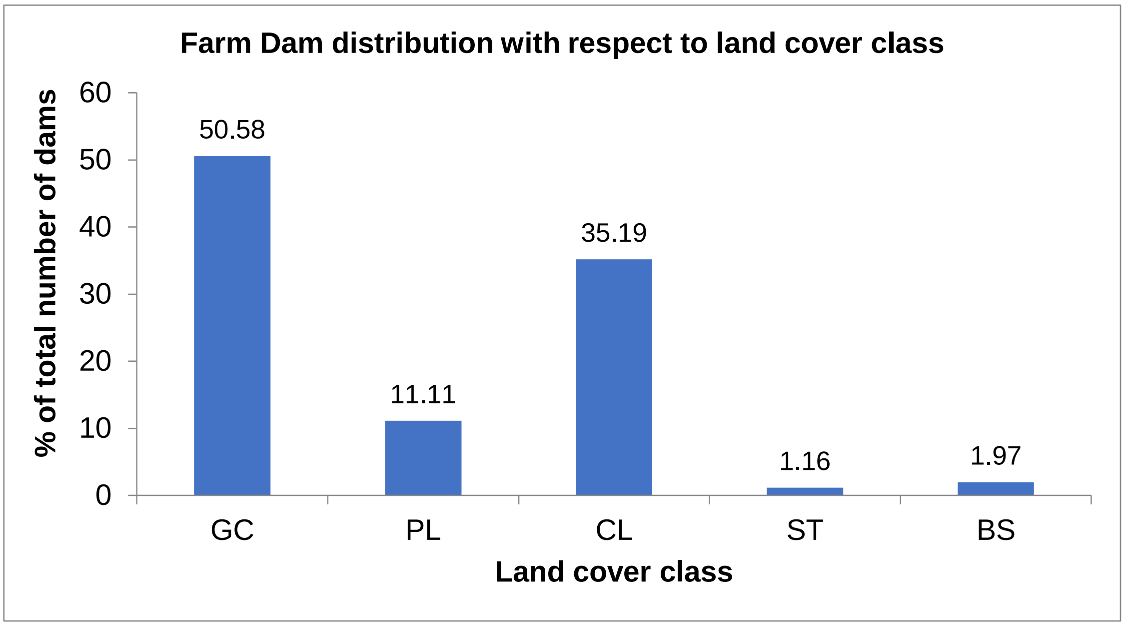

Figure 4-17. Distribution of farm dams with respect to slope

Figures 4-16 and 4-17 illustrate the distribution of farm dams constructed with respect to land use. Approximately 51% of U2 region farm dams were found in grasslands (GC), while

35% exist in areas near cropped land (CL). Plantations (PL) had 11%, while areas with bare soil (BS) and built-up areas (ST) had 1.97% and 1.16% respectively. Most farm dams are found in areas with grasslands.

4.4 Principal Component Analysis

The outcome of a principal component analysis for six physical factors affecting dam site selection is presented in this section. The principal component analysis is meant to compute the single variable that best summarizes all the factors affecting dam site selection.

Table 4-5: Correlation matrix

	Elevation	Slope	Aspect	Geology	Soil	Land use
Elevation	1.0000	-.0901	0.0087	-.3061	0.0557	-.2047
Slope	-.0901	1.0000	0.0878	0.0637	-.1143	0.0695
Aspect	0.0087	0.0878	1.0000	0.0161	0.0245	0.0781
Geology	-.3061	0.0637	0.0161	1.0000	-.3155	0.3376
Soil	0.0557	-.1143	0.0245	-.3155	1.0000	-.1959
Land use	-.2047	0.0695	0.0781	0.3376	-.1959	1.0000

The correlation matrix in Table 4-5 displays the correlation between the six variables. The variables are generally not exhibiting much dependence with each other, as most association are weak. Geology and elevation have a negative correlation of -0.3, meaning that there is a slight negative dependence between geology and topography. Geology and land use also have a slight positive correlation of 0.337 meaning that land use may change with rock type also changes.

Table 4-6: Eigenvalues of the Correlation Matrix

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	1.76922868	0.70227975	0.2949	0.2949
2	1.06694893	0.10730152	0.1778	0.4727
3	0.95964741	0.04178912	0.1599	0.6326
4	0.91785829	0.19873006	0.1530	0.7856
5	0.71912824	0.15193979	0.1199	0.9055
6	0.56718845		0.0945	1.0000

Table 4-7: The Eigen vectors

Eigenvectors						
	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5	PRIN6
Elevation	-.420332	0.107925	-.499534	0.533640	0.332737	0.408231
Slope	0.214074	0.568963	-.459896	-.602045	0.195930	0.134544
Aspect	0.069192	0.790195	0.316621	0.419343	-.303485	-.051030
Geology	0.581806	-.164675	0.059732	0.104032	-.217314	0.756815
Soil	-.426586	0.111800	0.628839	-.318132	0.335860	0.442806
Land use	0.502227	0.024670	0.199264	0.254738	0.773800	-.209274

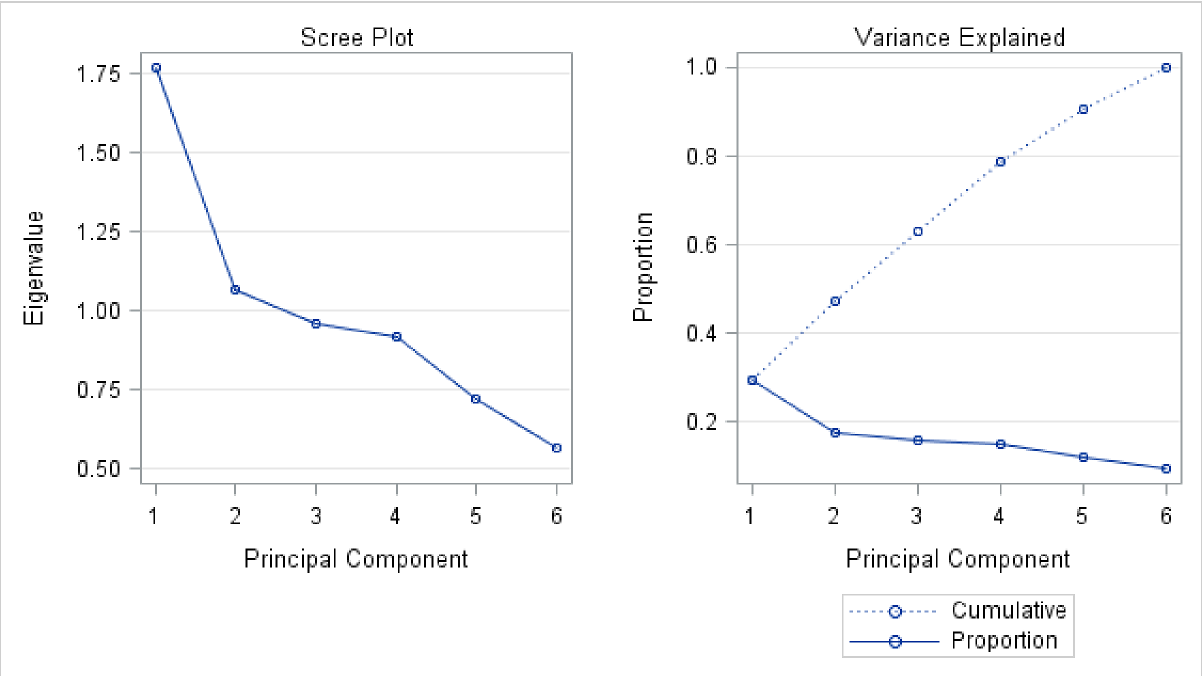

Figure 4-18. The scree plots showing how much variance is explained by each Principal Component

The eigenvalues and the scree plot in Table 4-6 and the scree plot in Figure 4-18 indicate that two or three components provide a good summary of the data, two components accounting for 47% of the total variance and three components explaining 63%. Subsequent components contribute 28%.

In Table 4-7 results of the loadings of the variables on the principal components are presented. Elevation and soil have negative loadings of -0.420 and -0.426. Principal component two has high positive loadings for two terrain variables, slope (0.568) and aspect (0.790) while the loadings of elevation are a lowly (0.107). In principal component 3, there are high positive loadings for soil (0.628) and reasonably high loadings for elevation (0.499) and slope (-0.459), while aspect loads 0.31. There is preponderance of terrain variables over other physical variables. Again this principal component seems to be a measure of the physical attributes of dam site location though the interpretation is not obvious.

4.5 Farm dam site location predictive modeling

This section presents the multivariate logistic regression model for prediction of farm dam distribution in the U2 region. Dam site location was the response variable, while elevation, slope, aspect, geology, soil and land use were the independent variables. The first step in the logistic regression procedure involved carrying out univariate analysis. The results for the univariate analysis are presented in Tables 4-8 to 4-13.

4.5.1 Univariate Analysis results

Univariate analysis gives a preliminary assessment of how useful a covariate may be as a predictor. In cases where the p-value is much greater than 0.05, it was discarded as the variable was seen as unlikely to be of use for prediction.

Table 4-8: Elevation model univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.9698101	0.4059990	-2.389	0.0169 *
elevation	0.0015471	0.0003777	4.096	<0.001 ***
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

Table 4-9: Slope model univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.76966	0.30500	9.081	<0.001 ***
Slope	-0.33076	0.04346	-7.610	<0.001 ***
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

Table 4-10: Aspect model univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.7948112	0.2464085	3.226	0.00126 **
aspect	-0.0008721	0.0012493	-0.698	0.48512
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

Table 4-11: Soil model univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.3433	0.7224	-3.244	0.00118 **
soil	0.8517	0.2030	4.195	<0.001 ***
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

Table 4-12: Geology univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.58194	0.27119	5.833	<0.001 ***
geology	-0.18754	0.04645	-4.038	<0.001 ***
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

Table 4-13: Land use univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.13980	0.27048	0.517	0.6053
Land use	0.18732	0.09078	2.063	0.0391 *
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

The univariate analysis gave significant outcomes for all variables at $\alpha = 0.05$ level of significance except for aspect. Aspect had a non-significant p value of 0.48512, which is greater than the significance level ($\alpha = 0.05$) and was preliminarily deemed unimportant for the multivariate model. However, the aspect variable was not completely thrown out as it was subdivided into smaller covariates which also underwent a univariate analysis.

A visual assessment of histograms plotted of the 6 variables suggested the possibility that breaking down the variables into carefully selected sub-groups could increase their predictive power. The process created new nominal-scale (categorical) variables. The histograms for the subdivided covariates are presented in Figures 4-19 to 4-23. There are two histograms for each variable, one for sites and one non-sites. The "site =1" histogram represents site presence, while the "site = 0" represents absence.

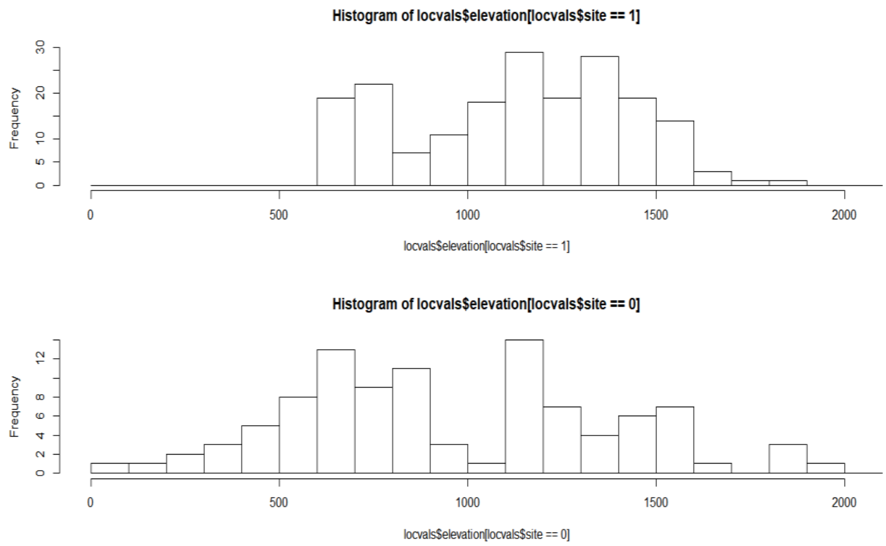

Figure 4-19. Histograms of the elevation of dam location points sites and none sites

Although elevation had a highly significant p value, it was further subdivided into groups to ensure that the most suitable elevation for dam sites could be picked based on the sample site locations. From the histogram of site points in Figure 4-19 it is clear to see that at least 90% of the dam sites are found in areas between 600 m asl and 1 600 m asl. The elevation was therefore subdivided into 3 groups low, medium and high and the range 700 – 1 600 m asl was put in the medium class.

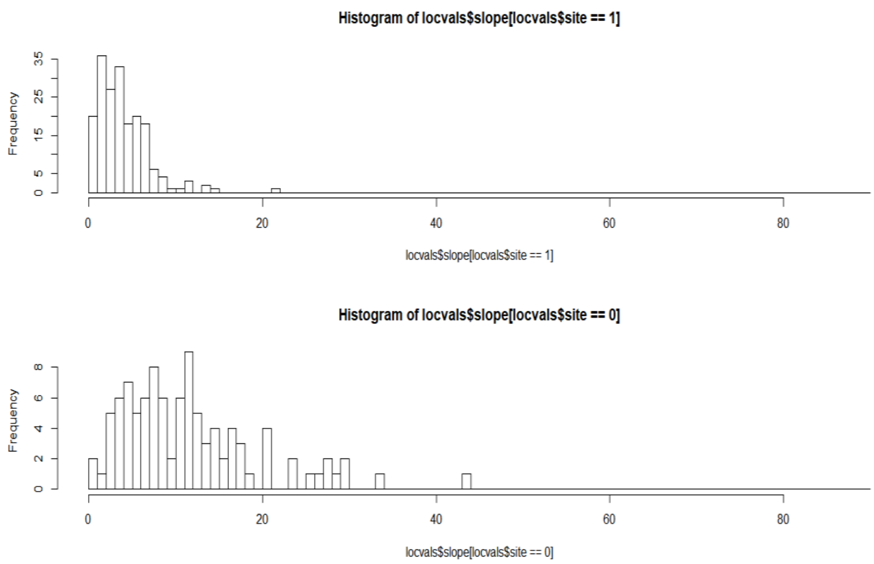

Figure 4-20. Histograms of the slope of dam location points sites and none sites

Figure 4-18 presents the histograms of slope for sample site and non-site points for the study area. In the site =1 histogram which represents site presence, it is evident that the range of slope for farm dam site presence is from between a slope of 0° and 21° slope. Most of the farm dams are found between 0–5°. Slope just like elevation also had a very highly significant p value in initial analysis. It was also subdivided to test the ranges of slope that were most significant and could be of importance in the multivariate model. The slope was divided into 3 classes, gentle (0°-5°), moderately steep (>5-15°) and steep (>15°).

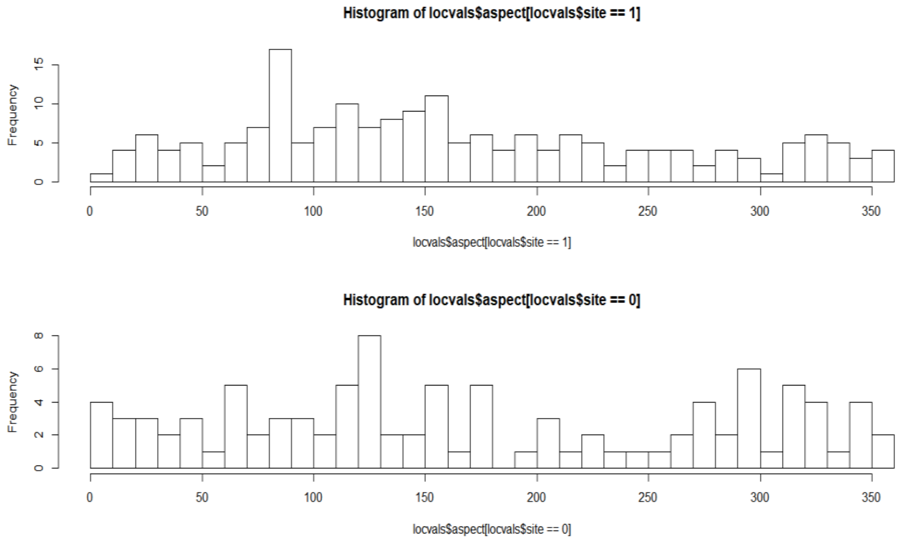

Figure 4-21. Histograms of the aspect of dam location points sites and none sites

Figure 4-21 shows a histogram of aspect for sample site and non-site points for the study area. In the site =1 histogram which represents site presence, it is the direction of slope seems to be equally spread for all farm dam sites. However 0°– 90° (north east facing slope) and 100° – 200° (south east facing slope) show slight dominance. The two subgroups were considered for the model.

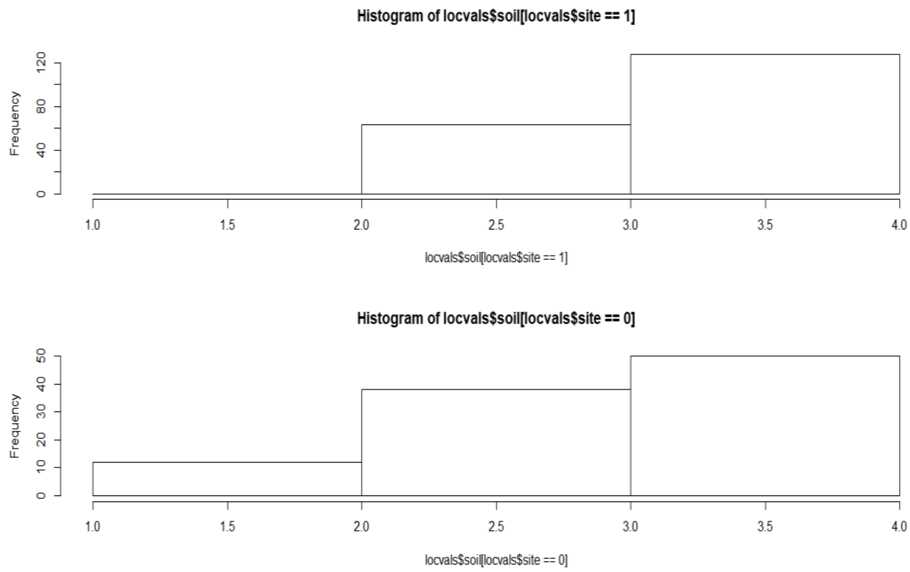

Figure 4-22. Histograms of the soil of dam location points sites and none sites

Figure 4-22 shows a histogram of soils for sample site and non-site points for the study area. The two soils that had the highest presence of farm dam sites were selected and considered for further assessment before fitting in the multivariate model. The two soils are Fr20 and Lo43.

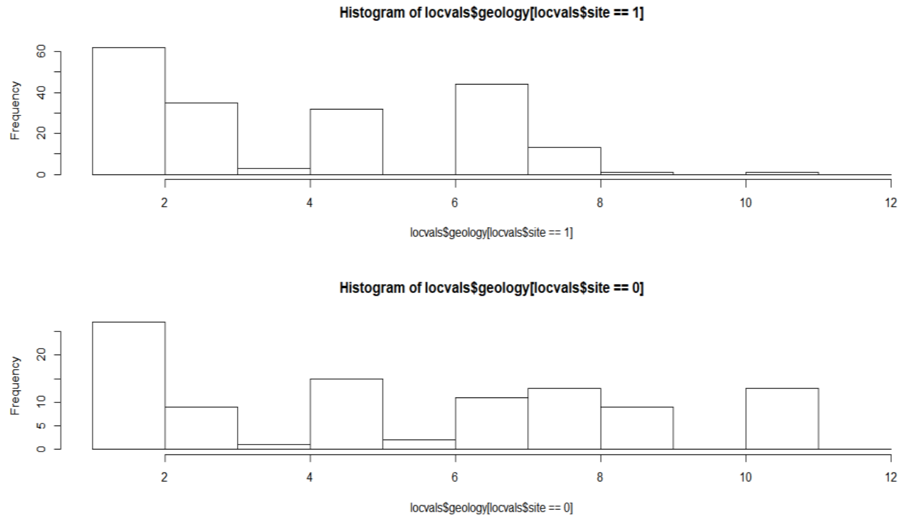

Figure 4-23. Histograms of the geology of dam location points sites and none sites

Selection of subgroups from the covariates was done. The subgroups are presented in Table 4-14

The histogram in Figure 4-23 shows sample site and non-site points for the study area with respect to geology. The rock types that had the largest proportion of farm dams per area were selected and assessed before being fitted into the multivariate regression model. Four rocks were selected and these are Adelaide, Karoo, Pietermaritzburg and Volksrust.

Table 4-14: Subgroups of covariates created to improve the model

Covariate	Sub group	Sub group range
Elevation	Low elevation model	0 – 700 m
	Medium elevation model	700 – 1 600 m
	High elevation model	1 600 – 2 100 m
Aspect	South East facing model	100° - 280°
	North East facing model	0° - 90°
Slope	Gentle model	0° – 5 °
	Steep model	5° – 15°
	Very steep model	15° – 88°
Soil	Ferralsol model	
	Luvisol model	
Geology	Adelaide model	
	Karoo model	
	Pietermaritzburg model	
	Volksrust model	
Land use	Grassland model	
	Agricultural land (cropped) model	

The subgroups were created by observing the variables that had high numbers of presence of sites from the histograms and expected to improve the model. A summary of the subgroups in Table 4-14 indicates upon creation of the subgroups. The variables underwent univariate analysis again.

4.5.2 Examination of sub groups (dummy variables)

Each of the subgroup's relationship with the dependent variable was tested to establish if the variables could be of use in the prediction of the dependent variable. Variables that had a significant p value at $\alpha = 0.05$ significance level were considered for the model, and those that were not significant were again discarded. The results of the covariates that had significant outcomes are presented in Tables 4-15.

4.6 Multivariate regression model

Table 4-15: Dummy variables that yielded significant p values in the univariate analysis

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
Gentle slope	2.2514	0.3240	6.948	<0.001 ***
Low elevation	-1.5091	0.3276	-4.607	<0.001 ***
Medium elevation	1.4563	0.3039	4.793	<0.001 ***
Luvisol	0.7089	0.2524	2.809	0.00497 **
Adelaide	0.8191	0.3963	2.067	0.0388 *
Volksrust	0.8845	0.3629	2.438	0.014789 *
Plantation	-0.9266	0.2567	-3.610	0.000306 ***
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1

The predictors in Table 4-15 were used in the multivariate regression model and many combinations were produced and measured by the AIC. The model with the smallest AIC value was considered the best. The step AIC function in R Studio suggested that the best combination of covariates to include in the final model are gentle slope, medium elevation, land use and geology. The rest of the variables had model combinations that had relatively higher AIC values than 285.56.

Table 4-16: Covariates included in the final model

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-2.3714	0.5045	-4.700	<0.001 ***
Gentle slope	2.2440	0.3422	6.557	<0.001 ***
Medium elevation	1.3087	0.3550	3.686	<0.001 ***
Landuse	0.3206	0.1106	2.898	0.003761 **
Luvisol	0.5991	0.3022	1.982	0.047433 *
Signif. codes:	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1
AIC				293.42

From the covariates the following expression was created

$$\begin{aligned} P\ (location|farm\ dams) \\ &= -2.3714 + (gentle\ slope * 2.2440) + \ (medium\ elevation * 1.3087) \\ &+ (land\ use * 0.3206) + (luvisol * \ 0.5991) \end{aligned}$$

The prediction surface was created using the raster map algebra and coefficients of the logistic regression.

4.6.1 Model performance

A total of 291 points were loaded in the model, and 191 points were sites while 100 were non-sites. Of the sites, 155 were found in areas with probabilities of 0.5 (set cut-off for site presence or absence) and above on the prediction surface while 36 sites had probabilities less than 0.5. Of the 100 none-sites 15% had probabilities of 0.5 or higher while the other 85% were less than 0.5. The best way to test the model is to use it to predict sites and non-sites in another catchment. Due to time constraints, the model was tested using other farm dam points not used in building the model for validation and national dams in the catchment. The validation process used 500 other sites to test the model and 87% of the points were found lying in areas with a probability greater 0.5.

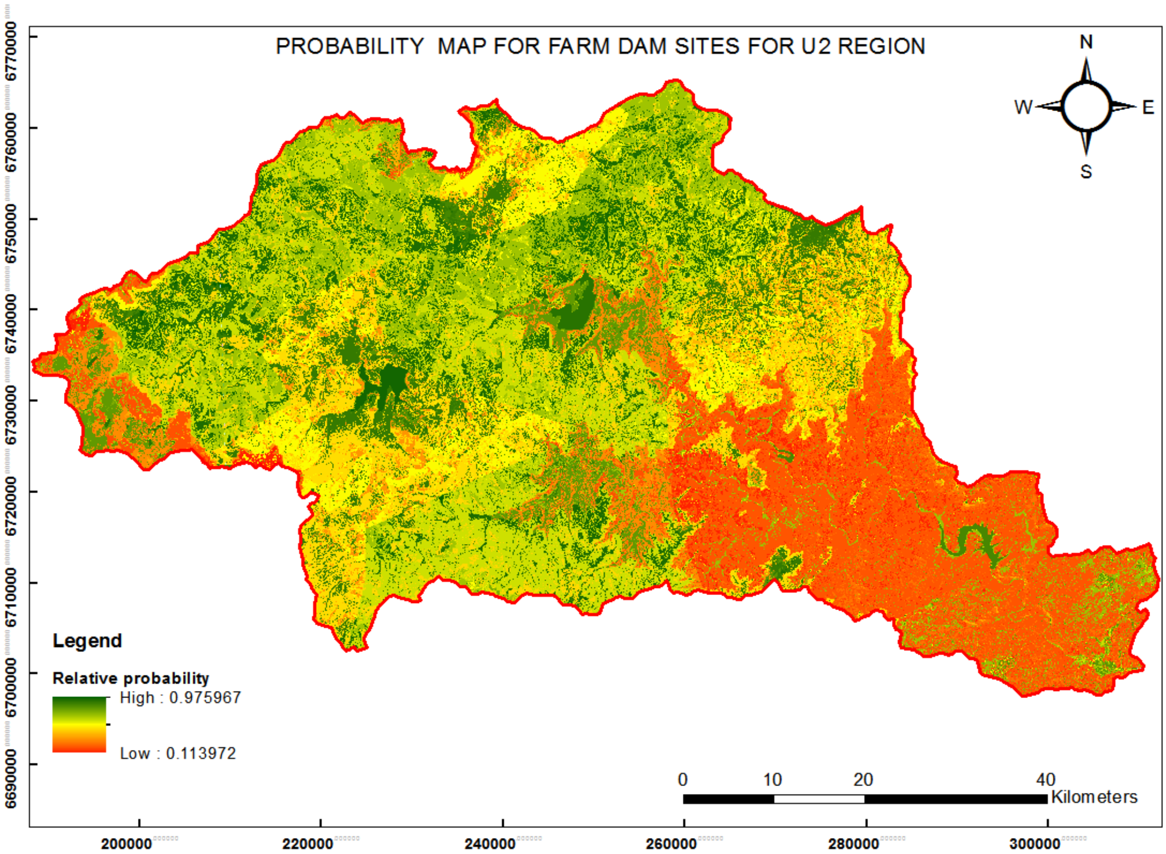

Figure 4-24. Probability map for farm dam site locations in region U2

Figure 4-24 presents a probability map for farm dam site locations for region U2. The green colour shows areas with the highest likelihood to be suitable sites for locating farm dams. The predictive map is a product of the combination of important variables in the model. The logistic output which is the default format, assumes that the sampling design is such that the typical presence localities have probability of presence of 0.5 and above. The output probability map gives values on a continuous scale ranging between 0 and 1, where pixels with a value of 1 are deemed to have highly suitable conditions for dam site location and thus represent a 50% probability of presence. Absence may still be due to many factors not present in the model, socio-economic conditions being one of them.

5. DISCUSSION

5.1 Overview

This chapter presents a discussion on the spatial distribution of farm dams in the U2 region. The study was based on the use of GIS and remote sensing techniques for farm dam data acquisition, processing and analysis. The findings of the study indicate clustering of farm dams because of good soils for agricultural crop production, favorable slope conditions for siting dams, and other physical factors that will be discussed in this section.

According to Stephens (2010), relying on stream flow when temperatures and evaporation are high is unrealistic and risky. It is important to construct farm dams on streams and rivers to ensure off-season storage of vital water supplies. The farm dams can be primarily used for irrigation; however, they can also be used for domestic purposes, fish farming, livestock watering, ground water recharge and conservation storage. South Africa is a semi-arid country with limited water resources. The increased demand of the water resource from the population and economic growth has mounted pressure to the limited resource (Letty et al., 2012).

5.2 Mapping farm dam distribution in the study area using GIS and Remote Sensing techniques

The objective was achieved by mapping the farm dam data acquired from Landsat imagery, and google earth. The classified Landsat 8 image for the month of April 2017 was used to identify all water bodies and thereafter google earth was employed for differentiating farm dams from other water bodies. There were approximately 2000 water bodies that were identified by the classification. These included large national dams, pools in golf courses, ponds and disused mine dumps. A total of 864 farm dams in the U2 region quaternary catchment were observed.

There are no known numbers of farm dams in the U2 region provided by literature. There are many unregistered farm dams in the region and this presents a challenge of knowing the exact quantities of the farm dams and this lack of information has negative impacts on water resources planning in the catchment. However, Letty et al. (2012) estimates that between 65 – 808 million m^3 are stored in the upper north western part and between 2 – 65million m^3 in the central and north eastern parts of U2. Employing medium resolution Landsat 8 satellite imagery and google earth for mapping farm dam distributions proved to be a time saving cost effective and reliable exercise. This compares well with the experiences of Liebe et al. (2005); Sawunyama et al. (2006); and Meigh (1995). They all used medium resolution Landsat satellite imagery to map farm dam distributions in their respective study areas.

The distribution of land use and land cover as shown in Figures 4-1 and 4-16 indicate that agricultural farms are largely located in the north-western parts spreading in an easterly direction towards the central areas and in the north-east of the study area. There are also relatively more water bodies in the areas where agricultural land is found since water is required for crop and livestock production. Moreover, the dominant soil types in the areas with most agricultural land are Ferralsols, which are iron-rich soils with high organic matter content, and are better suited for low input agriculture (Beinroth et al., 1996), and Orthic Luvisols, which have a clay fraction and moderately weathered, most favoured by smallholder rural farmers because of their ease of cultivation and no great impediments (Stocking and Murnaghan, 2001). There are tracts of land that have bare soil mainly in the northwest, west and southern parts of the study area. This may be due to overgrazing given the region practices livestock production. According to DAFF (2011), agriculture is a major economic and social force in the KZN province. The province is home to a large number of South Africa's small-scale farmers.

There are large built up areas in the central and eastern parts of the study area due to the presence of towns and major cities such as Howick, Pietermaritzburg and Durban. Plantations are prevalent mainly in the northern parts of the study area and are established on Orthic Luvisols as these soils are well drained and allow for the deep penetration of tree roots. There are also a number of farm dams in the areas where the plantations are sited. As for grass cover, it is found in all areas and some of the land that is covered by grass lies in farming areas and serves as pastureland for livestock. There are a number of farm dams in areas where there is grass cover.

5.3 Spatial distribution pattern of farm dams in the study area

The objective of determining the distribution of farm dams in the study area was achieved by using ArcGIS and R-Studio. The distribution of farm dams in the catchment is skewed evident from the histograms and maps in Figures 4-3 to 4-17. Figures 4-3 and 4-4 show a bull's eye effect in the in the north-western part that spreads towards the central parts of the study area signaling relatively higher density and intensity of farm dams. There are many factors that drive the cluster pattern in the area. The main reason is the land use (Crop production and livestock) of the area influenced mainly by soil type. The soils in the places where farm dams are located have many farm dams. There are large scale farms and communal farming settlements in the area, practicing both crop production and livestock production. Most of the farm dams are in farming areas with grasslands evident from Figures 4-16 and 4-17. These areas serve as pastureland for livestock. Furthermore, the dams may have been constructed in areas with grasslands because they are effective at intercepting runoff (Letty et al., 2012).

With respect to geology, the geology map for the U2 region in Figure 4-10 and Table 4-3 show that approximately 27% of U2 region farm dams were constructed on the Karoo dolerite rock formation which is a good rock material for dams. Although the Karoo rock formation covers the second largest area (606 km^2), it has a density of 38 dams/100 km^2 which is the highest. Karoo rock formations provide suitable sites as they provide a firm and foundation for construction of the dam. The soils (Luvisols with a clay fraction) that are produced from them through weathering are also suitable for agricultural crop production. Adelaide and Volksrust rank second and third, and account for 23% and 16% of the total number of dams in U2 region respectively. Adelaide has a density of 35 dams/100 km^2 while Volksrust has 31 dams/100 km^2. Pietermaritzburg rock formation, which is mainly shales (sedimentary), covers the largest proportion of the study area; however it ranks sixth in the density of farm dams with 16 dams/100 km^2 though shales are reasonably strong rocks. Vryheid and Mapumulo subgroups account for 1.2% and 0.6% of U2 region farm dams. They have very low farm dam densities of 5 and 2 dams/100 km^2 respectively. Oribi gorge granitoid and Maputaland rock formations have farm dam densities of 0 each.

Other factors that affect the distribution of farm dams in the U2 region are elevation and slope. The elevation map for the U2 region in Figure 11 and the histogram in Figure 4-12 illustrate the distribution of farm dams constructed with respect to elevation. Most farm dams were constructed on elevations of between 1 200 m asl and 1 500 m asl closely followed by those between 900 m and 1200 m found mostly in the north western and central parts of the U2 sub catchment where there is probably gentler slopes and relatively more feeder streams compared to the upper part of the catchment. The areas in the medium elevation range (900 m–1500 m) mostly have undulating topography and a significantly larger micro catchment which according to Roohi and Webb (2012) present good farm dam sites. There is also a large hectarage of cropped land and livestock farms that require water for irrigation and stock watering. The higher areas mainly found in the west close to the catchment boundary have steeper slopes of up to 20° as evident from Figure 4-13, presenting fewer suitable sites. Steep slopes are not suitable for farm dams as there is a high risk of siltation from eroded soil material in a short time period. Moreover, the high lying areas have a small catchment to harvest the water from. As such the dams may end up having poor yields. The low lying areas thus areas towards the east coast, on the other hand, have relatively flatter slopes than in the north western and central areas, and the main rivers and streams are much bigger than upstream. There are very few dams in the low lying areas probably because they do not need as much storage since the rivers are largest in this part of the catchment and perhaps because there is also relatively less agricultural production in the area.

Shao et al. (2012) highlighted that a 3° slope provided sites with good storage area for dams while anything less or more provided smaller capacity. Dams that were on flat land (close to 0° slope) were much smaller than those in the slope range of 3°. Approximately 70% of the larger farm dams, greater than 6 ha area are found in areas of slope close to 3°. From Figures 4-8 and 4-9, it can be seen that 90% of the farm dams are less than 6 ha in area. This may mean that they are small farm dams owned by individual smallholder farmers and are found on slopes less than 3° or much greater. Areas where the slope range is 0°–5° slope have 50% of the farm dams. Approximately 42% of the farm dams lie in areas with a moderately steep slope of range of >5° - 10°. The slope range of between 10°– 20° accounts for 8% of the total number of farm dams.

5.4 Important variables influencing the spatial distribution of farm dams

The correlation matrix in Table 4-5 shows that the six variables are poorly correlated. They have very little association between them. The only associations that can be picked from the Table are between geology and three other variables which are soil, elevation and land use. The associations that geology has with the three variables are however slight. Geology has a negative correlation with elevation (-0.3061), while it is positive correlations with soil (-0.3155) and land use (0.3376).

The Eigenvalues of the correlation matrix in Table 4-6 indicate that 2 principal components explain most of the variance. The scree plot in Figure 4-18 however shows that between 2 and 3 components explain most of the variance. Although the Kaiser criterion suggests selection of Eigen values above 1, the difference between the second and third component is small so the third component was selected. The three components account for 63% of the variance.

Table 4-7 shows Eigenvectors that were analysed to see which factors are of importance. The first principal component is correlated with four of the original variables. The first principal component increases with increasing land use (0.502227) and geology (0.581806) scores, and increases with decreasing elevation (-0.420332) and soil (-0.426586) scores. This suggests that geology and land use vary together meaning that if one increases the other increases. Elevation and soil also vary together, thus if one decreases, the other also decreases. If one increases, then the other one tends to increase as well. The first component is a measure of the physical attributes for the farm dam sites since the first eigenvector shows reasonably high loadings for most variables except aspect.

The second principal component increases with only two of the values, thus, increasing slope (0.568963) and aspect (0.790195) scores. This component can be viewed as a measure of flow direction of water. The third principal component increases with increasing soil (0.628839) and aspect (0.316621) scores and decreasing elevation (-0.499534) and slope (-0.459896). Generally, the outcome of the Principal Component Analysis (PCA) analysis shows the variables that are important in the predictive model for farm dam sites. Component one which measures of farm dam site location loads highly for soil, elevation, land use and geology which are the most important variables; the most important being soil.

5.5 Spatial predictive model for farm dam sites

The spatial predictive model was created using the binary multivariate logistic regression model. Table 4-16 shows the predictor variables that are important and were included in the final model. Of the six variables that were analysed, geology and aspect were dropped and not fitted in the model since their contributions was not important. It has to be noted however that, variables such as Luvisols (soil), land use, gentle slope and medium elevation were added to the model as they exhibited very strong significance when fitted in the final model as shown in Table 4-16.

The best way to test the model is to use it to predict sites and non-sites in another catchment. Due to time constraints, the model was tested using other farm dam points not used in building the model for validation. The validation process used 500 other sites to test the model and 86% of the points were found lying in areas with a probability greater 0.5.

According to Mc Donald (2009), statisticians have varying opinions on the best measure of fit for multiple logistic regressions. Deviance, D, is employed by some, for which smaller numbers represent better fit, and while others work with one of several pseudo-R^2 values, for which larger numbers represent better fit. The best model yielded an AIC value of 293.42 which was the smallest value obtained for all model fits.

The model is generally good at predicting suitable farm dam sites. Based on the model output, some farm dams were constructed on places that are not suitable. The need for sustainable crop and livestock production in the region could have driven farmers to construct farm dams in areas that may not have been ideal sites. Other than the physical factors, there may be other socio-economic factors that could have driven the farmers to construct the dams in the sites that are according to the model unsuitable with regards to the physical attributes. Factors such as dam cost, size of livestock units, ownership (communal or single) may have also contributed to the distribution of locations of the dams that are in areas of low probability. These socio-economic aspects are not considered in the model and that is one of the model's limitations.

5.6 Limitations of the study

Although the study yielded favorable findings and was completed in good time, there were challenges that were faced in the process. The challenges include:

- Lack of time to carry out validation of the model in another quaternary catchment.
- Inadequate access to different aerial photos with full coverage of the study area limited the visual interpretation and classification processes.
- Only physical factors affecting the distribution of farm dams were considered in the study and this left a relatively large variance score that was unexplained.
- There were no field measurements made to verify certain issues.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

From the inventory of farm dams and maps of the spatial distribution of farm dams for the study area, GIS and remote Sensing techniques are valuable tools for detecting, delineating and measuring the surface areas (estimating capacity) of small farm dams. The mapping of the farm dams will also help in giving better estimates of the amount of water storage the study area has, and this helps to better understand the water balance of the area and in turn help in effective water management planning.

It emerged from the study that soils drive land use and other factors such as slope, elevation and geology are critical influencing the distribution of farm dams in the U2 region. Land use in areas with farm dams was influenced by soils. Where good soils for crop production were found, farm dams were also found. Where there were grasslands with good palatable grass for livestock, farm dams were also found. Most of the farm dams in the study area are meant for livestock watering.

With respect to storage capacity, factors such as slope are also important as they determine the surface area that can be covered by a farm dam. Flat surfaces generally had smaller storage than areas with a slope angle of approximately 3°.

The predictive model produced in this study provides a tool for identification of suitable sites for locating farm dams based on physical factors. This enables planners to have a clear picture of the outlook of the catchment with respect to farm dam development in the future. They will know where to look for suitable sites which will effectively capture / intercept runoff.

The model also shows that some farm dams are located on sites that are not suitable from a physical attributes perspective.

Generally, from what was observed in the study, the farm dams in the U2 region have a very important role and have a significant effect in water resources planning and management in U2.

6.2 Recommendations

- Carrying out field work in order to validate the predictive model and also testing it in another catchment area. The model accuracy can also be improved by applying it on dams of the same size range in making predictions. When the farm dams are mixed some, thus the large ones and the extremely small ones, there may be disparities in the model leading to some parameters being non-significant to the model;
- There is also need to come up with a comprehensive model that incorporates physical and socio-economic factors; and
- There is need for further on assessment of evaporation rates and sediment yield in the farm dams as this has significant effect on the overall storage system.
- Section 21(b) of the NWA defines storage of water as a water use. Generally, if more than 10 000 m^3 of water is stored per registered property, then that water use must be registered on a form that will be supplied by the Regional Director for this purpose. The water act also requires users drawing than 50 m^3 of water from a surface water resource per day on average over a year on a property or piece of land in terms of this authorisation must register the water use with the responsible authority. Further work should be done to establish the farm dams that are registered and those that are not.
- Generally, for Thukela Water Management Area (WMA), the charge for agricultural water is 0.33 c/m^3. In order to establish how much water each farm requires per annum, there is need for future studies to collect farm dam areas either through questionnaires or by making measurements on the ground. The blanket amount required for 1 ha is between 10 000 m^3 – 12 000 m^3.
- There is also need for future studies to carry out field measurements for a sample of dams in the catchment in order to collect information on parameters for computation of farm dam capacities, which can be used for the theoretical derivation of the area – volume relationships. The area-volume relationships will then be used to compute volumes of all farm dams in the catchments.

Reference list

Abdi, H. and Williams, L.J., 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), pp.433-459.

Agrawala, S., 1998. Structural and process history of the Intergovernmental Panel on Climate Change. *Climatic change*, *39*(4), pp.621-642.

Adam, E., Mutanga, O., Odindi, J. and Abdel-Rahman, E.M., 2014. Land-use/cover classification in a heterogeneous coastal landscape using Rapid-Eye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, *35*(10), pp.3440-3458.

Akasapu, A.K., Rao, P.S., Sharma, L.K. and Satpathy, S.K., 2011. Density based k-nearest neighbors clustering algorithm for trajectory data. *International Journal of Advanced Science and Technology*, *31*(1).

Anderson, J.R., Roach, J.T and Witmer., R.E. (1976) *A Land Use and Land Cover Classification System For Use With Remote Sensor Data, Geological Survey Professional Paper,* Washington: US Government Printing Office.

Annor, F.O., Van De Giesen, N., Liebe, J., Van de Zaag, P., Tilmant, A. and Odai, S.N., 2009. Delineation of small reservoirs using radar imagery in a semi-arid environment: A case study in the upper east region of Ghana. *Physics and Chemistry of the Earth, Parts A/B/C*, *34*(4), pp.309-315.

Baleta, H. and Pegram, G., 2014. Water as an input in the food value chain. *Understanding the Food Energy Water Nexus.* WWF-SA, South Africa.

Bainbridge, W.R., Scott D.R. and Walker R. 1986. *Policy Statement for the Drakensberg State Forests*. Pietermaritzburg. Department of Environment Affairs.

Bainbridge, W.R., 2004. *Initiatives to conserve globally significant biodiversity resources in the Maloti-Drakensberg Transfrontier Conservation and Development Area, Lesotho and South Africa*. In: David Harmon and Graeme L. Worboys (eds.) Managing Mountain Protected Areas: Challenges and Responses for the 21st Century. Andromeda Editrice. 426 pp.

Beasely, D. B., Huggins, L. F., and Monke, E. J. 1980. ANSWERS: A model for watershed planning.Transactions of the American Society of Agricultural Engineers, 23(4), 938-944.

Beinroth, F.H., Eswaran, H., Palmieri, F. and Reich, P.F., 1996. *Properties, classification and management of oxisols*. USDA-National Resources Conservation Service-International Programs Division/University of Puerto Rico-Department of Agronomy and Soils/EMBRAPA-Centro Nacional de Pesquisa de Solos.

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, pp.24-31.

Berman, M. and Diggle, P. (1989) Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society, series B* **51**, 81–92.

Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, *24*(1), pp.43-69.

Bhagat, V.S. and Sonawane, K.R., 2011. Use of Landsat ETM+ data for delineation of water bodies in hilly zones. *Journal of Hydroinformatics*, *13*(4), pp.661-671.

Biswas, A.K., Tortajada, C. and Izquierdo-Avino, R. eds., 2009. *Water management in 2020 and beyond*. Berlin: Springer.

Boardman, J., Foster, I., Rowntree, K., Mighall, T. and Parsons, T., 2009. Small farm dams: a ticking time bomb?: farm dams. *Water Wheel*, *8*(4), pp.30-35.

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Bressers, H. and Lulofs, K. eds., 2010. *Governance and complexity in water management: Creating cooperation through boundary spanning strategies*. Cheltenham, United Kingdom. Edward Elgar Publishing.

Brouwer, C., Prins, K., Kay, M. and Heibloem, M., 1988. Irrigation water management: irrigation methods. *Training manual*, *9*.

Burrough, P.A., MCDONNELL. 1998. Principles of geographical information systems. *Economic Geography*, *75*(4), p.422.

Callow, J.N. and Smettem, K.R.J., 2009. The effect of farm dams and constructed banks on hydrologic connectivity and runoff estimation in agricultural landscapes. *Environmental Modelling & Software*, *24*(8), pp.959-968.

Cengiz, T. and Akbulak, C., 2009. Application of analytical hierarchy process and geographic information systems in land-use suitability evaluation: a case study of Dümrek village (Çanakkale, Turkey). *International Journal of Sustainable Development & World Ecology*, *16*(4), pp.286-294.

Chan, J.C.W. and Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*(6), pp.2999-3011.

Clark, P.J. and Evans, F.C. (1954) *Distance to nearest neighbour as a measure of spatial relationships in populations*. Ecology 35, 445-453.

Congalton, R.G. and Green, K., 2008. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.

Crema, E.R., Bevan, A. and Lake, M.W., 2010. A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record. *Journal of Archaeological Science*, *37*(5), pp.1118-1130.

DAFF (Department of Agriculture, Forestry and Fisheries), 2011. Trends in agricultural sector.

Dare, P., Fraser, C. and Duthie, T., 2002. Application of automated remote sensing techniques to dam counting. *Australian Journal of Water Resources*, *5*(2), pp.195-208

De Winnaar, G., Jewitt, G.P.W. and Horan, M., 2007. A GIS-based approach for identifying potential runoff harvesting sites in the Thukela River basin, South Africa. *Physics and Chemistry of the Earth, Parts A/B/C*, *32*(15), pp.1058-1067.

Diggle, P.J. (1985) A kernel method for smoothing point process data. *Applied Statistics* (Journal of the Royal Statistical Society, Series C) **34** (1985) 138–147.

Dorn, H., Tornros, T and Zipf, A. (2015) Quality Evaluation of VGI Using Authoritative Data: A Comparison with Land Use Data in Southern Germany, *International Journal of Geo-Information*, vol. 4, no. 3, pp. 1657-1671.

DSE, 2007. Your dam your responsibility, A guide to managing the safety of farm dams. Irrigation and Commercial Farm Dams, License to water. Department of Sustainability and Environment, Victoria, pp. 90.

ESRI, ArcGIS Network Analyst. "Redlands." *California (www. esri. com/software/arcgis/extensions/networkanalyst/index. html)* (2013).

Explorer, U.E., 2017. Earth Explorer. *Available from Internet: http://edcsns17. cr. usgs. gov/NewEarthExplorer*, *122*.

Ezemvelo KZN Wildlife, 2005. Integrated Management Plan: Ukhahlamba Drakensberg Park World Heritage Site (2006-2011), South Africa. Ezemvelo KZN Wildlife, Pietermaritzburg, pp 81.

Fey, M., 2010. *Soils of South Africa*. Cambridge University Press.

Fonji, S.F and Taff, G.N. (2014) Using Satellite Data to Monitor Land-use Land-cover Change in North-Eastern Latvia, *SpringerPlus*, 3:61, doi:10.1186/2193-1801-3-61.

Forkuor, G. and Maathuis, B., 2012. Comparison of SRTM and ASTER derived digital elevation models over two regions in Ghana-Implications for hydrological and environmental modeling. In *Studies on Environmental and Applied Geomorphology*. In Tech, pp 219 – 240

Fortin, J. P., Turcotte, R., Massicotte, S., Moussa, R., Fitzback, J., and Villeneuve, J. P. 2001. A distributed watershed model compatible with remote sensing and GIS Data, I: Description of model. Journal of Hydrologic Engineering, 6(2), 91-99. [doi:10.1061/(ASCE)1084-0699(2001)6:2(91)]

Global weather, 2012. NCEP Climate Forecast System Reanalysis (CFRS). http://globalweather.tamu.edu/, accessed September 2017.

Grabarnik, P. and Chiu, S.N., 2002. Goodness‐of‐fit test for complete spatial randomness against mixtures of regular and clustered spatial point processes. *Biometrika*, *89*(2), pp.411-421.

Habets, F., Philippe, E., Martin, E., David, C.H. and Leseur, F., 2014. Small farm dams: impact on river flows and sustainability in a context of climate change. *Hydrology and Earth System Sciences*, *18*(10), p.4207.

Hendriks, S., 2014. Food security in South Africa: Status quo and policy imperatives. *Agrekon*, *53*(2), pp.1-24.

Hosseinzadeh, S. R. 2011. Drainage Network Analysis, Comprise of Digital Elevation Models(DEMs) from ASTER and Remote sensing data. International Journal of Environmental Science and Development, Volume 2, Issue 3.

Huang, C., Homer, C. and Yang, L., 2003. Regional forest land cover characterization using medium spatial resolution satellite data. In *Remote Sensing of Forest Environments* (pp. 389-410). Springer US.

Hughes, D.A. and Mantel, S.K., 2010. Estimating the uncertainty in simulating the impacts of small farm dams on streamflow regimes in South Africa. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, *55*(4), pp.578-592.

Kahinda, J.M., Taigbenu, A.E., Sejamoholo, B.B.P., Lillie, E.S.B. and Boroto, R.J., 2009. A GIS-based decision support system for rainwater harvesting (RHADESS). *Physics and Chemistry of the Earth, Parts A/B/C*, *34*(13), pp.767-775.

Kaufman, Y.   J., Wald, A.   E., Remer, L.   A., Gao, B.-C., Li, R.-R. and Flynn, L. 1997. The MODIS 2.1 µm channel—correlation with visible reflectance for use in remote sensing of aerosol.. *IEEE Transactions on Geoscience and Remote Sensing*, 35: 1286–1298.

Kemp, M.U., Emiel van Loon, E., Shamoun-Baranes, J. and Bouten, W., 2015. RNCEP: global weather and climate data at your fingertips. *Methods in Ecology and Evolution*, *3*(1), pp.65-70.

Kiage, L.M., Liu, K.B., Walker, N.D., Lam, N. and Huh, O.K., 2007. Recent land-cover/use change associated with land degradation in the Lake Baringo catchment, Kenya, East Africa: evidence from Landsat TM and ETM+. *International Journal of Remote Sensing*, *28*(19), pp.4285-4309.

Knorn, J., Rabe, A., Radeloff, V.C., Kuemmerle, T., Kozak, J. and Hostert, P. (2009) Land cover mapping of large areas using chain classification of neighboring Landsat satellite Images, *Remote Sensing of Environment*, vol. 113, no. 1, pp. 957-964.

Knöpfle, W., Strunz, G. and Roth, A., 1998. Mosaicking of digital elevation models derived by SAR interferometry. *International Archives of Photogrammetry and Remote Sensing*, *32*, pp.306-313.

Letty, B., Shezi, Z. and Mudhara, M., 2012. Grassroots innovation as a mechanism for smallholder development in South Africa: can impact be measured?. *African Journal of Science, Technology, Innovation and Development*, *4*(3), pp.32-60.

Liang, S. L., Fang, H., Chen, M., 2001, Atmospheric Correction of Landsat ETM+ Land Surface Imagery—Part I: Methods, TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING,Vol, 39, No 11, pp 2490 – 2498.

Liu, D. And Cai, S. (2012) A Spatio-Temporal Modelling Approach to Reconstructing Land-Cover Change Trajectories from Multi-Temporal Satellite Imagery, *Annals of the Association of American Geographers*, vol. 102, no. 6, pp. 1329-1347

Liebe, J., 2002. ‘Estimation of water storage capacity and evaporation losses of small reservoirs in the Upper East Region of Ghana’. Diploma Thesis. *University of Bonn, Germany*.

Liebe, J., Van De Giesen, N. and Andreini, M., 2005. Estimation of small reservoir storage capacities in a semi-arid environment: A case study in the Upper East Region of Ghana. *Physics and Chemistry of the Earth, Parts A/B/C*, *30*(6), pp.448-454.

Lu, D., Mausel, P., Brondizios, E. and Moran, E. (2004) Change detection techniques, *International Journal of Remote Sensing*, vol. 25, no. 12, pp. 2365-2407.

Maaren, H. and Moolman, J. 1986. The effects of farm dams on hydrology. In: Second South African National Hydrology Symposium (Pietermaritzburg, 16–18 September 1985)

Malczewski, J., 2004. GIS-based land-use suitability analysis: a critical overview. *Progress in Planning*, *62*(1), pp.3-65.

Mantel, S.K., Hughes, D.A. and Muller, N.W., 2010. Ecological impacts of small dams on South African rivers Part 1: drivers of change-water quantity and quality. *Water SA*, *36*(3), pp.351-360.

Mas, J. F. (1999) Monitoring Land-cover Changes: A comparison of Change Detection Techniques, *International Journal of Remote Sensing*, vol. 20, no. 1, pp. 139-152.

Mas, J. F. and Ramirez, I. (1996) Comparison of Land use Classifications obtained by Visual Interpretation and Digital Processing, *ITC Journal,* vol. 3*,* no. 4, pp. 278-283.

Matteau, M., Assani, A.A. and Mesfioui, M., 2009. Application of multivariate statistical analysis methods to the dam hydrologic impact studies. *Journal of Hydrology*, *371*(1), pp.120-128.

Matthew, M.W., Adler-Golden, S.M., Berk, A., Felde, G., Anderson, G.P., Gorodetzky, D., Paswaters, S. and Shippert, M., 2002, October. Atmospheric correction of spectral imagery: evaluation of the FLAASH algorithm with AVIRIS data. In *Applied Imagery Pattern Recognition Workshop, 2002. Proceedings. 31st* (pp. 157-163). IEEE.

McDonald, J.H., 2009. Handbook of biological statistics (Vol. 2, pp. 173-181). Baltimore, MD: Sparky House Publishing.

Meigh, J., 1995. The impact of small farm reservoirs on urban water supplies in Botswana. In *Natural Resources Forum* (Vol. 19, No. 1, pp. 71-83). Blackwell Publishing Ltd. doi:10.1111/j.1477-8947. 1995.tb00594x.

Meissner, H.H., Scholtz, M.M. and Palmer, A.R., 2013. Sustainability of the South African livestock sector towards 2050 Part 1: Worth and impact of the sector. *South African Journal of Animal Science*, *43*(3), pp.282-297.

Mesev, V. (2010) Classification of Urban Areas: Inferring Land Use from the Interpretation of Land Cover in Rashed, T and C. Jurgens., eds., *Remote Sensing and Digital Image Processing: Remote Sensing of Urban and Suburban Areas,* New York: Springer, vol. 10, pp. 144-164.

Nathan, R., Jordan, P. and Morden, R., 2005. Assessing the impact of farm dams on stream flows, Part I: Development of simulation tools. *Australian Journal of Water Resources*, *9*(1), pp.1-12.

Nathan, R. and Lowe, L., 2012. The hydrologic impacts of farm dams. *Australian Journal of Water Resources*, *16*(1), pp.75-83.

Neal, B., Nathan, R.J., Schreider, S. and Jakeman, A.J., 2002. Identifying the separate impact of farm dams and land use changes on catchment yield. *Australian Journal of Water Resources*, *5*(2), pp.165-176.

Nikolakopoulos, K.G. and Chrysoulakis, N., 2006, October. Updating the 1: 50.000 topographic maps using ASTER and SRTM DEM: the case of Athens, Greece. In *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VI* (Vol. 6366, p. 636606). International Society for Optics and Photonics.

Norman, N. and Whitfield, G., 2006. Geological Journeys. Struik. 320 pp.

Oberholster, P.J. and Ashton, P.J., 2008. State of the nation report: An overview of the current status of water quality and eutrophication in South African rivers and reservoirs. *Parliamentary Grant Deliverable. Pretoria: Council for Scientific and Industrial Research (CSIR)*, pp1.

Ouerghi, S., ELsheikh, R.F.A., Achour, H. and Bouazi, S., 2015. Evaluation and Validation of Recent Freely-Available ASTER-GDEM V. 2, SRTM V. 4.1 and the DEM Derived from Topographical Map over SW Grombalia (Test Area) in North East of Tunisia. *Journal of Geographic Information System*, *7*(03), p.266.

Pramanik, M.K., 2016. Site suitability analysis for agricultural land use of Darjeeling district using AHP and GIS techniques. *Modeling Earth Systems and Environment*, *2*(2), p.56.

Pollard, S. and Du Toit, D., 2008. Integrated water resource management in complex systems: How the catchment management strategies seek to achieve sustainability and equity in water resources in South Africa. *Water SA*, *34*(6), pp.671-679.

Reshmidevi, T.V., Eldho, T.I. and Jana, R., 2009. A GIS-integrated fuzzy rule-based inference system for land suitability evaluation in agricultural watersheds. *Agricultural systems*, *101*(1), pp.101-109.

Richards, J.A. (2012) Remote Sensing Digital Image Analysis: An Introduction, Australia: Springer Science & Business Media

Rodrigues, L.N., Sano, E.E., Steenhuis, T.S. and Passo, D.P., 2012. Estimation of small reservoir storage capacities with remote sensing in the Brazilian Savannah Region. *Water Resources Management*, *26*(4), pp.873-882.

Roohi, R. and Webb, J., 2012. 'Landsat Image Based Temporal and Spatial Analysis of Farm Dams in Western Victoria'. MSc in Geospatial Science Thesis. RMIT University, Melbourne, Australia.

Saaty, T.L., 1988. What is the analytic hierarchy process? In *Mathematical models for decision support* (pp. 109-121). Berlin, Heidelberg: Springer.

Saaty, T.L., 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), pp.83-98. doi.org/10.1504/IJSSci.2008.01759

Savadamuthu, K., 2002. *Impact of farm dams on streamflow in the Upper Marne Catchment*. Resource Assessment Division, Department for Water Resources.

Sawunyama, T., Senzanje, A. and Mhizha, A., 2006. Estimation of small reservoir storage capacities in Limpopo River Basin using geographical information systems (GIS) and remotely sensed surface areas: Case of Mzingwane catchment. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(15), pp.935-943.

Schreider, S.Y., Jakeman, A.J., Letcher, R.A., Nathan, R.J., Neal, B.P. and Beavis, S.G., 2002. Detecting changes in streamflow response to changes in non-climatic catchment conditions: farm dam development in the Murray–Darling basin, Australia. *Journal of Hydrology*, *262*(1-4), pp.84-98.

Shao, Q., Chan, C., Jin, H. and Barry, S., 2012. Statistical Justification of Hillside Farm Dam Distribution in Eastern Australia. *Water Resources Management*, *26*(11), pp.3139-3151.

Singh, L.K., Jha, M.K. and Chowdary, V.M., 2017. Multi-criteria analysis and GIS modeling for identifying prospective water harvesting and artificial recharge sites for sustainable water supply. *Journal of Cleaner Production*, *142*, pp.1436-1456.

South Africa. Department of Water Affairs and Forestry (1990-2010), 2004. *National Water Resource Strategy*. Department of Water Affairs and Forestry.

South Africa. Department of Water Affairs and Forestry. 1986. *Management of the Water Resources of the Republic of South Africa.* Department of Water Affairs and Forestry, Pretoria, South Africa.

South Africa. Department of Rural Development and Land Reform. 2013. http://www.**ngi**.gov.za

South African Poultry Association, 2011. SAPA industry profile.

Stephens, T., 2010. *Manual on small earth dams: a guide to siting, design and construction*. Rome: Food and Agriculture Organization of the United Nations (FAO), 64, pp. vii – 144.

Stocking, M. and Murnaghan, N., 2001. Handbook for the field assessment of land degradation. Earthscan.

Troskie, D.P., 2013. Provinces and agricultural development challenge or opportunity. *Agrekon*, *52*(1), pp.1-27

Tingey-Holyoak, J.L., 2014. Water sharing risk in agriculture: Perceptions of farm dam management accountability in Australia. *Agricultural water management*, *145*, pp.123-133.

United States Geological Survey, 2017. Landsat Missions [online], available: http://landsat.usgs.gov/band_designations_landsat_satellites.php

Victoria, A. and Victoria, P., 2002. Department of Sustainability and Environment.

Verstraeten, G. and Prosser, I.P., 2008. Modelling the impact of land-use change and farm dam construction on hillslope sediment delivery to rivers at the regional scale. *Geomorphology*, *98*(3), pp.199-212.

Wold, S., Esbensen, K. and Geladi, P., 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), pp.37-52.

World Commission on Dams. 2000. Dams and Development: A New Framework for Decision-making. London: Earthscan Publications Ltd, pp 167-190.

Yalew, S.G., van Griensven, A., Mul, M.L. and van der Zaag, P., 2016. Land suitability analysis for agriculture in the Abbay basin using remote sensing, GIS and AHP techniques. *Modeling Earth Systems and Environment*, *2*(2), pp.1-14.

Yilmaz, H., Yilmaz, O.Y. and Akyüz, Y.F., 2017. Determining the factors affecting the distribution of Muscari latifolium, an endemic plant of Turkey, and a mapping species distribution model. *Ecology and evolution*, *7*(4), pp.1112-1124.

Zhang, H., Yao, Z., Yang, Q., Li, S., Baartman, J. E.M., Gai, L., Yao, M., Yang, X, Ritsema, C.J. and Geissen, V. 2017. An integrated algorithm to evaluate flow direction and flow accumulation in flat regions of hydrologically corrected DEMs. Catena, Volume 151 pp 174–181.