University of Nevada, Reno

An Evaluation of Spatial Data and Analysis for Identifying Potentially Favorable Areas for Manual Well Drilling: Zinder Region of Niger

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Hydrogeology

by

Sean A. Thomas

Dr. James M. Thomas / Thesis Advisor

August, 2010

University of Nevada, Reno
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

SEAN A. THOMAS

entitled

An Evaluation Of Spatial Data And Analyis For Identifying Potentially Favorable Areas For Manual Well Drilling: Zinder Region Of Niger

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Dr. James M. Thomas, Advisor

Dr. Kenneth McGwire, Committee Member

Dr. Scott Bassett, Graduate School Representative

Marsha H. Read, Ph. D., Associate Dean, Graduate School

August, 2010

Abstract

This thesis evaluated a variety of straightforward spatial data and analysis techniques for identifying potentially favorable areas for manual well drilling in the Zinder region of Niger. A key question was whether environmental variables derived from publically available spatial data had the capacity to augment groundwater depth data for mapping these potentially favorable areas. Some variables considered were: a new calculation of vegetation persistence derived from Moderate Resolution Imaging Spectroradiometer (MODIS) Enhanced Vegetation Index (EVI) data, MODIS night land surface temperature, and lineament properties, topographic convergence index, and landforms derived from the Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM). Regression tree analysis showed that geology and soils were the strongest variables for predicting groundwater depth in the study area. The results indicated and parsimony dictates that a geology map and adequate groundwater data are sufficient to map favorable areas for manual well drilling. However, the regression tree analysis also revealed that the combination of relatively high vegetation persistence and low night land surface temperature were related to shallow groundwater depth and can improve favorability mapping for manual well drilling. Additional research is needed to describe these relationships further. Among the output was a procedural outline for favorability mapping, which uses common hydrogeologic and terrain criteria to differentiate between topography and recharge controlled water tables, to direct the choice of variables used in future mapping efforts. Ultimately, several maps of favorable areas for manual well drilling for the Zinder region were created using geology, groundwater depth, and threshold values of environmental variables from the regression tree analysis.

Dedicated to Lindsey, Anya, and the new one

Jeremiah 17:7-8

Acknowledgements

Dr. James M. Thomas
Dr. Ken McGwire
Dr. Scott Bassett
Alan McKay
Dr. Alexandra Lutz
Christopher Kratt
The Manka Family
The Thomas Family
Jim and Darlene Uptain
Renovation Community
Tim Minor
Jamie Trammel
David McGraw
Mary Ohren
Dr. Chris Shope
Mark Lucas
Hal Voepel
Justin Huntington
Charles Morton
David Healy
Kirsten Danert
Dr. Todd Caldwell
Dr. Kenneth Hardcastle
Alio Aboubacar
Fabio Fussi
The United States Agency for International Development
The United Nations Children's Fund
Practica Foundation
Desert Research Institute
Direction Régionale Hydraulique, Zinder
World Vision Niger
Winrock International

Table of Contents

List of Tables

List of Figures

Abbreviations

ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer

CART – Classification and regression trees

CRU – Climate Research Unit of University of East Anglia

DEM – Digital Elevation Model

DRH – Direction Régionale Hydraulique

DRI – Desert Research Institute

ESRI – Environmental Systems Research Institute

EVI – Enhanced Vegetation Index

EVI_l – Stabilized low dry season EVI, subsequent to EVI_p

EVI_p – Peak wet season EVI

EVID – Enhanced Vegetation Index Difference

GDEM – ASTER Global Digital Elevation Model

GIS – Geographic Information Systems

GWD – Groundwater depth

HDF-EOS – Hierarchical Data Format-Earth Observing System file format

IAS – Iullemeden Aquifer System Project

IGRAC – International Groundwater Resources Assessment Centre

LF – Lineament factor

LP DAAC – Land Process Distributed Active Archive Center

LST – Land surface temperature

MDG – Millennium Development Goal

MODIS – Moderate Resolution Imaging Spectroradiometer

MOD13Q1 – MODIS Terra 16-Day EVI – 250m

MOD13A2 – MODIS Terra 16-Day EVI – 1000m

MRT – MODIS Reprojection Tool

NASA – National Aeronautics and Space Administration

NDVI – Normalized Difference Vegetation Index

NGO – Non-government organization

NLST – Night land surface temperature

RAP – Rapid assessment protocol in support of manual drilling

RS – Remote sensing

R(T) – Redistribution error or total sum of squared error for a regression tree

SDVI – Standard deviation of Normalized Difference Vegetation Index

SIGNER – Systeme d'Information Géographique du Niger

SRTM – Shuttle Radar Topography Mission

SSA – Sub-Saharan Africa

SSE – Sum of squared error

TAS – Terrain Analysis System

TCI – Topographic Convergence Index

TIR – Thermal Infrared

UN – United Nations

UNDP – United Nations Development Programme

UNEP – United Nations Environment Programme

UNICEF – United Nations Children's Fund

USAID – United States Agency for International Development

VPR – Vegetation Persistence Ratio

VPR* - Vegetation Persistence Ratio with optimized offsets

WHO – World Health Organization

WIST – NASA Warehouse Inventory Search Tool

WSSD – World Summit on Sustainable Development

WWF – World Water Forum

1 Introduction

1.1 The state of water access in sub-Saharan Africa

The violation of basic civil liberties in many regions of the world, including the human right to safe, accessible, and affordable water, led the United Nations (UN) General Assembly to craft the Millennium Development Goals (MDGs) in the year 2000. With respect to water, the MDGs include a target to halve the number of people who do not have sustainable access to clean drinking water and sanitation by the year 2015. Current data indicate that sub-Saharan Africa (SSA), the region ranking lowest on the UN Development Programme's (UNDP) human development index, will not achieve this goal (UNDP, 2006; UNDP, 2007; UNICEF and WHO, 2008).

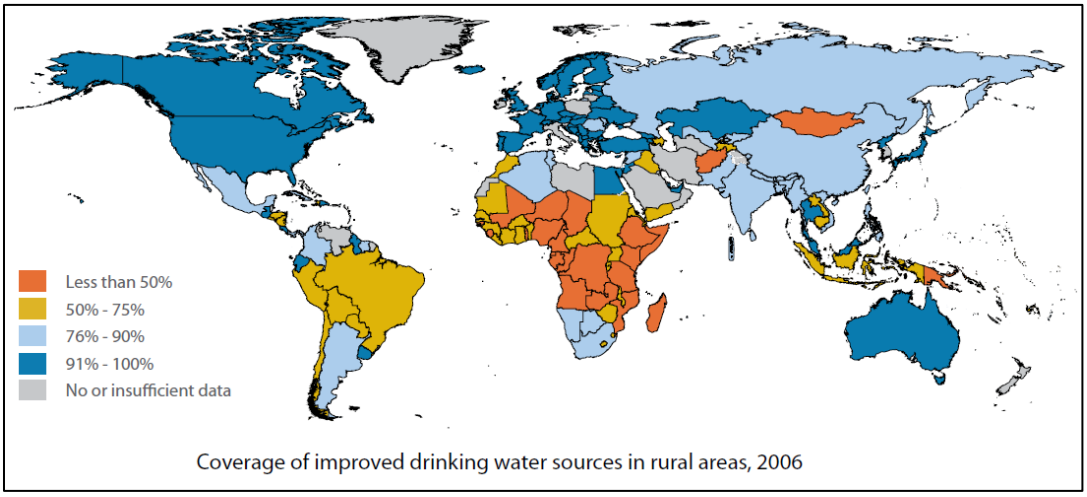


Figure 1: Improved drinking water coverage in rural areas (UNICEF and WHO, 2008).

An estimated 42% of the 2006 population of SSA, approximately 280 million people, require improvements to water supply to reach the MDG, and rural areas face the biggest challenge (UNICEF and WHO, 2008) (Figure 1). MacDonald et al. (2008) state that extensive rural groundwater development is the only sustainable and affordable means of reaching the MDG water target. It has been estimated that 40,300 boreholes will need to be developed

annually between 2006 and 2015 to reach the MDG (Danert, 2009). Also, the uncertainty of climate change and future drought in SSA adds motivation for groundwater development in terms of water and food security (UNDP, 2006).

1.2 UNICEF and manual well drilling in Africa

In response to the need for groundwater development in a cost-effective, sustainable, and widespread manner in SSA, UNICEF recently began a program to promote manual well drilling in the region (UNICEF, 2009). Manual drilling techniques can be applied to a variety of hydrogeologic conditions and provide viable alternatives to more costly, mechanically drilled wells (Danert, 2009; Elson and Shaw, 1995) (Figure 2).

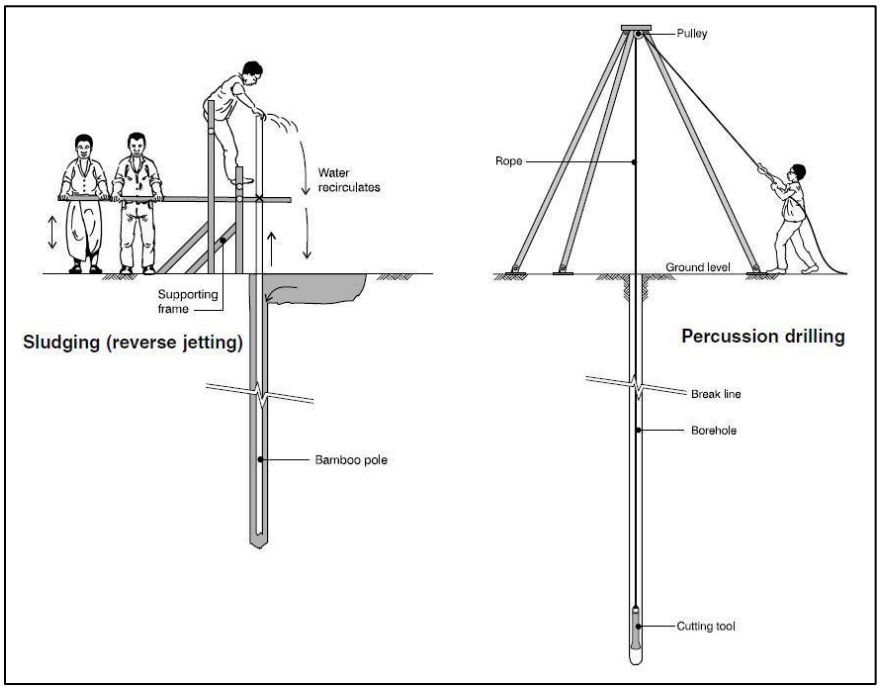


Figure 2: Examples of manual well drilling (Elson and Shaw, 1995).

There are four primary types of manual well drilling: augering, percussion, sludging, and jetting, with many variations of each (Danert, 2009). Several publications describe these methods in detail (Blankwaardt, 1984; Carter, 2005; Danert, 2009; Elson and Shaw, 1995; Naugle, 1996;

van der Waal, 2008). In general, unconsolidated and recently consolidated sediments are most promising for all types of manual well drilling (Blankwaardt, 1984). Auguring, sludging, and jetting, are mostly restricted to the softer geologic formations. Manual percussion drilling can be effective in hard rock environments, but is slow (Elson and Shaw, 1995).

In an effort to prioritize areas that would benefit most from manual well drilling and to improve the percentage of successfully drilled wet wells, UNICEF contracted Practica Foundation of the Netherlands to map priority areas for manual well drilling in several African nations. Practica's mapping effort in Chad involved classifying geology and groundwater depth (GWD) data obtained from various governmental and other organizations within country (Practica, 2005; UNICEF, 2009). Subsequently, a discussion arose at the United States Agency for International Development (USAID) about the possibility of using publically available data, such as satellite imagery, to increase the efficiency of mapping favorable areas for manual well drilling. In 2009, researchers at the Desert Research Institute (DRI) in Reno, Nevada were funded by USAID to assess the capacity of such data to map favorable areas for manual well drilling in Burkina Faso and Niger. The final goal of this project is the development of a rapid assessment protocol (RAP) in support of manual well drilling (McKay, November 3, 2008).

The RAP will be a procedural guide for the data and techniques used to develop maps of potentially favorable areas for manual well drilling in Burkina Faso and Niger. One key objective of this project is the identification and assessment of a variety of publically available environmental data that might be useful for identifying areas of relatively shallow groundwater. Examples of this data, which can be acquired without visiting the field, include: satellite acquired data, such as multispectral imagery or digital elevation models (DEMs), digital maps and geographic information system (GIS) layers, and any other relevant, electronic data that is disseminated freely via the internet or through other means. Also, through a comparison with

Practica's methods, this project will seek to improve understanding of the physical indicators and relationships that are important for identifying areas for manual drilling (McKay, November 3, 2008). This thesis is the key contribution to the RAP in terms of consolidation of information related to the improved understanding of the suitability of the identified data and methods for favorability mapping.

1.3 Research problem

The main purpose of this thesis was two-fold: 1) to support UNICEF's manual well drilling program by contributing to the RAP, and 2) to continue to probe the possibilities of using remotely sensed (RS) and other geographic data and spatial analysis techniques for hydrogeologic problems. This study focused on the Zinder region of Niger to address the following question: Can environmental variables, which may be derived from publically available data, act as proxies for groundwater depth data with respect to identifying potentially favorable areas for manual drilling? Specifically, can certain environmental variables be statistically related to groundwater depth for the purpose of favorability mapping for manual well drilling? The objectives of this research were: 1) to identify applicable publically available data sets; 2) to use and adapt previously researched RS and other spatial data and analysis methods; 3) to conduct a statistical analysis to evaluate the predictive power of these data for identifying shallow groundwater which would be favorable areas for manual well drilling; 4) to develop a simple procedural outline based on the study findings that may help direct favorability mapping for manual well drilling in other areas and; 5) to develop example maps of potentially favorable areas for manual well drilling in the Zinder region of Niger.

2 Basis and Background

2.1 Importance of topic

2.1.1 Mandate from the international community

In addition to the water MDG and the current statistics that indicate a water crisis in the developing world, there are several directives from the international community that form the foundation for this and other studies of its kind. Notably, the 2002 World Summit on Sustainable Development (WSSD) called on all countries to promote integrated water resources management in order to provide access to potable water, and to strengthen capacities for research and planning in water resources (UN, 2002). The WSSD and the 2009 World Water Forum (WWF) included similar appeals for the support of Africa in developing effective research that exploits new technology, and for assisting Africa in gathering resources for adaptation and mitigation relating to the possible negative impacts of climate change (WWC, 2009).

2.1.2 Calls for more research

As Hoffmann (2005) points out, though RS data cannot replace field data for thorough hydrogeologic investigation, its ever-increasing promise for hydrogeologic applications and the possibility that many hydrogeologists remain unaware of this potential, suggest the need for continued scientific investigation in this realm. Notably, a recommendation was made recently for a study to "assist in the identification of potential areas for hand-drilled wells in Niger" (Danert, 2007a). MacDonald et al. (2008) state that hydrogeologic knowledge transfer is fundamental to successful and sustainable rural water supplies. This transfer includes the increase of groundwater development skills in rural water supply projects through the creation of manuals specific to rural water supply, groundwater development maps showing availability of water resources, and techniques required to find and develop groundwater (MacDonald et al., 2008).

2.1.3 Data and technological expertise

Unlike many developing countries, Niger does not suffer from an acute lack of water resource data, but obtaining what data is available can be cumbersome according to David Healy, who gathered various GIS and other data in Niger during a 2001 poverty vulnerability project for the World Bank (Healy, June 3, 2009). The UN Environment Programme's Iullemeden Aquifer System Project (UNEP) (IAS) led a push to consolidate and maintain Niger's piezometric data, however the database is not yet available to the public (Dodo, June 9, 2009). Data scarcity is an added motivation for assessing the use of RS data and other widely available data. Also, Cavric (2002) argues that the main hindrances to GIS diffusion in Africa are human and organizational frameworks, not the assimilation of technological innovation. Therefore, the need exists for continued input to GIS structures in Africa, despite the other obstacles that will have to be overcome in realizing its full potential.

It seems apparent that hydrogeologic reconnaissance in Niger, like the rest of the developing world, stands to benefit from RS data, which can be valuable even in areas where data coverage is good (Hoffmann and Sander, 2007; Jha and Chowdary, 2007). However, the brief discussion in Kruck (1992) seems to be the only study that has addressed RS and GIS methods for groundwater exploration in Niger. Jha and Chowdury (2007) point out that RS and GIS technologies have not yet proved valuable to water resource managers in developing countries despite being essential tools, and suggest solutions such as better accessibility and distribution of RS data, and increasing RS and GIS aptitudes through training and support. Currently, RS data is becoming much more accessible, as is evident from the recent release of the entire Landsat satellite archive, which is valuable for the developing world even if internet infrastructure in some places may still be deficient.

2.1.4 Uncertainty of climate change

According to current research, drought-prone areas in SSA could grow by as much as 90 million hectares, approximately 4% of the total land area, because of climate change. This will lead to potentially negative impacts for water availability and subsequent agricultural productivity (UNDP, 2007). Interestingly, research in Niger has indicated increases in local water tables in recent years because of certain land use practices (Favreau et al., 2009). The International Groundwater Resources Assessment Centre's Global Groundwater Information System (IGRAC) (GGIS) indicates that Niger's annual per capita rate of groundwater extraction is less than annual recharge per capita (IGRAC, 2004). However, this is contradictory to the recent findings of the UNEP IAS, which found that estimated extraction exceeds estimated recharge by 25% in the Niger region (Hearns, 2009). Considering the data, the future of sustainable groundwater availability in Niger is uncertain at best, so a better understanding of this resource is crucial.

2.2 Site description

2.2.1 Geography of Niger

Niger is a landlocked country of the African continent that is primarily arid, with over 75% of its land area covered by the northerly encroaching Sahara desert (Figure 3). The southern breadth of Niger is located at the heart of the Sahel zone, which spans from the west coast of Africa at Senegal to Sudan on the east coast between approximately 14°N and 19°N latitude (Inset Figure 3). The Sahel is a tropical semi-arid to arid environment that forms a major ecological transition between the hyper-arid Sahara desert and the humid equatorial tropics to the south. Because of its geographical context as an edge environment, the Sahel is particularly vulnerable to externally forced climate change (Goutorbe et al., 1997). Annual rainfall in Niger is dominated by the monsoon that moves north off of the Gulf of Guinea (Lower left of inset Figure 3). Average annual precipitation ranges from 600 mm yr^{-1} in the south of the country to 200 mm

yr^{-1} in the north. Since the late 1960s the Sahel has received less than average annual rainfall (Figure 4). According to a recent UNEP evaluation, this climatic shift could negatively impact the region's groundwater supply (Hearns, 2009).

Figure 3: Niger reference map.

As of 2004, only 46% of Niger's total population of 15 million were using improved water sources and only 13% were using improved sanitation (UNDP, 2007). The majority of Niger's population lives and relies on the groundwater resource of the IAS (Figure 3). In addition to large areas of sedimentary rock, Niger's geology is comprised of some outcrops of crystalline basement rock in the south, west, and Aïr mountain range of the north (Figure 5). Most regional groundwater flow regimes and water tables in Niger are likely what Haitjema and Mitchell-

Bruker (2005) refer to as recharge controlled. Some areas could exhibit the topographically controlled water tables similar to the local regional flow systems described by Toth (1963).

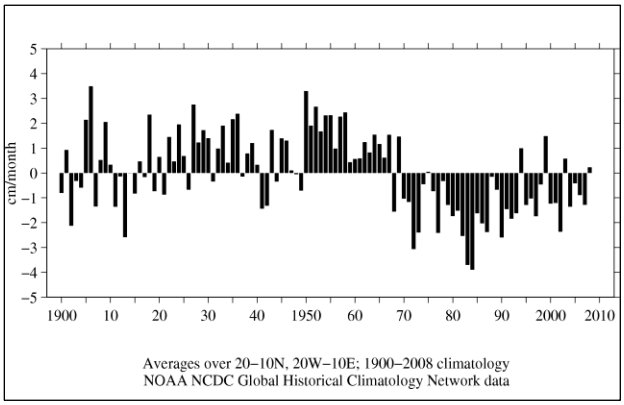


Figure 4: Mean Sahel precipitation anomalies, 1900-2009 (JISAO, 2010).

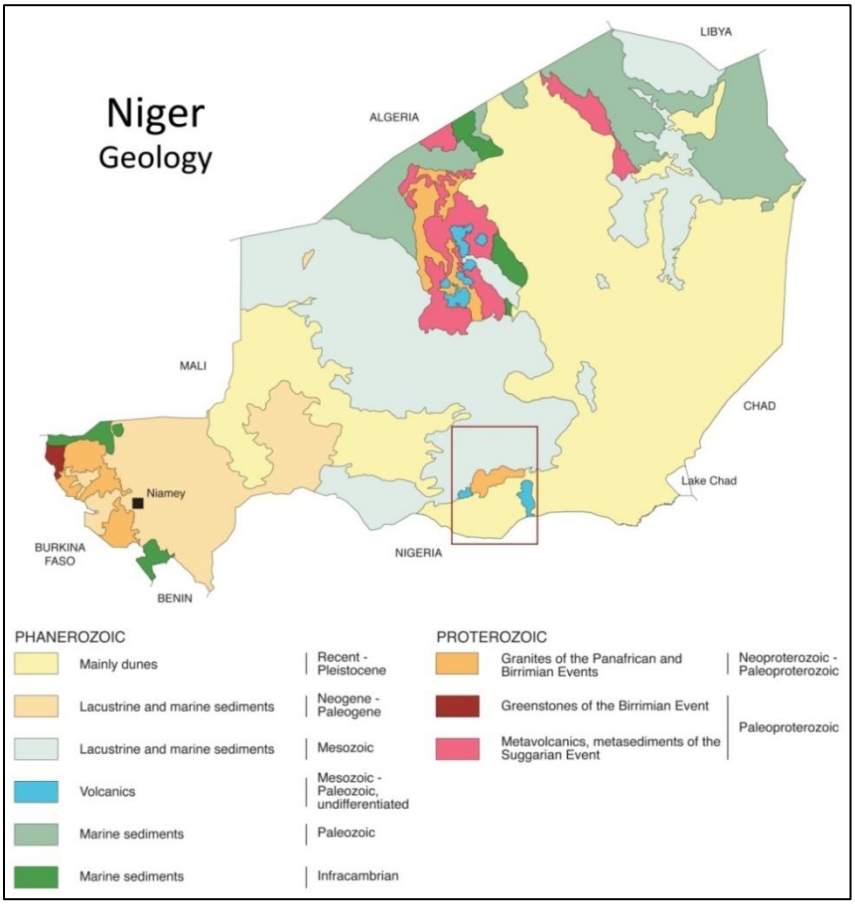


Figure 5: Geology of Niger (Schluter, 2006).

2.2.2 Geography of the Zinder region

This study focused on an approximate 45,000 km^2 rectangular area surrounding the city of Zinder, Niger (Figure 6). The Zinder region, which is a major population center in south-central Niger, is characterized by 2.5-7.5% of the population that is vulnerable to poverty because of a variety of physical, social, and economic indicators (Stone Environmental, 2001). Like the rest of Niger, the people in this region rely almost exclusively on groundwater.

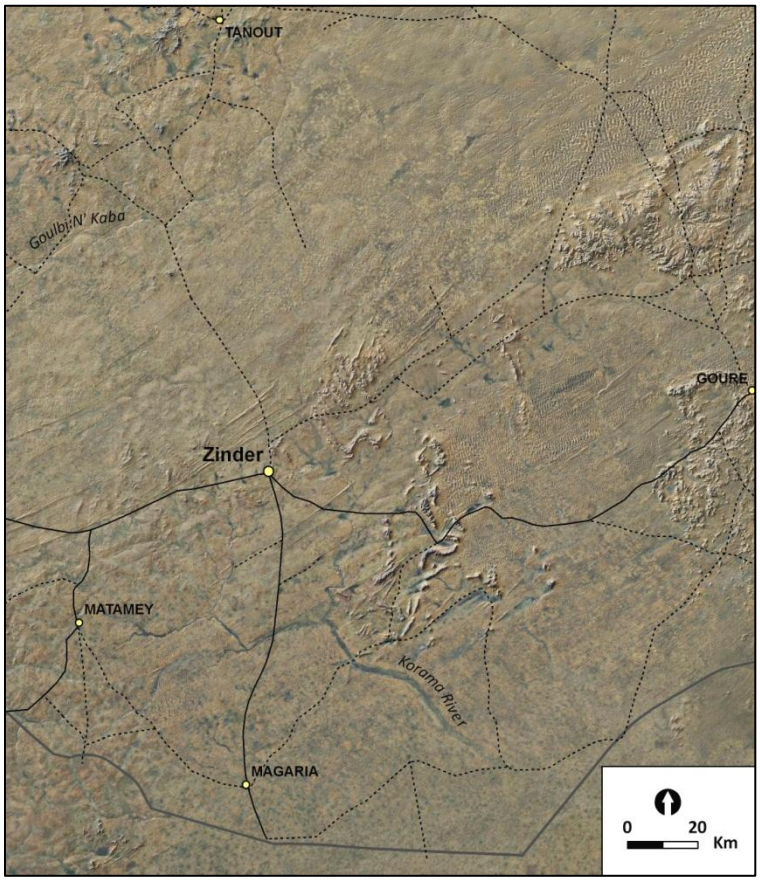


Figure 6: Reference map of Zinder region, Niger.

The Zinder area falls on the margins of the larger IAS and is more complex hydrogeologically because it is underlain by all the basic geologic formations found throughout the country (Figure 5). The northern part of this area is part of the Continental Intercaliare (Hamadien) lacustrine and marine sedimentary aquifer system (Figure 7). The discontinuous

aquifers of the Damagaram and Mounio granites and volcanic rocks, which lie in the center of this region, are surrounded in the south by outcrops of the more recent Continental Terminal sedimentary formation and Quaternary deposits associated with Lake Chad to the southeast (Figure 7).

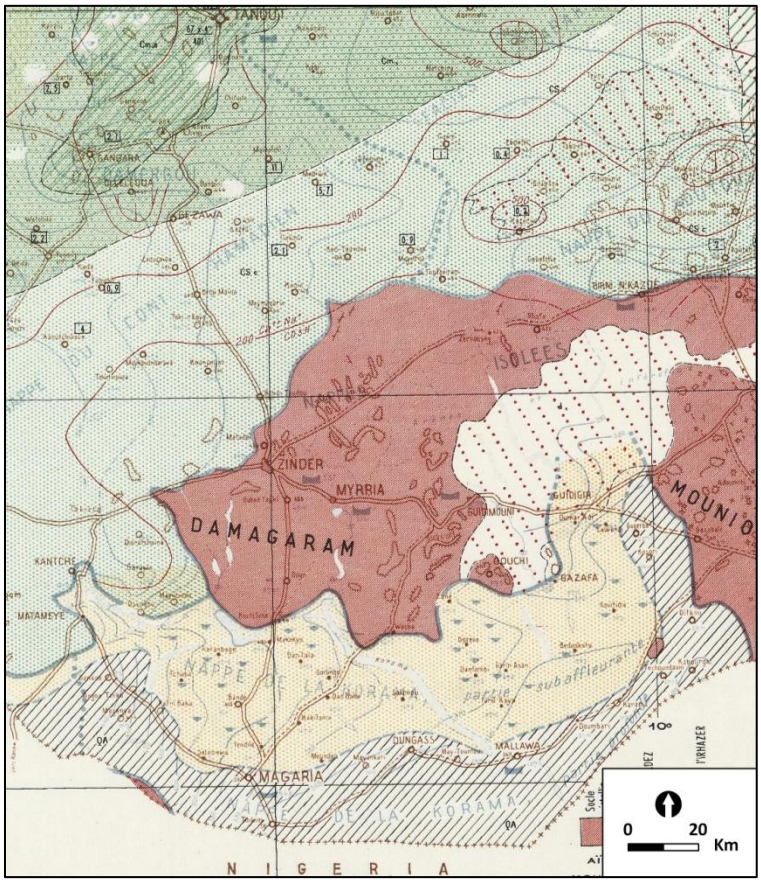


Figure 7: Hydrogeology of Zinder region, Niger from Pirard and Faure, 1964.

The varied hydrogeology of this area make it a good candidate for assessing the ability of various RS and GIS data for identifying potentially favorable areas for manual well drilling. Also, the groundwater regime of this region has been monitored by the Direction Régionale Hydraulique, Zinder (DRH) and various non-government organizations (NGOs) such as World Vision and Winrock International for several decades, thus a relatively dense database of wells was available (Danert, 2007a). Notably, implementation of manual well drilling in this region

has been ongoing since the 1970's, with drilling workshops being conducted as recently as 2007 (Danert, 2006; Danert, 2007b; Naugle, 1996).

3 Literature Review

3.1 Synopsis of spatial analysis for groundwater exploration

There have been many investigations in recent decades regarding the use of RS and GIS data and methods to explore for groundwater, and they have been summarized thoroughly (Bandyopadhyay et al., 2007; Becker, 2006; Hoffmann, 2005; Jha et al., 2007; Waters et al., 1990). This pool of research falls into four broad categories: 1) exploration for groundwater to increase efficiency of mechanized boreholes; 2) identification of regional groundwater recharge and discharge areas; 3) classification of groundwater dependent ecosystems; and 4) detection of shallow groundwater. It is ubiquitously noted in the scientific literature that RS data alone do not suffice to identify specific well drilling sites, but are better coupled with field methods, such as geophysical surveys. However, RS and GIS data and methods can be quite useful for assessing broad groundwater characteristics, and thus, aid in identifying areas for subsequent field based groundwater exploration, which relates to this study (Becker, 2006).

3.2 Spatial data and analysis techniques

3.2.1 Geologic lineaments

A widespread use of RS data for groundwater exploration is the development of geologic maps and the delineation of linear features or lineaments. Lineaments are thought to represent geologic structures such as fractures, faults, joints, or dykes that could be associated with secondary porosity and weathering in hard rock and regolith environments, which can potentially contain groundwater (Sander, 2007). A key assumption of lineament analysis is that linear features on the land surface are representative of, and vertically aligned with, subsurface features

(Sander, 2007; Waters et al., 1990). Lineaments are most often visually traced from false color composites of multispectral satellite imagery. Other RS data such as aeromagnetic surveys, radar, and aerial photography can augment a lineament analysis (Bailey and Halls, 2000; Galanos and Rokos, 2006; Ranganai and Ebinger, 2008; Salama et al., 1994; VanderPost and McFarlane, 2007). However, distinguishing actual geologic lineaments from human or other non-geologic linear features is problematic and can be minimized by using a shaded relief depiction of a DEM that shows little to no anthropogenic influence (Ghoneim and El-Baz, 2007; Masoud and Koike, 2006a). Also, Mabee (1994) and Sander et al. (1997) successfully identified only the most prominent geologic linear features through rigorous reproducibility methods in which lineaments that were interpreted by several analysts were compared for degree of overlap.

Often, lineament analysis involves the calculation of lineament densities from which thematic maps are created that represent areas of increased potential for groundwater development (Ahmed et al., 1984; Al Biely et al., 2000; Bala et al., 2005; Edet et al., 1998; Krishnamurthy et al., 1996; Kruck, 1992; Masoud and Koike, 2006a; Munch and Conrad, 2007; Mutiti et al., 2008; Rangzan et al., 2008; Sener et al., 2005). Other studies have statistically analyzed orientation of fracture traces to more objectively assess potential groundwater carrying features (Amesz, 1984; Boeckh, 1992; Galanos and Rokos, 2006; Madrucci et al., 2008; Masoud and Koike, 2006b; Okereke et al., 1998; Solomon and Ghebreab, 2008). Lineament analyses that combine multiple lineament attributes, such as what Teeuw (1995) and Madrucci et al. (2008) have done, seem most robust. For example, Hardcastle (1995) used an algorithm to generate a value derived from the total number of lineaments, number of directional families, and total length of lineaments, which successfully identified highly probable areas for wet wells.

3.2.2 Topography and geomorphology

Topographic relief strongly controls the energy and spatial distribution of water over a landscape. Therefore, elevation and geomorphology data are commonly used for groundwater exploration because of the clues they provide regarding runoff, recharge, and subsequent groundwater accumulation (Krishnamurthy et al., 1996). DEMs, such as NASA's Shuttle Radar Topography Mission (SRTM) DEM, have been used in some groundwater studies to identify and calculate various topographic features such as depressions, slope, drainage density, topographic indices, and hypsometric characteristics (Ghoneim, 2008; Masoud and Koike, 2006a; Munch and Conrad, 2007; Tweed et al., 2007). The relationships of these relief features to groundwater characteristics have only been studied at a cursory level. For example, Masoud (2006a) correlated the topographic convergence index (TCI) or wetness index developed by Beven and Kirby (1979) and the hypsometric integral, a measure of the erosional state of a basin, to well yield data in order to weight these variables in a groundwater potential model for the Sinai Peninsula.

DEMs are also useful for objectively calculating and classifying geomorphologic landforms, which can be useful to groundwater studies (Lindsay, 2005). The use of multispectral satellite imagery to visually interpret distinct geomorphologic units is common and effective (Chowdhury et al., 2009; Ghoneim, 2008; Jaiswal et al., 2003; Krishnamurthy et al., 1996; Madrucci et al., 2008; Rangzan et al., 2008; Ringrose, 1998). Paleo-features, such as sedimentary deposits from ancient rivers and lakes, which could constitute shallow aquifers, have also been successfully identified using multi-spectral imagery (Ghoneim and El-Baz, 2007; Mutiti et al., 2008; Robinson et al., 2007; Salama et al., 1994; VanderPost and McFarlane, 2007).

3.2.3 Vegetation

The field of ecohydrology deals with the complex relationship of surface and groundwater interaction and its impact on aquatic ecosystems, and offers good insights for groundwater exploration (Eamus and Froend, 2006; Elmore et al., 2008; Klijn and Witte, 1999; Naumburg et al., 2005). Many studies have successfully used statistical methods for analyzing groundwater dependent (phreatophytic) and non-groundwater dependent vegetation dynamics with respect to groundwater characterization (Batelaan et al., 2003; Bobba et al., 1992; Bukata et al., 1991; Munch and Conrad, 2007; Tweed et al., 2007). However, detailed information on phreatophytic species is not always available. Elmore et al. (2008) suggested that the insensitivity of vegetation to changes in precipitation could be a good indicator of its dependence on groundwater, and that this deserves more research. Tweed (2007) effectively used this principle to identify seasonal persistence of green leafed vegetation in southeastern Australia by applying a threshold to the standard deviation of the Normalized Difference Vegetation Index (NDVI), a measure of land surface greenness or biomass derived from several seasons of Landsat scenes.

Most groundwater exploration research with a vegetation component incorporates some kind of vegetation index, such as NDVI (Bailey and Halls, 2000; Chen et al., 2006; Munch and Conrad, 2007; Mutiti et al., 2008; Tweed et al., 2007). Baugh (2006) assessed the relative performance of a variety of vegetation indices in an arid environment by correlation to promotion of vegetation by antecedent precipitation and concluded that a variation of NDVI was best. The tasseled cap method, which gives measures of wetness and greenness, has also been used in the exploration for groundwater, but is most applicable to mesic areas (Bala et al., 2005; Munch and Conrad, 2007; Mutiti et al., 2008). Bukata et al. (1991) developed a “vigor of growth” curve, which worked well to separate vegetation found in regions of discharge, recharge, and transition zones for two relatively wet basins in southern Canada. For a shallow groundwater investigation

in semi-arid Botswana, Ringrose (1998) had mixed results correlating aquifer characteristics to vegetation classes developed from supervised classification based on field data.

3.2.4 Land surface temperature

Cartwright's (1974) theoretical work relating water table depth to surface soil temperatures was among the first studies to examine the relationship between soil heat and depth to groundwater. Building on Cartwright's work, Huntley (1978) assessed the capability of RS thermal infrared (TIR) imagery to detect shallow aquifers, and concluded that with the technology at that time it was not possible to estimate water table depth accurately. Some studies have continued to experiment with TIR for identifying shallow groundwater and areas of groundwater discharge (Banks et al., 1996; Bobba et al., 1992; Ghoneim, 2008). Ghoneim (2008) used Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature (LST) data to observe cooler anomalies after rainfall events in an arid plain of the United Arab Emirates, which suggested the prolonged evaporation of lingering soil moisture from groundwater discharge. Heilman and Moore (1982) noticed that cool areas in night time thermal imagery were correlated with shallow groundwater. Becker (2006) noted that pre-dawn imagery may hold the most promise for correlating surface temperatures to GWD.

3.2.5 Analysis techniques

A variety of techniques have been used to weight the importance of different environmental variables in groundwater exploration studies. Heuristic weighting and summation of different mapped variables is common in many groundwater potential studies (Krishnamurthy et al., 1996). Chowdury et al. (2009) used the Saaty hierarchy approach to weight many variables thought to be important for groundwater potential in West Bengal, India. Munch and Conrad (2007) used numerical weights for a variety of environmental variables used to identify groundwater dependent ecosystems in South Africa.

Statistical approaches are as varied as the number of groundwater exploration studies, most of them using linear models. For example, Chen et al. (2006) used multiple linear regression to correlate NDVI and potential evapotranspiration to GWD. Analysis of variance (ANOVA) was used by Mutiti et al. (2008) to evaluate the importance of a variety of variables thought to influence groundwater presence in Kenya. Ghoneim (2008) used principal components and dendrogram analyses to effectively group drainage basins characteristics as a component of a groundwater potential study.

4 Development of Variables

4.1 Choosing the data and development methods

Using the literature review as a guide, data mining was conducted over the initial several months of this study and continued through the data preparation phase as additional data leads and requirements materialized. This phase revealed a myriad of RS and other digital spatial data that were accessible through various portals on the internet (Appendix A). The data ranged from digital copies of paper maps to satellite imagery. A summary of the base data obtained for this study are found in Table 1. An abundance of data and methods necessitated the use of some guidelines for choosing the appropriate approach.

Relative ease of variable development was an important guiding principle. For example, raw Landsat imagery requires a variety of pre-processing, such as atmospheric correction, which could make it prohibitive to many novice end users, whereas MODIS data is preprocessed. All of the data identified for this study required some level of manipulation in order to be useful for research. Another example is the processing of the DEM to deal with artifact peaks and pits. Also, a straightforward workflow for the development of the variables was necessary in order to provide an example for adaptation of the techniques to other areas (Figure 8). All data pre-processing and development was done in Environment Systems Research's (ESRI) ArcGIS 9.3

with an ArcInfo license, Lindsey's (2005) free Terrain Analysis System (TAS), and Microsoft Access and Excel.

Table 1: Base data.

Category	Data Name	Data Type	Resolution	Organization	Availability	Website
Geology	Carte Géologique de la République du Niger	Map	1:2,000,000	Bureau de Recherches Géologiques et Minières (BRGM)	By request	http://library.stanford.edu/depts/branner/
Groundwater depths	Base de Données Piézométrique du Zinder, and Winrock – Groundwater database	Vector (Point)	Precise	Ministry of Water Resources of Zinder, Niger	By request	N/A
Hydrogeology	Carte de Reconnaissance Hydrogéologique du Niger Sud Oriental	Map	1:500,000	BRGM	Purchase	http://www.brgm.fr
Precipitation	Climate Research Unit (CRU) Average Monthly Precipitation	Raster	0.5°	CRU, University of East Anglia	Free	http://www.cru.uea.ac.uk/
Soils	Carte Pédologique de Reconnaissance de la République du Niger, Zinder	Map	1:500,000	ORSTROM	Free	http://eusoils.jrc.ec.europa.eu/esdb_archive/EuDASM/africa/
Thermal	MODIS Terra MOD11A1 Daily Land Surface Temperature	Raster	1000 m	NASA	Free	https://wist.echo.nasa.gov/~wist/api/imswelcome/
Topography and Geomorphology	Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM)	Raster	90 m	NASA and CGIAR-CSI	Free	http://srtm.csi.cgiar.org/
Vegetation	MODIS Terra MOD13Q1 16-day Enhanced Vegetation Index	Raster	250 m	NASA	Free	https://wist.echo.nasa.gov/~wist/api/imswelcome/
Waterbodies	MODIS Terra MOD44W Land Water Mask	Raster	250 m	NASA	Free	https://wist.echo.nasa.gov/~wist/api/imswelcome/
	Landsat Enhanced Thematic Mapper (ETM+)	Raster	30 m	NASA	Free	https://wist.echo.nasa.gov/~wist/api/imswelcome/

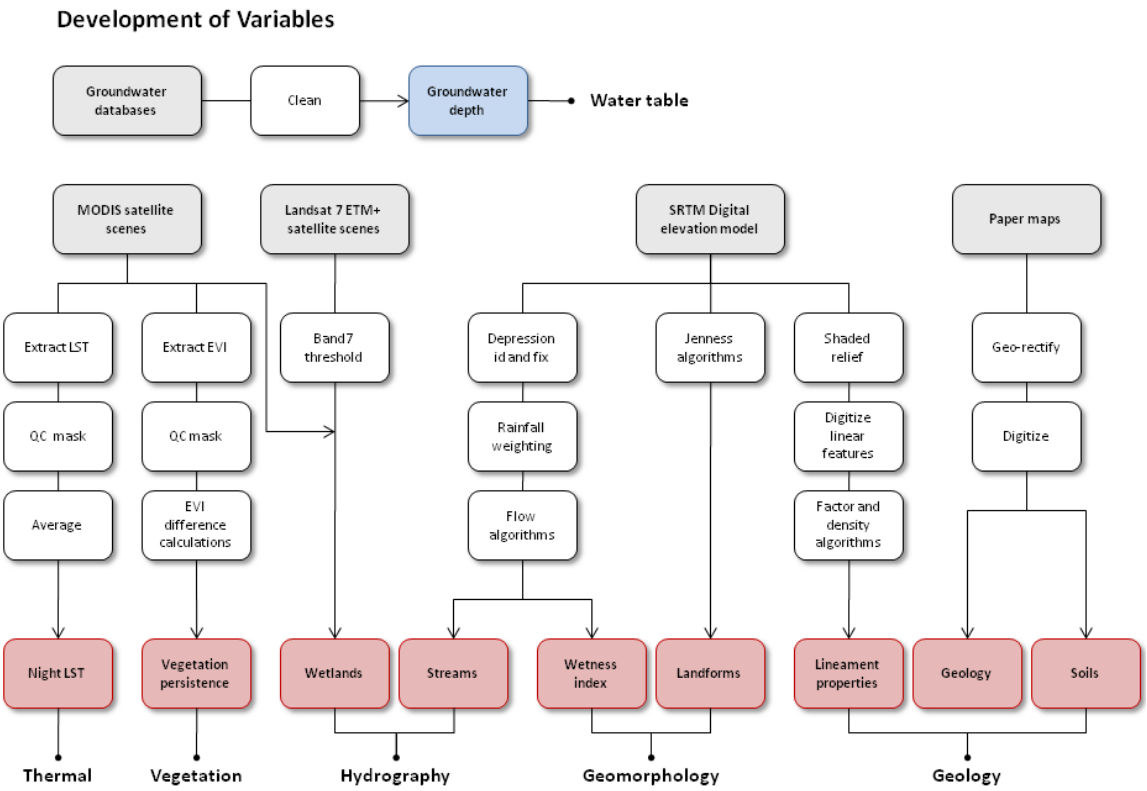


Figure 8: Workflow for the development of environmental variables.

The goal to produce broad maps of manual well drilling favorability was an important guide for selecting the suitable spatial resolution of the data. Relatively high spatial resolution, such as some Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data at 15 m resolution, was not appropriate for mapping broad attributes of the landscape where this detail would have been excessive. However, it was necessary that the spatial resolution of the data not be so coarse as to conceal important landscape attributes. Also, it was important to choose a data resolution that would adequately capture GWD variability without over-sampling or over-averaging. However, this is complicated by the likelihood that GWD has different scales of spatial correlation over the study area. For example, GWD could have relatively long distant correlations in sand formations compared to fractured bedrock environments. Semivariogram analysis of GWD data using 50 lags indicated that high resolution data such as 30 m Landsat data

would be redundant for predicting differences in GWD (Figure 9). At distances greater than 150 km, where the semivariogram begins to plateau (sill), GWD values would be averaged beyond utility. MODIS data at 250 m to 1 km resolution was the best balance between revealing landscape features of interest and distinguishing between shallow and deep GWD values without oversampling.

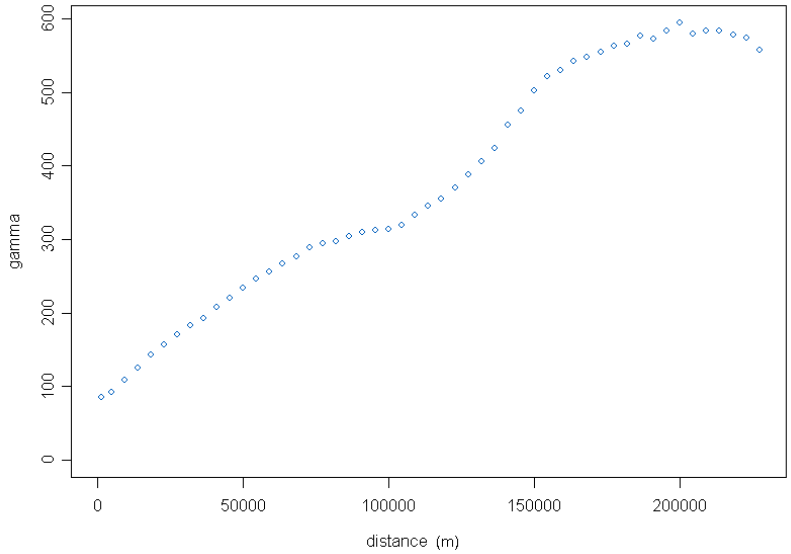


Figure 9: Semivariogram of groundwater depth – Zinder region.

4.2 Key limitations and assumptions

A primary limitation of this study was the fact that land surface attributes are, at best, secondary indicators of groundwater characteristics because they do not directly describe conditions underground. Another limitation was that low resolution land surface data has an inherent level of error that can be propagated throughout the analysis. Quantifying this uncertainty can be problematic. To address this issue a cross validation technique was used in the final statistical analysis. Also, confidence estimates of the study results can come from future data collection on the percentage of successfully drilled wet wells within identified target areas.

There were some overarching assumptions that are important to point out. Analysis of the DEM assumed that the elevation model was an accurate representation of the relief despite inherent error from low horizontal and vertical resolutions. Positive vegetation index values above approximately 0.1 were assumed to represent green leaf vegetation cover, despite issues such as atmospheric scattering. Another major assumption was that surface lineaments are vertically aligned manifestations of subsurface features such as faults and fractures that have associated secondary porosity and a greater degree of weathering. GWD data were assumed to be accurate and represent the average water table depth. This study also relied on geology and soil maps that were created at small spatial scales that generalize real landscape features.

4.3 Groundwater depth

Depth to groundwater was the dependent or response variable used in this study. These GWD data were obtained from the well databases of the DRH and Winrock International, which were generously shared by these organizations for this study. Both databases contained geographic coordinates for each well point, and these were projected to a custom Albers Equal Area Conic projection (Appendix C). The two databases together comprised over 3000 wells in the Zinder region (Figure 10). The GWD values from both databases were primarily measurements taken at a single point in time, most of which were recorded during the past decade. There were numerous records that lacked date information, which could possibly go back as far back as the mid-1980s.

The magnitude of error for each GWD point was unknown, and no effort was made to quantify this uncertainty. Where more than one database record existed for one well location the average and standard deviation of GWD were calculated. Only the wells with a standard deviation for GWD of less than or equal to 3 m were retained. This process eliminated duplicate records and produced an acceptable average GWD value for each of these sites. After the GWD

data were properly compiled, subsets of the entire database were created for the final statistical analysis. According to De'Ath and Fabricius (2000) a validation set comprising 33 to 50% of the entire data set is desirable. Therefore, 1284 points were randomly selected for validation, representing 35% of the full data set. The remaining 2397 points (65%) were used for statistical modeling.

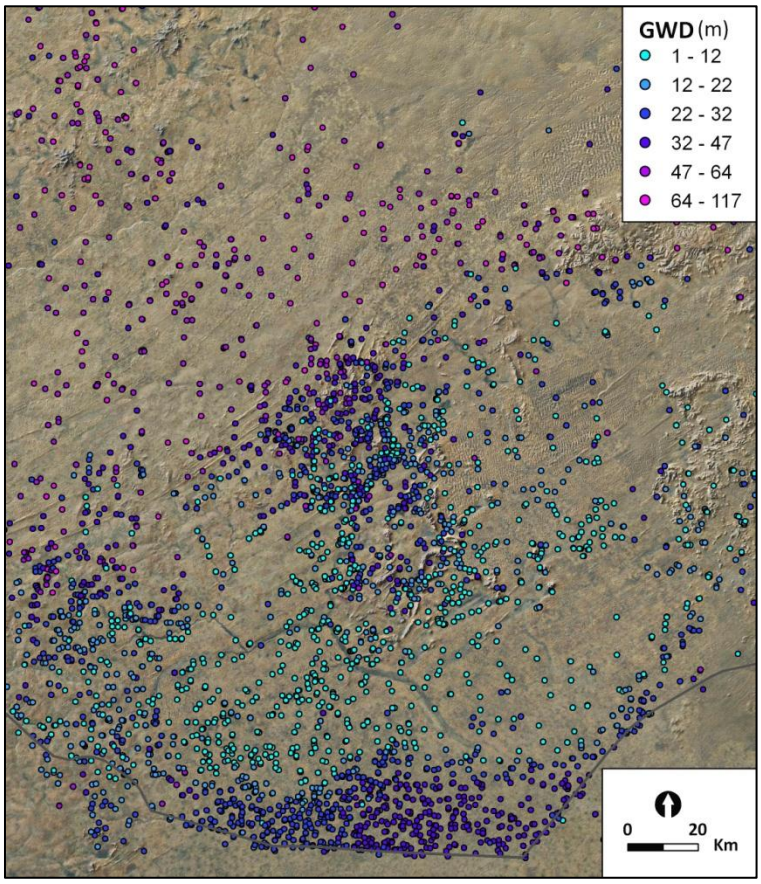


Figure 10: Well sites and associated groundwater depths - Zinder region.

4.4 Geology, soils, and hydrogeology

Geologic data are the foundation of any groundwater study. Medium scale GIS layers of geologic features, which are produced at 1:250,000 or larger, would have been ideal for this study. However, data with this level of accuracy did not exist for the entire study area. The official paper geologic map of the Republic of Niger produced by the Bureau de Recherches

Géologiques et Minières (BRGM) in 1967 at the scale of 1:2,000,000 was the basis for development of the geology layer for the study area (Figure 11) (Greigert and Pougnet, 1967). Note that this map is different than the highly generalized geology map of Niger shown in Figure 5. This paper map was acquired via interlibrary loan with Stanford University's Branner Earth Sciences library. It was assumed that maps such as this are generally accessible in the focus country, and thus, fit the description of publically available data.

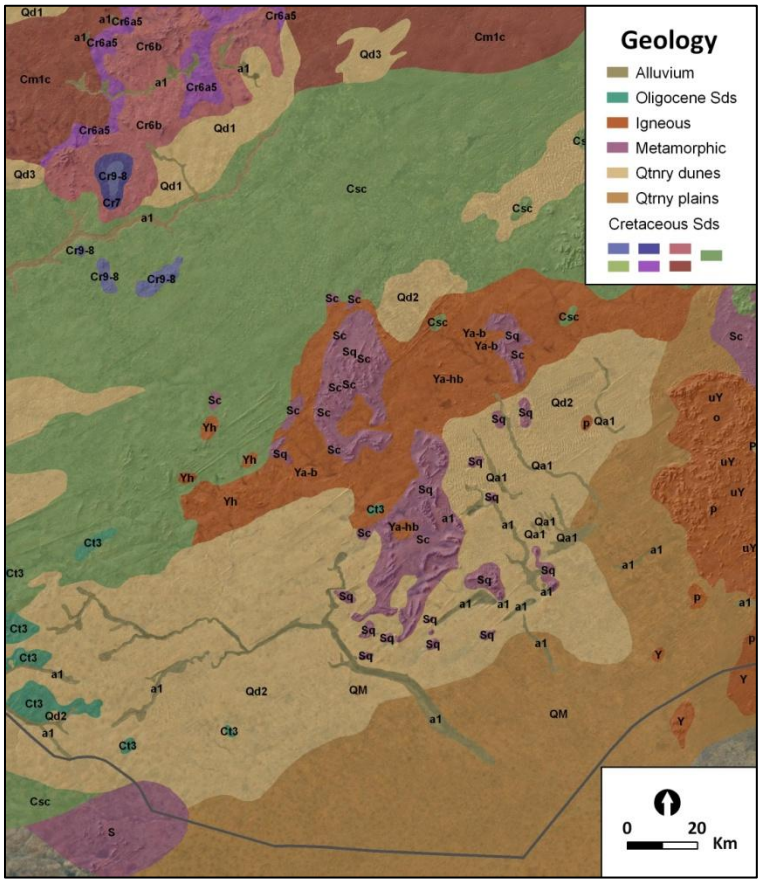


Figure 11: Geology of the Zinder region adapted from Greigert and Pougnet, 1967.

The map was geo-referenced to major land features evident on both the map and the free ESRI_Imagery_World_2D GIS layer, using over 20 spatially distributed points, and a 2nd order polynomial transformation with a root mean square error of 0.01 decimal degrees. The map was resampled using bilinear interpolation. The features on the map were then digitized into polygons

that covered the study area. These polygons were then compared to the ESRI imagery and re-digitized in some areas to correct discrepancies with obvious geologic features in order to improve the spatial accuracy of the final geology layer (Figure 11).

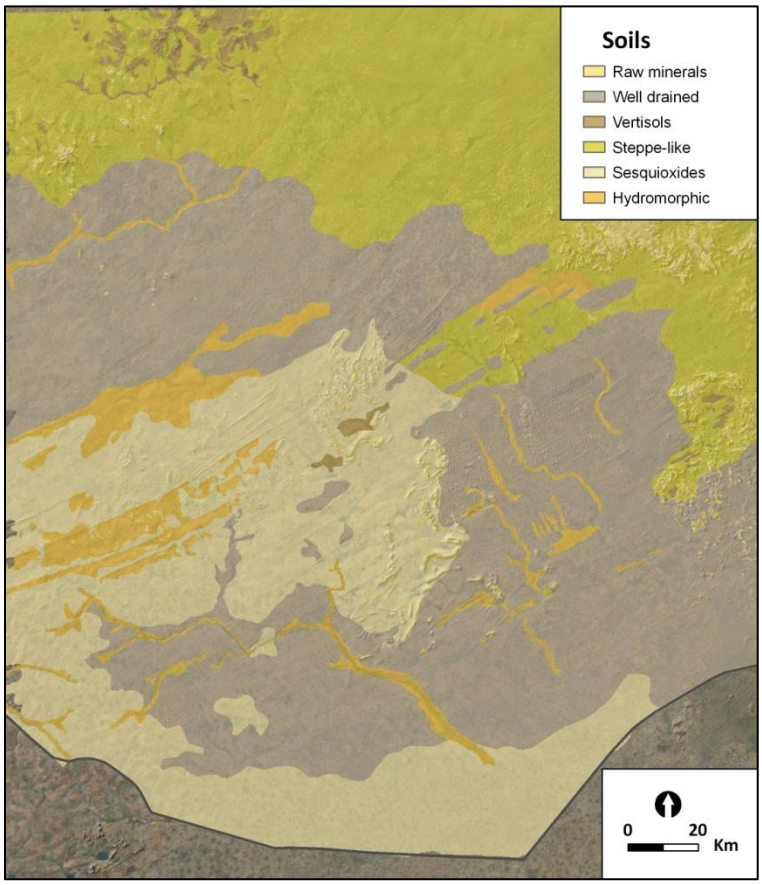


Figure 12: Soil groups of the Zinder region from Bocquier and Gavaud, 1967.

Most of the soil maps for the African continent have been digitally archived on the European Digital Archive of Soil Maps (EuDASM) website (Appendix A). Fortunately, the 1:500,000 scale soil map of the Zinder region, developed by the Office de la Recherche Scientifique et Technique Outre-Mer (ORSTROM) in 1967, had already been converted into a GIS layer by the Systeme d'Information Géographique du Niger (SIGNER) before this study began (Figure 12) (Bocquier and Gavaud, 1967). This GIS layer was acquired from a GIS consultant, and considered to be publically available.

Numerous hydrogeologic and groundwater development potential maps exist for Niger and the surrounding region (BRGM, 1976; Moussié 1986). A prime example is the exhaustive Atlas of Groundwater of the Republic of Niger complied by the BRGM, which includes many maps, drawings, and cross sections describing the stratigraphy and hydrogeology of Niger (Greigert and Bernert, 1978). However, after thorough investigation, it appeared that resources such as this are scarcer than typical geology maps and not within the realm of publically available data. Additionally, these resources are not easily repeatable, unlike the methods described in this study, so they were not considered. The hydrogeologic map of southeastern Niger at 1:500,000 available from the BRGM was only used for general hydrogeologic information of the study area and for qualitative comparisons (Figure 7) (Pirard and Faure, 1964).

4.5 Geologic lineaments

A lineament data set was developed for this study that minimized anthropogenic influence, and incorporated reproducibility in order to increase the confidence that only geologic features were identified (Figure 13). Digitizing of lineament features was done by three independent interpreters at the scale of 1:250,000 using a shaded relief rendering of the SRTM DEM. Adapting slightly the method described in Sander (1997), the three lineament layers , A, B, and C were overlaid and only the line features that were reproduced by two or more interpreters were retained in the final layer. This was done by a visual comparisons of A to B, B to C, A to C, and AB to C using the rules that identical or reproducible lines could only be separated by less than 2 mm and fall at an angle less than +- 10° on screen at the 1:250,000 scale. Results were compared to several alternate illumination angles of the DEM and it was found that no obvious linear features were missed because of the use of only one illumination angle while digitizing.

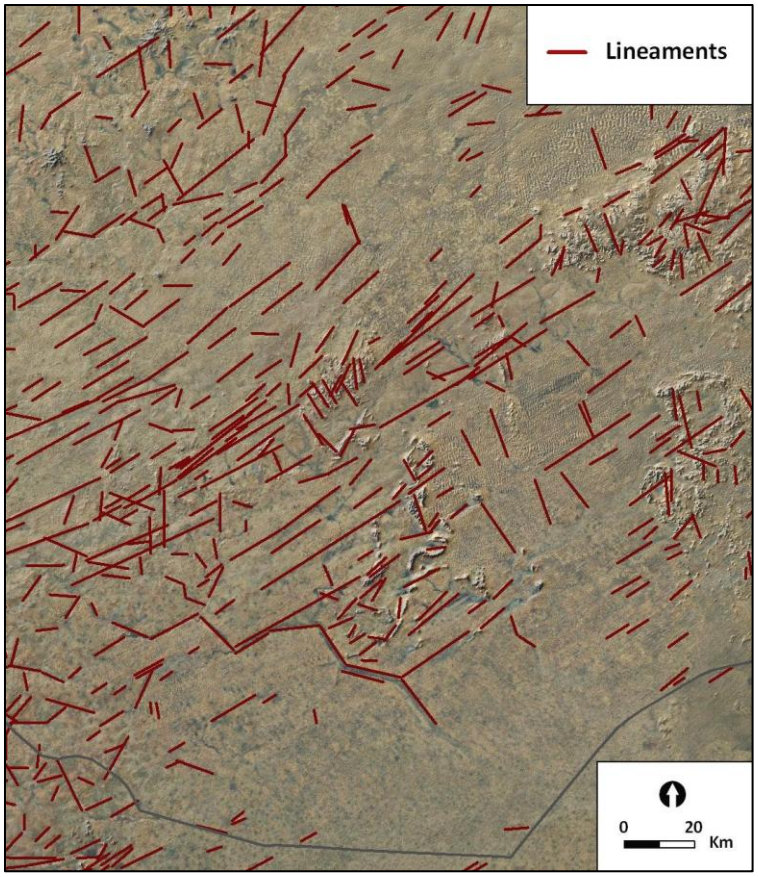


Figure 13: Reproducible lineaments derived from SRTM DEM – Zinder region.

4.5.1 Lineament factor determination

The final lineament layer was then used to develop an adaptation of the lineament factor (LF) presented by Hardcastle (1995). The LF analysis yielded continuous values across the study area that represented the relative degree of fracturing. The LF is a weighted value derived from the normalization of the total number of lineaments (TN_n), total length of lineaments (TL_n), and total number of directional families (DF_n) within 30% overlapping circular windows (Appendix B). Circular windows of 2000 km^2 generated an average of 15 lineaments per window for the statistical analysis, which is the minimum number suggested by Hardcastle (1995), to yield meaningful statistics. Azimuths for each lineament were calculated using the "polyline_Get_Azimuth_9x.cal" tool from Free ArcGIS Tools at http://www.ian-

ko.com/free/free_arcgis.htm. Statistics within each circular window were calculated using the "listintersectingfeatures" tool from Geospatial Modeling Environment at http://www.spatialecology.com/gme/index.htm. The number of lineament families per window was determined by applying a moving average on the normalized lineament histograms for each window, which was suggested by the author as a simplification of the Gaussian fitting algorithm used in Hardcastle (1995) (Hardcastle, February 3, 2010). For example, three directional families were determined for the data in Figure 14 by counting only the most prominent peaks in the moving average.

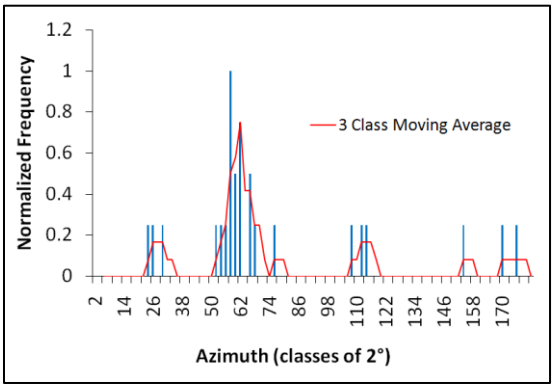


Figure 14: Moving average used to determine number of lineament directional families.

The following equation was used to calculate LF:

$$LF = (0.4 * TN_n + 0.4 * TL_n + 0.2 * DF_n)$$

Equation 1: Lineament factor (LF)

According to Hardcastle (1995), the weights used for each term are subjective and determined by what is considered to be the relative importance of each term for the study area. The DF_n term received the lowest weighting by 50% because the method used to determine this term was simplified and perhaps less meaningful than the Gaussian algorithm from Hardcastle (1995). The circle centers were then attributed with their respective LF values and these points were used to

interpolate LF values across the study area using ordinary kriging (Figure 15). Also, a simple line density calculation was performed on the lineament layer to compare with the LF data in the statistical analysis.

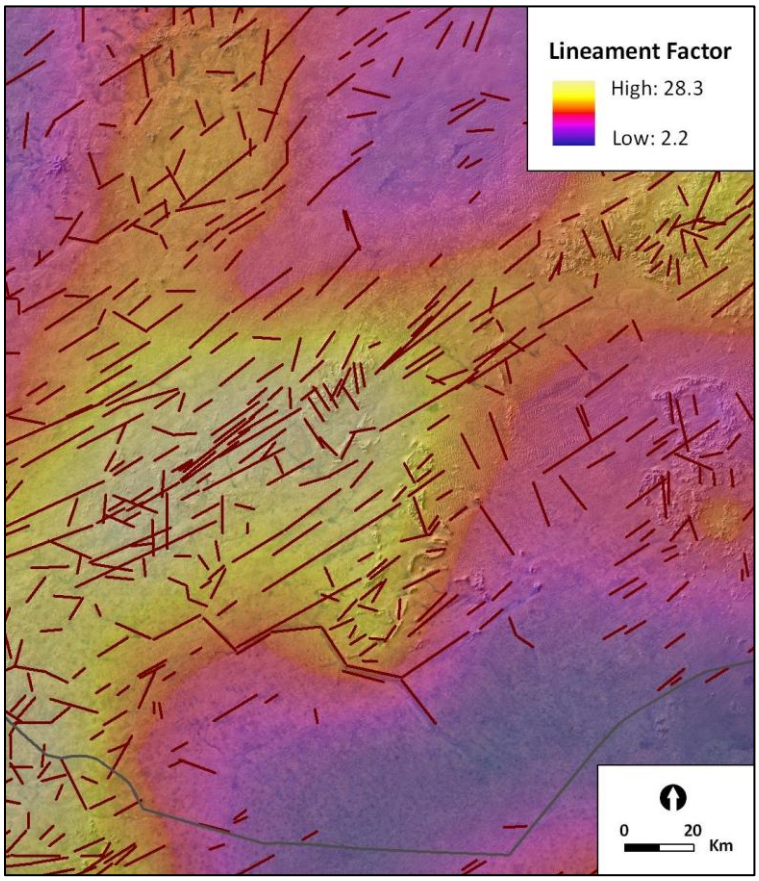


Figure 15: Lineament factor - Zinder region.

4.6 Topographic convergence index

The TCI or wetness index introduced by Beven and Kirby (1979), which is derived from a DEM using the hydrologic flow algorithms commonly available in GIS software, was included in this study. The following equation was used to calculate the TCI:

$$TCI = ln\left[\frac{contributing\ area}{\tan\ (slope\ in\ radians)}\right]$$

Equation 2: Topographic convergence index (TCI)

Values of TCI typically range from 0 to 20. Locations that have large contributing areas (basins) and are relatively flat exhibit high values of TCI, and are more likely to become saturated after precipitation than other areas of the landscape. Thus, areas with high TCI values can be expected to contribute more to groundwater recharge (Masoud and Koike, 2006a). The radar derived SRTM DEM at 90 m resolution was preferred to the newly developed ASTER Global DEM (GDEM) at 30 m resolution because of the visible noise in the GDEM, despite the fact Quinn et al. (1995) suggested that TCI values may become unreliable at resolutions near 100 m.

4.6.1 DEM preprocessing

The contributing area to each cell on the DEM is defined by the number of accumulated cells flowing to each cell, which can be easily calculated with the flow direction and flow accumulation algorithms available in the GIS software. In order for these hydrologic flow algorithms to work, the raw DEM has to be altered so that pits (low depressions) or peaks (tall flow obstructions) are eliminated. These features are commonly regarded as artifacts from DEM creation, and are dealt with by filling, breaching, or filter smoothing. Altering a DEM in this way can yield erroneous results where real landscape features are changed. This had implications for this study because, as Mark (1988) stated, low lying arid and semi-arid areas can likely have broad closed depressions that form ephemeral lakes or playas, and this is generally true in Niger.

To minimize the impact of DEM alteration, a stochastic shape analysis available in TAS was used to identify probable depressions (Lindsay, 2005). As stated in the TAS Help file, this algorithm adds a random error field to the DEM, and then finds the resulting depressions and/or flats. This process is then repeated iteratively to measure the probability that a cell will have a depression during the simulation. The simulation was run for over 300 iterations until the threshold root-mean-square-difference of 0.001 m between probability images was reached. DEM cells with a probability greater than or equal to 0.95 and a minimum area of 1 km^2 were

retained as real depressions (Figure 16). These depressions were preserved by setting null the lowest grid cells in each depression, which in effect embeds a flow outlet (Inset in Figure 16). Hydrologic flow algorithms were then able to route flow internal to each depression.

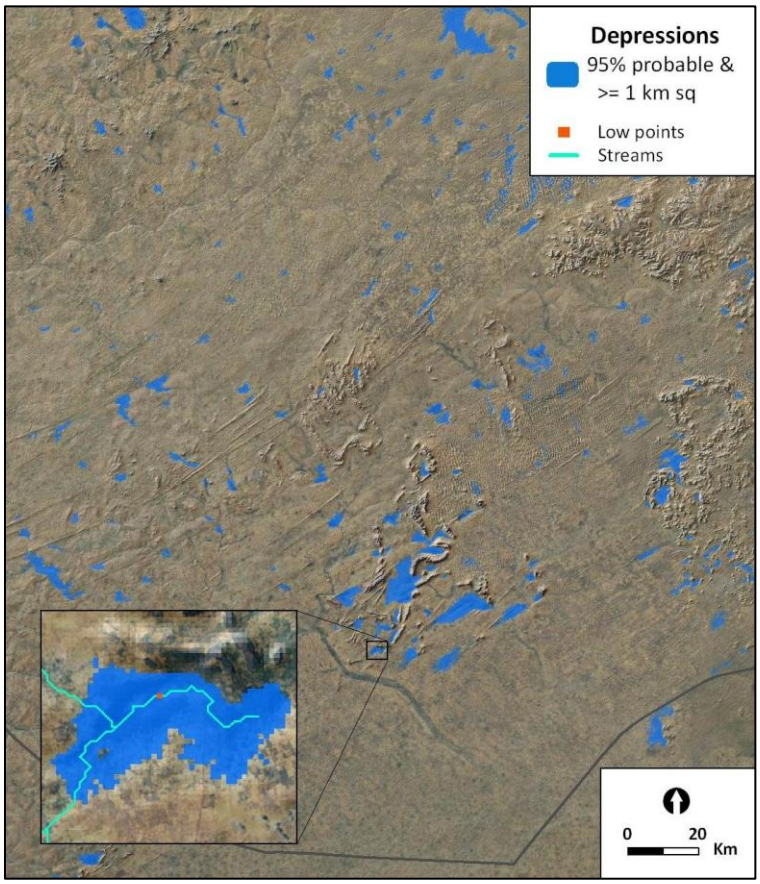


Figure 16: Topographic depressions and absolute low points – Zinder region.

4.6.2 Precipitation weighting

Precipitation-weighted flow accumulation to each cell was calculated for the filled DEM using the D8 algorithm of Jenson and Domingue (1988) included in the GIS software. The continuous precipitation grid used for this weighting was interpolated from the Climate Research Unit's (CRU) version 2.1, 0.5° (≈50 km) gridded rainfall data averaged for 1981 to 2002. The slope parameter was calculated using a tool included in the GIS software. Flow accumulation and slope were combined in Equation 2 to achieve the final TCI raster for the study area (Figure 17).

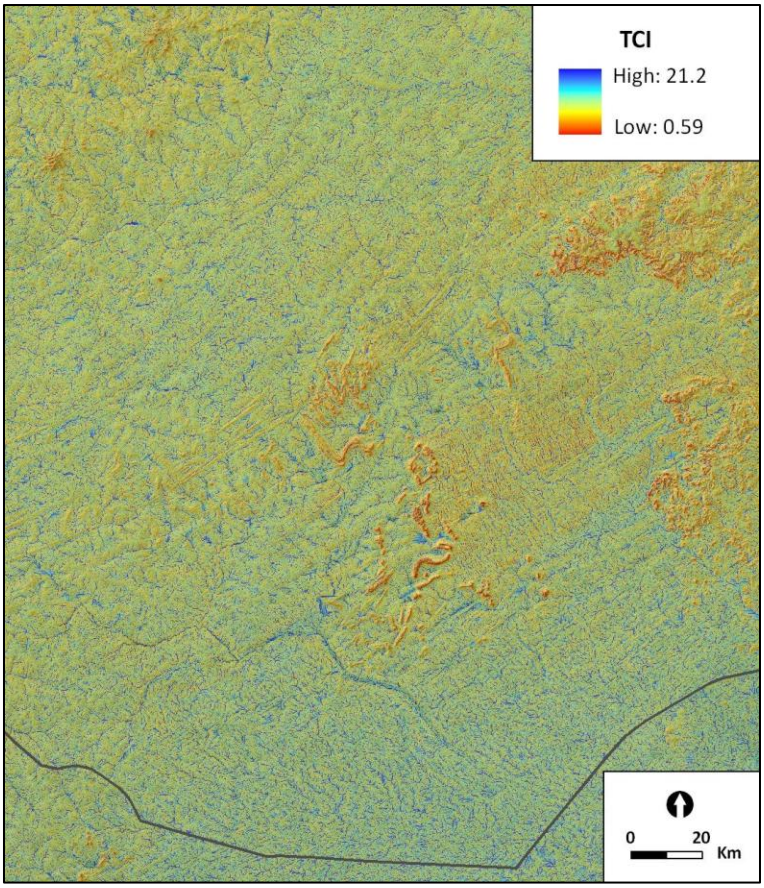


Figure 17: Topographic convergence index – Zinder region.

4.7 Landforms

Geomorphologic landforms that may be useful for identifying groundwater recharge and discharge areas are easily computed from DEMs. There are a variety of methods for calculating landform classes, Pennock et al. (1987) provides one example. The landform classifications from Jenness (2006) were generated from the DEM for the study area using Tom Dilt's Topography Tools (TT), which is available at ESRI's ArcScripts website http://arcscripts.esri.com/ (Dilts, 2009). Ten landform categories were identified for the study area: canyons, shallow valleys, upland drainages, u-shaped valleys, plains, open slopes, mesas, hills in valleys, small hills in plains, and high ridges. Two moving circular neighborhoods are required for the landform algorithm, and were chosen to have radii of 7 and 21-cells, approximately 600 m and 2000m,

respectively. The algorithm gathers statistics on the DEM within the given radii. Larger radii average relief features to a greater degree, which highlights larger features, and the converse is true for smaller radii. The radii values used were determined by measuring major features in the study area such as the width of ridgelines and drainages, all of which were less than or equal to approximately 1500 m.

4.8 Vegetation persistence

Information on phreatophytic vegetation of the Zinder region was not found, so the principle of insensitivity of vegetation to precipitation fluctuations through time as an indicator of groundwater utilization guided the vegetation component of this study (Elmore et al., 2008). An idea similar to the standard deviation of NDVI (SDVI) developed by Tweed (2007) was used to analyze vegetation persistence between wet and dry seasons. Instead of SDVI, the average difference between peak wet season Enhanced Vegetation Index (EVI) and low dry season EVI, which was termed EVI Difference (EVID), was calculated from 10 years of data for the study area. MODIS Terra 16-day, 250 m, version 5 EVI (MOD13Q1) scenes were the foundational data used for this analysis. Subsequent permutations on the EVID calculations led to the final vegetation persistence variable.

Vegetation indices, such as EVI and NDVI exploit the spectral properties of green leafed vegetation, which absorb in the red (0.65 μm to 0.70 μm) and strongly reflect in the adjacent near-infrared range of the electromagnetic spectrum (Figure 18). Most vegetation indices incorporate some adaptation of the ratio of red reflected radiant flux, ρ_{red}, to near-infrared radiant flux, ρ_{nir}, which yields an index value of greenness or green leaf cover (Jensen, 2007). The EVI (Equation 3), a modification of the NDVI (Equation 4), is a standard MODIS product that uses the blue band of MODIS to correct the red band for atmospheric aerosol scattering, and includes a soil adjustment factor, L, a gain factor, G, and coefficients, C_1 and C_2 (Jensen, 2007).

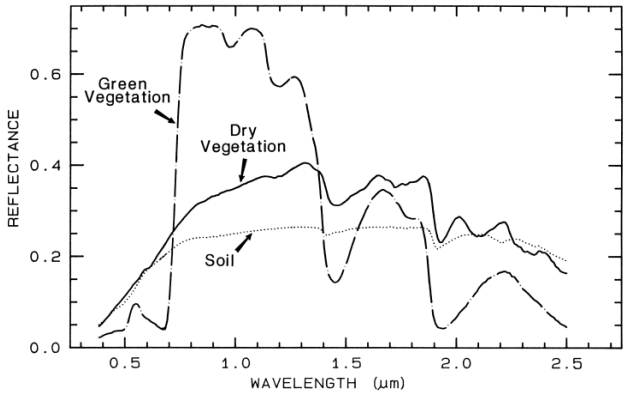


Figure 18: Vegetation reflectance curve (Clark et al., 2003).

$$EVI = G\,\frac{\rho^*_{nir} - \rho^*_{red}}{\rho^*_{nir} + C_1\rho^*_{red} - C_2\rho^*_{blue} + L}\,(1 - L)$$

Equation 3: Enhanced vegetation index (EVI)

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$$

Equation 4: Normalized difference vegetation index (NDVI)

MODIS was preferred to Landsat imagery for several reasons. In addition to the reasons stated in Section 4.1, MOD13Q1 scenes minimize the disturbance of clouds by averaging scenes acquired over a two-week period. Most importantly, MODIS allowed much better temporal coverage when compared to Landsat. MOD13Q1 scenes were available from early 2000 to the present. Issues with Landsat included the fact that Landsat 5 is long beyond its design life and scenes are not being retrieved by any international station for the study area. Also, since May 2003 the Scan Line Corrector (SLC) that compensates for the forward motion of Landsat 7 has been inoperable. This has caused a loss of data in Landsat 7 scenes acquired after this date. MOD13Q1 scenes are documented in detail on the National Aeronautics and Space Administration's (NASA) Land Process Distributed Active Archive Center (LP DAAC) website, and have been validated to be ready for scientific publication.

4.8.1 MODIS preprocessing

The Zinder region falls within MODIS tile 51018007 (column h=18, row v=7), which covers the majority of central and western Niger (Figure 19). Each MOD13Q1 granule (scene)

was downloaded from the NASA Warehouse Inventory Search Tool (WIST), which is linked to the LP DAAC website at https://lpdaac.usgs.gov/lpdaac/get_data. MODIS data are stored in the Hierarchical Data Format-Earth Observing System format (HDF-EOS), which is a stacked raster configuration that includes scientific data sets and metadata all in one file. Therefore, it was necessary to extract each data set of interest before subsequent analysis.

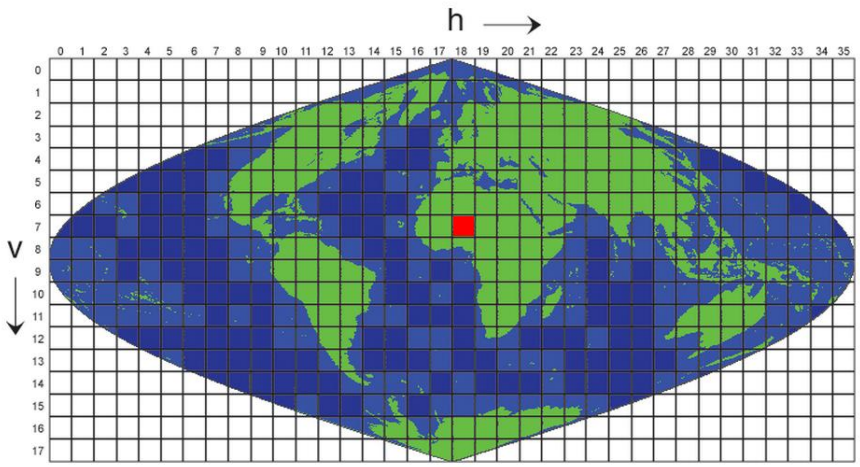


Figure 19: MODIS Sinusoidal grid with study tile highlighted in red.

MOD13Q1 version 5 scenes came in the Sinusoidal projection. All data were then batch extracted, projected to the custom Albers projection, and spatially clipped to the study area using a Perl script, which called the MODIS Reprojection Tool (MRT) (Appendix C). MRT was obtained from: https://lpdaac.usgs.gov/lpdaac/tools. Bilinear resampling was used to project the EVI data sets, and nearest neighbor resampling was used for Vegetation Index (VI) Quality metadata in order to preserve its thematic nature. Resampling was performed to 235 m cell resolution, which was the minimum integer value above the resolution of the Sinusoidal projected data. A multiplication factor of 0.0001 was applied to rescale the MOD13Q1 data to actual VI values of -1 to 1.

MOD13Q1 scenes come with metadata related to the quality of the VI data: Pixel Reliability and VI Quality. This information is embedded in the 16-bit decimal value for each

cell of the metadata sets. For this study, the VI usefulness information in the VI Quality metadata was used to mask out defective cells, so that only the highest quality data were used. Cells with VI usefulness levels (bits 2-5) equal to: highest 0000, high 0001 and good 0010 were retained. The bitwise AND operator, which is available in the GIS software, was used to extract the necessary VI usefulness bits from the decimal values in each metadata set. The resulting grids were then used to mask cells in the EVI grids that were defective. A check was made for known MODIS data quality issues, current maintenance, and fixes on the MODIS Land Data Discipline Team website at http://modland.nascom.nasa.gov/ as the final pre-processing step. No serious issues were found for the scenes used in this study.

4.8.2 Choosing peak and low EVI dates

In order to develop the EVID variable it was necessary to identify dates of average peak, wet season and stabilized low, dry season EVI levels for each year from 2000 to 2009. MODIS Terra 16-day, 1000 m, EVI data (MOD13A2) were used for this step. The average EVI value for every 16-day interval (one scene) from January 2000 to January 2010 was calculated for the study area by retrieving image statistics from only the study area extent of each scene in the GIS software.

The rainy season in southern Niger generally ranges from early June to mid-September. Consistent with the Nigerien wet season, EVI peaks for each year over the study area always coincided with the granule Julian dates of 225 or 241 (August 12th or 28th, respectively) (Figure 20). The average low, dry season EVI values were chosen to coincide with the dates that EVI began to stabilize, which always occurred before the absolute lows in April and May. Julian dates of 353 and 65 (December 18th and March 5th, respectively) were chosen for each year as the low EVI dates. Later comparison showed that the March scenes were cloudier than the December scenes, so the December 18th date was chosen for each year (Figure 20).

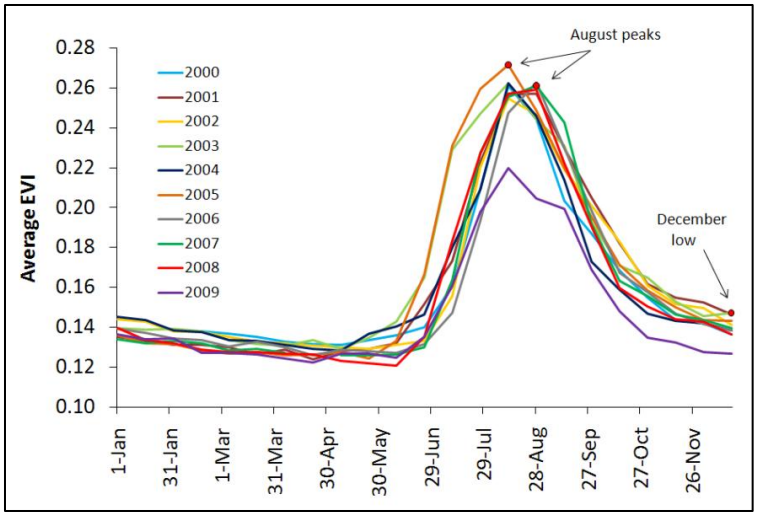


Figure 20: Plot of average EVI – Zinder region, 2000 to 2009.

4.8.3 Development of the EVI Difference (EVID)

The difference between peak, wet season EVI (EVI_p) and the subsequent low, dry season EVI (EVI_l) for each year was calculated (Equation 5), and then averaged for the 10 years. In order to do this the appropriate August (Julian dates 225 or 241) and December (Julian date 353) scenes for each year from 2000 to 2009 were acquired and pre-processed as described in Section 4.8.1. All subsequent calculations were done using several Python scripts to batch process the data (Appendix C).

$$EVID = \left(EVI_p - EVI_l\right)$$

Equation 5: EVI Difference (EVID)

Many scenes had patches of null cells because of defective data, so several extra steps were necessary in order to get the EVID average. The total count of valid cells over the ten images was calculated by reclassifying each EVID raster into 0's (nulls) and 1's (data present) and adding them together. The total sum EVID raster was then divided by the total count raster to get the average EVID at each cell. The results of the average EVID raster (Figure 21) highlighted a problem: that likely barren and vegetated groundwater utilizing areas (GW in

Figure 21) exhibited similar low values of EVID. The reason for this problem was that the relative magnitude of EVI for these different areas was not considered in the EVID calculations.

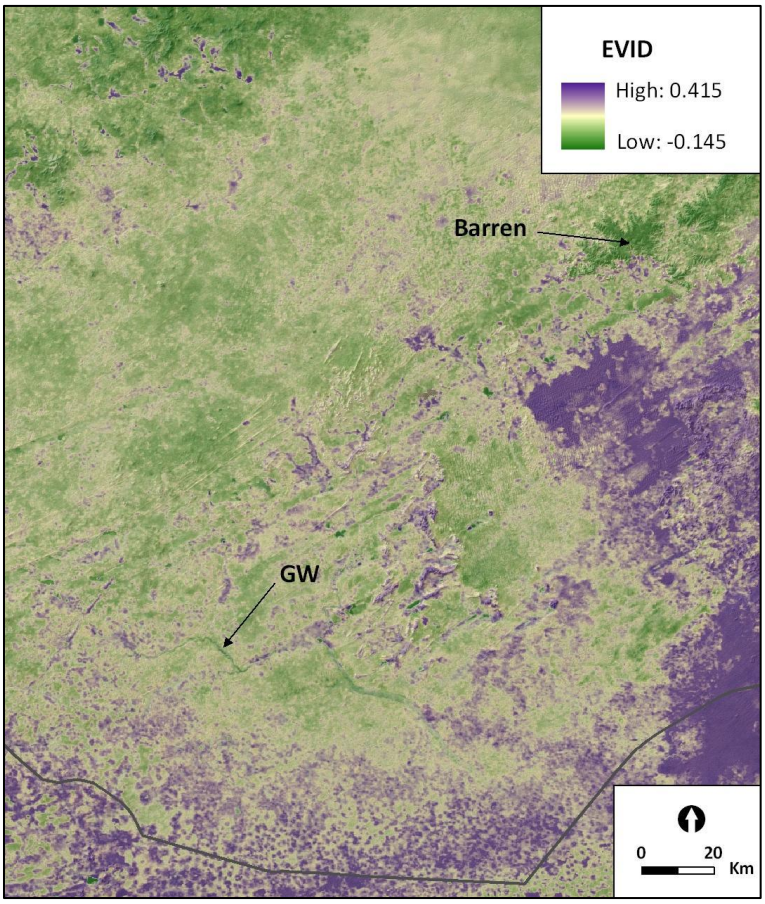


Figure 21: Average EVI Difference (EVID) – Zinder region, 2000 to 2009.

4.8.4 Development of the Vegetation Persistence Ratio (VPR)

To address this issue and discriminate low change in vegetation from low change in barren areas a ratio of EVI_p to EVID was developed, which was termed the Vegetation Persistence Ratio (VPR) (Equation 6). For example, a barren area should have a low magnitude EVI_p value and a low EVID value, yielding a relatively low VPR value compared with a vegetated groundwater utilizing area having a higher magnitude EVI_p and a similar low EVID. Summary statistics from training data gathered from likely barren, vegetated non-groundwater

utilizing, and vegetated groundwater utilizing zones in the study area were used to produce plots of VPR (Figure 22). These plots were then used to assess the effectiveness of the VPR to discriminate the different zones.

The zones of different vegetated conditions were digitized by comparing ESRI_Imagery_World_2D layer and dry season EVI to GWD values and choosing the most likely representative areas. These polygons were then used to calculate the mean and standard deviation of EVI_p and EVID within each zone for every year. The means were then used to calculate the VPR within each zone for every year. The VPR was then plotted as a slope radiating from the origin where y = EVI_p, and x = EVID (Figure 22). The plots revealed two remaining issues: 1) the ratio could be unstable near zero, possibly increasing exponentially, because of noise in the EVI values, and 2) VPR values for likely barren and vegetated groundwater utilizing areas remained similar for most years.

$$VPR = \frac{EVI_p}{EVID}$$

Equation 6: Vegetation Persistence Ratio (VPR)

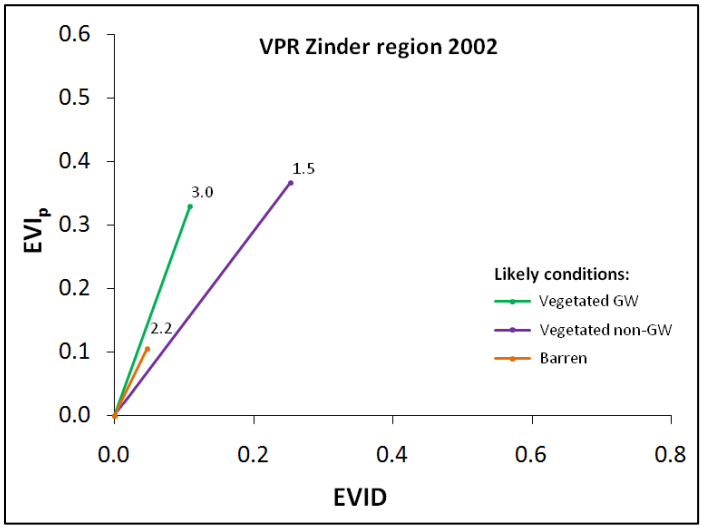


Figure 22: Plot of Vegetation Persistence Ratio (VPR) for three vegetation conditions – Zinder region 2002.

The first issue was dealt with by adding an offset ($offset_2$ in Equation 7) to the denominator, EVID, in order to shift the VPR slope away from the origin. Another offset ($offset_1$ in Equation 7) was added to the numerator in order to restore the original VPR slope. To address the second issue and separate VPR values of likely barren from likely vegetated groundwater utilizing areas, these offsets were optimized for each year. The concept was to find the smallest offsets that would cause the range of VPR values for barren areas to be less than or equal to the VPR of non-groundwater utilizing areas, while maintaining a positive VPR. In order to achieve this it was necessary to account for the variability around the means of EVI_p and EVID for the barren zones, specifically in the direction of the vegetated groundwater utilizing VPR slope.

$$VPR^* = \frac{(EVI_p + offset_1)}{(EVID + offset_2)}$$

Equation 7: Vegetation Persistence Ratio with optimized offsets (VPR*)

Within a spreadsheet, two standard deviations were added to the offset of EVI_p and subtracted from the offset of EVID for the barren zones. This shifted the VPR slope for the barren zones toward the vegetated groundwater utilizing VPR slope in order to assure that the optimization would equate or over-correct 98% of the barren VPR values with the non-groundwater utilizing VPR values (Figure 23). This optimization was performed using Solver in Microsoft Excel by changing either the numerator or denominator offset value with two conditions: 1) the absolute value of the difference between the non-groundwater utilizing and barren VPR values was equal to zero, and 2) the final VPR had to remain positive. Adding offsets to the VPR made the slopes less steep and the difference between VPR slopes of different zones smaller. Lesser offsets separated the slopes to a greater degree than greater offsets. The final equation developed for the Vegetation Persistence Ratio, including offsets, (VPR*) is shown in Equation 7. This routine performed well to numerically separate likely vegetated groundwater utilizing zones (GW in Figure 24) from likely barren zones (Figure 24).

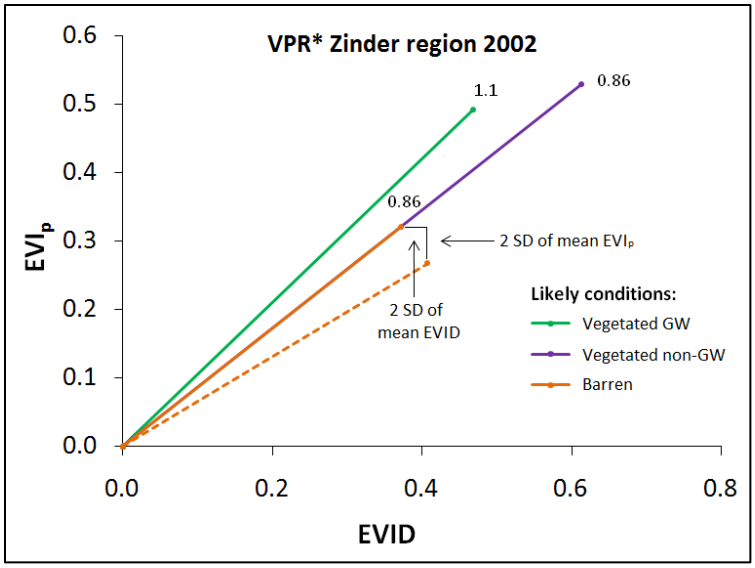


Figure 23: Plot of Vegetation Persistence Ratio with optimized offsets (VPR*) for three vegetation conditions – Zinder region 2002.

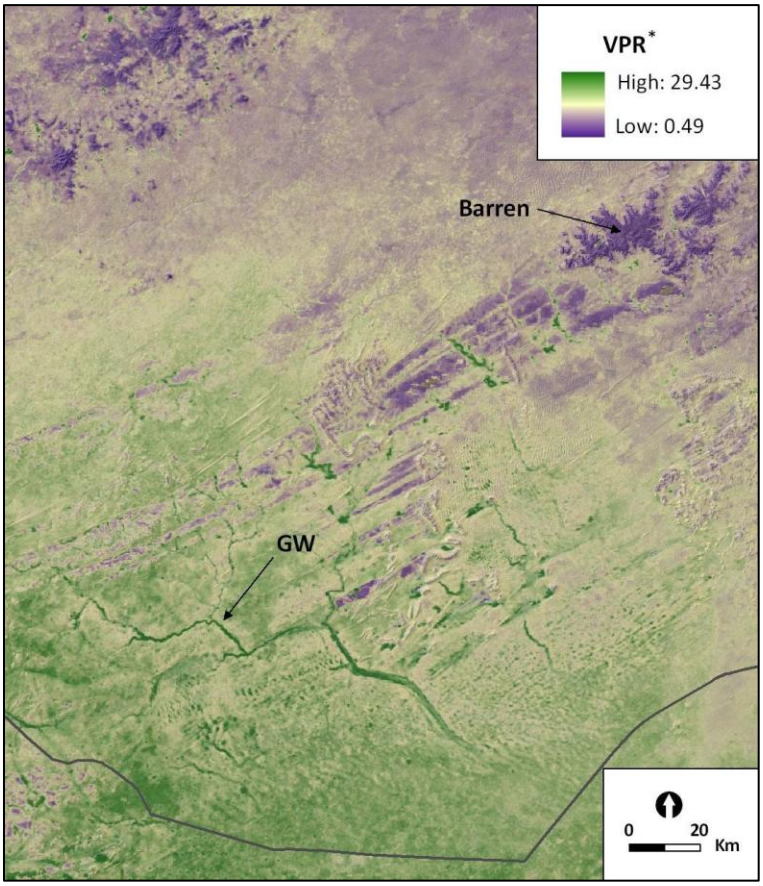


Figure 24: Average Vegetation Persistence Ratio with optimized offsets (VPR*) – Zinder region, 2000 to 2009.

4.9 Night land surface temperature

Cartwright (1974) found that regions of groundwater discharge in Illinois, USA were associated with increased soil temperatures in the winter time and decreased soil temperatures in the summer time (Figure 25). This is explained by the high specific heat of water, which dictates that more heat per unit mass is required to change the temperature of saturated soil than dry soil, so the temperature of wet soil changes less. Based on concepts from previous scientific research (See Section 3.2.4), a data set of average night land surface temperature (NLST) was created for a period immediately following the rainy season. Because Huntley (1978) concluded that shallow groundwater probably has minimal impact on diurnal temperature changes, an analysis such as this was not considered for this study.

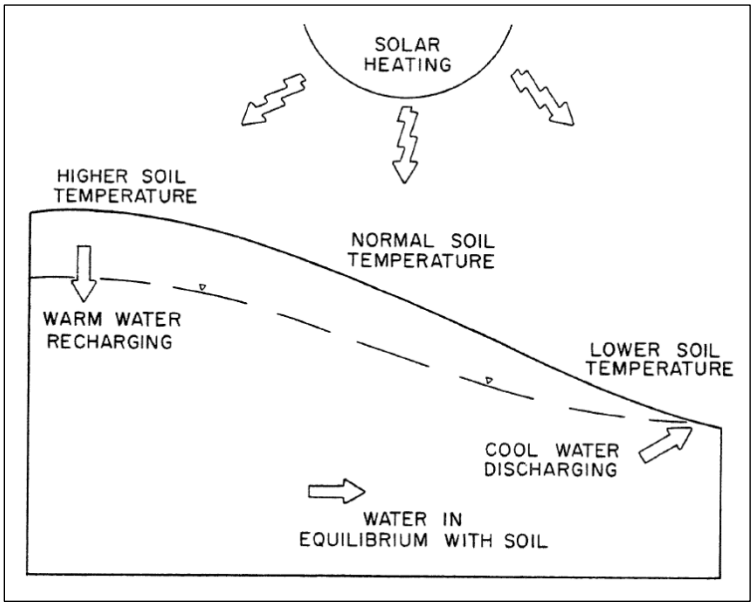


Figure 25: Generalized groundwater flow and temperature conditions during the hot season (Cartwright, 1974).

It was important to find scenes as close to the end of the wettest rainy season as possible, so that the groundwater tables in the study area could be assumed to be at their most shallow depths. Also, during October and November Niger experiences a short but intense hot season,

which was ideal to correspond with the conditions depicted in Figure 25. So, a series of MODIS Terra Daily Land Surface Temperature & Emmisivity, 1000 m, version 5 (MOD11A1) scenes, which spanned three days following the rainy season of 2005, were chosen (Julian dates 321-323 or November 17th to 19th) (Figure 26).

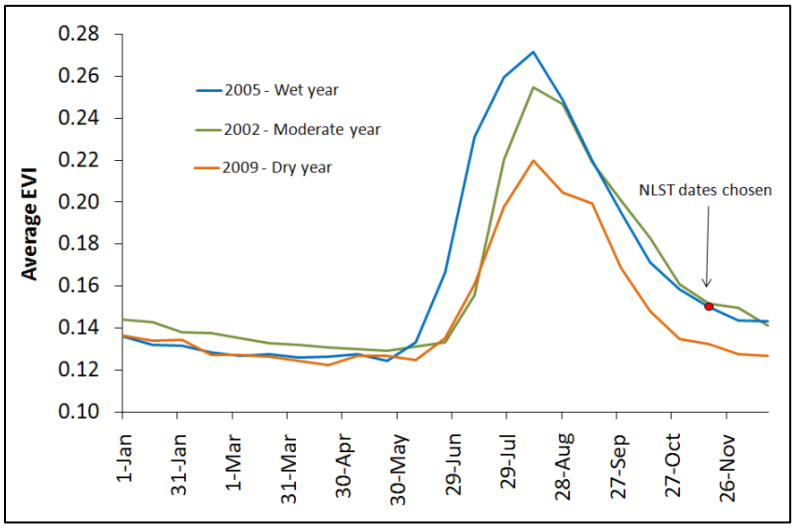


Figure 26: Night land surface temperature dates selected, November 2005.

All MOD11A1 scenes were pre-processed similar to the description in Section 4.8.1. However, the MODIS LST data and metadata have different multipliers and bit values. The raw data in °K was converted to °C by applying a multiplier of 0.02. The LST metadata included a range of error in °K for each cell. For all scenes the greatest error was found to be less than or equal to 3 °K, and the majority of each scene had less than or equal to 1 °K error. No effort was made to track this error. The NLST images all represent temperatures recorded near midnight. The final NLST variable was derived by averaging the temperatures over the three nights that were selected (Figure 27).

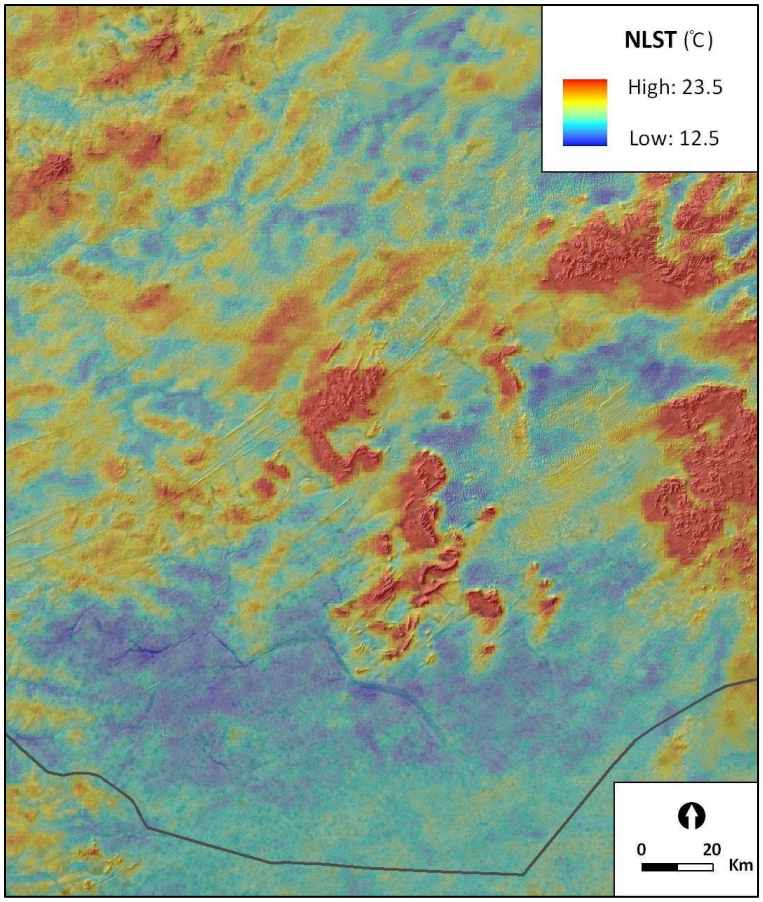


Figure 27: Average night land surface temperature for three nights – Zinder region, November 2005.

4.10 Proximity variables

A unique power of GIS is its ability to handle spatial relationships such as proximity, and this can be exploited in a study such as this to develop a host of spatially derived variables. Several proximity variables were produced for this study, which were considered to have a reasonable potential to predict GWD. Major ephemeral channels (Strahler 4th order and higher) were calculated for the study area by applying the hydrologic flow algorithms described in Section 4.6 to the SRTM DEM (Figure 28). Then the distance to major channels was calculated for each well point. Also, distance to lineament features was calculated for all well points.

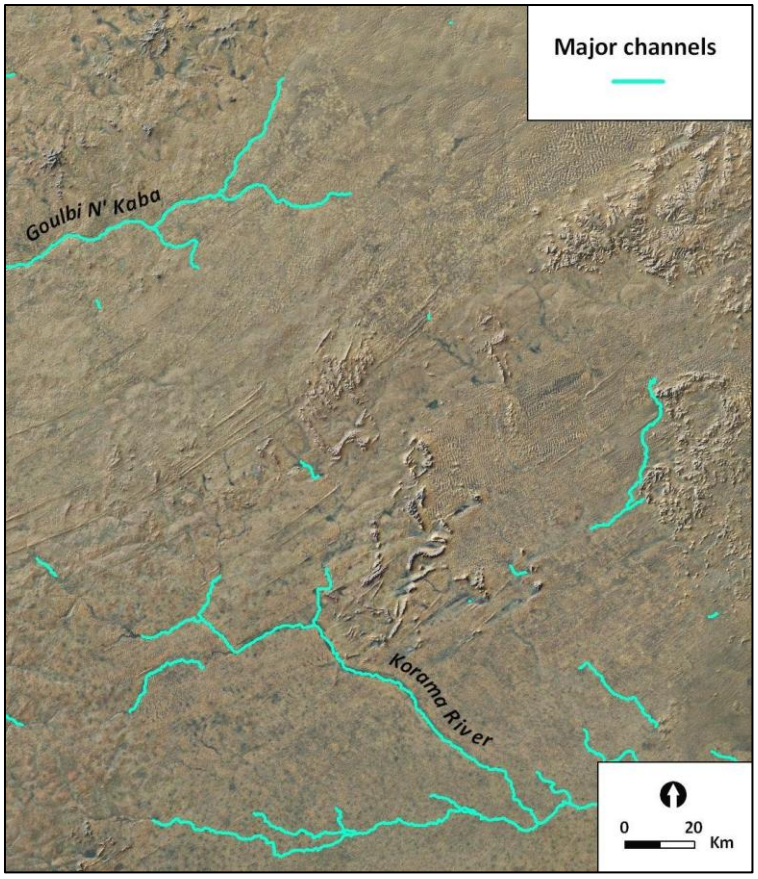


Figure 28: Major channels derived from the SRTM DEM – Zinder region.

Waterbodies could be important to consider in a study such as this, because persistent water surfaces are generally an outcrop of the local groundwater table. However, it is possible to have perched surface waters if low lying areas have underlying impermeable soil layers that separate the surface water from groundwater, which is possible for this study area. For this study, the MODIS Land-Water Mask data set (MOD44W) derived from 8+ years of MODIS Terra, 6+ years of MODIS Aqua 16-day composites, and the SRTM DEM were used to identify persistent waterbodies. Also, water strongly absorbs near and middle-infrared radiant flux and this can differentiated it in a satellite image. So, a threshold value corresponding to dark (wet) areas in Landsat band 7 (short wave near infrared) for the dry season of 2002 was used to identify ephemeral standing water or ponds. For each of these data sets a distance-to-waterbodies variable

was calculated for all the well points. More than one source for waterbodies was used in order to select the variable with the best linear correlation to GWD for the final analysis.

5 Analysis and Results

5.1 Choosing the analysis methods

The classification and regression tree (CART) approach discussed by Breiman et al. (1998) was used to evaluate the derived environmental variables because this model was able to handle both numeric and categorical data, was relatively easy to implement, and gave intuitive results. Tree models have been described as powerful for complex ecological studies that deal with a variety of data types and relationships which may not be linear and could exhibit higher order relationships (De'ath and Fabricius, 2000). Additionally, these models are insensitive to monotonic transformations of the data. These models can be used for descriptive or predictive purposes. Also, a method that would help to assess the statistical model for parsimony was desirable.

The backbone of this study was the statistical evaluation of the relationship of GWD data (the dependent or response variable) to the derived environmental variables (the independent or explanatory variables) described in Section 4 (Table 2). Exploratory and CART analyses provided the measure by which to evaluate the power of the environmental variables to predict GWD, and thus, be useful for mapping favorable areas for manual well drilling. The data analyses workflow is shown in Figure 29. Though the results of this study should only be applicable to the focus area and perhaps similar environments, these methods could be easily adapted. All data exploration and CART analysis was done in S-Plus 8.0 using the standard tree model tools and the rpart (recursive partitioning) library.

Once all the environmental variables were produced it was necessary to consolidate them with the GWD data. First, each environmental variable was converted to raster format and resampled to the lowest resolution, which was 90 m for the DEM. Then, GIS software was used to extract information from each variable layer to each GWD point. The GWD table was then split into validation and analysis sets as described in Section 4.3, using a random number generator and sorting in a spreadsheet.

Table 2: Environmental variables with variables selected for the final analysis shown in grey.

Variables	Data Type	Base Data
Dependent (response) variable		
Groundwater depth	Numeric (N)	Base de données piézométrique du Zinder
Independent (explanatory or environmental) variables		
Geology	Categorical	Carte Géologique de la République du Niger
Soil groups	Categorical	Carte Pédologique de Reconnaissance de la République du Niger, Zinder
Landforms	Categorical	SRTM DEM
Distance to waterbodies MOD	N - proximity	MODIS MOD44W Land Water Mask
Distance to waterbodies LS	N - proximity	Landsat 7 ETM+ band 7
Distance to 4th order channels	N - proximity	SRTM DEM
Distance to 5th order channels	N - proximity	SRTM DEM
Distance to lineaments	N - proximity	SRTM DEM
Lineament Factor	Numeric	SRTM DEM
Lineament density	Numeric	SRTM DEM
NLST - Average Nov 2005	Numeric	MODIS MOD11A1 Daily Land Surface Temperature
VPR* - Average 2000 to 2009	Numeric	MODIS MOD13Q1 16-Day Vegetation Indices
VPR* - Maximum 2000 to 2009	Numeric	MODIS MOD13Q1 16-Day Vegetation Indices
VPR* - Minimum 2000 to 2009	Numeric	MODIS MOD13Q1 16-Day Vegetation Indices
Topographic convergence index	Numeric	SRTM DEM

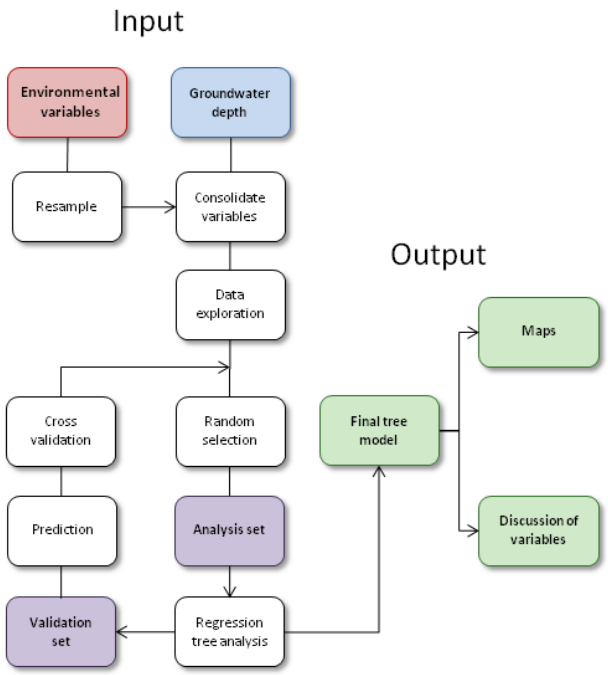


Figure 29: Workflow of data analysis.

5.2 Preliminary statistical exploration

5.2.1 Box plots

Preliminary statistical exploration was done using box plots to roughly assess the relationships between the derived environmental variables and GWD. This was the first step in distilling the list of most important variables to be included in the final CART analysis. Box plots are non-parametric graphical representations of groups of data based on the sample minimum, lower quartile, median, upper quartile, and the largest observation, and possible outliers (Khazanie, 1990).

The categorical variables of geology and soil groups showed the greatest heterogeneity when plotted against GWD, which was logical and indicated their potential usefulness for the final analysis (Figure 30 and Figure 31). The categories in Figure 30 and Figure 31 correspond to the geology and soil group data shown in Figure 11 and Figure 12, respectively, with slightly

different naming conventions for geology. The box plot of landform categories revealed a relative inability of this variable to discriminate GWD, and its potential weakness for the final analysis (Figure 32).

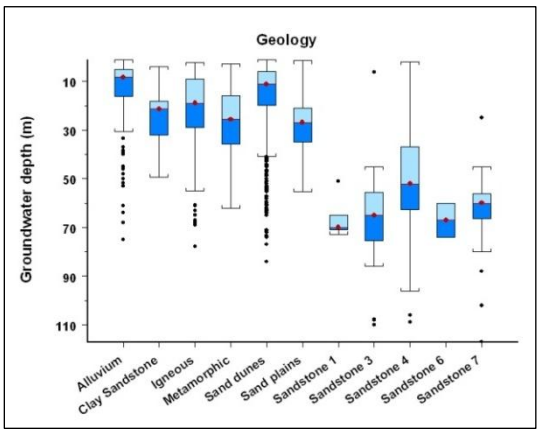


Figure 30: Box plot of groundwater depth and geology.

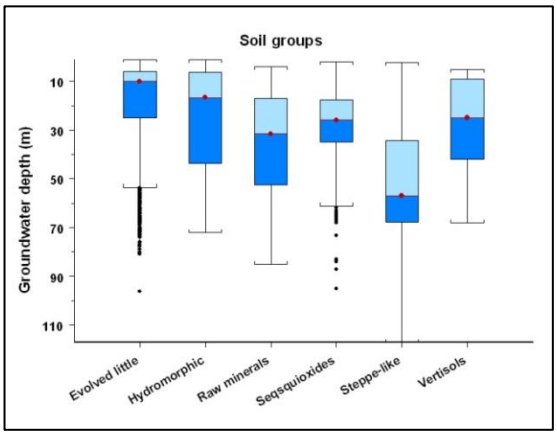


Figure 31: Box plot of groundwater depth and soil groups.

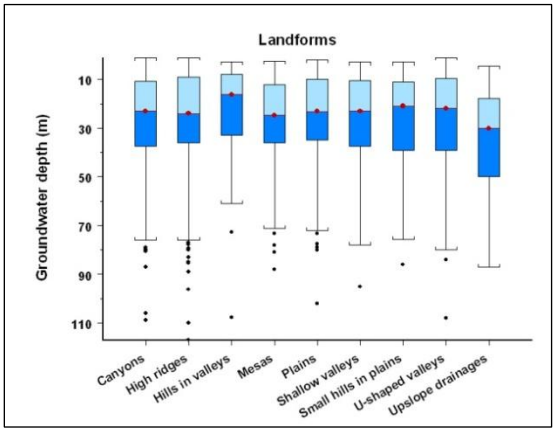


Figure 32: Box plot of groundwater depth and landforms.

All proximity variables were grouped into 200 m classes up to 5000+ m for the box plot analysis. None of these variables showed any noticeable discrimination of GWD in box plots, which suggested their possible lack of utility for the CART analysis (Figure 33 to Figure 35). These results seemed logical for the distance-to-major channels and waterbodies variables when considering the low relief terrain and lack of surface water typical of the study area. Box plots of

distance-to-5th order streams and distance-to-waterbodies created from the MOD44W data are not shown because they were similar to Figure 33 and Figure 34, respectively. The distance-to-lineaments variable was the weakest for differentiating GWD (Figure 35).

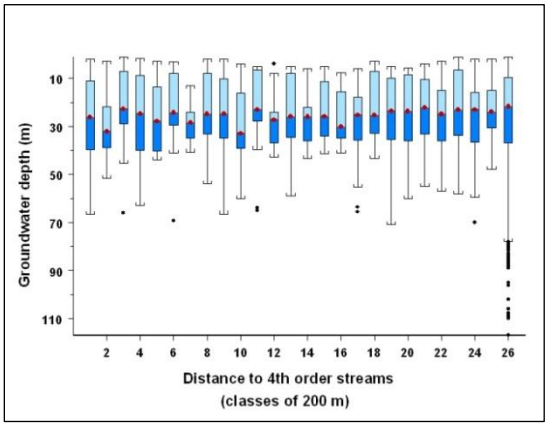


Figure 33: Box plot of groundwater depth and distance-to-4th order channels.

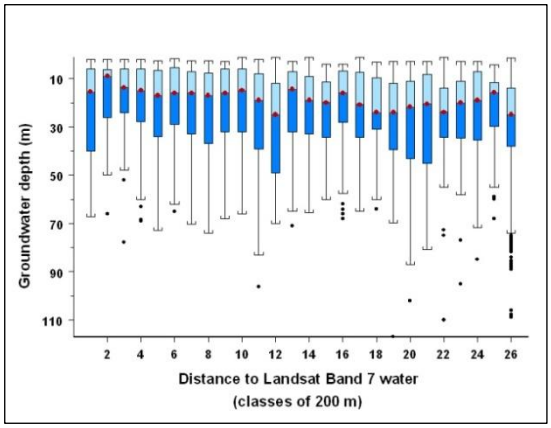


Figure 34: Box plot of groundwater depth and distance-to-Landsat band 7 waterbodies.

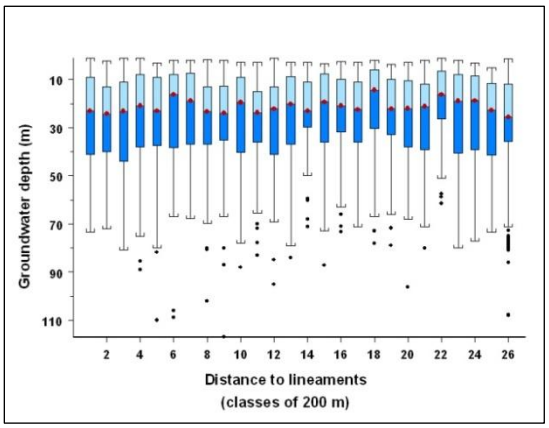


Figure 35: Box plot of groundwater depth and distance-to-lineaments.

The box plots of LF and lineament density (not shown) were very similar and distinguished little between GWD values with exception of a few tight plots (Figure 36). GWD can vary widely in a fractured rock environment, so it was expected that the lineament variables would generally have wide box plots. The box plot analysis of TCI showed little differentiation of GWD (Figure 37). The two tight plots in the high range of TCI were inconsistent for GWD,

which was another indication that TCI might be a relatively weak variable in the final analysis. The fact that these plots were relatively homogenous did not disqualify these variables for the final analysis, because a combination of environmental variables could have more power to predict GWD.

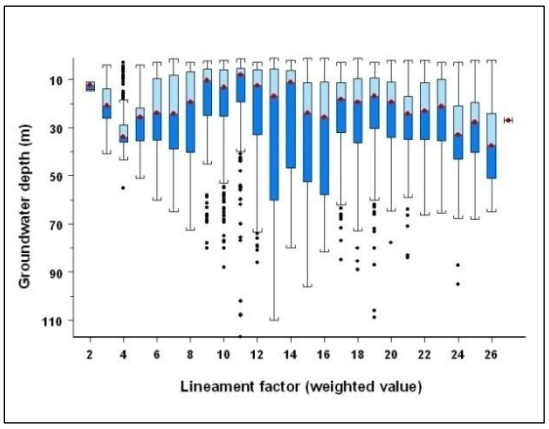


Figure 36: Box plot of groundwater depth and lineament factor.

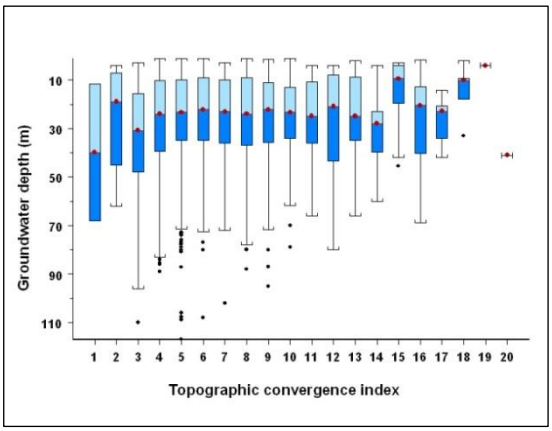


Figure 37: Box plot of groundwater depth and topographic convergence index.

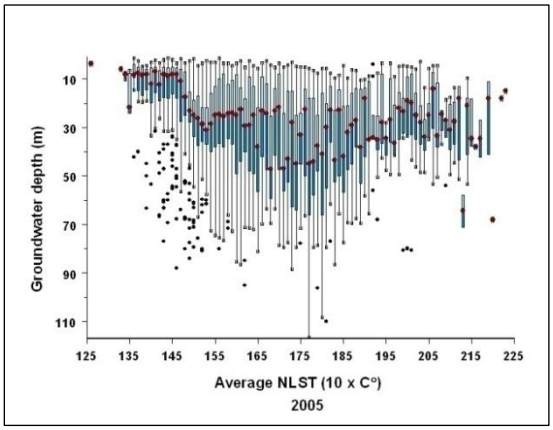


Figure 38: Box plot of groundwater depth and average night land surface temperature, 2005.

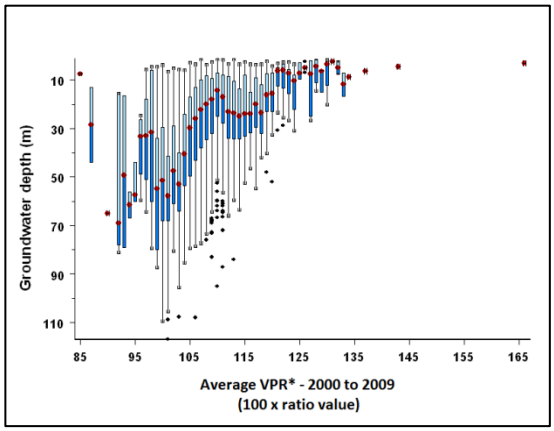


Figure 39: Box plot of groundwater depth and average vegetation persistence ratio, 2000-2009.

The box plot analysis of NLST showed that temperatures less than 14.5° C were associated with GWDs in a range equal to or shallower than 30 m (Figure 38). The VPR* box plot analyses were very similar for minimum, maximum and average VPR*, so only the average

VPR* plot is shown in (Figure 39). The average VPR* box plot revealed that high values of VPR* above approximately 1.20 also corresponded to GWD values less than or equal to 30 m. This discrimination of GWD in the shallow range, which is generally favorable for manual well drilling, indicated that NLST and VPR* could be important variables in the final analysis. Some preliminary CART analyses confirmed the relative strength of these two variables compared to the other environmental variables.

5.2.2 Correlation matrices

The numeric environmental variables were compared to each other in a correlation matrix to determine if any redundant variables could be eliminated from the final CART analysis (Table 3). To help with this selection and assess the degree of linear correlation with GWD, another correlation matrix was created with all the numeric variables and GWD (Table 4). A correlation matrix is a grid of correlation coefficients (0 = no correlation, 1 = perfectly correlated) that are associated with the two variables in a row/column combination.

Values of approximately 0.60 or greater were considered to be fairly strong correlations, and are outlined in Table 3. Redundancies among the numeric environmental variables became evident from the correlation matrix. All the environmental variables were strongly correlated within each group (See top row of Table 3), so only one variable from each group was sufficient to include in the final analysis. Redundant variables in the distance-to-major channels and VPR* groups with relatively low linear correlations to GWD (Table 4) were eliminated from the final analysis. The distance-to-waterbodies and lineament variables were the only variables from different groups that were strongly correlated. The LF variable was chosen for the CART analysis instead of the distance-to-waterbodies and other lineament variables despite having a lower linear correlation to GWD than these variables. This was done because persistent surface

water in the study area is almost non-existent, and the other lineament variables had similar near-zero correlations with GWD.

Table 3: Correlation matrix of numeric environmental variables.

	Distance to waterbodies		Distance to major channels		Lineaments			VPR*			NLST	TCI
	DSTWBMOD	**DSTWBLS**	**DST4**	**DST5**	**DSTLIN**	**LF**	**LINDNS**	**VPR***	**VPR*MAX**	**VPR*MIN**	**NLST**	**TCI**
DSTWBMOD	1.00	0.65	-0.22	-0.46	0.58	-0.84	-0.81	0.14	0.08	0.17	-0.17	0.08
DSTWBLS	0.65	1.00	-0.25	-0.35	0.63	-0.70	-0.69	0.21	0.15	0.26	-0.23	0.05
DSTSTR4	-0.22	-0.25	1.00	0.68	-0.27	0.35	0.33	-0.43	-0.30	-0.46	0.39	-0.10
DSTSTR5	-0.46	-0.35	0.68	1.00	-0.44	0.59	0.54	-0.43	-0.31	-0.41	0.38	-0.11
DSTLIN	0.58	0.63	-0.27	-0.44	1.00	-0.70	-0.69	0.27	0.18	0.31	-0.28	0.07
LINFCT	-0.84	-0.70	0.35	0.59	-0.70	1.00	0.96	-0.31	-0.22	-0.32	0.32	-0.08
LINDNS	-0.81	-0.69	0.33	0.54	-0.69	0.96	1.00	-0.31	-0.21	-0.35	0.34	-0.08
VPRAVG	0.14	0.21	-0.43	-0.43	0.27	-0.31	-0.31	1.00	0.87	0.88	-0.44	0.13
VPRMAX	0.08	0.15	-0.30	-0.31	0.18	-0.22	-0.21	0.87	1.00	0.64	-0.32	0.10
VPRMIN	0.17	0.26	-0.46	-0.41	0.31	-0.32	-0.35	0.88	0.64	1.00	-0.42	0.13
NLST05	-0.17	-0.23	0.39	0.38	-0.28	0.32	0.34	-0.44	-0.32	-0.42	1.00	-0.09
TCI	0.08	0.05	-0.10	-0.11	0.07	-0.08	-0.08	0.13	0.10	0.13	-0.09	1.00

Table 4: Correlation matrix of groundwater depth and numeric environmental variables.

	Distance to waterbodies		Distance to major channels		Lineaments			VPR*			NLST	TCI
	DSTWBMOD	**DSTWBLS**	**DST4**	**DST5**	**DSTLIN**	**LF**	**LINDNS**	**VPR***	**VPR*MAX**	**VPR*MIN**	**NLST**	**TCI**
GWD	0.23	0.18	0.24	0.20	0.02	0.02	0.04	-0.37	-0.30	-0.36	0.26	-0.03

Table 4 revealed that all the variables had weak linear correlations with GWD. Average VPR* exhibited the strongest correlation with GWD of -0.37. The fact that none of the variables had a strong linear correlation with GWD supported the use of CART, which can deal with more complex, non-linear relationships. Some cursory regression tree analyses supported the general indications of the box plots and correlation matrices with respect to the relative power of the variables to predict GWD. The variables selected for the final analysis are highlighted in grey in Table 2.

5.3 Regression tree analysis

5.3.1 Trees and pruning techniques

CART analysis, as stated in De'Ath and Fabricius (2000), explains the variance of a response variable by iteratively splitting the data set into increasingly less variable (more homogenous) groups using mixtures of numeric and/or categorical explanatory variables. In the case of regression trees, during one iteration the algorithm searches all the possible splits for each explanatory variable until it finds the split that minimizes the sum of squared error (SSE) or what is commonly referred to as deviance. The algorithm then continues on to the next iteration in both of the new groups, and so on. The resulting groups or branches of the tree are characterized by a threshold value of the response variable, the number of observations making up that branch, and the mean values of the explanatory variables that define that group (De'ath and Fabricius, 2000). The result of such an analysis is a hierarchical grouping that radiates out from nodes (split points), much like nested branches do in a tree, hence the name. The tree can be viewed graphically and provides a synopsis of the important relationships and threshold values. Despite the seeming applicability of tree models for a study such as this one, no evidence was found in the scientific literature regarding their use in groundwater exploration studies with the exception of the dendrograms used by Ghoneim (2008).

An important caveat to tree models is that they have the ability to over or under predict the data by growing trees that are either too large or too small. A heuristic approach to deal with this issue would be to manually prune back the tree to eliminate nested branches that do not make sense or do not explain enough variation in the response variable. Pruning can also be done automatically by setting a minimum node deviance, which stops the tree from growing more branches at this threshold value. Another technique involves minimal cost complexity pruning as described in Breiman et al. (1998). This method builds upon the use of a random subset of

approximately 33% to 50% of all the data for validation. A series of trees are developed from the larger (50% to 67%) portion of the dataset, which are then used to predict the response variable for the validation set, and errors are calculated (Breiman et al., 1998). The tree size with the lowest error can then be chosen. A similar technique was used in this study.

Minimal cost complexity pruning involves the use of the redistribution error R(T), which, for a regression tree, is the total SSE (Breiman et al. 1998). As De'ath and Fabricius (2000) clarify, for any value of the cost complexity parameter α (≥ 0) there exists only one smallest tree that minimizes $R(T) + \alpha|\text{Tree size}|$. Tree size is simply the number of terminal nodes in a tree. So, by sequentially increasing the value of α from 0 a series of nested sub-trees is grown, each being the best of its size. Each of these trees is then used to predict the response variable using the random sub-set of the data that was set aside for cross validation. For each of these predictions the total SSE is calculated and plotted against tree size. The smallest tree size within one standard error (1SE) of the tree with the minimum SSE is chosen (Breiman et al., 1998). Ideally, this routine is performed many times, perhaps 50 or more, in order to generate a frequency distribution of tree sizes, of which the modal tree size can be chosen. This can be computationally intense, so for this study only two series of cross validations were performed from which the best tree was selected using the 1SE rule.

5.3.2 Tree development

To assess the power of the final explanatory variables to predict GWD for the study area, a regression tree analysis with cross validation was performed. Using the minimal cost complexity pruning methods described above the best sized tree was selected. The final regression tree model was used to create several maps of favorability for manual well drilling in the study area.

Selection of the best sized tree was performed using the cross validation results shown in Figure 40. The relative error shown in Figure 40 is equal to the total SSE of a particular tree divided by the total SSE of the root or un-split dataset. A series of approximately 30 nested sub-trees were built with minimal cost complexity pruning to ensure that each tree was the best (least deviance) of its size. Certain tree sizes were unable to be built because the difference in deviance to the next greater tree sizes was too small (See tree sizes 12, 16, 21, and 23 in Figure 40). Each tree was then used to predict GWD using the validation set described in Section 4.3. The minimum error occurred at a tree size of 18 nodes. Then using the 1 SE rule, an optimal tree of 7 terminal nodes was chosen as the final model (Figure 40). This entire process was performed again with a different random set of the data, and this also revealed a 7-node tree as the best size.

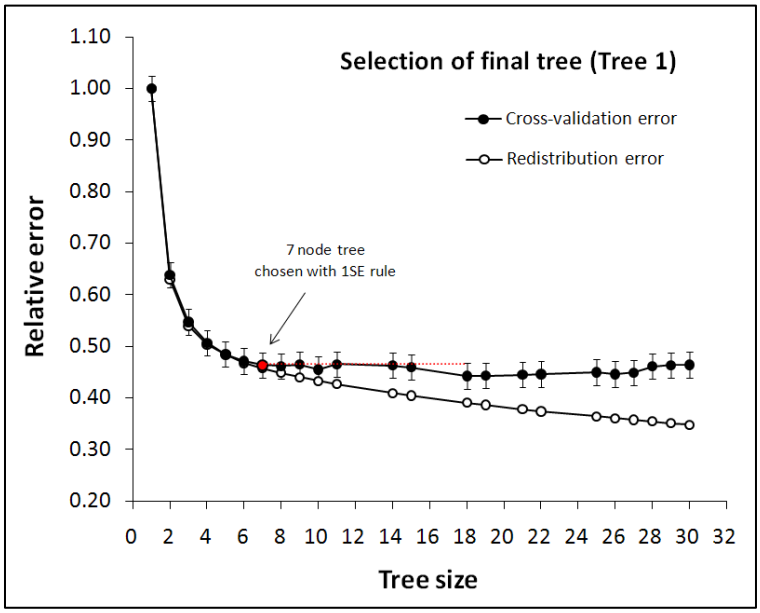


Figure 40: Cross validation relative error for selection of final regression tree (Tree 1).

This final 7-terminal node tree, Tree 1, is shown in Figure 41. The geology and soils codes used in Tree 1 are shown in Table 5. Trees are read by considering the value of the explanatory variable at each split and following the branch to the left if the condition is true and to the right if it is false. The terminal nodes (Denoted by green text in Figure 41) show the

average value of GWD (Denoted by blue text in Figure 41) predicted by the corresponding path of the tree, which is a combination of environmental variable conditions. Notably, all the split criteria of Tree 1 were logical in a physical sense. Also, shown at the terminal node is the number of data points within that node (Small black numbers in Figure 41).

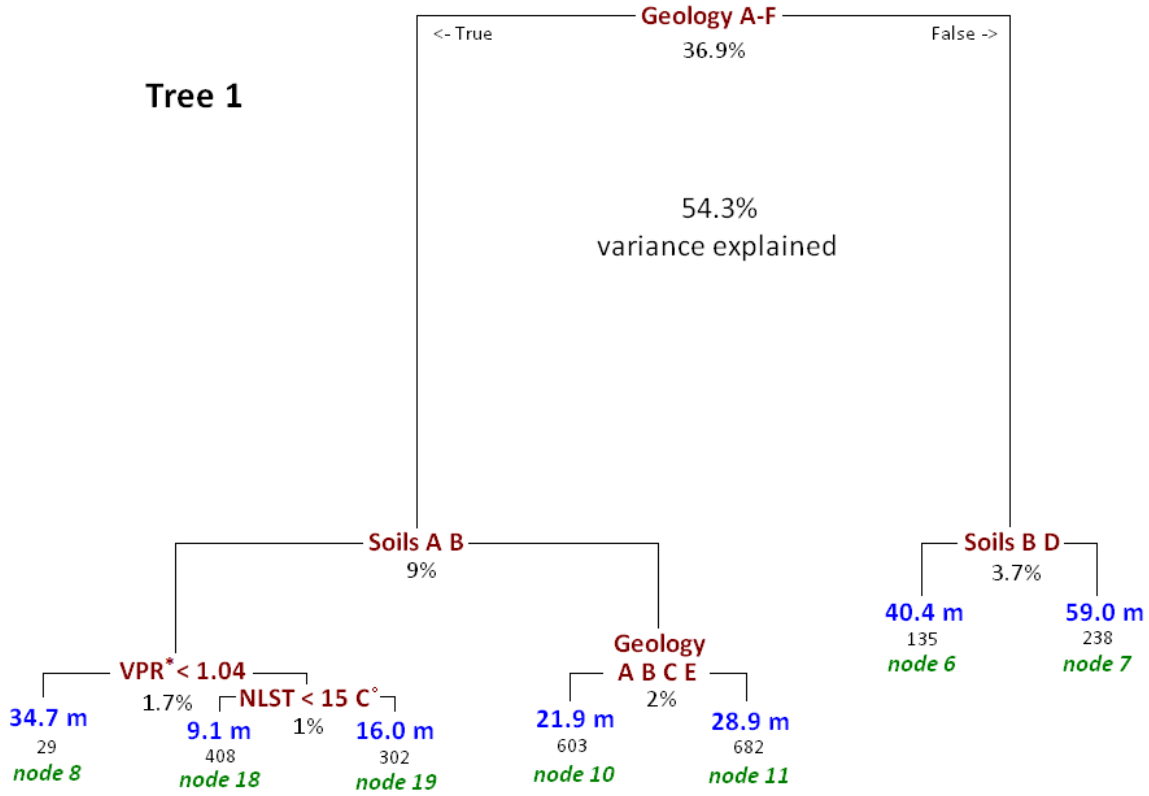


Figure 41: Final 7-node regression tree developed from all environmental variables (Tree 1).

Table 5: Codes for environmental variables in Tree 1.

Geology	Code	Soils
Alluvium	A	Evolved little / Well drained
Oligocene sediments	B	Hydromorphic
Igneous	C	Raw minerals
Metamorphic	D	Sesquioxides
Quaternary dunes	E	Steppe-like / Iso-humic
Quaternary plains	F	Vertisols
Cretaceous sediments	G-K	

The relative length of each tree branch represents the amount of variance (deviance) explained by each split. The six splits of Tree 1 explained a total of 54.3% of the root deviance of

the data set. This tree indicated that geology and soils were the dominant environmental variables for predicting GWD in the study area. Splits based on these variables comprised 51.6% out of the total 54.3% of the variance explained. Detailed results for Tree 1 are shown in Appendix D.

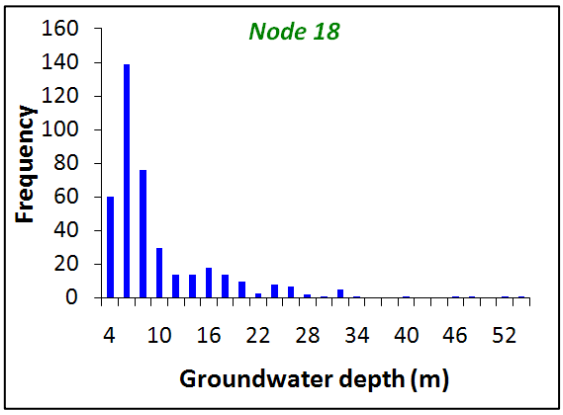


Figure 42: Histogram of groundwater depth – node 18 of Tree 1.

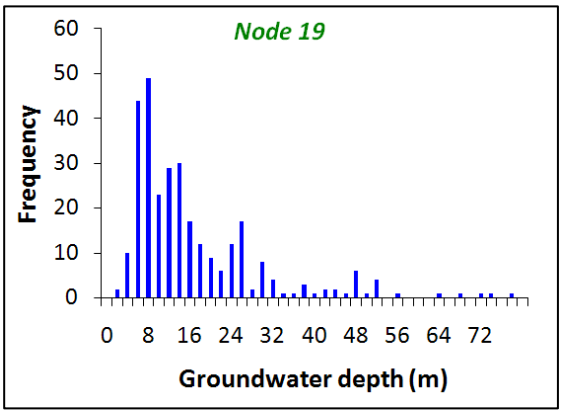


Figure 43: Histogram of groundwater depth – node 19 of Tree 1.

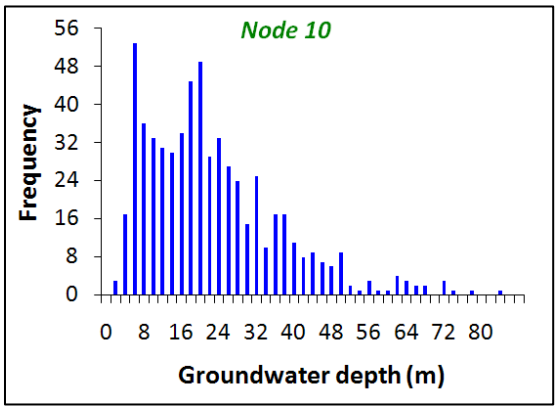


Figure 44: Histogram of groundwater depth – node 10 of Tree 1.

It was important to assess the frequency distributions of GWD at each terminal node to understand how the tree thresholds related to favorable conditions for manual well drilling. Generally depths shallower than 30 m are reasonable for manual well drilling, especially in softer geology. Notably, nodes 18 and 19 of Tree 1 had distributions of GWD strongly in the range that is highly favorable for manual well drilling (Figure 42 and Figure 43). These two terminal nodes were associated with relatively high VPR* values. The distribution of GWD for terminal node 10

was in the deeper range of favorable for manual well drilling, with one standard deviation above the mean equal to 36 m (Figure 44).

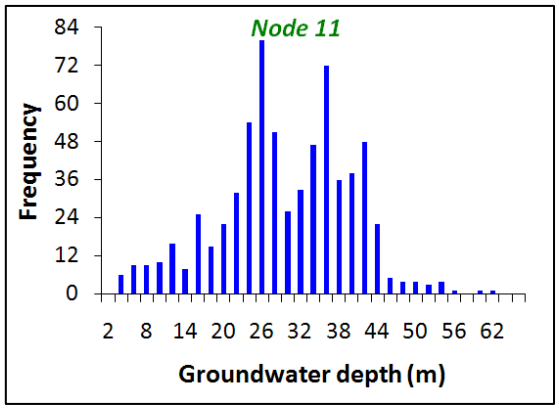


Figure 45: Histogram of groundwater depth – node 11 of Tree 1.

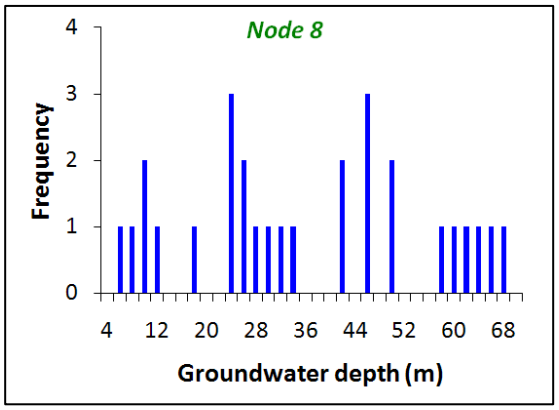


Figure 46: Histogram of groundwater depth – node 8 of Tree 1.

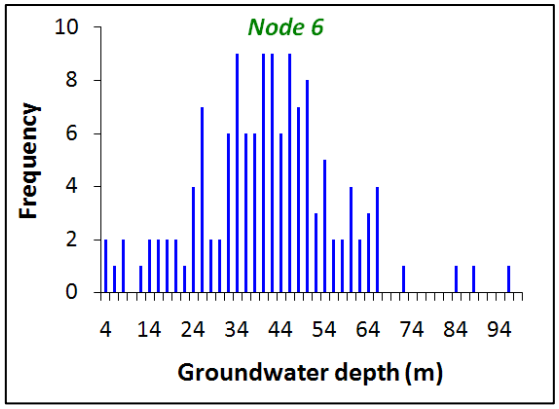


Figure 47: Histogram of groundwater depth – node 6 of Tree 1.

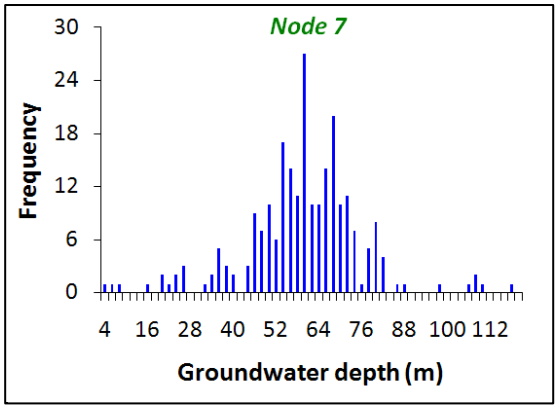


Figure 48: Histogram of groundwater depth – node 7 of Tree 1.

Terminal node 11 had a distribution of GWD with a range of 18 m to 39 m within 1 standard deviation around the mean, which is on the deeper end of favorable for manual well drilling (Figure 45). The remaining terminal nodes had GWD histograms mostly outside the range of favorable for manual well drilling (Figure 46 to Figure 48). Node 8 of Tree 1 was comprised of only 29 data points distributed across a large range of depth, which was weak for the purpose of this study (Figure 46).

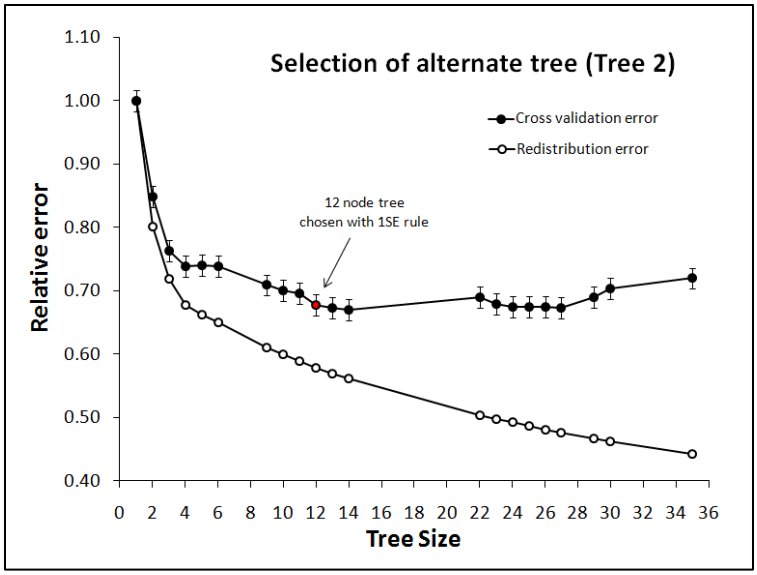


Figure 49: Cross validation relative error for selection of alternate regression tree excluding geology and soils (Tree 2).

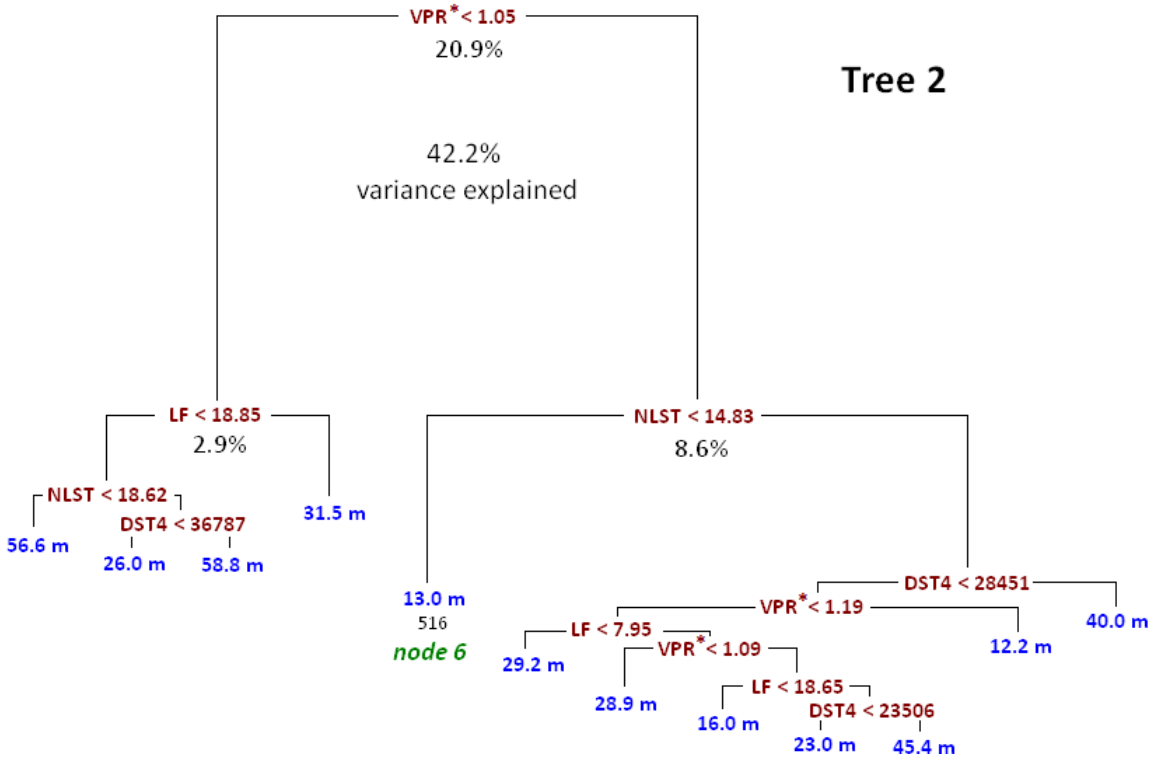


Figure 50: Alternate 12-node regression tree excluding geology and soils (Tree 2).

Another regression tree analysis excluding geology and soils was also performed. This was done to test the case where environmental variables derived from maps, which are not easily reproduced, are not available. All the other environmental variables are easily reproducible from satellite data. A best sized tree of 12 terminal nodes was selected using the same methods as for Tree 1 (Figure 49). The alternate tree, Tree 2 (Figure 50), explained 42.2% of the root variance, which was less than Tree 1. However, Tree 2 confirmed that the most important variables, aside from geology and soils, for prediction of GWD were VPR* and NLST. The same combination of these two variables again predicted a favorable range of GWD, with a range of 2 m to 25 m within 1 standard deviation around the mean of 13 m for node 6 (Figure 50).

6 Discussion

6.1 Identification of data and methods

The data mining phase of this study was a vital objective and a considerable effort that was accomplished using only a computer connected to the internet. All the data identified in Table 1 are freely available over the internet, on CD, or assumed to be available from governmental and other organizations within the area of focus. These data represent the full extent of data types that are publically available for projects of this kind. The list of spatial data portals compiled in Appendix A highlights most, if not all, of the key internet gateways for obtaining these and other data, especially with respect to African countries. Also a comprehensive list of the main types of spatial analysis techniques and the corresponding base data that have been used for groundwater exploration studies are listed in Appendix A. This list obviously does not represent the full realm of possibility of techniques, because with current GIS software an analyst is only limited by the deficiencies of the base data and his or her ideas, but it presents solid examples to use or build upon.

6.2 Strengths and weaknesses of the data and methods

6.2.1 Topography, geomorphology, and proximity to surface water

Haitjema and Mitchell-Bruker (2005) state that it is possible to have topographically controlled water tables in nearly flat terrain, but it does not seem likely that the Zinder region fits nicely into this mold. The relief of the Zinder region lies mostly between 350 and 450 m in elevation, with some deep unconfined aquifers. Preliminary data exploration (See Section 5.2) suggested that the TCI and landforms variables, those derived directly from the SRTM DEM, would not be helpful for identifying zones of shallow groundwater in this area. Also, these variables never factored into the output of the two tree analyses even at large tree sizes. Other variables derived from a DEM might have been useful in terrain such as this. For example, identification of paleo-drainages, which can form shallow buried alluvial aquifers, such as was observed in the Sudan, can be done using a DEM (Ghoneim and El-Baz, 2007). However, this was not done in this study because the hydrogeology map (See Figure 7) suggested that these features do not exist or are inconsequential in the Zinder region. These DEM derived variables could be quite useful in areas with more topographically controlled water tables, where relative impermeable geology and greater relief could lead to a stronger manifestation of shallow groundwater in topographic lows. As DEM resolution improves, the applicability of variables such as the TCI may improve (Quinn et al., 1995). Additionally, the DEM derived variables were among the quickest and most easily developed, which suggests they can be experimented with in any setting with minimal cost in time and effort.

The distance-to-4th order channels (DST4) variable was similarly as weak as the topographic variables in terms of its relationship to GWD in the study area. The Korama River drainage, the jagged, Z-like stream channel evident in the middle of the lower third of the study area that has shallow GWD near it, was one possible exception (See Figure 28). The DST4

variable was included in the final analysis in part because it best captured the prominence of this drainage and the Goulbi N' Kaba stream channel in the northwest of the study area. Neither of these drainages has perennially flowing water, but they transmit water in the rainy season. The DST4 variable did factor into the alternate regression tree analysis, but explained little of the deviance in the GWD as indicated by its relatively short branch lengths (Figure 50). It was apparent for this study area that distance-to-surface water variables had little to no utility for mapping favorable areas for manual well drilling.

Distance-to-surface water variables might be more useful in an area which strongly exhibits topographically controlled water tables, where perennial streams and lakes intersect the local groundwater table. It is true that even in such a setting groundwater tables are not well correlated with local scale topography because groundwater hydraulic gradients are controlled by other factors as well (Haitjema and Mitchell-Bruker, 2005). However, streams and waterbodies are usually manifested in topographic lows so, because these variables are very simply derived, they should be considered for manual well drilling favorability studies in certain areas.

6.2.2 Geologic lineaments

The development of the LF variable was perhaps the most time consuming and computationally demanding aspect of this study despite its conceptual simplicity. It is likely that the initial development of a data set like this will be more resource intensive compared to subsequent applications. The LF variable was not represented in Tree 1 and described a fair amount of the deviance in Tree 2 despite performing inconsistently (Figure 41 and Figure 50). For example, a close examination of Tree 2 revealed that lesser LF values described both greater and lesser GWD values.

The intention of the LF variable for this study was to highlight areas of greater fracture density, which in turn could indicate areas of greater weathering of basement rock and the

potential for increased secondary porosity, both of which can be favorable for manual well drilling. For this study, the only means available to assess these assumptions was to statistically describe the relationship of high fracture density to GWD, but this method was deficient. Fractured bedrock aquifers by nature are highly discontinuous, so an area of high fracture density can and will exhibit widely varying GWDs, not only because of the heterogeneity of the geology, but because of the bias of well types and their locations. This was another reason that the distance-to-lineaments variable was eliminated from the final analysis.

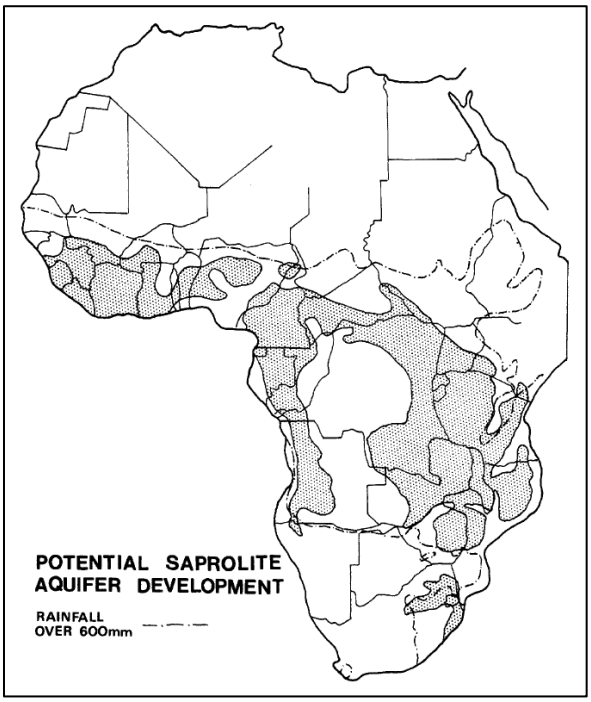


Figure 51: Potential saprolite aquifer development in Africa shaded in grey (Jones, 1985).

The spatial scale of lineament studies is an important consideration. Using a combination of data at multiple scales such as was presented in Hardcastle (1995), can better emphasize the local scale variation of a hard rock environment that could be favorable for manual drilling. At small spatial scales, such as 1:250,000, lineament analyses like the LF are most useful for broadly identifying harder, more brittle geologic formations. This information can be useful especially in

climatically wetter regions that have greater chemical weathering and resultant saprolite aquifers in areas of high fracture density (Jones, 1985) (Figure 51). For example, areas of high fracture density found in appropriate climatic conditions could be simply outlined on a favorability map.

It is also important to note that the time spent producing the LF variable was excessive with respect to its performance relative to more simply derived variables. The correlation matrix (Table 4) indicated that a simple lineament density calculation, which is easily implemented in the GIS software, could potentially work just as well as the LF. Visual results from the LF and lineament density were virtually identical. When these variables were normalized 73% of all pixel values were within 10% of each other, which supported the qualitative observations. When combined in several test regression tree analyses these lineament variables were consistently mingled at similar levels in the tree, no one being clearly more powerful than the other. Thus, for broad scale mapping of favorability for manual well drilling a simple lineament density variable should suffice. If the mapping is proposed for a smaller area, and sufficient high resolution data are available, the LF could highlight favorable areas better than lineament density, but at a higher cost in time. Ultimately, this study revealed no more information than what has already been noted about similar lineament studies, the short-comings and strengths of which are examined in numerous studies especially Waters et al., (1990).

6.2.3 Night land surface temperature and vegetation persistence

The NLST variable showed some striking visual correlations with GWD for the study area, particularly for the shallow aquifer surrounding the Korama drainage in the south (Figure 52). However, NLST had a low quantitative correlation with GWD of 0.26 (Table 4). When compared to the geology and soils in Tree 1, NLST showed similarly weak predictive power. The cross validation error showed that with the inclusion of NLST to Tree 1 the relative error improved only 0.08%, compared to 51.5% for geology and soils (Figure 53). Also, NLST only

comprised 1% of the variance explained in Tree 1 (Figure 41). However, NLST did show strongly in Tree 2 (Figure 50), which revealed that it has some power for predicting GWD. More importantly, the combination of NLST and VPR* in the tree analyses was promising for predicting shallow GWD, and is discussed further below.

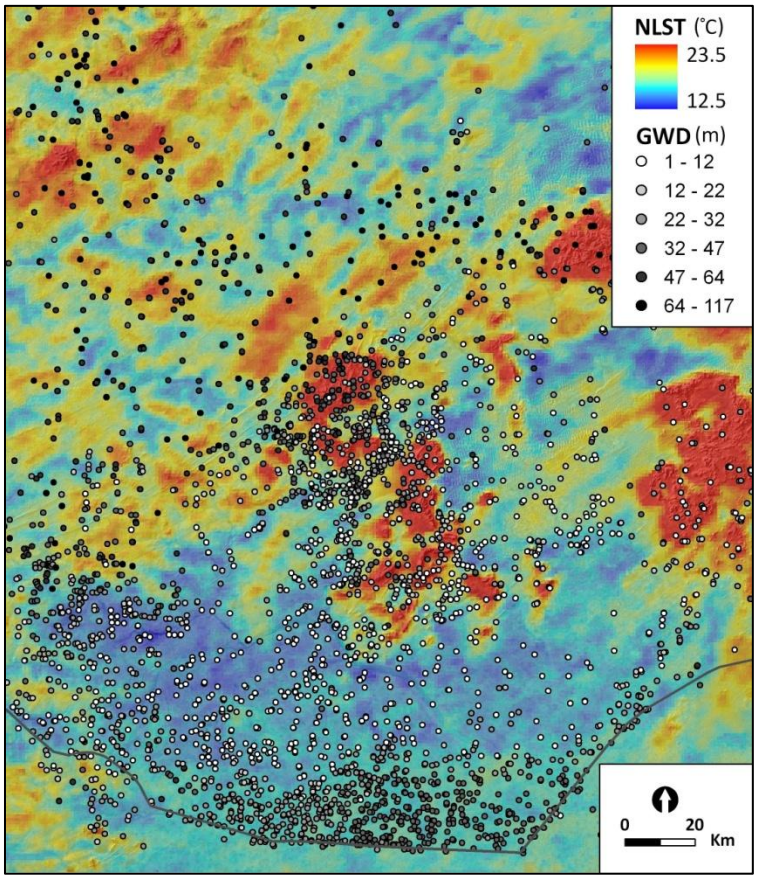


Figure 52: Night land surface temperature (Nov 2005) and groundwater depth – Zinder region.

A distinct shelf in the GWD data in the southern part of the study area marks the transition between separate parts of the Korama aquifer, and this is evident on the NLST data (Figure 27 and Figure 52). This suggested that NLST might actually be more correlated with certain geologic conditions. Figure 54 shows an overlay of Quaternary geology (primarily sands) and NLST, which is intriguing. However, more research is necessary to investigate these patterns, which was beyond the scope of this study.

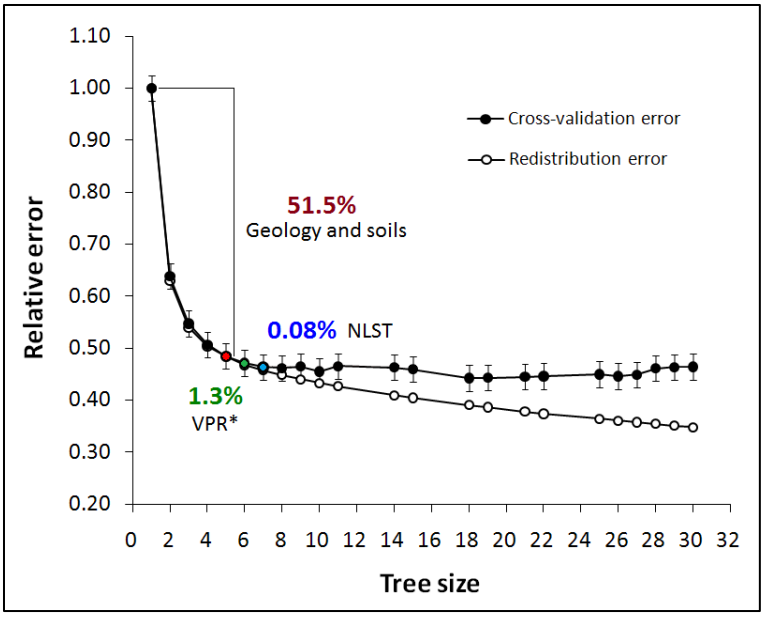


Figure 53: Relative error improvement for environmental variables in the final regression tree (Tree 1).

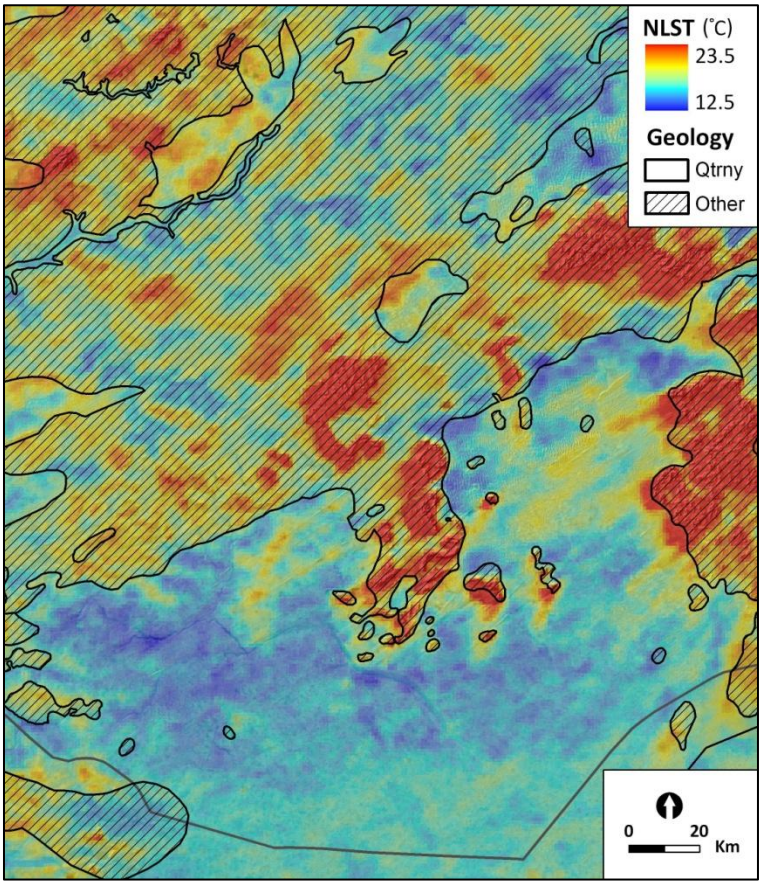


Figure 54: Night land surface temperature (Nov 2005) and Quaternary geology – Zinder region.

For the vegetation variable of this study, it would have been ideal to identify the phreatophytic vegetation common to the study area, and take the approach of classifying particular species using multispectral imagery. However, information on the groundwater dependent species in the Zinder region is probably non-existent. This scarcity of data was a primary reason for developing a more general vegetation variable like the VPR*. Also, the cost to benefit ratio of obtaining the necessary data on phreatophytes in a given study area could be high and should be assessed for other favorability mapping studies. The software expertise and computational power to do spectral signature classifications with the necessary high resolution data such as Landsat ETM+ could also be prohibitive for rapid assessments for favorable areas for manual well drilling. These reasons lend support to the use of a variable such as the VPR* for favorability mapping, especially if it can be applied in different areas. The VPR* could also have other potential uses, such as distinguishing areas of potential recharge and discharge.

From the data exploration and tree analyses it was apparent that average VPR* outperformed the other variables, with exception of geology and soils, in its ability to discriminate GWD. The box plots of VPR* versus GWD were tight and in the favorable range of GWD for high values of VPR*. Tree 2 revealed that in the absence of geology and soils average VPR* comprised the largest portion of the GWD variance explained by the tree at approximately 23% (Figure 50). However, linear correlation between GWD and VPR* was low at -0.37 (Table 4). Also, cross validation for Tree 1 revealed that the relative error improved only 1.3% with the addition of VPR* (Figure 53). Similarly, beside geology and soils in Tree 1, VPR* comprised only 1.7% of the variance explained (Figure 41). The lack of a linear relationship with GWD and the relatively weak contribution to Tree 1does not invalidate the utility of VPR* as an indicator of shallow groundwater.

Assuming the VPR* has good utility for favorability mapping for manual well drilling, experimentation needs to be done in the area of interest and then proper thresholds determined. If no GWD data is available, this could be problematic. The regression tree analyses from this study could offer some insight for determining approximate thresholds. The initial split for VPR* in Tree 2 (Figure 50) was approximately 1.05. This value is approximately at 64% percent of the maximum average VPR* value, which might be a threshold that relates to shallow GWD more universally. Perhaps a combination approach with NLST and VPR* data could better reveal areas of shallow groundwater in other areas like the relationships at node 18 of Tree 1 (Figure 41) and node 6 of Tree 2 (Figure 50) suggest.

As shown in Tree 1 (Figure 41) and the histogram for node 18 of Tree 1(Figure 42), the combination of relatively high average VPR* values and low NLST values strongly predicted shallow GWD. These results were valuable for creating favorability maps for manual well drilling for the study area, especially when combined with the geology and soils data. Despite the relatively weak contributions of VPR* and NLST to the variance explained and relative error improvement for the tree analysis, it is clear the relationship of high VPR* and low NLST could be the most important in terms of locating shallow groundwater areas. This is supported by node 6 of Tree 2 (Figure 50), which has a distribution of relatively shallow GWD values with the exclusion of the more powerful geology and soils variables. Additionally, these two variables relate to physical conditions more accurately than generalized geology and soils maps, and for this reason are potentially able to highlight pockets of shallow groundwater that would otherwise be missed. These results indicate a need for additional research regarding these relationships.

6.2.4 Geology and soils

Geology and soils explained the greatest proportion of the variance in the GWD data for the final tree at 51.6% out of the total 54.3% (Figure 41). This is logical because geology and soil

conditions are the major controls on infiltration and groundwater accumulation. For example, softer, younger, unconsolidated formations, such as the unconfined Quaternary aquifer in the Korama River basin, are permeable and if underlain by less permeable formations may exhibit shallow GWDs (less than 20 m). Also, older sandstone formations, such as the Continental Hamadien aquifer system, which underlies the northern extent of the Zinder region, are less permeable and may contain water that was recharged many thousands of years ago and have GWDs that generally exceed 30 m.

In terms of identifying favorable areas for well drilling of any kind, the regression tree results from this study reinforced what is apparent, that geologic data is the most robust for this purpose. All other land surface attributes seem capable of only marginally adding information content, albeit this information might be vital for locating pockets of shallow groundwater that might otherwise be missed as the VPR* and NLST combination indicated. It was beyond the scope of this study to perform in depth mapping of surficial geology using multi-spectral imagery. However, this approach has been used in other groundwater exploration studies (Al Biely et al., 2000). Although regions in the developing world could probably profit from high resolution updates to currently available geology maps, a side benefit of which would be improved favorability mapping for well drilling, this approach was considered to be too specialized, and thus, not a viable choice for the Zinder region. Geologic maps are widely available and, as this study showed, sufficiently useful.

6.2.5 Regression tree statistical approach

The regression tree approach was very useful for this study where no variables exhibited a linear relationship with GWD. All tree analyses were done in S-Plus statistical software, which requires a license. However, with some training these analyses can be implemented using the open source statistical software “R”, which is available on the internet. CARTs are relatively

simple to implement and in a synopsis provide a wealth of information about how each independent variable explains the variance of the response variable. Sometimes the top variables in a tree do not remain consistent over multiple tree runs with the same data set, which was not the case with this study. If so, the use of a random forest model can help rank the best variables. With respect to favorability mapping for manual well drilling, it was beneficial to have threshold values for the most important variables, so that the data could be easily structured into a map. Even if the thresholds were not used for the creation of a map, the trees gave a good sense of the relative importance of each variable, which would help with a heuristic weighting approach for mapping.

The fact that future favorability mapping efforts in other areas of Africa may not have access to adequate GWD data is problematic. Without a way to assess the relative importance of environmental variables with respect to favorable areas for manual well drilling, the weights and/or thresholds used to create favorability maps have to be based on expert opinion or possibly gleaned from a study such as this. The relative importance of the data sets and variables developed for this study should be applicable in similar climatic and geologic settings, but this may be optimistic. If even just a minimum of well distributed GWD points exist for an area to be mapped, then CART analysis can be used to assess the relative predictive power of variables similar to what was presented in this study. In this way, a map covering the whole study area can be generated using the continuous environmental variables as proxies for GWD. This is a good approach because the environmental variables are related directly to physical conditions in the area of interest as opposed to interpolated GWD values, which are not real. In fact, the relative error from interpolating 100 random GWD points for the Zinder region was greater than using the environmental variables as a surrogate for GWD (Figure 55), which supports the regression tree approach. If the GWD data set is dense for the area of interest interpolation may have less error

than the regression tree approach, as in this study (Figure 55), but again without the relationship to real physical characteristics of the landscape.

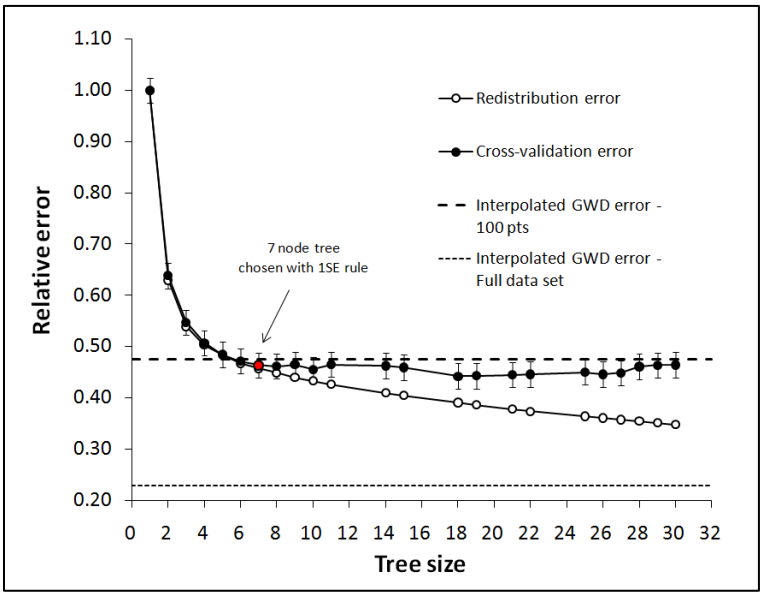


Figure 55: Relative errors for interpolated groundwater depth and final regression tree (Tree 1).

6.3 Favorability mapping procedure

An important observation from this study was the lack of quality information contained in the DEM derived variables with respect to GWD. The fact that topographic attributes would not work well in the relatively flat Zinder region is not surprising and seemed obvious even from the beginning, but could not be confirmed without testing. It is clear that if distinctions such as this can be made at the beginning of a favorability mapping effort, then time and effort can be saved by focusing on the variables that will be most meaningful in the project area.

A base layer of geology equal to or preferably larger than the scale used in this study of 1:2,000,000 is a good starting point for favorability mapping for manual well drilling. GWD data is an equally important component. These two core data sets, as shown in this study, will contain the majority of information that is useful for favorability mapping for manual well drilling. So, it is proposed that the simple criteria put forth by Haitejema and Mitchell-Bruker (2005) to

categorize topography versus recharge controlled water tables (Figure 56) using commonly available hydrogeologic parameters be used as the first step in determining what type of environmental variables to produce as complementary or proxy data for the favorability mapping.

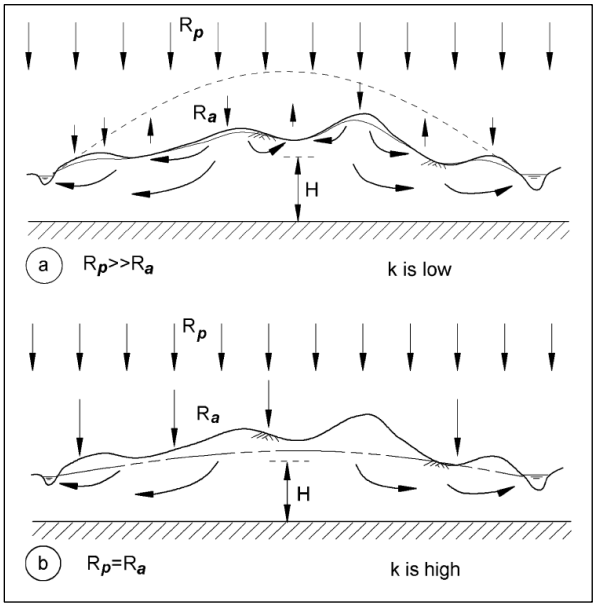


Figure 56: a) Low-permeable aquifer with topography controlled water table; b) highly permeable aquifer with a recharge controlled water table (Haitjema and Mitchell-Bruker, 2005).

$$\frac{RL^2}{mkHd} > 1 \;\; topography\ controlled$$

Equation 8: Topography controlled water table

$$\frac{RL^2}{mkHd} < 1 \;\; recharge\ controlled$$

Equation 9: Recharge controlled water table

Equation 8 and Equation 9 from Haitejema and Mitchell-Bruker, (2005) can be used for the determination of topography and recharge controlled water tables. As stated in Haitejema and Mitchell-Bruker (2005), R [m d^{-1}] is the average annual recharge rate, L [m] is the average distance between surface waters, m [-] is a factor between 8 and 16 for aquifers that are rectangular or circular, respectively, k [m d^{-1}] is the horizontal hydraulic conductivity of the aquifer, H [m] is the aquifer thickness (kH is also known as transmissivity, T), and d [m] is the elevation difference between the average surface water level and the maximum terrain elevation.

Parameters *L*, and *d* can be calculated easily using GIS software and all other variables are common hydrogeologic parameters that are often available. As an example, for the study area recharge was estimated to be 3.56×10^{-5} m d^{-1} (13000 m^3 km^{-2} y^{-1}) (Edmunds and Gaye, 1994). Transmissivity in the south-central part of the study area was estimated to be 345 m^2 d^{-1} (4×10^{-3} m s^{-1}) (BRGM, 1976). The *d* parameter was estimated at approximately 100 m. The last two parameters were guesses, $m = 10$ and $L = 80000$ m (80km), which is about one third the length of the study area and seemed appropriate because perennial surface water is scarce. Using these figures for the southern part of the study area the value obtained was 0.66, which falls within the recharge controlled category and seems logical given the general knowledge of hydrology for this area.

Of course, not all environments will fit satisfactorily into one of these two categories, in which case variables could simply be chosen if adequate knowledge of the hydrogeology of the area is known. Similarly, if the data are not available to make these calculations, the categorization can be done heuristically and then appropriate variables chosen. Another approach would be to break up the focus area into logical hydrogeologic zones and then apply these categories to each zone separately. Once this initial categorization has been made, then the environmental variables would be broken up into their applicability to softer, younger geologic formations and older, harder geologic formations. The flowchart shown in Figure 57 gives an overview of the proposed variable development procedure for manual well drilling favorability mapping including some potentially important variables that were not considered in this study. Also, Appendix A and the bibliography of this study are useful guides to choosing, finding, and generating the environmental variables of choice. The final steps in the procedure, which are not shown in Figure 57, are to establish appropriate thresholds for the variables either statistically, as done in this study, or by another method, and create the final favorability map(s), which could be done in a variety of ways.

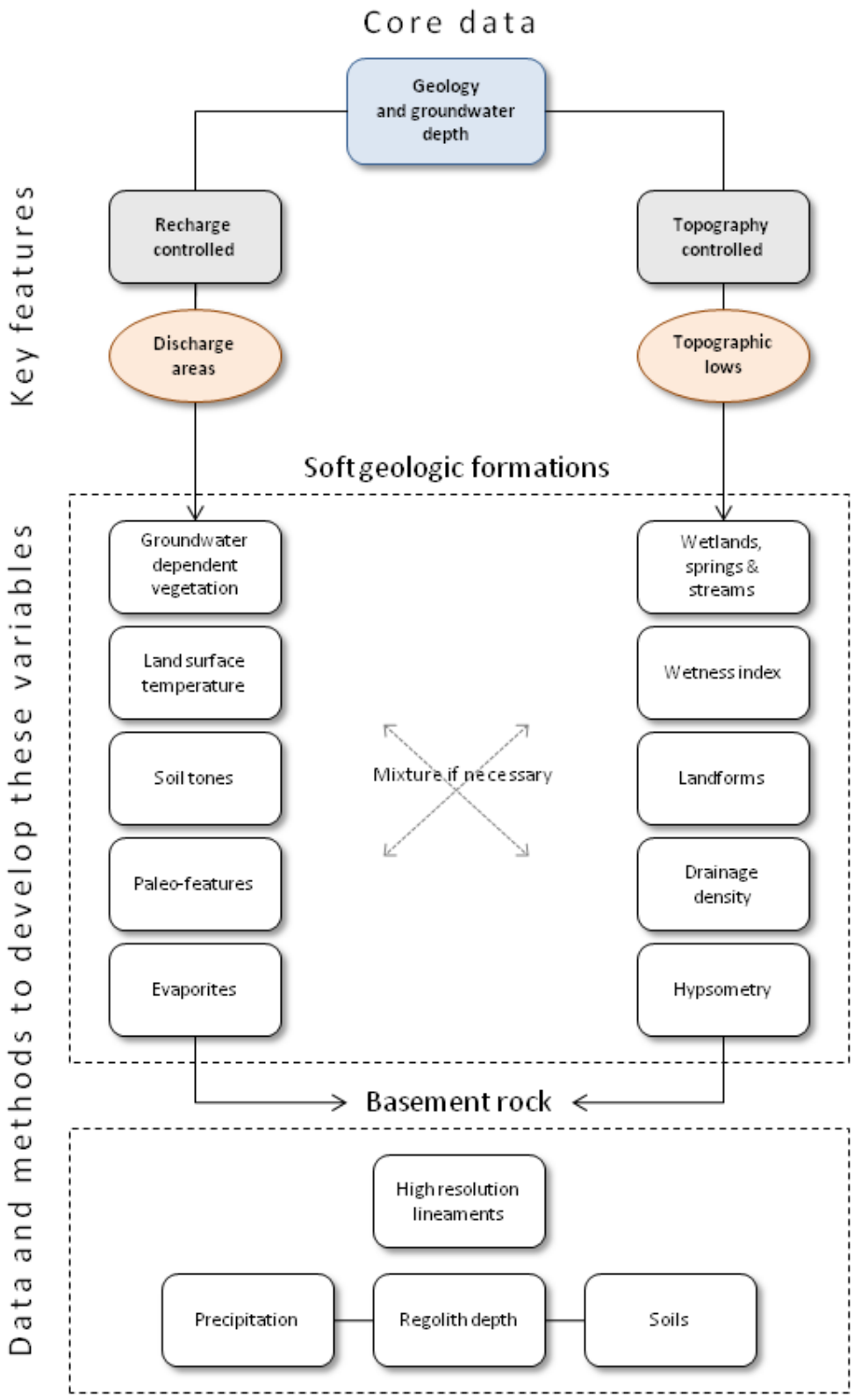


Figure 57: Procedural outline for variable development for manual well drilling favorability mapping.

6.4 Map creation and comparisons

Through the use of GWD data this study sought to present environmental variables that could potentially be complementary or used alternatively to scarce GWD data. However, this required that a fairly dense and accurate GWD database exist in order to develop a good test of the variables, which was the case for this study. A comparison of several favorability maps was performed to qualitatively assess some different approaches. Map creation was straightforward once all the variables were developed and the tree model had been built. All vector layers were converted to raster format. Also, all the variables were structured in the same spatial projection and coordinate system, cell size, and spatial extent. A cell size of 235 m was used to correspond with the MODIS data, which also seemed an appropriate minimum mapping unit, despite the inclusion of 1000 m resolution NLST. Basic raster math in the GIS software was used to create maps from the appropriate threshold values.

To provide an example of the typical favorability mapping done by Practica Foundation a map was created of the Zinder region using roughly the same classes of geology and GWD that were used for Practica's mapping effort in Chad (Figure 58). As stated earlier, if GWD data are sparse, but they are sufficiently well distributed over the project area, these data can be used to create a regression tree. This makes possible the use of threshold values of continuous explanatory variables from the tree as surrogates for the GWD data where they are missing within the study area, which may improve the error as indicated in Figure 55. As an example, the classes used in Figure 58 were employed for another map that incorporated the thresholds values from Tree 1 (Figure 41) as proxies for GWD (Figure 59). The qualitative results of these two maps are similar, suggesting again that the regression tree approach could work well as an alternative to interpolated GWD data. This method accentuated smaller favorable areas that were missed with the typical geology and interpolated GWD approach, which is to be expected because the environmental variables represent real landscape attributes. For example, some

highly favorable areas are distinct in the Goulbi N'Kaba drainage in the northwest of the study area in Figure 59, which do not exist in Figure 58. This also suggests that environmental variables such as VPR* and NLST could be used for favorability mapping for manual well drilling if there is no extant GWD data, again assuming appropriate thresholds can be determined without a statistical model related to GWD.

A final map was produced using the results of Tree 1 (Figure 41) without lumping geology and GWD into classes (Figure 60). This approach preserved all the information content of the variables without any loss because of grouping into classes. This final map allows the interpreter to make a more informed decision about where to attempt manual well drilling. For example, in the basement rock geology in the center of the Zinder region the final map reveals a few pockets of relatively shallow groundwater depth (mean of 22 m). The distribution of GWD around the mean value (Figure 44) in those areas shows it is probable to find even shallower groundwater there. Consider that these pockets might be near communities with a high priority need for well development, and the final map seems superior to the other two maps. These hypothetical communities might be overlooked using the low and poor favorability rankings in this area shown in the first two maps. The final map (Figure 60) should be used alongside the histograms of GWD (Figure 42 to Figure 48) from Tree 1, which can provide quantifiable probabilities for GWD. These probabilities coupled with unaltered geologic data (not grouped into classes) result in a useful tool for the hydrogeologist trying to target areas to attempt manual well drilling.

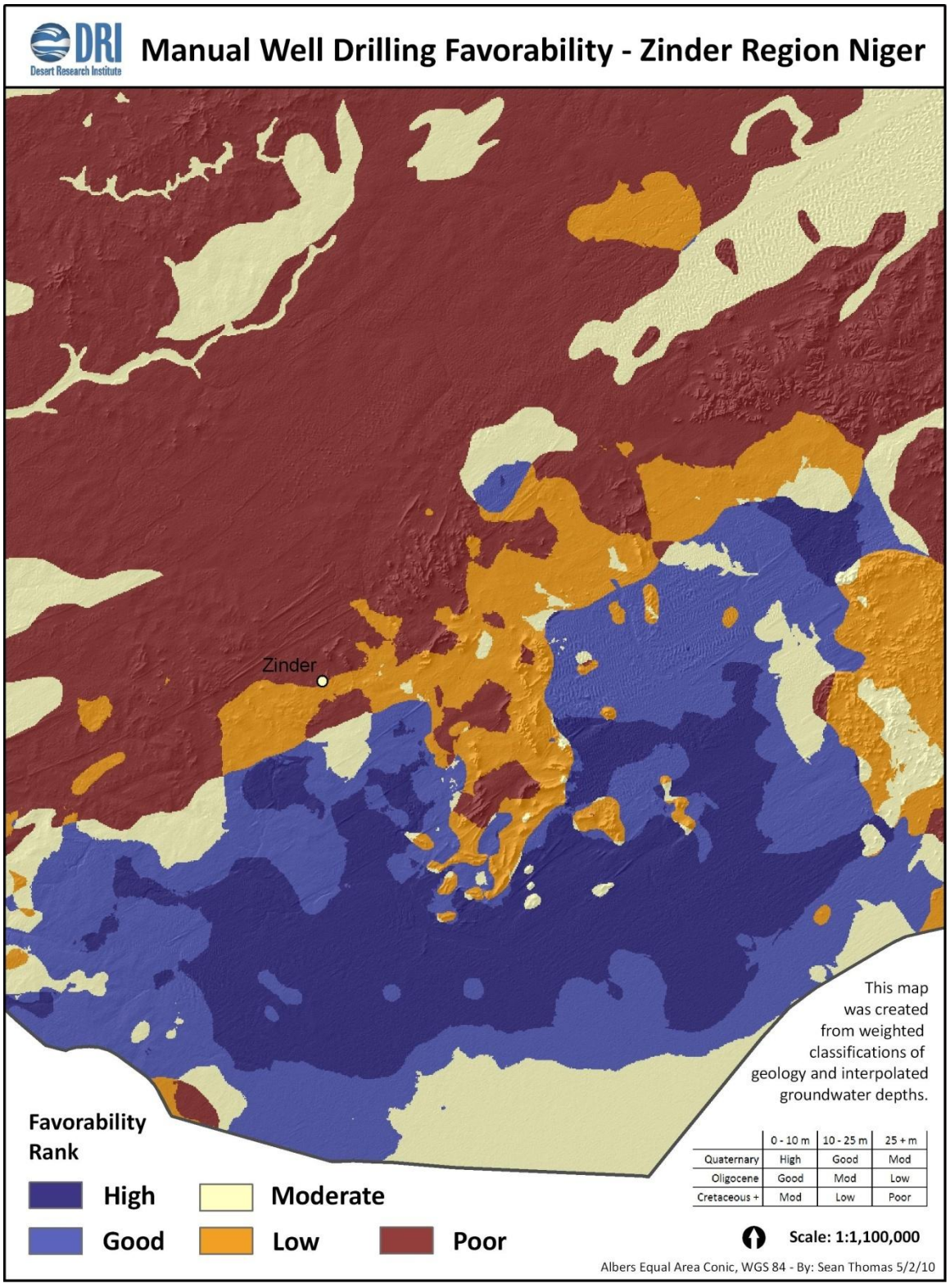


Figure 58: Map of manual well drilling favorability created from classes of both geology and interpolated groundwater depth – Zinder region.

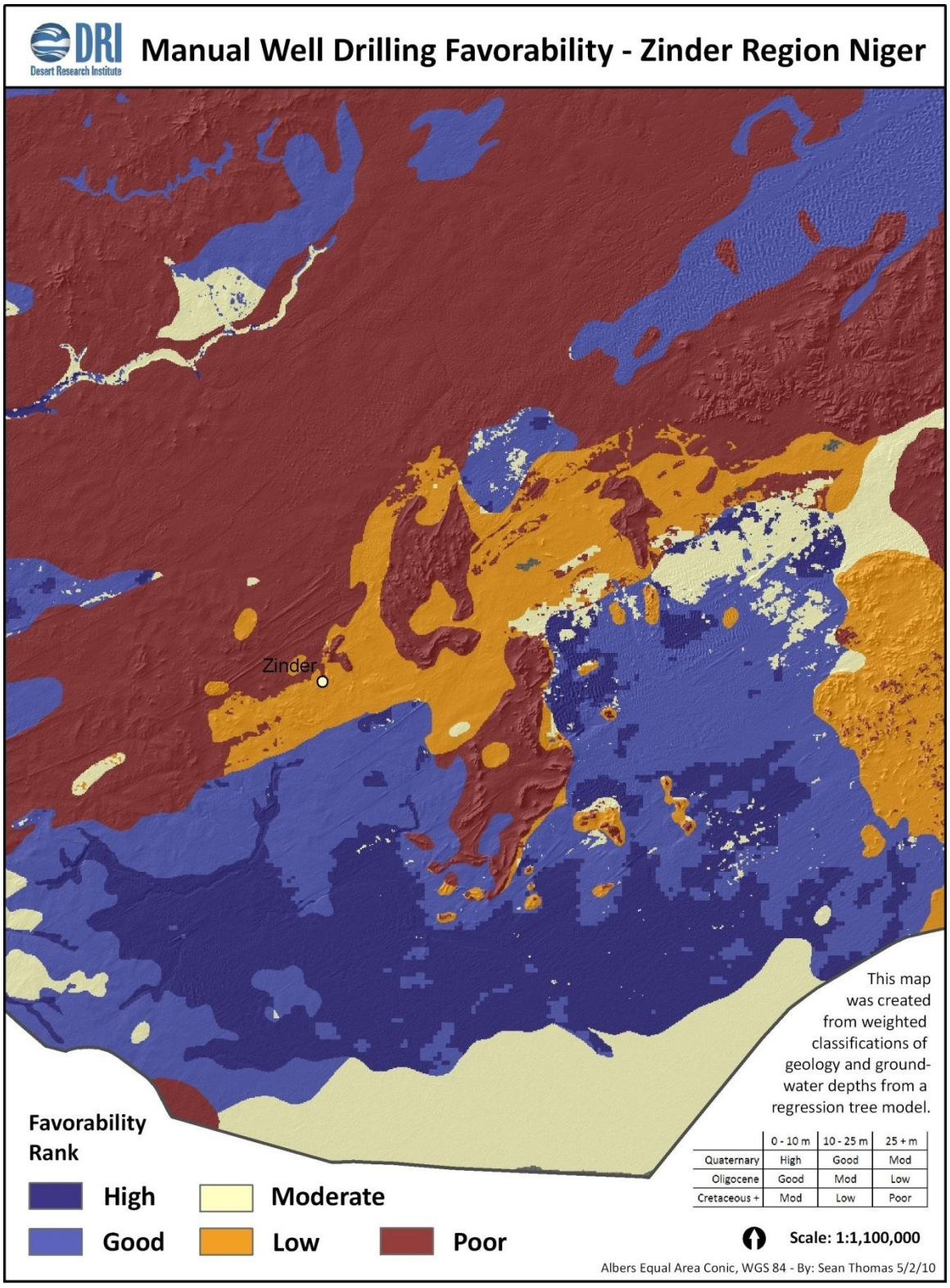


Figure 59: Map of manual well drilling favorability created from classes of both geology and groundwater depth from the final regression tree (Tree 1) – Zinder region.

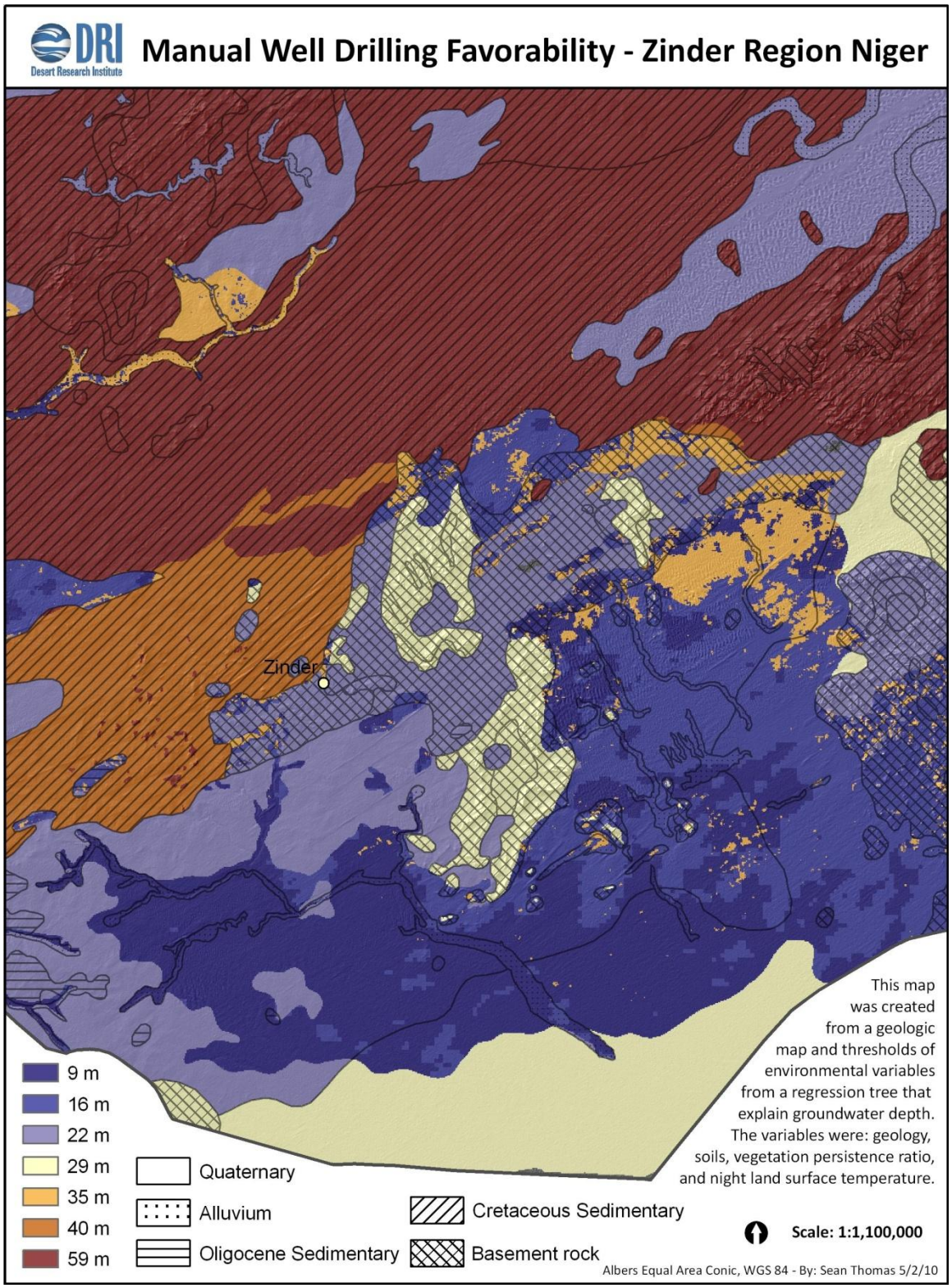


Figure 60: Map of manual well drilling favorability created from geology and groundwater depths from the final regression tree (Tree 1) – Zinder region

7 Conclusions

At the core of this study was the question of whether current favorability mapping, like the methods of Practica Foundation using geology and GWD data (Practica, 2005), could be enhanced through the use of publically available data and spatial analysis techniques, with an emphasis on ease of application and utility over large spatial extents. Because manual well drilling is strongly constrained by the factors of geology and GWD, it is obvious without any investigation that these pieces of information are crucial for identifying good areas to drill. A thorough literature review and regression tree analysis of the identified data confirmed this fact. Therefore, it is the primary finding of this study, and parsimony would dictate, that a basic geologic map and decent GWD data should provide the majority of useful information for efficiently targeting favorable areas for manual well drilling. Ultimately, these data should not be lumped into classes for the final mapping in order to preserve as much information in the final map as possible.

Additional data that are accessible and fairly easy to develop do show some promise to add a degree of valuable refinement to maps created with the two core data sets of geology and GWD. The VPR* variable was weakly correlated to GWD individually, but high values were robust to predict shallow GWD, particularly in combination with low NLST values. These results indicated a need for additional research to investigate the relationship of these data to GWD and possible utility for similar favorability mapping in other areas or other purposes such as differentiating recharge and discharge areas. Notably, the use of environmental variables, such as NLST and VPR*, as proxies for GWD can yield results that are truer to physical conditions, unlike interpolated GWD data. These two variables could even be valuable when GWD data does not exist if appropriate thresholds can be determined.

Scale and resolution of the data were important considerations for this study. Coarse resolution data, like the MODIS products used in this study, preclude highly accurate targeting for manual well drilling, but are better for mapping broad landscape characteristics. They are also less computationally intensive than higher resolution data such as 30 m Landsat ETM+ satellite imagery. Once broad areas of favorability are identified it would be best to do local scale investigations, which could be performed rapidly, depending on the available data, technology, and expertise. This hierarchical approach should be advantageous for improving well drilling efficiencies. As indicated by this study, even at a scale that is only a fraction of the total area of an entire country, like the Zinder region, it is not possible to pinpoint well drilling sites.

A final key observation was that if data in addition to geology and GWD are desired to be used to enhance an effort at favorability mapping for manual well drilling, then an initial discrimination should be made that narrows the list of potentially valuable environmental variables. This approach will improve on time and resource efficiency. This cut can be based on the concepts of topography and recharge controlled water tables. According to the suggested procedural outline for variable development (Figure 57), the focus area could be classified into one of these two categories by performing some basic calculations using common hydrogeologic parameters and easily determined spatial characteristics of the terrain. Recharge controlled water table regimes would most likely not benefit from the use of DEM derived or other topographically related variables, but instead variables like vegetation characteristics, such as VPR*, and/or land surface temperature data, such as NLST, could be useful for highlighting groundwater discharge areas. Topographically controlled water table areas could make use of the DEM derived variables. Areas in either category could conduct a lineament analysis in hard rock geology to highlight potential for saprolite aquifers, or fractured areas that have been weathered.

UNICEF's push to bring SSA up to speed in terms of the water MDG through the promotion of manual well drilling, provided the impetus and direction for this study. The intention of this research was to examine and make known a range of easily accessible and applicable data and methods for the purpose of enhancing favorability mapping for manual well drilling efforts in SSA. In so doing, the value of this thesis lies primarily in the compilation and documentation of data sources, related scientific publications, and applicable spatial analysis and statistical techniques that are potentially useful for this purpose. The basic evaluations of data and methods that were performed in this study provide examples of what is possible with some adaptation in different settings, and represent only a first attempt at highlighting some useful approaches. Therefore, more research is needed for similar attempts at favorability mapping for manual well drilling in other areas. The methods presented here can improve from additional study. More importantly, the people of SSA who still lack adequate access to clean water stand to benefit from such work.

Appendix A

- Internet portals for obtaining base data applicable for groundwater exploration studies

Spatial Data	
ACP Données - Niger Geology	http://mines.acp.int/html/NE_GEO_fr.html
Africa Atlas for Dynamic Planet	http://www.dynamicplanet.com/AfricaAtlas/AfricaAtlas.htm
Africa Data Dissemination Service	http://earlywarning.usgs.gov/a
AfricaMap - Harvard Site	http://cga-3.hmdc.harvard.edu/africamap/
African Maps Links | Map Africa	http://www-sul.stanford.edu/depts/ssrg/africa/map.html
African Population Database Documentation	http://na.unep.net/globalpop/africa/
African Regional Water Resources Information System	http://www.uneca.org/awich/
AGRHYMET Regional Database	http://193.251.227.200/index.php?op=edito
AMMA-CATCH - Database and GIS	http://itheln21.hmg.inpg.fr/catch/index.php?page=database&lang=en
AMMA-SAT & AMMA-MOD Databases	http://bddamma.ipsl.polytechnique.fr/
Atlas natural agronomic ressources Niger Benin	https://www.uni-hohenheim.de/~atlas308/startpages/page2/english/content/title_en.htm
BGR WHYMAP Network	http://www.whymap.org/cln_092/nn_354274/whymap/EN/Network/network__node__en.html?__nnn=true
BIU Scientifique Jussieu - Catalogue de la cartothèque - BRGM Afrique	http://www.bius.jussieu.fr/web/carto/afrique/indexafr/brgm-afr.html
BURKINA FASO - CARTOTHEQUE DE L'IRD : Sciences de la Terre, Sciences humaines	http://www.cartographie.ird.fr/sphaera/liste_cartes.php?iso=BFA&nom=BURKINA+FASO
CGIAR-CSI SRTM 90m DEM Digital Elevation Database	http://srtm.csi.cgiar.org/index.asp
CISL Research Data - Climate ds570.0 PREPARED DATA SUBSETS	http://dss.ucar.edu/datasets/ds570.0/data/data.html
CRU Dataset	http://badc.nerc.ac.uk/data/cru/
CRU TS 2.0: grids of climate observations	http://www.cru.uea.ac.uk/~timm/grid/CRU_TS_2_0.html
eoPortal - Sharing Earth Observation Resources	http://maps.eoportal.org/
ESA - TIGER Initiative	http://www.tiger.esa.int/home.asp
ESA AQUIFER - Earth Observation for Shared Aquifers in Africa	http://www2.gaf.de/aquifer/index.htm
EuDASM - The soil maps of Africa	http://eusoils.jrc.ec.europa.eu/esdb_archive/EuDASM/africa/index.htm
FAO - GIS in Sustainable Development	http://www.fao.org/sd/eidirect/gis/Elgis000.htm
FAO - Spatial Information Management and Decision Support Tools	http://www.fao.org/spatl/gateway_en.asp

Spatial Data - Continued	
FAO Country Profiles and Mapping Information System -- Home	http://www.fao.org/countryprofiles/default.asp?lang=en
FAO GeoNetwork- The portal to spatial data and information	http://www.fao.org/geonetwork/srv/en/main.home
FAO Water , Development and Management Unit	http://www.fao.org/nr/water/
FEWS NET Data	http://earlywarning.usgs.gov/?l=en
gData - Berkley GIS Server	http://biogeo.berkeley.edu/bgm/gdata.php
GEO Data Portal - The Environment Database	http://geodata.grid.unep.ch/
GEOPortal - GEO-Portal Home	http://www.geoportal.org/web/guest/geo_home?cache_control=0
GGIS-MAP	http://igrac.nitg.tno.nl/ggis_map/start.html
GIEWS - Global Information and Early Warning System on Food and Agriculture	http://www.fao.org/giews/english/index.htm
GLCF: Global Land Cover Facility	http://glcf.umiacs.umd.edu/index.shtml
GLOWA Volta Geoportal	http://131.220.109.6/Geoportal/index.jsp
Gridded Population of the World - GPW v3	http://sedac.ciesin.columbia.edu/gpw/global.jsp
HAPEX SAHEL Information Ground data CD-ROM	http://www.ird.fr/hapex/
Harmonized World Soil Database	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html
HSM - Système d'Informations Environementales sur les Ressources en Eaux et leur Modélisation	http://www.hydrosciences.fr/sierem/
IGBF - Institut Geographique du Burkina Faso	http://www.igb.bf/
IGN.FR - Portail	http://www.ign.fr/index.do
International Groundwater Resources Assessment Centre - Igrac	http://www.igrac.net/
IRD Niger	http://www.ird.ne/
ISARM - Internationally Shared Aquifer Resources Management - Isarm	http://www.isarm.net/
ISARM Africa - Isarm	http://www.isarm.net/publications/174
Iullemeden Maps & GIS	http://iullemeden.iwlearn.org/gis
L'IRD AU BURKINA FASO	http://www.ird.bf/
Le Portail du Secteur de L'eau au BF	http://www.eauburkina.bf/

Spatial Data - Continued	
L'Observatoire du Sahara et du Sahel	http://www.oss-online.org/index.php?option=com_content&task=view&id=22&Itemid=57
Madmappers.com - Maps and GIS for Africa	http://www.madmappers.com/
MAHRH - Ministere Hydraulique BF	http://www.agriculture.gov.bf/SiteAgriculture/index.jsp
Map Library - Burkina Faso	http://www.maplibrary.org/stacks/Africa/Burkina%20Faso/index.php
Map Library - Niger Data	http://www.maplibrary.org/stacks/Africa/Niger/index.php
MARA Climate CD	http://www.mara.org.za./climatecd.htm
National Mapping Agencies in Africa	http://nma.agirn.org/index.php/agencies/registered_agencies#NE
NIGER - CARTOTHEQUE DE L`IRD : Sciences de la Terre, Sciences humaines	http://www.cartographie.ird.fr/sphaera/liste_cartes.php?iso=NER&nom=NIGER
NOAA NCEP CPC FEWS	http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/
Photographic Library - Découvrir l'IGN	http://www.ign.fr/institut/28/activites/phototheque-et-cartotheque.htm
Project Environment - Isarm	http://www.isarm.net/publications/125
SEAMIC - GIS Africa Home	http://www.seamic.org/gisafrica/
SERVIR AFRICA - Home	http://www.servir.net/africa/index.php?option=com_frontpage&Itemid=1
SIG Afrique - Geology GIS Data for Africa	http://www.sigafrique.net/Default.aspx
Site Internet de SIGNER - Systeme d'Information Geographique du Niger	http://www.pnud.ne/signer/index.htm
Transboundary Freshwater Spatial Database | OSU	http://www.transboundarywaters.orst.edu/database/transfreshspatdata.html
United Nations Economic Commission for Africa (UNECA)	http://www.uneca.org/disd/geoinfo/
USGS EROS EarthExplorer	http://edcsns17.cr.usgs.gov/EarthExplorer/
USGS Global Ecosystems	http://rmgsc.cr.usgs.gov/ecosystems/dataviewer.shtml
USGS Global Visualization Viewer	http://glovis.usgs.gov/
VALPEDO - MIRURAM : Accueil	http://miruram.mpl.ird.fr/valpedo/miruram/index.html
Valsol Webmapping Burkina Faso	http://miruram.mpl.ird.fr/valpedo/miruram/burkina/index.html
WHYMAP - Webmapping Application	http://www.whymap.org/
World Data Center (WDC) Portal	http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=wdc&MetadataType=0&lbnode=mdlb2

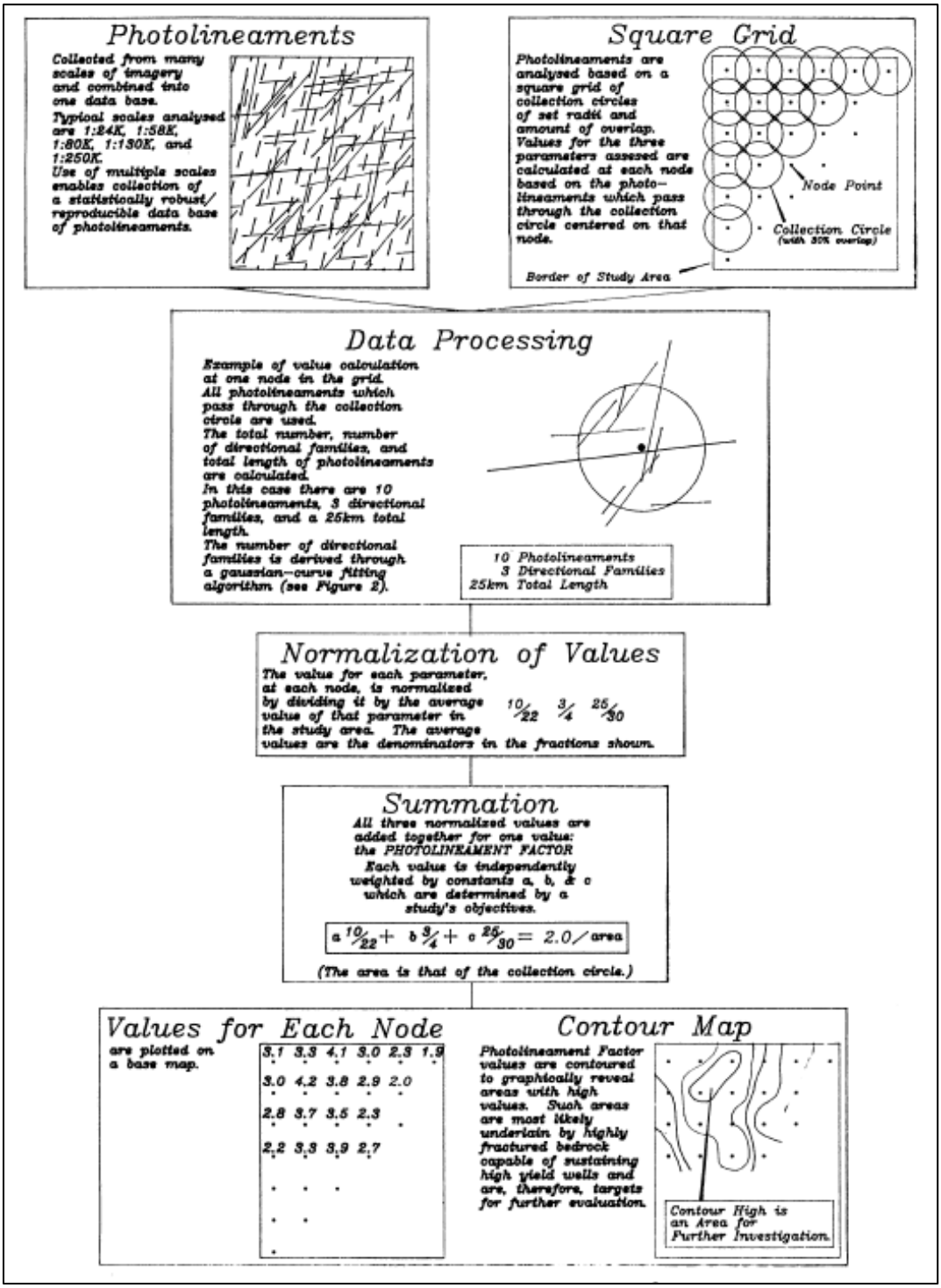
Information	
African Development Bank	http://www.afdb.org/en/home/
AMMA - AMMA Science Website	http://amma.mediasfrance.org/
AMMA - US Home Page	http://www.eol.ucar.edu/projects/amma-us/
ESA Portal - Satellites help locate water in Niger	http://www.esa.int/esaCP/SEMRA6VPXPF_index_0.html
FAO - Low-cost shallow tube well construction in West Africa	http://www.fao.org/docrep/W7314E/w7314e0v.htm
GRACE - Gravity Recovery and Climate Experiment	http://www.csr.utexas.edu/grace/
GWCLIM.ORG - Groundwater & Climate in Africa	http://www.gwclim.org/presentations.php
International Water Management Institute (IWMI)	http://www.iwmi.cgiar.org/index.aspx
JMP > WHO & UNICEF Joint Monitoring Programme	http://www.wssinfo.org/en/welcome.html
Les espèces végétales et leurs utilités - Niger	http://ne.chm-cbd.net/biodiversity/la-diversite-biologique-vegetale/les-especes-vegetales-et-leurs-utilites
Les ressources forestières naturelles au Niger	http://www.fao.org/DOCREP/004/X6813F/X6813F14.htm
MIT Research - Ecohydrology	http://hydrology.mit.edu/index.php/Research/Ecohydrology
NWT Uranium Corporation - Niger - Wed Mar 4, 2009	http://www.nwturanium.com/s/Niger.asp
Poverty Analysis - Poverty Mapping	http://www.worldbank.org/poverty/
Poverty Mapping - Home	http://www.povertymap.net/
Practica Foundation | Drilling & Wells	http://www.practicafoundation.nl/products/drilling-wells/
RWSN community site	http://www.rwsn.ch/about/prarticle_view
Stanford U - Branner Library and Map Collections	http://library.stanford.edu/depts/branner/research_help/brief_map.html
STONEHAMMER DRILLING METHOD	http://www.daenvis.org/stonehammer6.htm
StoneISD Report - Ref for Niger Spatial Data	http://www.stone-env.com/docs/reports/StoneISD-PhaseIPrelimPovWBrpt.pdf
The Geological Setting in west Niger	https://www.uni-hohenheim.de/~atlas308/b_niger/projects/b2_1_1/html/english/ntext_en_b2_1_1.htm
WorldBank Groundwater and Rural Dev	http://www-esd.worldbank.org/esd/ard/groundwater/Home.htm
ZEF Homepage: ZEF Homepage	http://www.zef.de/index.php?id=zefhomepage

- Spatial analysis techniques for deriving variables for groundwater exploration

Category	Type of technique	Type of base data used				
		DEM	MS Satellite	Map	Air photo	Radar
Geology	ID geologic formations		X	X	X	
	ID soils		X	X	X	
	Lineaments	X	X		X	X
Geomorphology	Drainage density	X				
	Hypsometric indices	X				
	ID geomorphologic features	X	X		X	X
	ID paleo-features	X	X		X	X
	Landform indices	X	X			
Hydrography	Flow algorithms	X				
	ID waterbodies	X	X	X	X	
	ID springs		X	X	X	
	ID streams	X	X	X	X	
Topography	Depressions	X				
	Topographic convergence indices	X				
	Slope indices	X				
Thermal	Land surface temperature		X			
Vegetation	Vegetation indices		X			
	Vegetation classification		X		X	
	Land cover classification		X	X	X	

Appendix B

- Lineament factor workflow from Hardcastle (1995)



Appendix C

- Albers Equal Area Conic projection parameters (centered on Niger and Burkina Faso):

```
False easting            0.0
False northing           0.0
Standard parallel 1      21.0
Standard parallel 2      12.0
Central meridian         5.0
Latitude of origin       9.0
Datum                    WGS84
Linear unit              meter
```

- Perl script that calls MRT, the projection parameter file, and the file with a list of scenes:

```
#! /usr/bin/perl –w

# A script for batch reprojecting modis data
# 1)This file should be copied to the c:/mrt directory and run
    from there
# 2)Reference the correct directory the scenes are in and the
    type name below
# 3)Create a file called modis_scenes.dat with a list of the
    files in the directory and place this in that directory
# 3)Obtain the correct substrings off the Modis files
# 4)Reference the correct *.prm file for the desired projection
# 5)Print out the substrings before running system command as check

$file1 = "d:/thesisgis/rasters/temp/modis/test/modis_scenes.dat";
$type = "_ZIN_LST.HDF";
open FILENAMES, "$file1" or die $!;
while (<FILENAMES>) {
    $outid1 = substr($_, 0, 37);
    $outid2 = substr($_, 45, 8);
    $outname = $outid1 . $outid2 . $type;
    #this runs the Modis Reprojection Tool with proper override
    arguments
```

```
    system("resample -p modis_lst_alb.prm -i $_ -o $outname");
    print "$_\n";
    print "$outname\n\n";
}
close FILENAMES;
```

- Example of projection parameter file (modis_lst_alb.prm in the script above):

```
INPUT_FILENAME = MODIN.hdf
SPECTRAL_SUBSET = ( 1 0 0 0 1 0 0 0 0 0 0 0 )
SPATIAL_SUBSET_TYPE = OUTPUT_PROJ_COORDS
SPATIAL_SUBSET_UL_CORNER = ( 349900.0 677500.0 )
SPATIAL_SUBSET_LR_CORNER = ( 569000.0 416000.0 )
OUTPUT_FILENAME = MODOUT.HDF
RESAMPLING_TYPE = NN
OUTPUT_PROJECTION_TYPE = AEA
OUTPUT_PROJECTION_PARAMETERS = ( 0.0 0.0 21.0 12.0 5.0 9.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 )
DATUM = WGS84
OUTPUT_PIXEL_SIZE = 930
```

- Example of Python script to batch process MODIS images (in this case EVID):

```
# Script to subtract peak vs lows EVI
# Create the geoprocessor module
import arcgisscripting
gp = arcgisscripting.create(9.3)
# Check out Spatial Analyst extension license
gp.CheckOutExtension("Spatial")
# Sets environment
gp.workspace = "d:/thesisgis/rasters/veg/modis/250m_2"
gp.pyramid = "NONE"

# Get a list of rasters in the workspace.
maps = gp.ListRasters("*241_ZIN_EVI.img", "ALL")
# Loop for 1)Subtracting peak from low1 and low2
for map in maps:
```

```
    year = map[1:5]
    getyr = int(map[1:5])
    nxtyr = str(getyr + 1)
    low1 = map[0:5]+"353_ZIN_EVI.img"     #December date
    low2 = "A"+nxtyr+"065_ZIN_EVI.img"    #March date
    minus1 = map[0:8]+"-A"+year+"353_ZIN_EVI.img"
    minus2 = map[0:8]+"-A"+nxtyr+"065_ZIN_EVI.img"
    # 4 Calc EVID
    gp.Minus_sa(map, low1, minus1)
    gp.Minus_sa(map, low2, minus2)
    # Print to check for errors
    print map
    print year
    print nxtyr
    print low1
    print low2
    print minus1
    print minus2
```

- Example of Python script to batch process MODIS images (in this case average EVID):

```
# Script to avg EVI Difference ratios over all images (years)
# Instructions:
# 1) Change ListRasters args to get files you want to sum
# 2) Run first only with the print len(maps) to get how many files
# 3) Make sure the expr variable has 0 thru (# of files - 1) elements
# 4) Run
# Create the geoprocessor module
import arcgisscripting
gp = arcgisscripting.create(9.3)
# Sets environment
gp.workspace = "d:/thesisgis/rasters/veg/modis/250m_2"
gp.pyramid = "NONE"

# Get lists of rasters in the workspace
maps = gp.ListRasters("*-*353_ZIN_EVI_DIFFNONLS.img", "ALL")
map2s = gp.ListRasters("*-*353_ZIN_EVI_PRSABS.img", "ALL")
```

```
# Output names
sum = "A2000225-to-2009353_ZIN_EVI_DIFFSUM.img"
cnt = "A2000225-to-2009353_ZIN_EVI_DIFFCNT.img"
avg = "A2000225-to-2009353_ZIN_EVI_DIFFAVG.img"
# Expressions to do sum and count
expr1 = "("+maps[0]+" + "+maps[1]+" + "+maps[2]+" + "+maps[3]+" +
"+maps[4]+" + "+maps[5]+" + "+maps[6]+" + "+maps[7]+" + "+maps[8]+" +
"+maps[9]+")"
expr2 = "("+map2s[0]+" + "+map2s[1]+" + "+map2s[2]+" + "+map2s[3]+" +
"+map2s[4]+" + "+map2s[5]+" + "+map2s[6]+" + "+map2s[7]+" +
"+map2s[8]+" + "+map2s[9]+")"
# Sum, Count and Average
gp.SingleOutputMapAlgebra_sa(expr1, sum)
gp.SingleOutputMapAlgebra_sa(expr2, cnt)
gp.Divide_sa(sum, cnt, avg)
# Print to check for errors
for map in maps:
    print map
print len(maps)
for map2 in map2s:
    print map2
print len(map2s)
print sum
print cnt
print avg
print expr1
print expr2
```

Appendix D

- Detailed results of the final 7-node regression tree (Tree 1):

```
Regression tree:

snip.tree(tree = tree(formula = NSAVG ~ GEOLTYP + SOILSTYP + DSTSTR4 +
LINFCT + VPR.AVG + NLST.05 + LDFM721 + TCI, data = GW.FNLa4.97,
    na.action = na.exclude, mincut = 0, minsize = 1, mindev = 0), nodes
= c(18., 6., 11., 8., 7., 19., 10.))

Variables actually used in tree construction:
[1] "GEOLTYP"  "SOILSTYP" "VPR.AVG"  "NLST.05"

Number of terminal nodes:  7

Residual mean deviance:  159.1 = 380300 / 2390

Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.
Max.
 -5.575e+001 -6.917e+000 -2.144e+000 -5.680e-015  6.947e+000
6.214e+001

node), split, n, deviance, yval
      * denotes terminal node

 1) root 2397 831600 25.850
   2) GEOLTYP:Alluvium,Clay Sandstone,Igneous,Metamorphic,Sand
      dunes,Sand plains 2024 396200 20.990
     4) SOILSTYP:Evolved little,Hydromorphic 739 108600 12.950
       8) VPR.AVG<1.04255 29  10420 34.670 *
       9) VPR.AVG>1.04255 710  83910 12.060
        18) NLST.05<15.005 408  22950  9.144 *
        19) NLST.05>15.005 302  52800 16.000 *
     5) SOILSTYP:Raw minerals,Seqsquioxides,Steppe-like,Vertisols 1285
        212500 25.610
      10) GEOLTYP:Alluvium,Clay Sandstone,Igneous,Sand dunes 603 125400
          21.860 *
      11) GEOLTYP:Metamorphic,Sand plains 682  71170 28.920 *
   3) GEOLTYP:Sandstone 1,Sandstone 3,Sandstone 4,Sandstone 6,Sandstone
      7 373 127500 52.250
     6) SOILSTYP:Hydromorphic,Seqsquioxides 135  34400 40.350 *
     7) SOILSTYP:Evolved little,Raw minerals,Steppe-like,Vertisols 238
        63200 59.000 *
```

- Detailed results of the alternate 12-node regression tree excluding geology and soils (Tree 2):

```
Regression tree:

snip.tree(tree = tree(formula = NSAVG ~ DSTSTR4 + LINFCT + VPR.AVG +
NLST.05 + LDFM721 + TCI, data = GW.FNLa.97, na.action = na.exclude,
    mincut = 0, minsize = 1, mindev = 0), nodes = c(19., 18., 29.,
463., 5., 230., 462., 15., 56., 114., 8., 6.))

Variables actually used in tree construction:
[1] "VPR.AVG" "LINFCT"  "NLST.05" "DSTSTR4"

Number of terminal nodes:  12

Residual mean deviance:  203.5 = 485400 / 2385

Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.
Max.
 -5.334e+001 -8.369e+000 -1.917e+000 -1.271e-014  7.601e+000
6.096e+001

node), split, n, deviance, yval
      * denotes terminal node

  1) root 2397 840200 25.79
    2) VPR.AVG<1.0495 334 171700 46.55
      4) LINFCT<18.85 231 116500 53.26
        8) NLST.05<18.62 193  86520 56.59 *
        9) NLST.05>18.62 38  16970 36.33
         18) DSTSTR4<36787.5 26   5724 25.98 *
         19) DSTSTR4>36787.5 12   2422 58.76 *
      5) LINFCT>18.85 103  21490 31.50 *
    3) VPR.AVG>1.0495 2063 501300 22.43
      6) NLST.05<14.835 561  63670 12.92 *
      7) NLST.05>14.835 1502 367900 25.99
       14) DSTSTR4<28451 1448 332300 25.47
         28) VPR.AVG<1.192 1394 318200 25.98
           56) LINFCT<7.95 455  40650 29.17 *
           57) LINFCT>7.95 939 270700 24.44
            114) VPR.AVG<1.08815 433 152100 28.86 *
            115) VPR.AVG>1.08815 506 102900 20.65
              230) LINFCT<18.65 235  27770 16.00 *
              231) LINFCT>18.65 271  65600 24.69
                462) DSTSTR4<23506 251  50270 23.04 *
                463) DSTSTR4>23506 20   6051 45.41 *
         29) VPR.AVG>1.192 54   4144 12.16 *
       15) DSTSTR4>28451 54  24580 39.99 *
```

References

Ahmed, F., Andrawis, A., Hagaz, Y., 1984. Landsat model for groundwater exploration in Nuba mountains, Sudan. Advances in Space Research, 4(11): 123-131.

Al Biely, A., Sanhouri, I., El Khidir, S., Anonymous, 2000. Delineation of potential target zones for groundwater accumulations using geologic remote sensing and GIS techniques, case study; En Nuhud area, West Kordofan State, Sudan. Proceedings of the International Conference on Applied Geologic Remote Sensing, 14: 462-469.

Amesz, B., 1984. Satellite survey boosts drilling programme success. World Water, 7(4): 24-25.

Bailey, M., Halls, H., 2000. Use of remote sensing data to locate groundwater trapped by dykes in Precambrian basement terrains. Canadian Journal of Remote Sensing, 26(2): 111-120.

Bala, A., Ike, E., Usman, Y., 2005. An evaluation of the groundwater resources in the basement complex rocks of Kano area, Nigeria, using Landsat-5 TM imagery. Africa Geoscience Review, 12(1-2): 25-35.

Bandyopadhyay, S., Srivastava, S., Jha, M., Hegde, V., Jayaraman, V., 2007. Harnessing earth observation (EO) capabilities in hydrogeology: an Indian perspective. Hydrogeology Journal, 15(1): 155-158.

Banks, W., Paylor, R., Hughes, W., 1996. Using thermal-infrared imagery to delineate ground-water discharge. Ground Water, 34(3): 434-443.

Batelaan, O., De Smedt, F., Triest, L., 2003. Regional groundwater discharge: phreatophyte mapping, groundwater modelling and impact analysis of land-use change. Journal of Hydrology, 275(1-2): 86-108.

Baugh, W., Groeneveld, D., 2006. Broadband vegetation index performance evaluated for a low-cover environment. International Journal of Remote Sensing, 27(21-22): 4715-4730.

Becker, M., 2006. Potential for satellite remote sensing of ground water. Ground Water, 44(2): 306-318.

Beven, K., Kirby, M., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences - Bulletin - des Sciences Hydrologiques, 24(1): 43-49.

Blankwaardt, B., 1984. Hand Drilled Wells A Manual on Siting, Design, Construction and Maintenance. Rwegarulila Water Resources Institute, Dar Es Salaam, Tanzania.

Bobba, A., Bukata, R., Jerome, J., 1992. Digitally processed satellite data as a tool in detecting potential groundwater-flow systems. Journal of Hydrology, 131(1-4): 25-62.

Bocquier, G., Gavaud, M., 1967. Carte Pédologique de Reconnaissance de la République du Niger. Office de la Recherche Scientifique et Technique Outre-Mer (ORSTROM), pp. 1:500,000.

Boeckh, E., 1992. An exploration strategy for higher-yield boreholes in the West African crystalline basement. In: Wright, E., Burgess, W. (Eds.), Hydrogeology of Crystalline Basement Aquifers in Africa. The Geological Society, London.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1998. Classification and Regression Trees. Chapman & Hall/CRC Press, Boca Raton, Florida, 358 pp.

BRGM, 1976. Carte de Planification des Ressources en Eau Souterraine de l'Afrique Soudano-Sanelienne. Bureau de Recherches Géologiques et Minieres (BRGM), Bureau Central d'Études pous les équipements Outre-Mer (BCEOM), and Comité Inter-African d'Études Hydrauliques (CIEH), pp. 1:5,000,000.

Bukata, R., Bruton, J., Jerome, J., 1991. The state of vegetation cover as an indicator of groundwater, based on observations from space. Soviet Journal of Remote Sensing, 9(2): 328-343.

Carter, R., 2005. Human-powered drilling technologies, Cranefield University, Sisloe, Bedford, UK.

Cartwright, K., 1974. Tracing shallow groundwater systems by soil temperatures. Water Resources Research, 10(4): 847-855.

Cavric, B., 2002. Human And Organizational Aspects of GIS Development in Botswana, GSDI 6 Conference, Budapest, Hungary.

Chen, P., Arnold, J., Srinivasan, R., Volk, M., Allen, P., 2006. Surveying ground water level using remote sensing: An example over the Seco and Hondo Creek watershed in Texas. Ground Water Monitoring and Remediation, 26(2): 94-102.

Chowdhury, A., Jha, M., Chowdary, V., Mal, B., 2009. Integrated remote sensing and GIS-based approach for assessing groundwater potential in West Medinipur district, West Bengal, India. International Journal of Remote Sensing, 30(1): 231-250.

Clark, R. et al., 2003. Imaging Spectroscopy: Earth and Planetary Remote Sensing with the USGS Tetracorder and Expert Systems. Journal of Geophysical Research, 108(E12): 5-1 to 5-44.

Danert, K., 2006. A brief history of hand drilled wells in Niger. Waterlines, 25(1): 4-6.

Danert, K., 2007a. Niger RWSN Focus Country Report 1, RWSN, Rural Water Supply Network.

Danert, K., 2007b. West Africa Water Initiative (WAWI) Hand Drilling in Niger Expert Drillers Workshop, USAID, ARD.

Danert, K., 2009. Realizing the potential of hand-drilled wells for rural water supplies. Waterlines, 28(2): 108-129.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology, 81(11): 3178-3192.

Dilts, T., 2009. Topography Tools. http://arcscripts.esri.com/details.asp?dbid=15996%20, pp. Geoprocessing Model tools.

Dodo, A., June 9, 2009.

Eamus, D., Froend, R., 2006. Groundwater-dependent ecosystems: the where, what and why of GDEs. Australian Journal of Botany, 54(2): 91-96.

Edet, A., Okereke, C., Teme, S., Esu, E., 1998. Application of remote-sensing data to groundwater exploration: A case study of the Cross River State, southeastern Nigeria. Hydrogeology Journal, 6(3): 394-404.

Edmunds, W., Gaye, C., 1994. Estimating the spatial variability of groundwater recharge in the Sahel using chloride. Journal of Hydrology, 156: 47-59.

Elmore, A., Kaste, J., Okin, G., Fantle, M., 2008. Groundwater influences on atmospheric dust generation in deserts. Journal of Arid Environments, 72(10): 1753-1765.

Elson, R., Shaw, R., 1995. Technical Brief No.43: Simple Drilling Methods. Waterlines: Journal of Appropriate Technologies for Water Supply and Sanitation, 13(3): 15-18.

Favreau, G. et al., 2009. Land clearing, climate variability, and water resources increase in semiarid southwest Niger: A review. Water Resources Research, 45.

Galanos, I., Rokos, D., 2006. A statistical approach in investigating the hydrogeological significance of remotely sensed lineaments in the crystalline mountainous terrain of the island of Naxos, Greece. Hydrogeology Journal, 14(8): 1569-1581.

Ghoneim, E., 2008. Optimum groundwater locations in the northern United Arab Emirates. International Journal of Remote Sensing, 29(20): 5879-5906.

Ghoneim, E., El-Baz, F., 2007. DEM-optical-radar data integration for palaeohydrological mapping in the northern Darfur, Sudan: implication for groundwater exploration. International Journal of Remote Sensing, 28(22): 5001-5018.

Giancoli, D., 1991. Physics. Prentice Hall, Englewood Heights, New Jersey, 962 pp.

Goutorbe, J.P. et al., 1997. An overview of HAPEX-Sahel: A study in climate and desertification. Journal of Hydrology, 189(1-4): 4-17.

Greigert, J., Bernert, G., 1978. Atlas Des Eaux Souterraines Du Niger, Ministère Des Mines et De l'Hydraulique.

Greigert, J., Pougnet, R., 1967. Carte Géologique de La République du Niger. Bureau de Recherches Géologiques et Minières, pp. 1:2,000,000.

Haitjema, H., Mitchell-Bruker, S., 2005. Are water tables a subdued replica of the topography? Ground Water, 43(6): 781-786.

Hardcastle, K., 1995. Photolineament Factor - A new computer-aided method for remotely sensing the degree to which bedrock is fractured. Photogrammetric Engineering and Remote Sensing, 61(6): 739-747.

Hardcastle, K., February 3, 2010.

Healy, D., June 3, 2009.

Hearns, G., 2009. Terminal Evaluation of UNEP/GEF Project GF/1030-03-06 (4728) Managing Hydrogeological Risk in the Iullemeden Aquifer System (IAS), United Nations Environment Programme.

Heilman, J., Moore, D., 1982. Evaluating depth to shallow groundwater using Heat-Capacity Mapping Mission (HCMM) data. Photogrammetric Engineering and Remote Sensing, 48(12): 1903-1906.

Hoffmann, J., 2005. The future of satellite remote sensing in hydrogeology. Hydrogeology Journal, 13(1): 247-250.

Hoffmann, J., Sander, P., 2007. Remote sensing and GIS in hydrogeology. Hydrogeology Journal, 15(1): 1-3.

Huntley, D., 1978. On the detection of shallow aquifers using thermal infrared imagery. Water Resources Research, 14(6): 1075-1083.

IGRAC, 2004. Global Groundwater Information System. International Groundwater Resources Assessment Centre.

Jaiswal, R., Mukherjee, S., Krishnamurthy, J., Saxena, R., 2003. Role of remote sensing and GIS techniques for generation of groundwater prospect zones towards rural development - an approach. International Journal of Remote Sensing, 24(5): 993-1008.

Jenness, J., 2006. Topographic Position Index (TPI) v.1.2 Manual.

Jensen, J., 2007. Remote Sensing of the Environment. Pearson Prentice Hall, Upper Saddle River.

Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information-system analysis. Photogrammetric Engineering and Remote Sensing, 54(11): 1593-1600.

Jha, M., Chowdary, V., 2007. Challenges of using remote sensing and GIS in developing nations. Hydrogeology Journal, 15(1): 197-200.

Jha, M.K., Chowdhury, A., Chowdary, V.M., Peiffer, S., 2007. Groundwater management and development by integrated remote sensing and geographic information systems: prospects and constraints. Water Resources Management, 21(2): 427-467.

JISAO, 2010. Mean Sahel Precipitation Anomalies 1900-2009, http://jisao.washington.edu/data/sahel/. Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA.

Jones, M., 1985. The weathered zone aquifers of the basement complex areas of Africa. Quarterly Journal of Engineering Geology, 18: 35-46.

Khazanie, R., 1990. Elementary Statistics in a World of Applications. HarperCollinsPublishers Inc., 716 pp.

Klijn, F., Witte, J., 1999. Eco-hydrology: Groundwater flow and site factors in plant ecology. Hydrogeology Journal, 7(1): 65-77.

Krishnamurthy, J., Kumar, N., Jayaraman, V., Manivel, M., 1996. An approach to demarcate ground water potential zones through remote sensing and a geographical information system. International Journal of Remote Sensing, 17(10): 1867-1884.

Kruck, W., 1992. Interpretation of satellite data with special regard to hydrogeology. Hydrogeologie(1-2): 113-117.

Lindsay, J., 2005. The Terrain Analysis System: a tool for hydro-geomorphic applications. Hydrological Processes, 19(5): 1123-1130.

Mabee, S., Hardcastle, K., Wise, D., 1994. A method of collecting and analyzing lineaments for regional-scale fractured-bedrock aquifer studies. Ground Water, 32(6): 884-894.

MacDonald, A.M., Davies, J., Calow, R.C., 2008. African hydrogeology and rural water supply. In: Adelana, S.M.A., MacDonald, A.M. (Eds.), Applied Groundwater Studies in Africa. CRC Press, London, pp. 127-148.

Madrucci, V., Taioli, F., de Araujo, C., 2008. Groundwater favorability map using GIS multicriteria data analysis on crystalline terrain, Sao Paulo State, Brazil. Journal of Hydrology, 357(3-4): 153-173.

Mark, D., 1988. Network models in geomorphology. In: Anderson, M. (Ed.), Modelling Geomorphological Systems. John Wiley & Sons Ltd., pp. 73-97.

Masoud, A., Koike, K., 2006a. Constructing a ground water potential model for an arid environment integrating remote sensing and GIS, Proceedings of the International Association for Mathematical Geology, Toronto, Canada, pp. 53-68.

Masoud, A., Koike, K., 2006b. Tectonic architecture through Landsat-7 ETM+/SRTM DEM-derived lineaments and relationship to the hydrogeologic setting in Siwa region, NW Egypt. Journal of African Earth Sciences, 45(4-5): 467-477.

McKay, A., November 3, 2008.

Moussié , B., 1986. Map of Groundwater Resource Potentiality of West and Central Africa. In: (ICHS), I.-A.C.f.H.S. (Ed.). L'Institute Géographique National, Paris and BRGM, pp. 1:5,000,000.

Munch, Z., Conrad, J., 2007. Remote sensing and GIS based determination of groundwater dependent ecosystems in the Western Cape, South Africa. Hydrogeology Journal, 15(1): 19-28.

Mutiti, S., Levy, J., Mutiti, C., Gaturu, N., 2008. Assessing ground water development potential using Landsat imagery. Ground Water.

Naugle, J., 1996. Hand augered garden wells. Lutheran World Relief.

Naumburg, E., Mata-Gonzalez, R., Hunter, R., McLendon, T., Martin, D., 2005. Phreatophytic vegetation and groundwater fluctuations: A review of current research and application of ecosystem response modeling with an emphasis on Great Basin vegetation. Environmental Management, 35(6): 726-740.

Okereke, C., Esu, E., Edet, A., 1998. Determination of potential groundwater sites using geological and geophysical techniques in the Cross River State, southeastern Nigeria. Journal of African Earth Sciences, 27(1): 149-163.

Pennock, D., Zebarth, B., De Jong, E., 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. Geoderma, 40: 297-315.

Pirard, F., Faure, H., 1964. Carte de Reconnaisance Hydrogéologique du Niger Sud Oriental. Bureau de Recherches Géologique et Minières, pp. 1:1,000,000.

Practica, 2005. Report Phase 1: Assessment of the feasibility of manual drilling in Chad, Practica Foundation, Netherlands; United Nations Children's Fund, Chad; Oxfam, GB, Delft, Netherlands.

Quinn, P., Beven, K., Lamb, R., 1995. THE ln(a/tan-beta) index - How to calculate it and how to use it within the TOPMODEL framework. Hydrological Processes, 9(2): 161-182.

Ranganai, R., Ebinger, C., 2008. Aeromagnetic and Landsat TM structural interpretation for identifying regional groundwater exploration targets, south-central Zimbabwe Craton. Journal of Applied Geophysics, 65(2): 73-83.

Rangzan, K., Charchi, A., Abshirini, E., Dinger, J., 2008. Remote Sensing and GIS Approach for Water-Well Site Selection, Southwest Iran. Environmental & Engineering Geoscience, 14(4): 315-326.

Ringrose, S., 1998. Evaluation of vegetative criteria for near-surface groundwater detection using multispectral mapping and GIS techniques in semi-arid Botswana. Applied Geography, 18(4): 331-354.

Robinson, C. et al., 2007. The nubian aquifer in southwest Egypt. Hydrogeology Journal, 15(1): 33-45.

Salama, R., Tapley, I., Ishii, T., Hawkes, G., 1994. Identification of areas of recharge and discharge using Landsat-TM satellite imagery and aerial-photography mapping techniques. Journal of Hydrology, 162(1-2): 119-141.

Sander, P., 2007. Lineaments in groundwater exploration: a review of applications and limitations. Hydrogeology Journal, 15(1): 71-74.

Sander, P., Minor, T., Chesley, M., 1997. Ground-water exploration based on lineament analysis and reproducibility tests. Ground Water, 35(5): 888-894.

Schluter, T., 2006. Geological Atlas of Africa. Springer-Verlag, Berlin, Germany, 272 pp.

Sener, E., Davraz, A., Ozcelik, M., 2005. An integration of GIS and remote sensing in groundwater investigations: A case study in Burdur, Turkey. Hydrogeology Journal, 13(5-6): 826-834.

Solomon, S., Ghebreab, W., 2008. Hard-rock hydrotectonics using geographic information systems in the central highlands of Eritrea: Implications for groundwater exploration. Journal of Hydrology, 349(1-2): 147-155.

Stone Environmental, I., 2001. Republic of Niger Preliminary Atlas of Poverty/Vulnerabilty, Montpelier, VT, pp. Prepared for The World Bank Technical Department, Africa Region, March 2001.

Teeuw, R., 1995. Groundwater exploration using remote sensing and a low-cost geographical information system. Hydrogeology Journal, 3(3): 21-30.

Toth, J., 1963. A theoretical analysis of groundwater flow in small drainage basins. Journal of Geophysical Research, 68(16): 4795-4812.

Tweed, S., Leblanc, M., Webb, J., Lubczynski, M., 2007. Remote sensing and GIS for mapping groundwater recharge and discharge areas in salinity prone catchments, southeastern Australia. Hydrogeology Journal, 15(1): 75-96.

UN, 2002. Report of the World Summit on Sustainable Development, United Nations, New York.

UNDP, 2006. Human Development Report 2006. Beyond scarcity: Power, poverty and the global water crisis, United Nations Development Programme, New York.

UNDP, 2007. Human Development Report 2007/2008. Fighting climate change: Human solidarity in a divided world, United Nations Development Programme, New York.

UNICEF, 2009. Technical Notes 1-5: Manual Drilling in Africa. In: Fund, U.N.C.s., EnterpriseWorks/VITA, Foundation, P. (Eds.), www.unicef.org/wash/files/1_case_EN.pdf.

UNICEF, WHO, 2008. World Health Organization and United Nations Children's Fund Joint Monitoring Programme for Water Supply and Sanitation (JMP). Progress on Drinking Water and Sanitation: Special Focus on Sanitation, United Nations Children's Fund, New York; World Health Organization, Geneva.

van der Waal, A., 2008. Understanding Groundwater & Wells in manual drilling, Delft, Netherlands.

VanderPost, C., McFarlane, M., 2007. Groundwater investigation in semi-arid developing countries, using simple GIS tools to facilitate interdisciplinary decision making under poorly mapped conditions: The Boteti area of the Kalahari region in Botswana. International Journal of Applied Earth Observation and Geoinformation, 9(4): 343-359.

Waters, P., Greenbaum, D., Smart, P., Osmaston, H., 1990. Applications of remote sensing to groundwater hydrology. Remote Sensing Reviews, 4(2): 223-264.

WWC, 2009. 5th World Water Forum: Istanbul Ministerial Statement, Ministry of Foreign Affairs of Turkey; World Water Council, Instanbul, Turkey.