

2016

Using Remote Sensing and Biogeographic Modeling to Understand the Oak Savannas of the Sheyenne National Grassland, North Dakota, USA

Mandira Sigdelphuyal

South Dakota State University, mandira.sigdelphuyal@sdstate.edu

Follow this and additional works at: <http://openprairie.sdstate.edu/etd>

 Part of the [Physical and Environmental Geography Commons](#)

Recommended Citation

Sigdelphuyal, Mandira, "Using Remote Sensing and Biogeographic Modeling to Understand the Oak Savannas of the Sheyenne National Grassland, North Dakota, USA" (2016). *Theses and Dissertations*. Paper 679.

This Thesis - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact michael.biondo@sdstate.edu.

USING REMOTE SENSING AND BIOGEOGRAPHIC MODELING TO
UNDERSTAND THE OAK SAVANNAS OF THE SHEYENNE NATIONAL
GRASSLAND, NORTH DAKOTA, USA

BY

MANDIRA SIGDELPHUYAL

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Geography

South Dakota State University

2016

USING REMOTE SENSING AND BIOGEOGRAPHIC MODELING TO
 UNDERSTAND THE OAK SAVANNAS OF THE SHEYENNE NATIONAL
 GRASSLAND, NORTH DAKOTA, USA

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Geography degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Darrell Napton, Ph.D Date

Major Advisor

Niall Hanan, Ph.D Date

Thesis Advisor

George White, Ph.D Date

Head, Department of Geography

Dean, Graduate School Date

ACKNOWLEDGEMENTS

First, I would like to thank to South Dakota State University and the Geography Department, where I got the opportunity to pursue my goal. I would like to express sincere gratitude to my thesis advisor, Dr. Niall Hanan, who trusted me and my qualification, accepted my capacities, advised me, and elaborated of my thesis research. Besides, he also accepted my problems and weakness, and guided me in the direction of achieving the best goal in life.

In addition, I am grateful with my major advisor, Dr. Darrell Napton, and would like to thank for his support, guidance, advice, and encouragements throughout the academic program. Further, I would like to thank my committee members, Dr. George White and Dr. Bruce Bleakley for their suggestions, advice, and valuable time to make this research successful.

Personally, I would also like to express my thanks to Aashis Lamsal, Bijaya Upadhyaya, Christoffer Axelsson, Lara Prihodka, Armel Kaptue, and all fellow graduate students, researchers, and co-workers, including all GSCE/Geography and SDSU family for all the memorable and valuable time spent and the experience shared together.

I am very grateful to my family for all their sacrifices and support. I am pleased with Mr. Khem P. Phuyal for his incredible love, support, and guidance being as the best husband, a good friend, and a co-worker. Finally, I would like to express an unconditional love to our little charming baby, Aarya Phuyal, and say sorry for not providing her enough time when she needed, while working on this research.

CONTENTS

LIST OF ABBREVIATIONS	viii
LIST OF FIGURES	xi
LIST OF TABLES	xiii
ABSTRACT	xiv
CHAPTER 1: INTRODUCTION	1
1.1. PROBLEM STATEMENT AND DESCRIPTION	1
1.2. THESIS STATEMENT AND RESEARCH OBJECTIVES	2
CHAPTER 2: LITERATURE REVIEW	5
2.1. BIOGEOGRAPHY	5
2.1.1. <i>Introduction</i>	5
2.1.2. <i>Biogeographical understanding of species distribution</i>	6
2.1.3. <i>Importance of scale for species diversity and distribution</i>	7
2.1.4. <i>Oak Savanna</i>	8
2.1.4.1. Definition	8
2.1.4.2. Importance of oaks and oak savannas	9
2.2.4.3. Oak Savanna in the Woodland-Prairie Ecoregion of the Midwestern USA	9
2.2.4.3.1. Extent of oak savannas.....	9
2.2.4.3.2. History and management factors.....	14
2.2.4.4. A case study of the Sheyenne National Grassland Oak Savanna.....	15
2.1. SPECIES DISTRIBUTION MODELING AND USES.....	17
2.2.1. <i>Development of SDMs and Model Selection</i>	17
2.2.2. <i>Model Comparison and Evaluation of Drivers</i>	20
2.3. SUMMARY OF CHAPTER 2	20
CHAPTER 3: METHOD	22

3.1.	GEOGRAPHICALLY FRAMING THE TOPIC	22
	3.1.1. <i>Ecoregion: A broad scale understanding of ecosystem function</i>	22
3.2.	STUDY AREA	25
	3.2.1. <i>Geographic Location</i>	25
	3.2.2. <i>History</i>	27
	3.2.3. <i>Topography and features</i>	29
	3.2.4. <i>Land Use Practice and Grassland Management</i>	30
3.3.	METHODOLOGY	32
	3.3.1. <i>Data Collection</i>	32
	3.3.1.1. The Ortho Imagery	32
	3.3.1.2. Digital Elevation Dataset	32
	3.3.1.3. Soil Datasets	33
	3.3.1.4. Fire Data and Other Datasets	35
	3.3.2. <i>Data Analysis</i>	35
	3.3.2.1. Dependent variable	35
	3.3.2.1.1. Woody Canopy cover	35
	3.3.2.1.2. Presence/ Absence data (sample selection)	38
	3.3.2.2. Environmental Variables	41
	3.3.2.2.1. Soil Variables	41
	3.3.2.2.2. Topographic Variables	41
	3.3.2.2.3. Fire Disturbance Map	44
	3.3.3. <i>Data Interpretation</i>	48
	3.3.3.1. Empirical Variable Analysis	48
	3.3.3.2. Model Selection	49
	3.3.3.3. Modelling Techniques	49
	3.3.3.4. Model Evaluation	49
	CHAPTER 4: RESULTS	51
4.1.	LAND COVER CLASSIFICATION	51
	4.1.1. <i>Woody canopy estimate of the SNGI</i>	51

4.1.2. <i>Woody canopy estimate of the SNG2</i>	57
4.2. THE EMPIRICAL VARIABLES ANALYSIS	60
4.3. SPECIES DISTRIBUTION MODEL RESULTS	64
4.3.1. <i>Model Evaluation Statistics</i>	64
4.3.2. <i>Variable Importance</i>	66
4.3.3. <i>Projection Plots</i>	68
4.3.4. <i>Predictive Response Curve Analysis</i>	71
CHAPTER 5: DISCUSSION	75
5.1. DATA SELECTION AND ITS EFFECT ON DECISION-MAKING	75
5.2. EMPIRICAL VS PREDICTED RESULT COMPARISON	76
5.2.1. <i>Role of Soil types in oak distribution</i>	76
5.2.2. <i>SDM offered contrasting results at landscape level prediction and between two homogeneous environments</i>	76
5.2.3. <i>Superiority of RF Approach</i>	77
5.2.4. <i>Slope: The major driver of oak savanna existence</i>	78
5.2.5. <i>Disturbances effect</i>	78
5.3. FIELD VISIT: A VALIDATION TOOL	79
CHAPTER 6: CONCLUSION	81
6.1. SUMMARY OF THE RESEARCH AND RESULTS	81
6.1.1. <i>Influence of data quality in classification accuracy</i>	81
6.1.1. <i>SDMs offered important understanding of grassland oaks biogeography</i>	82
6.2. LIMITATIONS IN RESEARCH AND RESULTS	82
6.3. RESEARCH IMPLICATIONS	83
6.4. FUTURE WORK	84
APPENDICES	85
APPENDIX I: THE ORIGINAL SOIL POLYGON OF THE SNG1 REPRESENTING DIFFERENT SOIL	85

APPENDIX II: THE NRCS SOIL TEXTURE MAP OF THE SNG1 CREATED IN ARCMAP.....	86
APPENDIX III: THE NRCS SOIL DRAINAGE CLASSES MAP OF THE SNG1	87
APPENDIX IV: THE NRCS SOIL WATER SATURATION ZONE (WSZONE) MAP OF THE SNG1.....	88
APPENDIX V: THE 1-M SPATIAL RESOLUTION ORIGINAL DEM MAP OF THE SNG1.	89
APPENDIX VI: OAK VERSES ENVIRONMENTAL VARIABLES RESPONSE PLOT FOR THE SNG1	90
APPENDIX VII: OAK VERSES ENVIRONMENTAL VARIABLES RESPONSE PLOT FOR THE SNG2.....	91
APPENDIX VIII: OAK SAVANNA DISTRIBUTION IN THE SNG.....	92
APPENDIX IX: OAK SAVANNA DISTRIBUTION IN THE SNG.....	93
APPENDIX X: KUCHLER’S POTENTIAL NATURAL VEGETATION MAP (ORIGINAL)	94
REFERENCES	95

LIST OF ABBREVIATIONS

AAA:	Agricultural Adjustment Administration
AIC:	Akaike Information Criterion
ANN:	Artificial Neural Networks (Model)
ANOVA:	Analysis of Variance (one of the statistical tests)
AOIs:	Area of Interests
AUC:	Area Under Curve
BIC:	Bayesian Information Criterion
CTA:	Classification and Regression Trees (Model)
DEM:	Digital Elevation Model (data)
DPG:	Dakota Prairie Grassland
ERDAS Imagine:	One of the remote sensing software
FDA:	Flexible Discriminant Analysis (Model)
GAM:	Generalized Adaptive Model
GBM:	Generalized Boosted Models
GHF:	Grassland Heritage Foundation
GIS:	Geographical Information System

GLMs:	Generalized Linear Models
IA:	Indiana State
IL:	Illinois State
KS:	Kansas State
K-statistics:	Cohen's Kappa statistics
MAP:	Mean Annual Precipitation
MARS:	Multivariate Adaptive Regression Splines (Model)
MAXENT (MaxEnt):	Maximum Entropy (Model)
MrSID:	Multiple Resolutions Seamless Database
MUSYSM:	Soil Map Unit System by the Web soil Survey
NAIP:	National Agricultural Imagery Program
NFS:	National Forest Service
NLCD:	National Land Cover Database
NPP:	Net Primary Productivity
NPWRC:	Northern Prairie Wildlife Research Center
Pre/Abs:	Presence and Absence
PNV:	Potential Natural Vegetation

RF:	Random Forest (Models)
RRBMI:	The Red River Basin Mapping Initiative Project (website)
SCS:	Soil Conservation Service
SDMs:	Species Distribution Models/Modeling
SNG:	Sheyenne National Grassland
SNG-1 (SNG1):	Sheyenne National Grassland -1, a larger unit of the study area
SNG-2 (SNG2):	The Sheyenne National Grassland-2, a smaller unit or the study area
SRE:	Surface Range Envelop (Model)
TSS:	True Skills Statistics
US:	United States
USA:	United States of America
USDA:	United State Department of Agriculture
USGS:	United States Geological survey
WSS:	Web Soil Survey

LIST OF FIGURES

FIGURE 1: THE POTENTIAL NATURAL VEGETATION VERSION 2000 (v2000) MAP	11
FIGURE 2: MAP LEGENDS FOR THE FIGURE 1.	12
FIGURE 3: OAK AND OAK SAVANNA DISTRIBUTION IN THE MIDWESTERN.	13
FIGURE 4: OMERNIK LEVEL-III AND IV ECOREGIONS CLASSIFICATION MAP	24
FIGURE 5: GEOGRAPHICAL LOCATION OF TWO DIFFERENT UNITS OF THE SHEYENNE	26
FIGURE 6: PICTURES OF THE SNG FEATURES,	31
FIGURE 7: LAND COVERS CLASSIFICATION FLOW CHART.....	36
FIGURE 8: TREE (OAK SAVANNA) PRESENCE AND ABSENCE MAP OF THE SNG1	39
FIGURE 9: TREE PRESENCE AND ABSENCE MAP OF THE SNG2	40
FIGURE 10: ASPECT MAP OF A SMALL PART OF THE SNG1	42
FIGURE 11: BLOCK STATISTIC CALCULATION FLOW USING <i>MAXMIMUM</i>	43
FIGURE 12: FIRE FREQUENCY MAP OF THE SNG1.....	46
FIGURE 13: FIRE FREQUENCY MAP OF THE SNG2.....	47
FIGURE 14: THE MOSAICKED AERIAL PHOTO CONTAINING THREE SPECTRAL BANDS	53
FIGURE 15: LAND COVER CLASSIFICATION MAP OF THE SNG1	54
FIGURE 16: CHART SHOWING THE COMPARATIVE PERCENTAGE AREAS COVERED	55

FIGURE 17: LAND COVER CLASSIFICATION ACCURACIES COMPARISON CHART.....	56
FIGURE 18: LAND COVER CLASSIFICATION MAP OF THE SNG2	58
FIGURE 19: CHART SHOWING THE COMPARATIVE PERCENTAGE AREAS.....	59
FIGURE 20: THE EMPIRICAL VARIABLES ANALYSIS PLOTS OF THE SNG1.	62
FIGURE 21: THE EMPIRICAL VARIABLES ANALYSIS PLOTS OF THE SNG2.....	63
FIGURE 22: <i>BIOMOD</i> ₂ MODEL PREDICTIVE PROBABILITY OF OAK SAVANNA.....	69
FIGURE 23: <i>BIOMOD</i> ₂ MODEL PREDICTIVE PROBABILITY OF OAK SAVANNA PRESENCE IN SNG2.....	70
FIGURE 24: THE SNG1 PROBABILITY RESPONSE PLOTS OF OAK	73
FIGURE 25: THE SNG2 PROBABILITY RESPONSE PLOTS OF OAK	74

LIST OF TABLES

TABLE 1: HISTORY OF THE SNG ESTABLISHMENT PROCESSES	28
TABLE 2: MERGING AND RE-RANKING THE LIKELY SIMILAR NRCS SOIL TEXTURE.....	34
TABLE 3: LAND COVERS OF THE SNG1 WITH THEIR CLASSIFICATION ACCURACY IN PERCENTAGE.....	56
TABLE 4: LAND COVERS OF THE SNG2 WITH THEIR CLASSIFICATION ACCURACY IN PERCENTAGE.....	59
TABLE 5: REPRESENTATION OF DIFFERENT SDMS AND THEIR EVALUATION STATISTICS. ..	65
TABLE 6: THE VARIABLE IMPORTANCE SCORES OF THE SNG1 AND SNG2.	67

ABSTRACT

USING REMOTE SENSING AND BIOGEOGRAPHIC MODELING TO
UNDERSTAND THE OAK SAVANNAS OF THE SHEYENNE NATIONAL
GRASSLAND, NORTH DAKOTA, USA

MANDIRA SIGDELPHUYAL

2016

Oak savannas are valuable and complex ecosystems that provide multiple ecosystem goods and services, including grazing for livestock, watershed regulation, and recreation. These ecosystems of the woodland-prairie ecoregion of the Midwestern United States are, however, in danger of disappearing. The Sheyenne National Grassland, North Dakota, a protected Prairie grassland-savanna, is a representative of such rare habitats, where oak savanna is found at the landscape scale. In this research, I map the distribution patterns of oak savanna in the Sheyenne using a combination of remote sensing and geospatial datasets, including landscape topography, soils, and fire disturbance. Further, I interpret the performance of a suite of advanced Species Distribution Modeling approaches including Maximum Entropy, Random Forest, Generalized Boosted Model, and Classification Tree to analyze the primary environmental and management factors influencing oak distributions at landscape scales. Woody canopy cover was estimated with high classification accuracy (80-95%) for two study areas of the Sheyenne National Grassland. Among the four species distribution modeling approaches tested, the Random Forest (RF) approach provided the best

predictive model. RF model parameters indicate that oak trees favor gently sloping locations, on well-drained upland and sandy soils, with north-facing aspect. While no direct data on water relationships were possible in this research, the importance of the topographic and soil variables in the SDM presumably reflect oak preference for locations and soils that are not prone to water saturation, with milder summer temperatures (i.e. northern aspects), providing conditions suitable for seedling establishment and growth. This research increases our understanding of the biogeography of Midwestern tall-grass oak savannas and provides a decision-support tool for oak savanna management.

Key words: Oak Savanna, Woodland-Prairie, Midwestern Ecoregion, Sheyenne National Grassland, Species Distribution Model, Biogeography

CHAPTER 1: INTRODUCTION

1.1. Problem Statement and Description

Savannas are ecosystems with mixtures of trees or shrubs and herbaceous species (mostly grasses) (Hill, Román, and Schaaf 2011; Hanan & Lehmann 2011). They tend to occur in seasonally heterogeneous environments, with distinct wet and dry, or warm and cold, seasons. Prior work (e.g. Grundel, Pavlovic, and Bollinger 2007) suggested that globally, savannas once occupied about 1.6 billion hectares, or approximately 20% of the global land area (Hill, Román, and Schaaf 2011). Savannas provided some 30% of terrestrial net primary production (NPP) and supported diverse flora and fauna (Lehmann et al. 2014).

Savannas globally, including oak savannas in the central United States, are vulnerable to both climate and land use change (Apfelbaum and Haney 1987; Nielsen, Kirschbaum, and Honey 2003; Peterson, Reich, and Wrage 2007). Prior to Anglo-European settlement in North America, savannas covered more than 50 million hectares (e.g. Nielsen et al. 2003; Coop and Givnish 2007; Grundel, Pavlovic, and Bollinger 2007). However, in the Midwest, the original savannas have been altered or degraded through human activities such as transformation to agriculture, grazing, fire suppression, and increases in woody vegetation. Thus, savannas, including oak savannas, are increasingly rare native ecosystems, and are in need of protection and restoration. Ecosystem restoration provides richer biodiversity and enhances ecosystem services (Benayas et al. 2009). Hence, managers of the remaining US savannas need improved

understanding of the drivers and constraints on oak establishment and survival to protect these vulnerable systems, particularly at landscape scale (Schröder and Seppelt 2006).

Our understanding of species-environment relationships is challenging, at point to global scales, because of the limitations in data and technology. Geographic information system (GIS) and remote sensing techniques, however, may help mapping the savannas and improve our understanding of the biogeography of oaks and oak savannas, the impact of disturbances (Luoto and Heikkinen 2003; Roy et al. 2011; Schetter et al. 2012), and their relationship with different environmental variables. In addition, the development of Species Distribution Models (SDMs) may help us understand the drivers of current species distribution patterns, and thereby allow us to predict the future distribution of species in space and time as climate and anthropogenic pressures change (Thuiller 2003; Elith and Leathwick 2009). SDMs employ several distinct statistical approaches, with varying predictive abilities (Thuiller 2004; Elith and Graham 2009). Thus, choosing an appropriate SDM for a particular application (e.g. predicting where oak savannas might occur), and deriving biographical insight from those models is a complex task (Santana et al. 2010).

1.2. Thesis statement and Research objectives

This research is a case study of a protected Prairie grassland and oak savanna ecosystem in the woodland-prairie ecoregion of the Midwestern United States: the Sheyenne National Grassland (SNG) of North Dakota. The SNG is representative of the remaining Midwestern oak savanna ecosystems. However, there is not an accurate record of current distribution of oaks in the SNG as well as the physical and biotic factors that

control the presence of oak in this protected landscape. Based on research in other savannas (e.g. Sankaran et al. 2008) and in other Tallgrass prairie regions (e.g. Knapp et al. 1999), the density and tree cover in the Sheyenne is likely to vary with local environmental factors including climate, fire, topography, soil, and biological factors such as grazing and Native American activities (Leitner et al. 1991; Wolf 2004). These environmental factors that controlled oak distributions and created the savanna ecosystems of the SNG are the focus of this research.

In order to gain biogeographical understanding of oak savanna in the SNG, I developed the main hypothesis for this research that the non-random distribution of oaks in the SNG is constrained by the combination of soil types and topography, and modulated by disturbances (particularly fire). In order to test this hypothesis, this research also developed two main objectives:

1. To understand the particular driving factors that created non-random distribution of trees in the landscape.
2. To compare and contrast SDMs fitted using distinct statistical approaches.

In addition, to test the proposed hypothesis and to meet the objectives, in this research, I try to answer the following research questions:

1. What is the spatial distribution pattern of grassland-savannas and, in particular, woody canopy cover in the Sheyenne National Grassland?
2. What are the main biogeographic factors that determine tree distribution in the grassland?

3. To what extent do different statistical SDM techniques contribute new or contrasting understanding of the controls on species distributions?

I conducted the biogeographical analysis of current oak distribution using freely available statistical software *BIOMOD₂* (BIODiversity MODelling; Thuiller 2003) written for the R-programming language to answer my research questions. *BIOMOD₂* facilitates the calculation of multiple SDMs using different statistical approaches. I used *BIOMOD₂* to compare fitted models and the physical and biological covariates that provide most explanatory power for observed tree cover distributions within the Sheyenne study area.

The analysis provides key insights into the physical and biotic drivers impacting oak-savanna distribution at the landscape scale. It also delivers foresight for ecosystem conservation and managers, as well as new insights for scientists and students interested in biogeography and management of US mid-continental oak savannas.

CHAPTER 2: LITERATURE REVIEW

2.1. Biogeography

2.1.1. *Introduction*

Biogeography is a branch of geography that studies the distribution of different species including plants, animals, and others by examining how the physical environment affects them and shapes their distribution in space and through geological time. The definition of biogeography stated here is a general definition: definitions vary based on individual research interest, background, and understanding (Millington et al. 2011). Accordingly, biogeography has developed various sub-disciplines through time, including historical, conservation, ecological, analytical, and applied biogeography (Millington et.al. 2011). These sub-disciplines have contributed significantly to a variety of biogeographic and ecological theories, including speciation and evolution, and the role of glaciation, continental drift, extinction, dispersal, and other processes in controlling regional and global species distributions. The research reported here combines ideas from ecological and analytical biogeography to analyze the distribution of species (i.e. oak trees) in terms of their relation with physical and biotic environments using spatial and ecological niche modeling techniques (Santana et al. 2008; 2010; Blumler et al. 2011; Millington et al. 2011).

Malanson (2011) stated that in recent decades geographers have adopted complexity theory or “complex adaptive system theory” in their research to understand ecosystem function. Complexity theory assumes that the combined effect of multiple environmental factors determines the distribution of certain species, and thus those

factors need to be considered together to understand species distribution patterns. This research also assumes that multiple environmental factors (physical, biotic, and human-induced) are potentially important for oak distribution in the Sheyenne National Grassland.

2.1.2. Biogeographical understanding of species distribution

Biogeographers study biophysical factors that determine species range limits and how species change over time. Biogeographic research in vegetation dynamics, vegetation-environment patterns, and human impacts on biota has long been conducted (Veblen 1989; Young et al 2003). Species distributions at various geographic levels (e.g., global, regional, landscape, local, and micro (point) level) have fascinated biogeographers for centuries (e.g. Alexander Von Humboldt), and different biogeographical theories for species distribution patterns have emerged. However, it is still challenging to understand the factors that determine the distribution and dynamics of species diversity because of the limitations in data and technology (Pennisi 2005; Eiserhardt et al. 2011).

The encroachment of woody plants into grassland and conversion of open woodlands and savannas into shrublands, have been reported as a global phenomenon during the past decade (Maestre et al. 2009). Schlesinger and Pilmanis (1998) suggested that, in general, the transition from one ecosystem to another is accompanied by changes in the spatial pattern of soil resources and topography of the location. Further, Veblen (1989) and Young et al (2003) synthesized savanna studies and concluded that savanna dynamics could change over both short and long time intervals in relation to variable weather and long-term climate as a function of fire suppression, hydro-geomorphic

conditions, drought, insect outbreaks, and soil moisture availability. However, the effect of each variable is location, scale, and time dependent.

Geographic studies on species distributions and diversity provide important insights into the roles played by different environmental variables, including abiotic and biotic environments (Mistry 2011), as well as dispersal and biogeographic history of a particular species in a certain area (Gill and Beardall 2001; Eiserhardt et al. 2011). These studies increase the understanding of the drivers of species distribution at different spatial and temporal scales.

2.1.3. Importance of scale for species diversity and distribution

Species distribution and diversity relationships are often scale dependent (Diniz Filho et al. 2010; Tamme et al. 2010; Mutke 2011). Climate is the major driver for distribution of species at different scales, particularly, at global or regional scale (Bailey 2002; Schickhoff 2011). Broad-scale climatic patterns limit the extent of biomes (Bailey 2002), including the savanna biomes (Ritchie & Haynes 1987). Abiotic factors, particularly, temperature, precipitation, soil characteristics, and topography (which are directly or indirectly influenced or modified by climate) also determine the size and shape of species distributions at global, continental, and regional scales (Willis and Whittaker 2002; Coblentz and Riitters 2004).

Regional scale studies of African savannas suggest that woody cover is determined by the availability of resources (e.g. Sankaran et al. 2005; Bucini & Hanan 2007; Sankaran et al. 2008; Hanan & Lehmann 2011), such that maximum tree cover is constrained by mean annual precipitation (MAP) (Hanan & Lehmann 2011). However,

the combination of other disturbances such as fire frequency, grazing intensity, and soil texture or other nutrients also have direct and indirect impact on growth and regeneration via seed of woody vegetation (Sankaran et al. 2004). Research also suggests that savanna tree density can be limited by regional changes in herbivore abundance (e.g. Ritchie, Tilman, and Knops 1998).

At finer scale, local topography, hydrology, and geology can affect the species composition and tree density in savannas (Mistry 2011). At these fine spatial scales, the influence of biotic factors such as interactions among species (competition for resources, predation, and mutualism; Turner 2005) and combinations of abiotic factors, including local variation in soil properties and hydrology, orientation and exposure to direct sunlight, nutrient distributions, and other microclimate conditions, become progressively more important (Hortal et al. 2010; Eiserhardt et al. 2011). However, these factors may become less influential as the spatial extent increases.

2.1.4. *Oak Savanna*

2.1.4.1. Definition

Oak savannas are typically comprised of oak species (Bur Oak, *Quercus macrocarpa*, and other *Quercus* species) growing within the larger matrix of diverse prairie grasses and forbs, with tree canopy cover generally < 50% (Bradley et al. 2006; Coop and Givnish 2007). Oak seedlings are shade tolerant, and in certain circumstances oaks can form dense stands because of their longevity, resistance to fire, and ability to recolonize following disturbances (McShea and Healy 2002).

2.1.4.2. Importance of oaks and oak savannas

Oak woodlands have various ecological, environmental, and economic values. They are important for sustaining diverse groups of the North American floras and faunas (McWilliams et al. 2002). Because of their open structure (i.e. low canopy cover), these ecosystems are attractive to many species of wildlife and birds (game and non-game), while providing additional ecosystem services including provision of fodder for wild and domestic herbivores (Folliott, McShea, and Healy 2002; Sere et al. 2008). In addition, well-developed oak savannas promote high infiltration and groundwater recharge during the wet season, and a gradual release of water during the dry season, and thus, are important for watershed maintenance (Folliott, McShea, and Healy 2002). Further, oak timber and acorns are commercially important (McWilliams et al. 2002). They also have various cultural, aesthetic, and recreational values (Ffolliott, McShea, and Healy 2002).

2.2.4.3. Oak Savanna in the Woodland-Prairie Ecoregion of the Midwestern USA

2.2.4.3.1. Extent of oak savannas

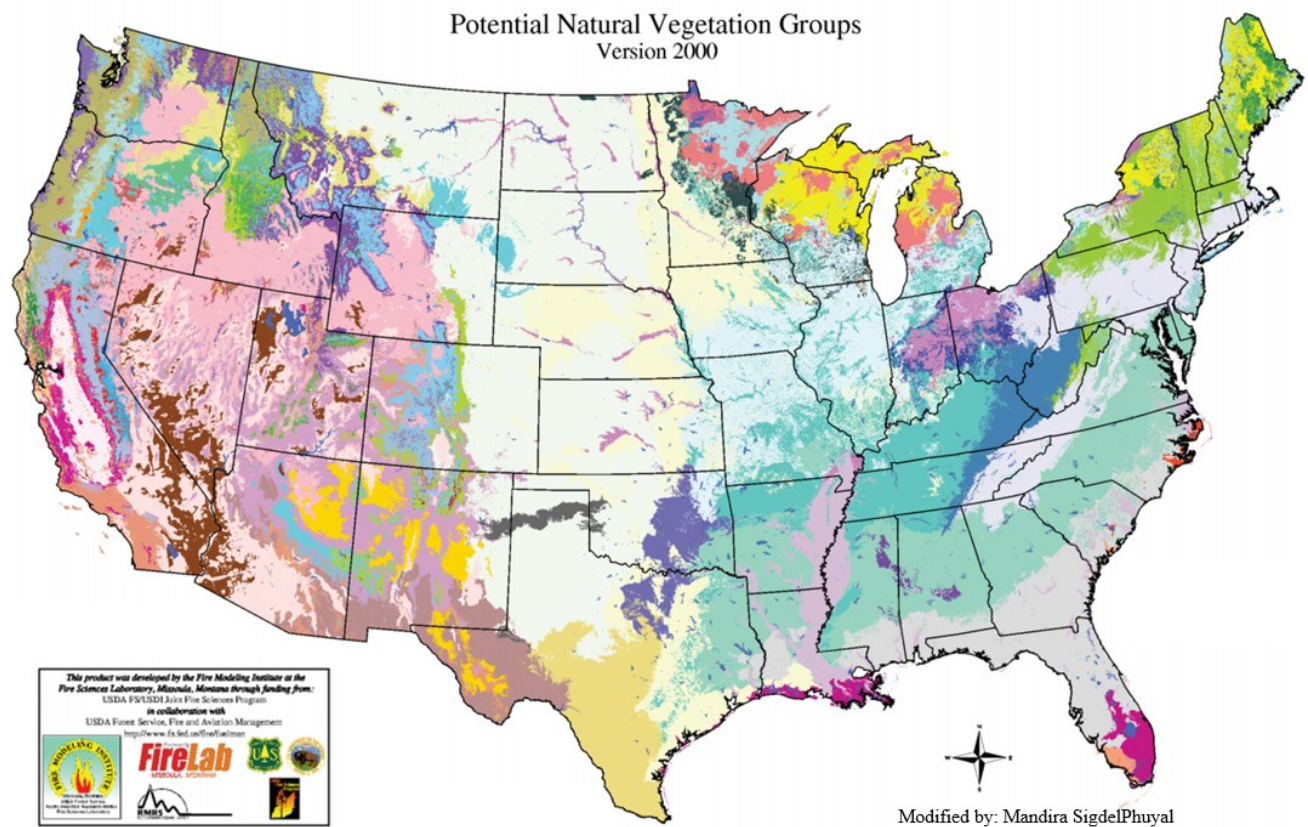
The Midwestern oak savanna was once an extensive ecosystem that covered some 11-13 million hectares stretching from Texas to Manitoba, Canada (Nielsen, Kirschbaum and Honey 2003). However, Nuzzo (1986) estimated only about 0.02% (2,600 ha) of undisturbed Midwest oak savanna remains (Restated in Brudvig and Asbjornsen 2005 ; Schetter et al. 2013). This highly dynamic vegetation community developed along ecotone between the grasslands and prairies.

According to Kuchler's (1964) original "Potential Natural Vegetation (PNV¹) map (Appendix X) and the later modified version (Version-2000, see Figure 1; USDA Forest Service 2000), oaks are the major vegetation types, particularly of the eastern USA. Further, the Midwestern region oak map (see Figure 3; extracted from the PNV v2000 map data) represents the potential dominant areas of oaks and oak savannas existence. These maps indicate that oak savannas were widespread in the grasslands and prairies of the Great Plains and woodlands to the east (details in Figure 3).

Figure 1 and Figure 3 also show that oak and oak savanna in the Midwestern regions has a mixed distribution, with other vegetation species, particularly, pine, hickory, and bluestem. The distribution of "Northeastern oak-pine" in the Midwest is negligible (see Figure 3; some small patches are at near to the Lake Michigan). Other vegetation types in the region include 'Plains grassland', 'Prairie' in SD, ND, NE, KS, and also in western and south-western part of MN and northern part of Iowa and 'Northern hardwoods', 'Great lakes pine forest, and 'Spruce-Fir' in eastern and northeastern parts of MN, WI, and MI respectively.

¹ PNV (or PNV map) is the "climax" vegetation of an area, projected considering certain environmental constraints but without human and natural disturbances (USDA Forest service 2014).

Figure 1: The Potential Natural Vegetation version 2000 (v2000)² map of the USA representing different projected vegetation types. The different colors in the map represent the locations of the different vegetation species. (Note: Because of the large map and map legends size, map legends are separated in Figure 2; source: USDA Forest Service, Fire, Fuel and Smoke Science Program, Rocky Mountain Research Station 2014 online).



² “This map is based on a terrain-matched refinement of Kuchler’s Potential Natural Vegetation (PNV) map. Kuchler’s PNV map was digitized for the conterminous United States, then adjusted to match terrain using a 500 meter Digital Elevation Model, 4th Code Hydrologic Unit delineations, and Ecological Subregions (Bailey’s Sections). These biophysical data layers were integrated with current vegetation layers, Resource Planning Act’s Forest Types and Forest Densities of the United States, and USGS EROS Data Center’s Land Cover Characterization database, to develop generalized successional pathway diagrams. Expert regional panels refined the PNV map based on these successional pathways” (USGS Forest Service, Fire, Fuel and Smoke Science Program, Rocky Mountain Research Station 2014 (online)).

Figure 2: Map Legends for the Figure 1.

Legend**Western Forests and Woodlands**

- 1: Pine forest
- 2: Great Basin pine (NV, UT)
- 3: Pine –Douglas –fir
- 4: Douglas –fir
- 5: Mixed conifer
- 6: Silver fir –Douglas –fir
- 7: Grand fir –Douglas –fir
- 8: Red fir (CA)
- 9: Spruce –fir –Douglas –fir
- 10: SW mixed conifer (AZ, NM)
- 11: Redwood (CA)
- 12: Cedar –hemlock –pine (WA)
- 13: Cedar –hemlock –Douglas –fir
- 14: Spruce –cedar –hemlock (WA, OR)
- 15: Fir –hemlock (WA, OR)
- 16: Spruce –fir
- 17: Lodgepole –subalpine (CA)
- 18: California mixed evergreen
- 19: Oakwoods (CA)
- 20: Mosaic cedar –hemlock –Douglas –fir & oak (OR)
- 21: Alder –ash (OR, WA)
- 22: Juniper –pinyon
- 23: Juniper steppe

Grasses, Shrubs, & Alpine

- 24: Mesquite bosques (NM)
- 25: Sagebrush
- 26: Chaparral
- 27: Southwest shrub steppe
- 28: Desert shrub
- 29: Shinnery
- 30: Annual grassland
- 31: Mountain grassland
- 32: Plains grassland
- 33: Prairie
- 34: Desert grassland
- 35: Texas savanna
- 36: Wet grassland
- 37: Alpine meadows –barren

Eastern Forests

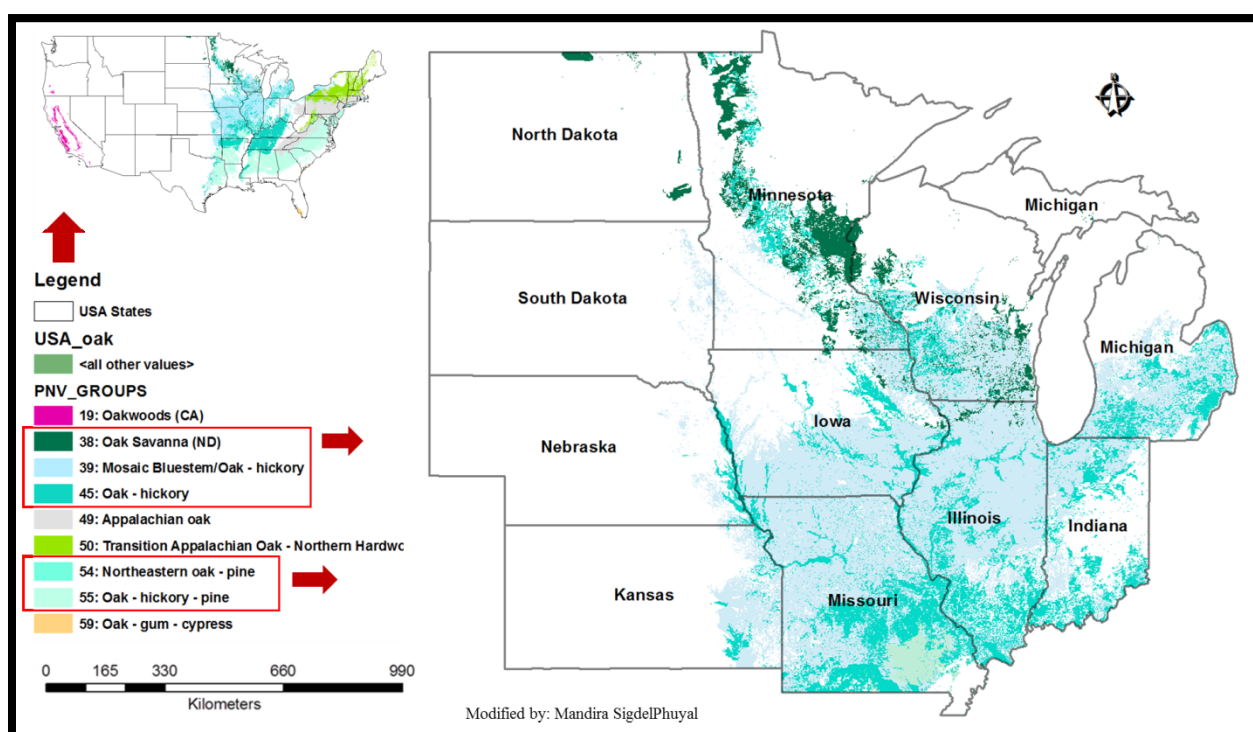
- 38: Oak savanna (ND)
- 39: Mosaic bluestem/oak –hickory

- 40: Cross timbers
- 41: Conifer bog (MN)
- 42: Great Lakes pine forest
- 43: Spruce –fir
- 44: Maple –basswood (MN, WI, IL)
- 45: Oak –hickory
- 46: Elm –ash forest
- 47: Maple –beech –birch
- 48: Mixed mesophytic forest
- 49: Appalachian oak
- 50: Transition Appalachian oak –northern hardwood
- 51: Northern hardwoods
- 52: Northern hardwoods –fir (MA, NH, NY)
- 53: Northern hardwoods –spruce
- 54: Northeastern oak –pine
- 55: Oak –hickory –pine
- 56: Southern mixed forest
- 57: Loblolly –shortleaf pine
- 58: Blackbelt
- 59: Oak –gum –cypress

Other

- 60: Northern floodplain
- 61: Southern floodplain
- 62: Barren
- 63: Water

Figure 3: Oak and oak savanna distribution in the Midwestern United States³ created by extracting the Potential Natural Vegetation version 2000 (v2000) data (Inset by: Mandira SigdelPhuyal, data source: USGS Forest Service, Fire, Fuel and Smoke Science Program, Rocky Mountain Research Station, Fire Science Laboratory 2014 online). The legend colors were created to match with the Figure 1 legends, and this map excluded other projected vegetation except oaks that are shown in the Figure 1.



³ Note: in this map (Figure 3), the presence of “Northeastern oak-pine” species is negligible; some small patches are occupying at the far eastern and northeastern parts (near to the Lake Michigan) of Wisconsin and Illinois respectively, which are difficult to see in this large scale image.

2.2.4.3.2. History and management factors

In the tallgrass prairie regions, the historical development and long-term expansion of oaks has been reported since the arrival of Europeans (Gibbens et al. 2005, 652). Several disturbance factors including frequency and intensity of fire, extensive land-use practices, and insects and disease outbreaks have been identified as having significant influence on oak establishment (Kellman, Miyanishi, and Hiebert 1985; Abrams, McShea, and Healy 2002; Nielsen, Kirschbaum, and Honey 2003). Such changes in the landscape, including the transition of grassland to shrub and tree dominated ecosystems were historically explained by considering the influence of factors such as climate, topography, soils, and large mammal grazing (including the interaction among those mentioned factors; Gleason 1922). However, bio-geographers, ecologists, naturalists, and other nature conservation and management scientists (Bader 2001) still debate the specific factors that determine the species composition and vegetation structure in these mixed woody-herbaceous (“tree-grass”) systems.

Coop and Givnish (2007) argued that the presence of woody species in US grassland in recent decades might depend upon the interactions among spatio-temporal variables including soil moisture availability and livestock grazing. Climate, however, has both direct and indirect influence in determining soil properties, topographic relief, vegetation structure, and fire frequency (Camill et al. 2003). Those heterogeneous landscape factors might have a significant influence on the distribution of woody species, including oak.

In addition, other research (e.g. Kellman and Sanmugadas 1985; Apfelbaum and Haney 1987; and Bock and Bock 1997) examined the impact of fire frequency and burning practices in restoring and maintaining the grassland oak savannas of the USA. These studies emphasize the role of fire in maintaining oak savannas in the woodland-prairies eco-tones of the Central United States. Other studies found that, in the absence of fire, oak regeneration is prevented by mesic species and characteristics of herbs (e.g. Apfelbaum and Haney 1987; Briggs et al. 2005; Brudvig and Asbjornsen 2005; Peterson, Reich, and Wrage 2007; George and Alonso 2008; Harrington and Kathol 2009; and Considine et al. 2013). Other disturbances such as drought, grazing, and tree harvesting can also alter oak stand structure both indirectly through altering fire behavior and directly by removing trees and creating gaps (Hayes and Holl 2003; Considine et al. 2013). In recent years, sensor and data management technologies have made possible satellite observation of wildfire frequency and extent in the grassland-savanna region (Roy et al. 2011), improving our ability to identify the role of fire in ecosystem processes at regional scales.

2.2.4.4. A case study of the Sheyenne National Grassland Oak Savanna

The undulating sand-dune landscape of oak savanna contributes to the remarkable scenic quality of the Sheyenne National Grassland. Bur oak savanna is a characteristic habitat of tall grass prairies including the SNG. It represents the remaining 1% of the original (i.e. ~ 13 million hectares; Domek 1998) oak savanna of the USA. However, Domek (1998) stated that, oak savanna and their quality are declining, and one of the main reasons for such decline in the SNG is that, oak savannas are not considered within the National Wilderness Preservation System conservation and management plan. This

statement is still valid, and thus, the management focus is only to Tallgrass Prairie (B. Stotts, US Forest Service, Lisbon, North Dakota, personal interview on October 10, 2014). The United States Geological Survey (USGS) Northern Prairie Wildlife Research Center (NPWRC) and the Grassland Heritage Foundation (GHF) also found several similar causes of oak savanna decline, including loss and fragmentation of habitat by agricultural conversion, fires, intensive livestock grazing, recreational activities, and the lack of spatially resolved information on oak savanna distributions (Apfelbaum and Haney 1991; Domek 1998).

Bryan Stotts (ranger of the SNG; US Forest Service, Lisbon, North Dakota, personal interview on October 10, 2014) also agreed that one of the major causes for declining oak savanna quality and quantity in the Sheyenne might be lack of information on appropriate management strategies. He added that low appreciation of the ecosystem goods and services provided by the oak savannas might also lead to oak savanna degradation in the SNG and elsewhere.

While, there is anecdotal evidence for the loss of oak-savanna, no prior work has mapped oak distributions in the SNG. Furthermore, no scientific research has been conducted to determine the physical and biotic factors that control the presence of oak-savannas in this protected landscape. Without understanding the major drivers that determine the presence and absence of oak-savannas in the grassland by the land managers and conservationist, it is difficult to apply proper management practices.

2.1. Species Distribution Modeling and Uses

Species Distribution Models (SDMs) are the tools that establish relationships between species occurrences and biophysical and environmental conditions. These modeling tools are becoming increasingly popular in many biogeographical and ecological applications (Anderson et al. 2006; Peterson et al. 2006). Numerous researchers (e.g. Guisan and Thuiller 2005; Garzon et al. 2006; Kelly et al. 2007; Gastón and García-Viñas 2011; van Gils et al. 2012) have used varieties of species distribution modeling methods to understand how physical and biotic drivers impact species distributions and community assembly. Based on these studies, SDMs can be used to estimate past, present, and future probability of species occurrence based on environmental conditions. These tools are also useful to predict species abundance, identify potential locations for rare and endangered species habitats, and estimate the potential spatial patterns of biological invasion.

2.2.1. Development of SDMs and Model Selection

Lazo (2013, 14) described that how SDMs are associated with the following three classes of techniques:

- 1) Profile techniques: requiring only species presence data
- 2) Discriminative techniques: requiring presence and absence data, and
- 3) Mixed-modeling approaches: using a combination of profile and discriminative methods mentioned above (1 and 2).

SDM algorithms are also classified as regression methods, machine-learning methods, classification methods, and enveloping methods. Thuiller, Georges, and Engler (2013; revised in 2015) made available the *BIOMOD₂* package in R-programming language that implements multiple SDM approaches, including the following:

1. Artificial Neural Networks (ANN)

ANN is a machine-learning method for predicting response vs explanatory variable relationships (Franklin 2010).

2. Surface Range Envelope (SRE)

SRE predicts the impact of climate on species in different environmental condition. This model works for broad scale (regional scale) species probability distribution modeling.

3. Generalized Boosted Regression Model (GBM):

GBM implements Boosted Regression Trees (BRT) and Gradient Boosted Regression models (Friedman et al. 2000; Friedman 2001; Moisen et al. 2006) which provides efficient non-parametric methods for fitting data with reliable predictive abilities (Thuiller et al. 2006).

4. Classification Trees Analysis (CTA) Model

CTA is a standard regression-tree method, which tests and evaluates multiple decision-trees with cross-validation on subsets of the training data (Lewis 2000).

5. Generalized Linear Models (GLM)

GLMs utilize linear techniques to predict species geographical distributions based on independent environmental data layers. GLM uses both presence and absence data, with alternative models evaluated based on AIC and BIC criterion.

6. Random Forest (RF) Model:

Breiman and Cutler's random forest approach is a powerful machine-learning method, and is considered to provide maximum prediction accuracy (Breiman 2001; Cutler et al. 2007). It has an ability to cope with non-linear relationships between response variables and driver variables and potential interactions between driver variables (Rodriguez-Galiano et al. 2012). Further, it can be used with both presence only, and presence-absence data sets (e.g. Thullier et al. 2014), and is flexible with both categorical and continuous data sets.

7. Generalized Adaptive Model (GAM):

This is a non-parametric model that provides better prediction performance if both presence and absence data are available. Instead of using linear or quadratic relationship to fit the predictor, it uses independent smoothing functions (Zaniewski, Lehmann, and Overton 2002).

8. Maximum Entropy (MaxEnt):

MaxEnt is another machine learning method, which works with presence only data to predict species distributions in environmental space based on a limited sample (Phillips, Dudík, and Schapire 2004).

This research tests the performance of the RF, CTA, GBM, and MaxENT modeling approaches (i.e. those models most suited to the presence/absence data available for oak distributions in the SNG).

2.2.2. *Model Comparison and Evaluation of Drivers*

Model comparison and evaluation of environmental drivers that cause the distribution of species in a particular study area, and model accuracy assessments are based on comparative analysis of the predictive performance of the models (Hanspach et al. 2010; Guo and Liu 2010; Santata et.al. 2010). In *BIOMOD*₂, several statistics are provided to compare model types and assess goodness-of-fit, including Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), Area Under the Curve (AUC) or Receiver Operating Characteristics (ROC), Cohen's Kappa (K), and True Skills Statistics (TSS) test (Thuiller 2003; Thuiller et al. 2003). Generally, model comparison is performed based on AUC (herein *BIOMOD*₂, ROC) value. If a model has AUC or ROC > 0.85, such model has strong predictive performance (Lazo 2013).

2.3. Summary of Chapter 2

Many of the woodland-prairie ecotone studies discussed above have tried to identify the causes of oak-savannas distributions in North America and causes for their decline, and most relate their work to discussions of the factors that determine the distribution of trees in the savannas at regional and global scales. However, few if any of these studies used the species distribution modeling approach proposed here to examine the role of physical environmental factors to determine the distribution of trees at landscape scales in North American oak savannas.

A localized analysis of the drivers of oak-savannas distribution will allow a better understanding of the current- status of these systems, identify the drivers of oak ability to persist in Tallgrass prairie, and develop the potential for management of tree density in the woodland-prairie ecotone region. By developing a fuller understanding of oak savanna ecology and biogeography, and selection of improved predictive models, we will provide managers with tools for better management of the remaining oak savanna woodlands of the region.

CHAPTER 3: METHOD

3.1. Geographically Framing the topic

3.1.1. Ecoregion: A broad scale understanding of ecosystem function

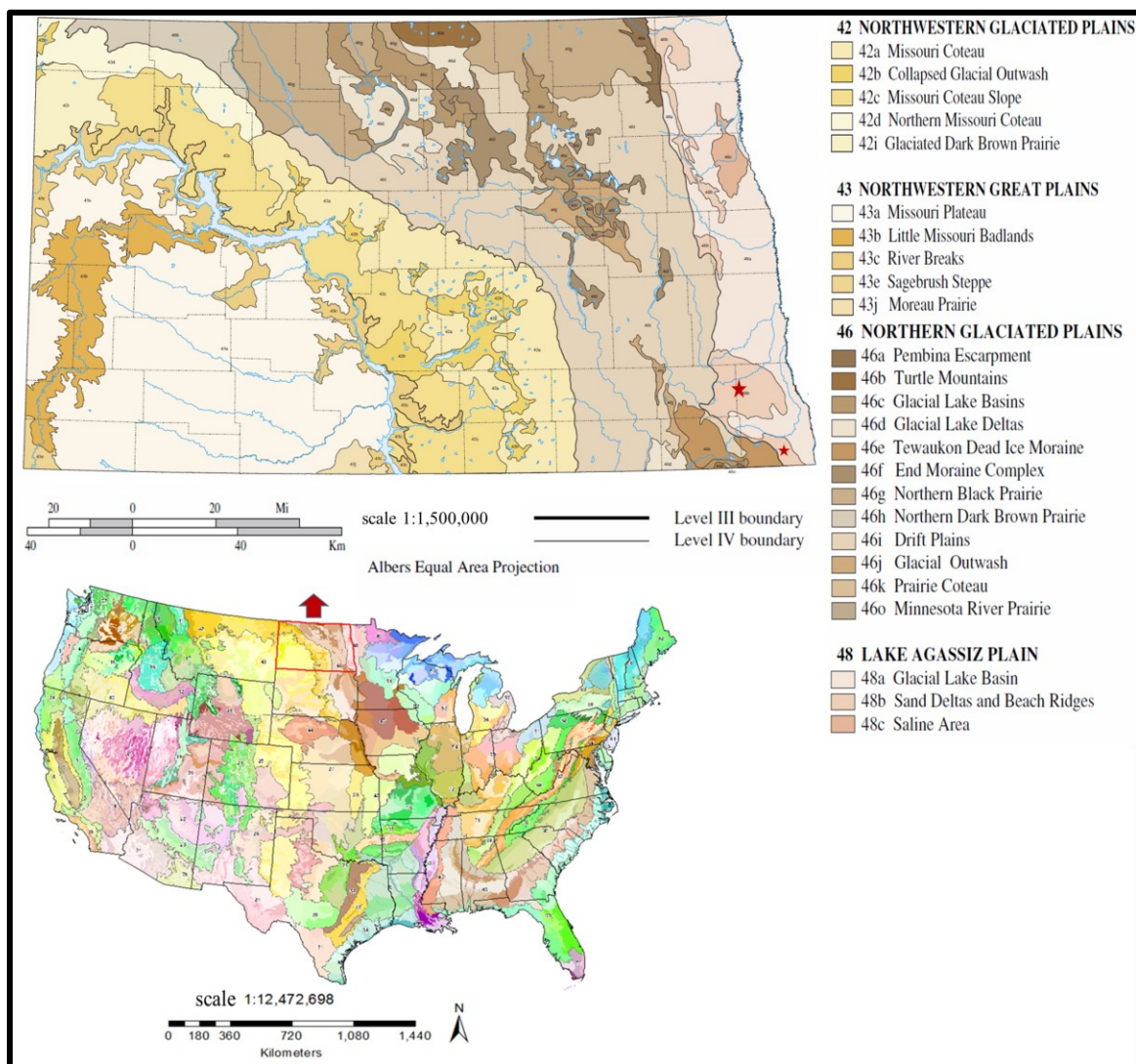
Ecoregions are the “areas of general similarity in ecosystem and in the type, quality, and quantity of environmental resources” (Omernik 1995). They relate to environmental characteristics and provide a geographically coherent context for studying the change of environmental features including land cover and land use (Omernik 1995; Loveland and Merchant 2004). Thus, they provide a spatial framework for understanding ecosystems and to guide ecosystem research, assessment, monitoring, and management (Loveland and Merchant 2004; Omernik 2004; Matlock and Morgan 2011).

According to Bailey (2002; 2004), at broad scales (i.e. “mesoscale” or regional level), hierarchies of ecosystems are in the form of all small ecosystems residing within the large ecosystems, thus, the regional scale ecosystem approach is useful for planning and management of ecosystems. Further, the proposed ecoregion concepts encourage studying ecosystem functions considering the relation of different environmental factors such as climate, topography, soil, vegetation, culture, and so forth (Omernik 2004).

Midwestern oak savannas appeared in various ecological regions based on individual ecoregions delineation maps (e.g. Bailey 1976 and 1995 maps by USDA Forest service; Omernik 1987 and 1995 (updated); and the Olson et al. 2001 (World Wildlife Fund (WWF) system)). Omernik ecoregions delineation corresponded to different levels of precision of climatic conditions and vegetation characteristics, where,

patterns and composition of biotic and abiotic factors (including physiography, geology, climate, soils, vegetation, land use, hydrology) are important in determining an ecological region. However, Bailey's ecoregion concept is climatology based, where variability in climate and climate driven factors are considered as the primary controllers over more localized ecosystems. In this research, I will use Omernik Level IV ecoregions in order to build general understanding of critical ecosystem aspects of the SNG, because Omernik Level IV (see Figure 4) is a local level ecoregion created by modifying and subdividing the Level III ecoregion. This is a more detailed map, and provides more specific local characteristics than other ecoregion maps. It helps me develop biogeographical understanding of oak savanna distribution in the protected landscape while considering together the other environmental factors.

Figure 4: Omernik Level-III and IV ecoregions classification map of North Dakota state including with USA ecoregions map⁴ (Source: USA map; US EPA Office of Environmental Information (OEI), Data: US EPA office of Research and Development (ORD) 2015; ND map and legends; Bryce et al. 1996 (color poster map); map modified by: Mandira SigdelPhuyal).



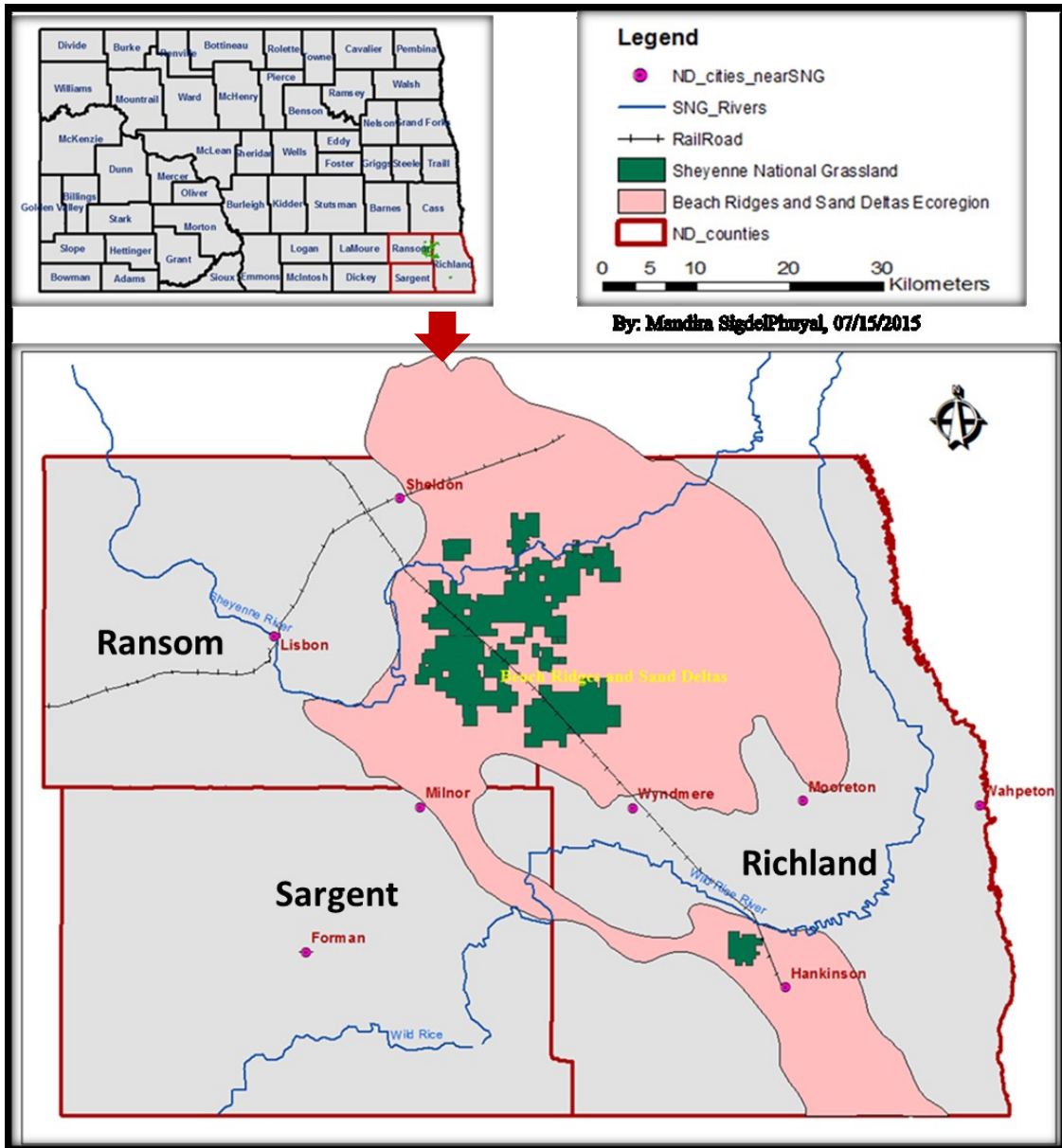
⁴ Numbers in the map represent different (Level III) ecoregions and sub ecoregion (Level IV), and different colors in the map indicated various level of ecoregions and sub ecoregions boundaries. Two red stars inside the Sand Deltas and Beach Ridges ecoregion (southeast corner of ND) are the location of the SNG study areas.

3.2. Study Area

3.2.1. *Geographic Location*

The Sheyenne National Grassland (SNG), the major part of the Sheyenne Ranger District, which is a part of Dakota Prairie Grassland (DPG), is in North Dakota, USA. The area is located approximately 46^o North latitude and 97^o West in the two counties of Ransom and Richland. The SNG comprises about 28,400 ha of public and about 26,200 ha of privately owned land, and the US Forest Service manages the public lands (Scheiman, Bollinger, and Johnson. 2003; Svedarsky and Van Amburg 1996; Manske and Barker 1988). It consists of two units (see Figure 5): the main north unit near the village of McLeod, comprises ~27,000-hectare, while the south unit is located near to the city of Hankinson, occupies a smaller (~1,000-hectare) area (Cunningham, Johnson, and Svingen 2006). Both units are analyzed in this study, which are represented as Sheyenne-1 and Sheyenne-2 (or SNG1 and SNG2) respectively. The grassland regions are located in the 'Beach Ridges and Sand Deltas' ecoregion (Figure 4 and Figure 5) of Omernik level IV ecoregions map. This ecoregion is on the geological formation i.e. delta formation, featuring some flat deltic plains to numerous choppy sand dune landscape structures (Bryce et al. 1996).

Figure 5: Geographical location of two different units of the Sheyenne national Grassland study area representing a part of the Level IV ecoregion (Figure 4): Beach Ridges and Sand Delta Ecoregion boundary and study area features.



3.2.2. *History*

The SNG is located on the Sheyenne Delta, which was formed during the Wisconsin glaciation where the Sheyenne River flowed into the southwest corner of the glacial Lake Agassiz leaving the area with sedimented layers of sandy soil (Stroh 2002) (see Figure 6c). The nickname of the area is “Sand Hill”, also called “Dakota Sand Hills”. Following extensive farming and the great drought of the mid-1930s, the area was in ‘Dust Bowl’ conditions and the extreme hardships caused many landowners to abandon the land and migrate west. To overcome or to mitigate the adverse consequences of the ‘Dust Bowl’, and to establish the SNG, several actions were implemented as shown in the Table 1 below.

Table 1: History of the SNG establishment processes

Actions	Year	Objectives or Consequences
Sheyenne River Land Utilization Project (under Agricultural Adjustment Administration (AAA))	1935	Resettlement plan completed
The Bankhead-Jones Farm Tenant Act The Land Utilization Project passed from AAA to the Resettlement Administration	1937	For the acquisition of the sub-marginal farm lands
Administration of the Project was transferred to the Soil Conservation Service (SCS)	1938	To ensure rehabilitation of the drought-devastated grasslands
Formation of the Sheyenne Valley Grazing Association (Non-profit organization of local landowners)	1941	To graze cattle on the surrounding federal land by Association members
Land transferred to National Forest Service (NFS)	1954	Land use practice changed to reestablish vegetative cover
The Sheyenne River Land Utilization project was formally named the Sheyenne National Grassland	1960	Further controls on grazing practices led to improvements in grassland condition.

(Source: Ransom County online)

3.2.3. *Topography and features*

The SNG has a gently rolling eolian topography, shaped by windblown sand (Stroh 2002). The area has several shallow wetlands, hills, roads, walking trails, river channels, and a railroad. The meandering Sheyenne River flows through the northwestern and middle of the northern part of the grassland (Based on Aerial photo 2012; field visit 2014) (see Figure 5 above). The area has a long, snowy, cold winter and short, dry, and hot summer.

The SNG is the only National Grassland in the Tallgrass prairie region of the United States. It comprises native tall prairie grasses, native forest, and non-native (e.g. Cropland) plant species including the beautiful oak savanna habitat. The grassland provides habitat for 850 endemic plant species out of 1,200 endemic plant species of North Dakota, among them, 40 of the present plant and animal species are considered rare and endangered (Svedarsky and Van Amburg 1996; Sieg and Bjugstad 1994). Rare species include the western prairie fringed orchid, greater prairie chicken, sharp-tailed grouse badgers, Dakota Skipper Butterfly, and so forth. Several different tree species, including Bur oak, Aspen, American elm, Basswood, Cottonwood and Willow are the major tree species of the grassland (Stroh 2002; field visit 2014). The presence of bur oak, often mixed with aspen, defines the oak savanna habitat.

The grassland is a home of several other flora including adders, tongue fern, lady-slippers, prairie rose, blazing stars, leadplant, and purple coneflowers, and so forth (Stroh 2002). Other fauna such as coyote, wild turkey, white-tailed deer, skunks, and on some

rare occasion, moose, elk, wolves, and so forth are also present there (Ransom County online).

3.2.4. *Land Use Practice and Grassland Management*

According to Bryan Stotts, a district Park Ranger (US Forest Service, Lisbon, North Dakota; personal interview in 10 October 2014), prescribed grazing (see Figure 6a) by Rancher Association leaseholders is the most common land use in the area. Prescribed burning (see Figure 6b) and mowing are also common practices to restore the native vegetation in the grassland. Other management treatment to reduce scrub encroachment includes herbicide application, bio-control, and use of sheep to control noxious weeds.

Figure 6: Pictures of the SNG features, land use, disturbances, and a common soil type: a. cattle grazing in the SNG1 with standing cottonwood trees, b. Prescribed fire practice under the Bur oak trees, c. Sandy Soil Profile. (Photo by: Mandira SigdelPhuyal, field visit 10 October 2014)

a.



b.



c.

3.3. Methodology

3.3.1. *Data Collection*

3.3.1.1. The Ortho Imagery

The major data for this research is the county-level Aerial (Ortho) photos of the two study areas of the SNG available for the year 2012. The United States Department of Agriculture, National Agricultural Imagery Program (USDA NAIP) (<http://gisdata.nd.gov/NAIP/>) collected these aerial photos. The images are provided in the Multiple Resolution Seamless Database (MrSID) image file format (raster); the file format specially designed for processing and mapping satellite image in portable format (Hovanes, Deal and Rowberg 1999). The images have 1-m spatial resolution with three spectral bands (3-Red, 2-Green, and 1-Blue) with spatial reference and datum, NAD83 UTM zone_ 14N and North America_ 1983 respectively. These images were used to estimate land cover and, in particular, woody canopy cover in the two study areas (Sheyenne-1 and 2). The woody canopy cover maps are the primary response variable for each study area.

3.3.1.2. Digital Elevation Dataset

The Digital Elevation data (DEM) were obtained from the Red River Basin Mapping Initiative (RRBMI) web site (<http://gis.rrbdin.org/lidarapps.htm>). The site provides high-resolution DEM raster datasets. They were collected using the airborne LiDAR, and available in two different spatial resolution (3 m and 1 m). The DEM has a similar spatial resolution and datum as the Ortho images, but were provided in ESRI grid Geographic Information system (GIS) raster format (AAIGrid but original files are in the

ASCII raster format) with 32-bit pixel depth and signed integer as pixel types. For this research, I used the higher resolution (1m) data. I used this DEM data to generate topological derivatives, including slope, aspect, curvatures, and local DEM to provide a detailed understanding of tree distribution patterns with respect to each topographic variable.

3.3.1.3. *Soil Datasets*

Soil data were available from the Web Soil Survey (WSS) (<http://websoilsurvey.nrcs.usda.gov/>). Data comes in spatial and tabular format. This dataset also has similar spatial reference and datum as the aerial images; however, they come in vector format, requiring pre-processing into 1 m raster format using GIS software. These datasets were used to generate soil variables, including soil classes, drainage, and depth of the seasonal water saturation zone (hereafter WSZone, which estimates the depth of the water table below the surface). The different soil variables were used as environmental layers for fitting the SDMs. These variables are categorized (i.e. categorical variables) in different soil classes by the United States Department of Agriculture, Natural Resource Conservation Service (USDA NRCS) based on physical properties. In this study, the NRCS classes were re-ranked based on soil texture with likely similar properties (see Table 2, under the title ‘NRCS Classes’; also see the map in Appendix I for SNG1) into numeric classes for ease of analysis in *BIOMOD*₂. (See Table 2 below):

Table 2: Merging and re-ranking the likely similar NRCS soil texture and Drainage property classes. Here, two or more nearly similar NRCS soil properties are merged, made a one class, and re-ranked them by number for ease of analysis in the modelling.

Categorical Variables	NRCS Classes	Reclassified
Soil Texture	Fine Sand/Moderately Decomposed Plant Materials (FS/MDPM)	1
	Loamy Fine Sand/Loamy Sand (LFS/LS)	2
	Fine Sandy Loam/Loam/Sandy loam (FSL/L/SL)	3
	Silty Clay/Silty Clay loam/Silt loam (SC/SCL/SL)	4
Drainage	Excessively Drained (ED)	1
	Moderately well drained/Well Drained (MWD/WD)	2
	Poorly drained/very poorly Drained (PD/VPD)	3
	Somewhat poorly drained (SPD)	4

3.3.1.4. *Fire Data and Other Datasets*

Several other GIS feature shape files of the study area, including States and counties boundaries, roads, railroads and water (as shown in Figure 5) were downloaded from the USGS Tiger-line data repository ([https:// www.census.gov/geo/maps-data/data/tiger-line.html](https://www.census.gov/geo/maps-data/data/tiger-line.html)). Management data on grazing and prescribed fire practices (vector data) were obtained from the SNG management and the United State Department of Agriculture, Forest Service (USDA FS) (<http://fsgeodata.fs.fed.us/vector/lrsr.php>); however, because of the insufficiency of grazing records, I only used fire data for the analysis.

3.3.2. *Data Analysis*

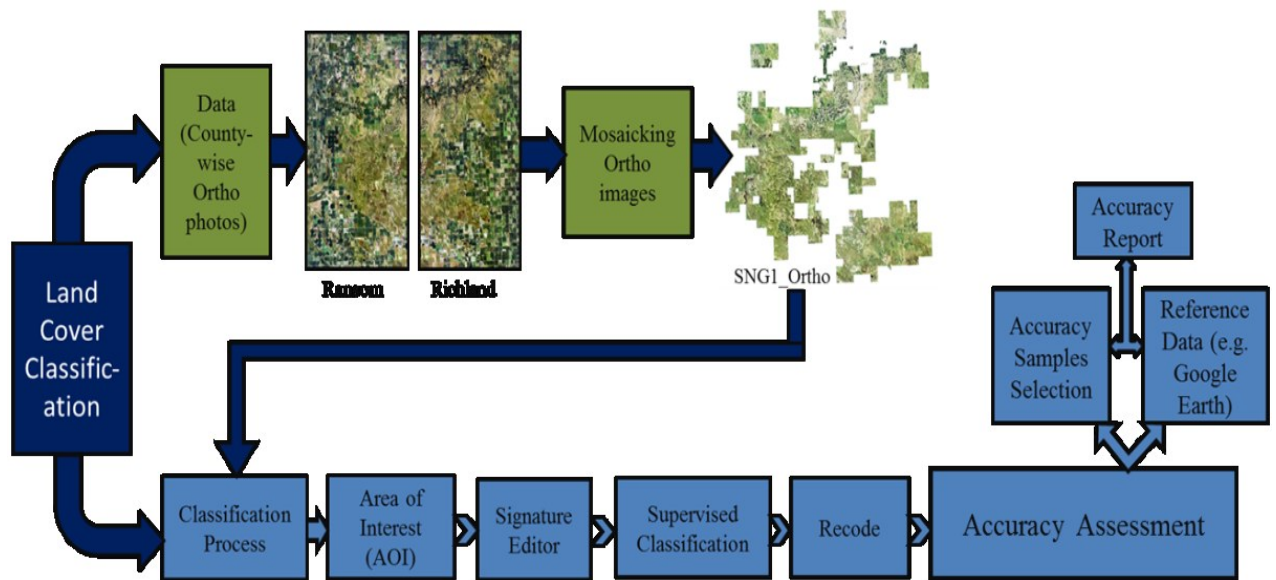
For the processing of data, this research used a combination of two different geospatial software applications including GIS software (ArcGIS 10.1/ 10.2) and ERDAS Imagine (2013-14), remote sensing image analysis software.

3.3.2.1. *Dependent variable*

3.3.2.1.1. *Woody Canopy cover*

In this research, the major land cover types including the woody canopy cover of the study areas were classified using supervised classification techniques (Long and Srihann 2004) as implemented in ERDAS-Imagine and illustrated in Figure 7.

Figure 7: Land covers classification flow chart using supervised classification techniques in ERDAS Imagine program.



In ERDAS Imagine, I made a single study area map by combining aerial imagery data for the two counties, a process called image mosaicking (Vinod et al. 2007). After mosaicking, landscape features, in particular, the oak and aspen tree canopies of the savannas, relative to open grassland and the distinctive willow canopies of the seasonally flooded areas were separated using a supervised classification process.

The supervised classification approach allowed me to identify five major land cover classes: oak savanna, grassland, water bodies, bare ground, and other tree canopies. First, signature editor files containing training data (pixels) for each land cover type were created using the *Area of Interest (AOI) tool*. The signature editor tool helps to select several training samples in polygon (AOIs) forms and merge them making a single

representative class in the final classification. The *Recode function* was then used to remove biases in the classification through the combination of similar classifications to a single classification (Demharter et al. 2011).

Finally, I performed accuracy assessments with ground truth samples from independent sources (Google earth images). For the accuracy assessment, 100 (SNG1) and 80 (SNG2) random points were projected in the classified image and assessed relative to ground truth images. These random accuracy points provided the percentage of classification accuracy. For the land cover classification accuracy, other reference data sources were used, including the GAP 2003 and NLCD 2006 land cover maps, personal contact with park personnel (interview with the Sheyenne National Grassland District Ranger - Bryan Stotts on October 10, 2014), and direct field observation. The final classification accuracies were determined based on accuracy statistics. Generally, three different types of classification accuracy measures are discussed for the supervised classification, including the Producer and User agreements, and Kappa coefficient (Herold et al. 2008). The producer and user errors relate to the omission and commission error statistics (Muller et al. 1998), where, producer agreements equal to 1-omission error and the user's agreement equals to 1-commission error. The classification accuracy (Kappa) also indicates the overall agreements in the classification. The higher the values of all three measures the better the classification agreement is considered.

The final classified image was used to generate an image showing the presence and absence of oak trees represented by number 1 and 0 respectively. The oak presence-absence data layer provided input (dependent variable) to run the SDM models.

3.3.2.1.2. Presence/ Absence data (sample selection)

Oak presence and absence data were randomly sampled to provide input to run the SDMs in *BIOMOD₂*. The high spatial resolution (1 m) tree-cover (presence-absence) data has more than 270 million pixels for the SNG1 and more than 11 million pixels for the SNG2 (see Figure 8 and Figure 9). To reduce sample size and speed up analysis, sub-sampling was necessary. Following experimentation with different numbers of data points, I selected a sample size of 200,000 random presence and absence points for the larger area (SNG1), and 100,000 random points (Presence & Absence) for the small study area (SNG2).

Figure 8: Tree (Oak Savanna) Presence and Absence Map of the SNG1 generated using the R programming language in *BIOMOD₂* package, showing a clear visual distribution of oak in different parts of the grassland. The number scale in X-axis represents longitude and Y-axis latitude of the study area; unit is in decimal degree (m).

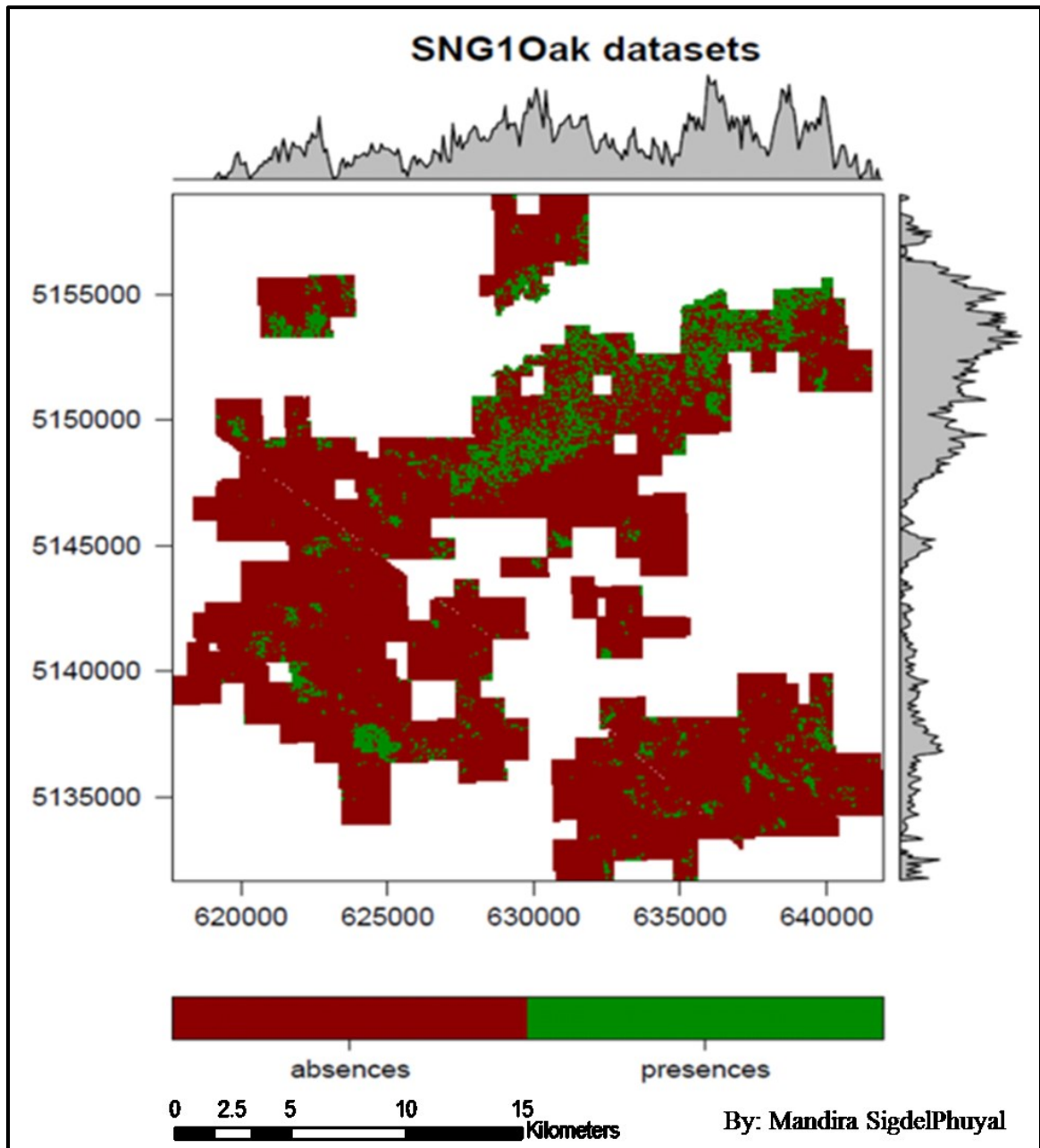
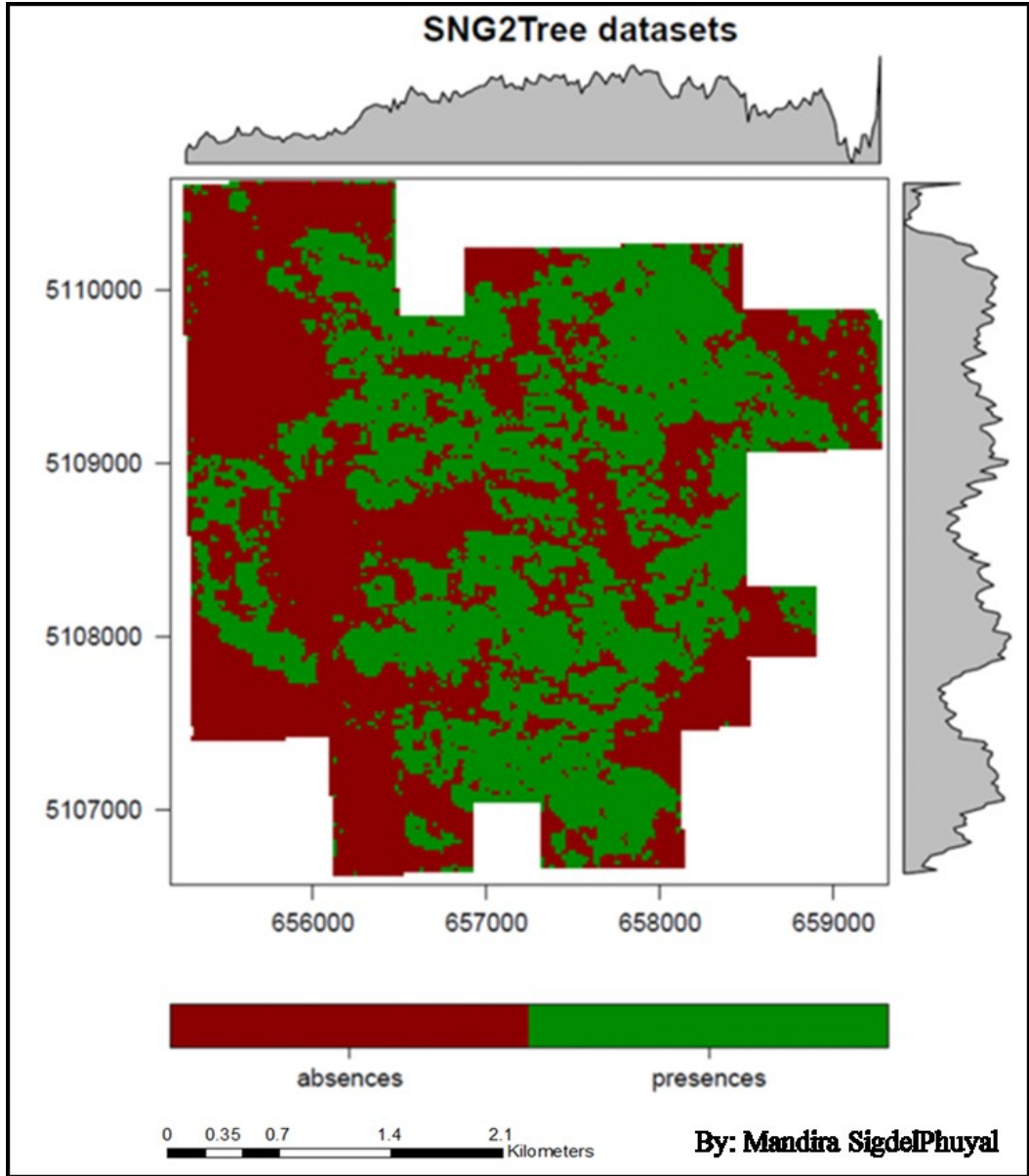


Figure 9: Tree Presence and Absence Map of the SNG2 generated using the R programming language in *BIOMOD₂* package, showing a clear visual distribution of oak in different parts of the grassland. The number scale in X-axis represents longitude and Y-axis latitude of the study area; unit is in decimal degree (m).



3.3.2.2. Environmental Variables

3.3.2.2.1. Soil Variables

I created a soil classification table using the NRCS soil type information to provide attribute information on the different soil classes for the selected soil polygons of both study areas. The soil polygon map only contains different soil polygon names in symbolic form (e.g. see Appendix I for the SNG1) but it does not provide the actual soil properties. Thus, by spatially joining (using *Join* function in ArcMap) the soil properties table with a vector data layer, I prepared vector maps containing different soil properties. Later, I generated different soil variable maps in raster (grid) format (e.g. see Appendix II–IV for the SNG1) that are required for running the final analysis (modelling). The maps (raster map) were generated in 1-m spatial resolution *Geotiff* grid format, with consistent metadata (e.g. spatial reference, datum, and so forth). These maps represented the independent environmental variables of oak savanna.

3.3.2.2.2. Topographic Variables

I used the digital elevation dataset (DEM; see Appendix V for the SNG1) to create the following four topographic derivative maps: slope, aspect, curvature, and local DEM using ArcMap version 10.1 and 10.2.

1. Deriving a Slope raster from a DEM

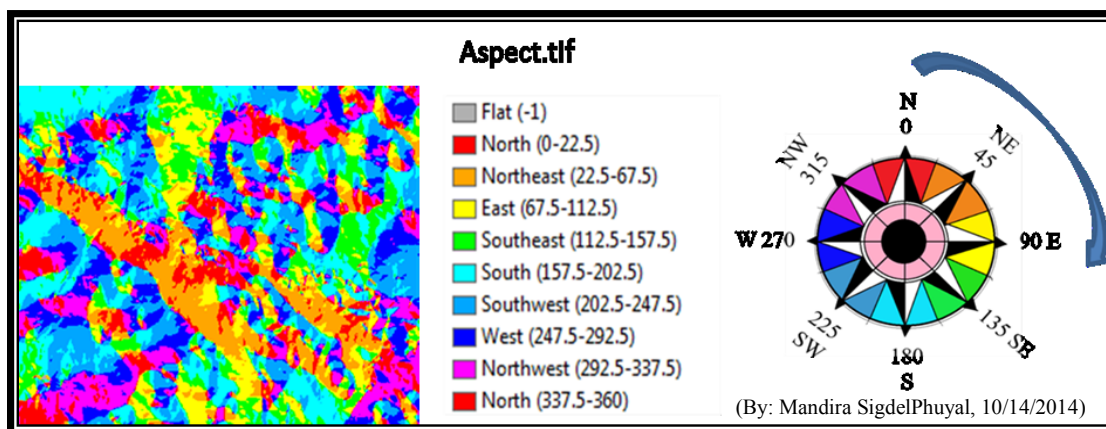
Slope identifies the vertical change in elevation. In a raster DEM, the slope represents the rate of maximum change in elevation (z- value) between cells.

For this analysis, slope was derived from the original DEM using the *DEM to Slope tool* in ArcMap retaining the original spatial resolution and providing output in degrees. In this research, slope represented the terrain steepness, which may affect oak tree establishment and survival processes, and thus influence the distribution of oaks in the SNG.

2. Deriving an Aspect raster from a DEM

Aspect represents the polar orientation of a slope from different angles such as zero (North), 90 (East); 180 (South), 270 (West), and the 360 (North) (see Figure 10). In this research, Aspect is calculated using the *DEM to Aspect tool* in ArcMap making all properties consistent with the original DEM. In my research of understanding the species verses environmental variables relationship (i.e. Oak vs Aspect), aspect could provide information about the directional pattern of oak distribution in the grassland.

Figure 10: Aspect map of a small part of the SNG1 representing the aspect (unit = degree). The different colors represent aspect directions. The flat areas do not have any downslope direction, which is generally represented by “-1”.



3. Deriving Local DEM from a DEM

Local DEM quantifies elevation changes relative to the local mean elevation. It is calculated over a defined distance, where the DEM itself is relative to mean sea level.

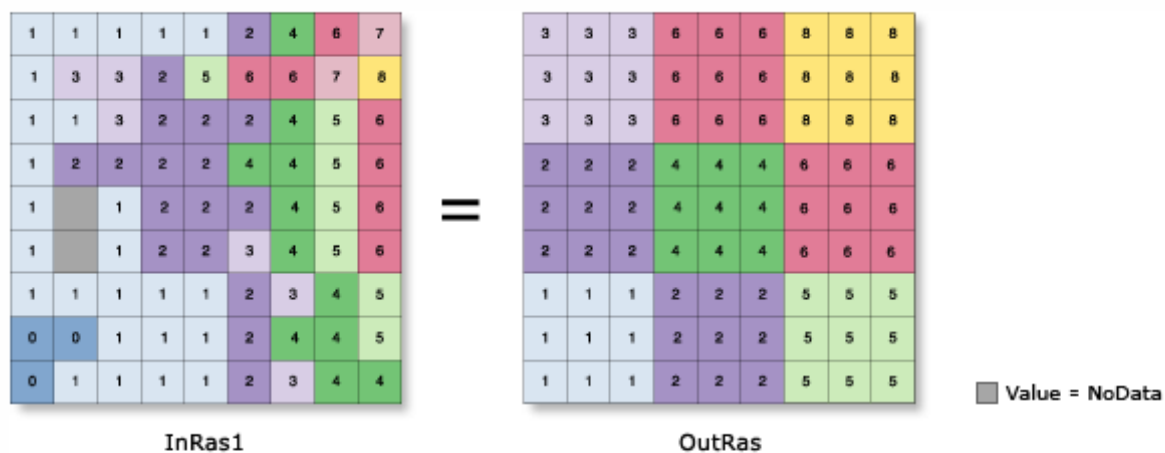
In this research, I calculated Local DEM using the following equation in ArcMap.

$$\text{LocalDEM} = (\text{DEM} - \text{Block Stat})$$

Where, the Block Statistic is a neighborhood function that partitions the input cells (raster pixel) into specified non- overlapping blocks such as 3x3, 4x4, 100x100, and so forth with values in each block (e.g. see Figure 11 below).

Figure 11: Block Statistic calculation flow using *MAXIMUM* as the statistics type.

(Source: ArcGIS Help Online).



`OutRas = BlockStatistics(InRas1, NbrRectangle(3,3,MAP), "MAXIMUM", "")`

The blocks can be defined in different shape including rectangular or square (used here), and circular; and the output can be adjusted to provide maximum, mean, minimum and other statistics depending on research needs. For this research, I calculated average (i.e. statistics type is 'MEAN') block statistics of each input cell by defining 500mx500m neighboring blocks, where the final product will provide the average elevation of the defined location. The analysis of mean neighborhood elevation was performed in ArcMap using the *Neighborhood Statistics tool*.

4. Deriving Curvature from a DEM

Curvature is defined as the second derivative of the DEM surface and can be calculated in ArcMap. ArcMap allows creation of three different curvature output rasters, including profile curvature raster (optional), plain curvature (optional), and a combined curvature raster using the *Curvature Tool*. In this research, I used the combined curvature raster, where curvature value ranges from negative, zero, and positive. A negative value in the cell indicates the upward convex and a positive profile indicates upward concave surface, and a value zero ('0') indicates a linear surface profile. This map will allow us to quantify oak distribution patterns relative to curvature.

3.3.2.2.3. Fire Disturbance Map

The fire frequency, disturbance maps were prepared using fire shape files obtained from the SNG for both study areas. The shape file has 15 years (2000-2014) of overlapping fire polygons. The fire frequency (presence and absence) raster map was obtained by stacking all the layers (fire polygons) in ArcMap. Fire frequency is the total number of fires observed in each pixel location during a 15-year period (hereafter

FireFreq; Figure 12 and Figure 13). In this research, I interpreted the fire frequency numbers as shown below

Fire frequency value,

0 = Fire free

1 and 2 = Low fire

3 and 4 = Intermediate fire

5 and 5+ = High fire

Figure 12: Fire frequency map of the SNG1 indicating the numbers of fires occurring during 15 years (2000-2014).

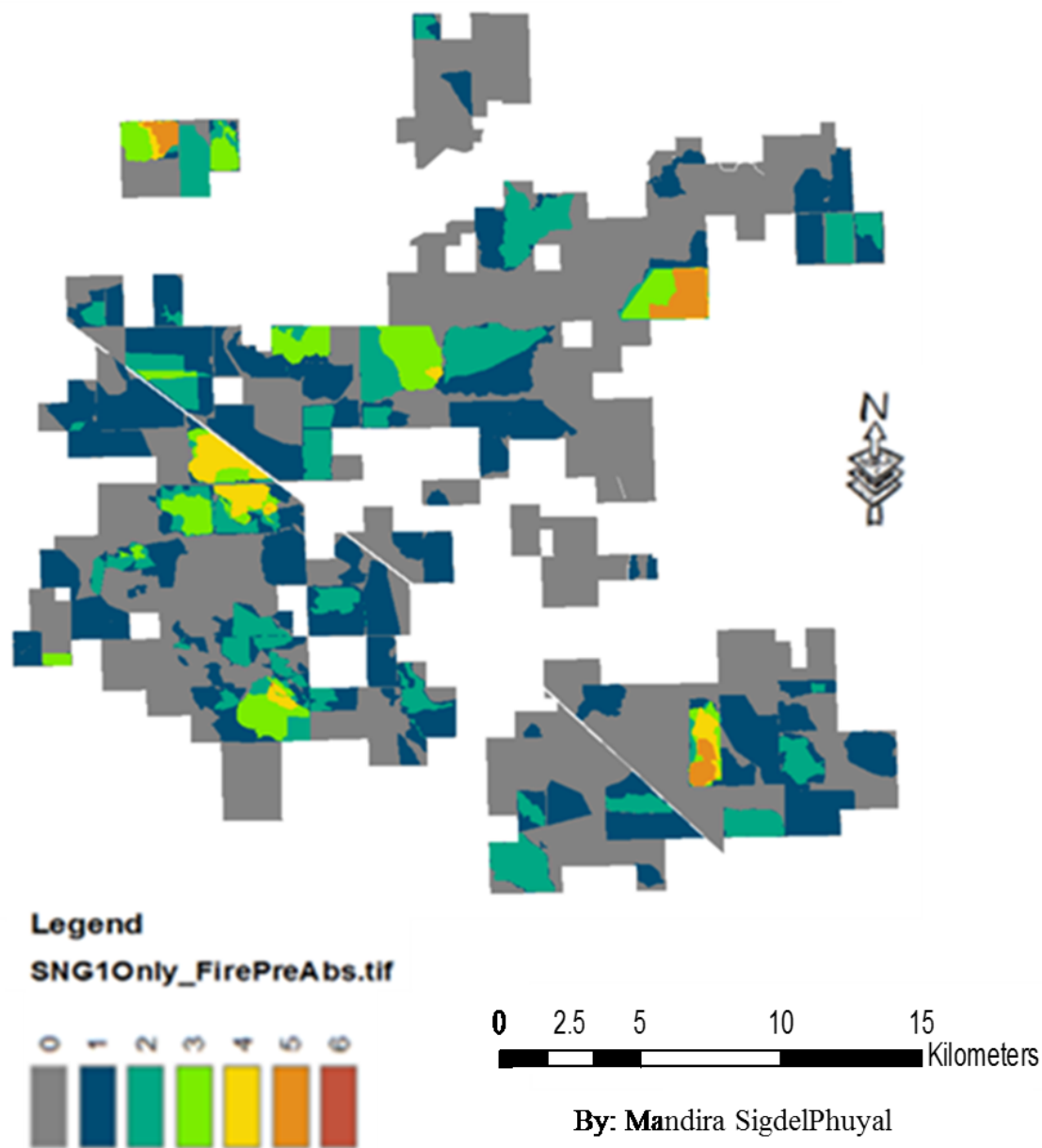
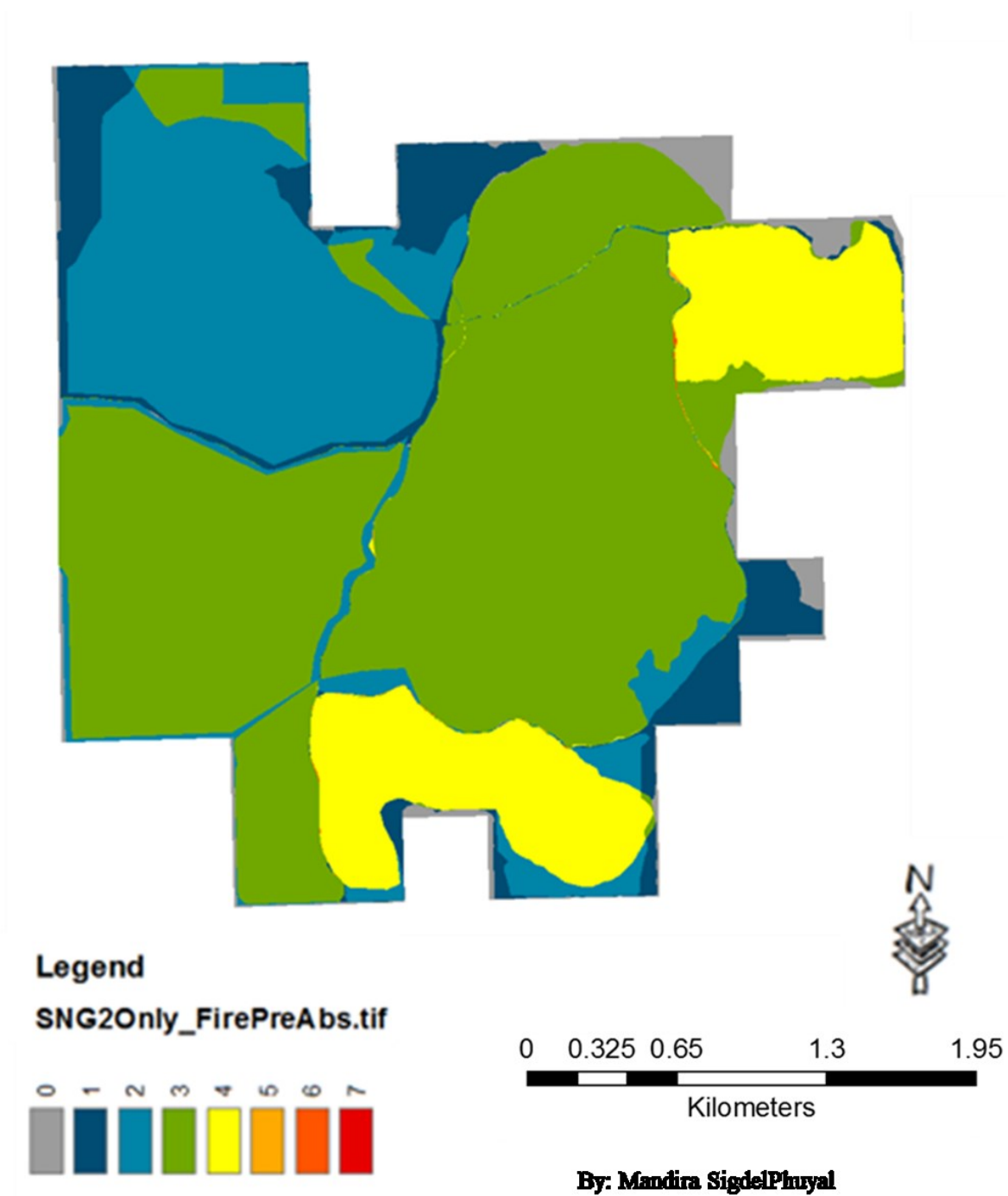


Figure 13: Fire frequency map of the SNG2 indicating the numbers of fires occurring during 15 years (2000-2014).



3.3.3. Data Interpretation

The high spatial resolution (1 m) of the tree cover and associated environmental data sets results in a total number of potential samples (i.e. Pixels) in the SNG1 domain of more than 270 million. For computational reasons, I randomly sub-sampled across the entire domain to obtain a dataset on presence-absence of oak trees, together with corresponding data on soil types, topography and fire history.

3.3.3.1. Empirical Variable Analysis

Empirical variable analysis is the simple relationships between oak occurrence and individual environmental variables, where, the individual relationships between the dependent variable (i.e. fraction of the oak canopy cover) and independent variables (environmental variables) plotted on the Y-axis and X-axis respectively. Here, the dependent variable is a fraction of the total oak savanna of the study area, calculated as:

$$\frac{\text{Oak Savanna Pixels}}{\text{Total Pixels}}$$

Where “Oak savanna pixels” are cumulative oak presence pixels in the random sample separated into regular interval bins across each environmental dataset, and “Total pixels” quantifies the total number of random points in the bin.

The fraction of oak presence versus environmental variables relationships are shown to provide an initial assessment of the potential importance of each variable. However, the effect of individual environmental variables on oak occurrence may be masked or exaggerated by correlation among variables. The detail about the trends in tree

occurrence relative to environmental variables is discussed in chapter 4 (Figure 20 and Figure 21) of this research.

3.3.3.2. Model Selection

Four SDM modeling approaches: MAXENT, GBM, RT, and CRT, were selected for this analysis based on their ability to work with both categorical and continuous variables, and flexibility with non-linear relationships (see Figure 20 and Figure 21 in the results section) and correlation between independent variables.

3.3.3.3. Modelling Techniques

For the modelling, all the environmental variables were prepared with a similar extent, resolution, geographical projection, and format (i.e. Geotiff grid, or other GIS format) for ease of analysis in *BIOMOD*₂. The R-script, techniques of modelling, and parameter settings were based on recommendations described by Georges and Thuiller (2013) and Thuiller, Georges, and Engler (2013 and 2015). Model fitting and cross-validation was carried out using an 80 to 20 split of the dataset (setting `DataSplit = 80` in *BIOMOD*₂, such that models were fit with an 80% sub-set of the data and cross validated with an independent 20% sub-sample. Model results are presented for all calibrated models as well as full model predictions, where the performance of each model is also analyzed based on the entire dataset.

3.3.3.4. Model Evaluation

To evaluate the predictive performance of each model, I used the predictive response curve (predictive response of tree cover with selected environmental variables),

predictive projection plots, and model performance statistics (i.e. ROC, TSS, and KAPPA respectively). Further, the ranking of each independent variable's importance in the fitted models ("VarImp") is used to assess each variable's contribution to fitted models.

However, variable importance is only useful to understand the direct effect of variables to the model (not the combined effect) (Thuiller, Georges, and Engler 2012).

CHAPTER 4: RESULTS

4.1. Land Cover Classification

4.1.1. *Woody canopy estimate of the SNG1*

In this research, I classified the major five land cover types of the SNG1 using the 1-meter spatial resolution aerial imagery (Figure 14). These classes included oak savanna, water, bare ground, grassland, and other trees. Here, the ‘other tree’ class includes some other tree species besides oak. Field visit helped me to identify the location of different vegetation types, which ultimately, helped with generation of training data to classify the distinct tree classes.

The SNG1 is the larger and more heterogeneous than the SNG2 and thus relatively more difficult to classify using the spectral bands available in the aerial imagery (which has just three bands) (see Figure 14). This limitation created some spectral confusion (Hu and Wang 2013) between land cover types (e.g. two species of tree with nearly similar spectral reflectance). Thus, in some cases, separating (spectrally and visually) objects with similar reflectance properties during classification are difficult. In my classification, the oak canopy class includes some aspen, since these often co-occur in the areas typically considered as oak savanna. Further, the classification between grassland and bare soil was often less successful because of spectral similarity of bare soil and sparse herbaceous vegetation.

Overall classification accuracy was ~85% (see Table 3), with a 0.82 Kappa statistic score. The individual Kappa scores (see Table 3 and Figure 17) vary between

0.94 for the oak tree class to as little as 0.65 for bare soil. However, since the focus of this study is on oak distributions in the grassland, the relatively high accuracy of oak detection suggests the classification can be used for subsequent analysis. The classification result for the SNG1 estimated oak tree cover of ~5% with other trees (mostly lowland willow and cottonwood) at ~2% (see Figure 16 and Table 3).

Based on those classification results, the majority of oaks in the SNG1 are in the Northeastern part of the grassland, with distinct patches of oak in the south-west of the SNG1 and smaller patches (or individual trees) scattered across the area (See Figure 15 and for better visual see Figure 8). The patchy distribution of oaks may reflect locations with favorable environmental condition for seedling establishment and growth.

Figure 14: The mosaicked Aerial photo containing three spectral bands used in the classification of the SNG1 (data source: United States Department of Agriculture (USDA) Farm Service Agency, National Imagery Program (NAIP); acquisition date July 2012).

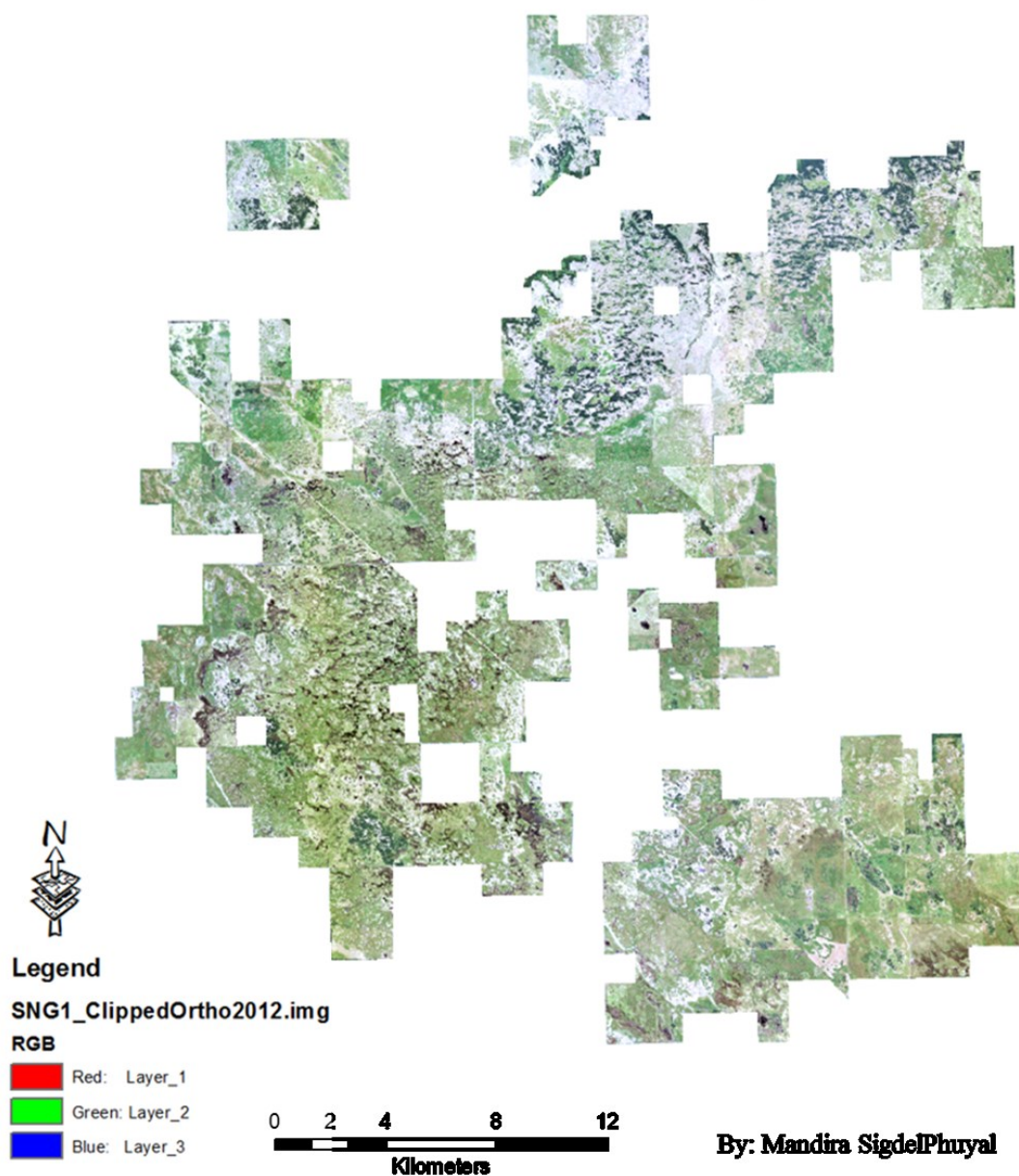


Figure 15: Land cover classification map of the SNG1 created using the supervised classification technique in ERDAS Imagine. The land cover classification in this map was generated using 1-meter spatial resolution aerial photo, and thus it is a one-meter resolution raster image.

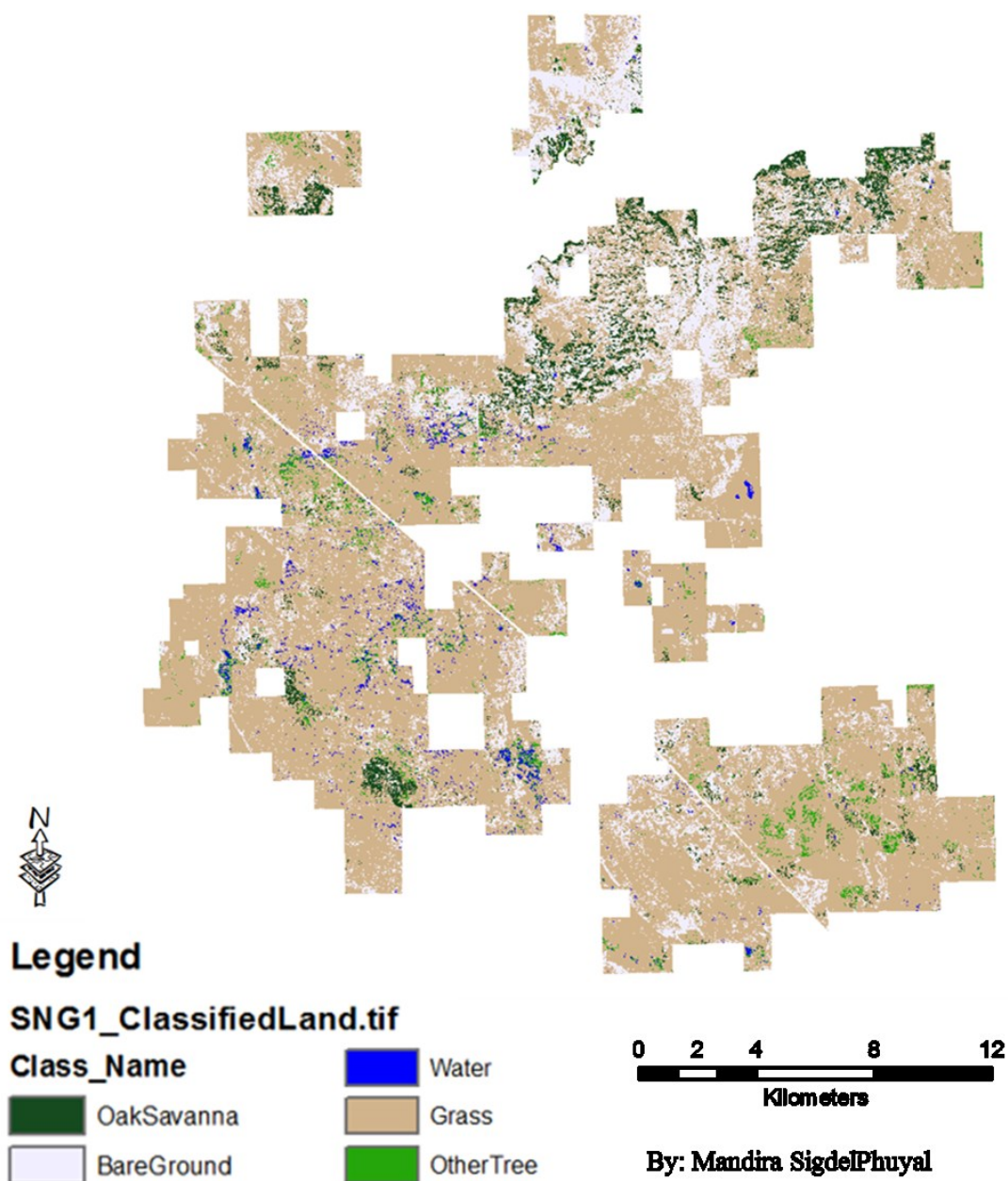
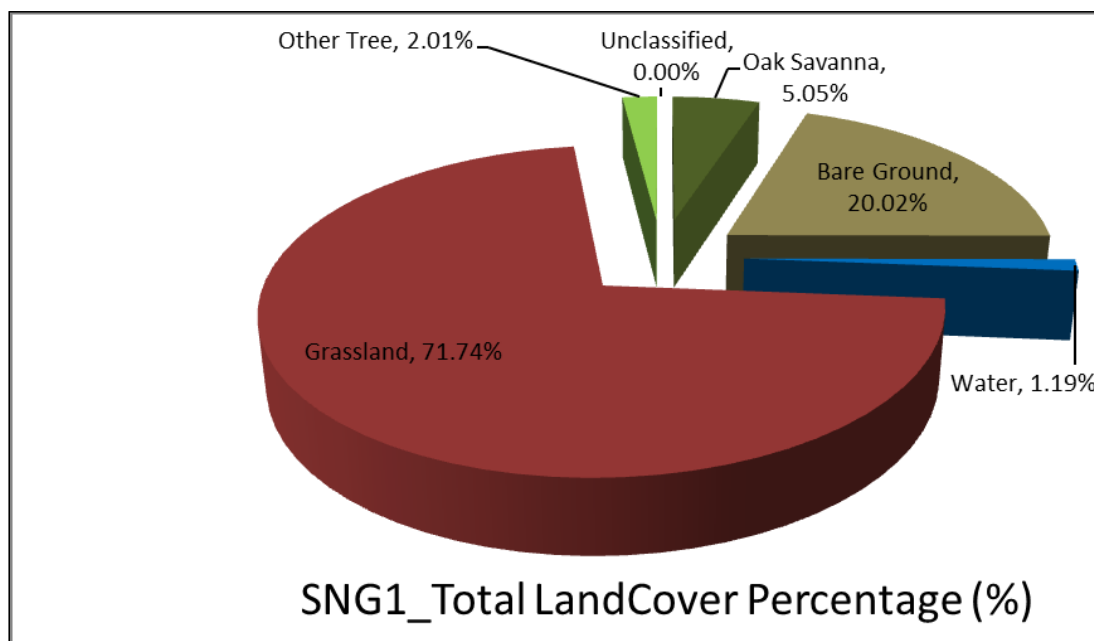


Figure 16: Chart showing the comparative percentage areas covered by five different land covers of the SNG1.



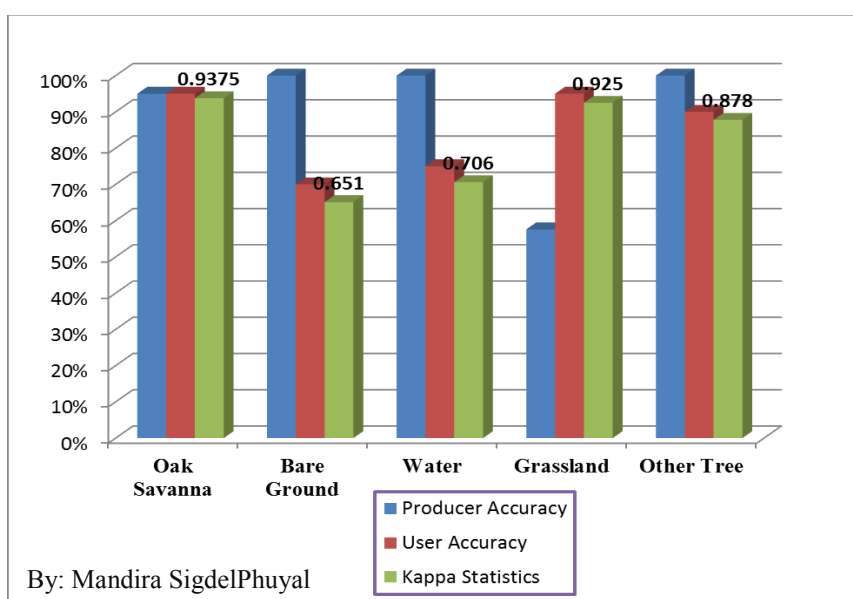
By: Mandira SigdelPhuyal

Table 3: Land covers of the SNG1 with their classification accuracy in percentage

The SNG1 Classification Accuracy Statistics				
Land Cover Classes	Area covered %	User Accuracy	Producer Accuracy	Kappa
Oak Savanna	5.05%	95%	95%	0.9375
Bare Ground	20.02%	70%	100%	0.651
Water	1.19%	75%	100%	0.706
Grassland	71.74%	95%	58%	0.925
Other Tree	2.01%	90%	100	0.878
Unclassified	0.00%			
Overall Classification Accuracy = 85.00%				
Overall Kappa Statistics = 0.8125				

By: Mandira SigdelPhuyal

Figure 17: Land cover classification accuracies comparison chart of the SNG1



4.1.2. *Woody canopy estimate of the SNG2*

The SNG2 is a small area and the aerial image that I used for the classification was clearer compared to the SNG1. The clear visual quality of the image provided better classification accuracy than at the SNG1. Land cover for this area was classified into four different classes: trees, water, grassland, and bare ground. At the SNG2, oak trees mixed with aspen (as at the SNG1) are widespread with few other tree species present, thus we didn't add a second tree class.

Overall, the land cover classification of the SNG2 is excellent, with 97.5 % classification and 0.94 Kappa coefficient (Table 4). Further, the individual classification accuracy, particularly of woody species (i.e. tree) is very high at the SNG2 (i.e. about 95 percentage, see Table 4).

The overall percentage covers of the four different land cover types of the SNG2 were calculated and shown in the Figure 18 and Figure 19. Tree cover of the SNG2 is markedly higher than at SNG1, with about one-fourth (25 %) tree cover (see also Figure 9) at SNG2 (Table 3 and Table 4) which made an average 6 % oak canopy cover across the entire SNG (combining SNG1 and SNG2).

The higher woody cover of the SNG2 was reflected in the number of presence data points in the random samples selected for subsequent modeling compared to the SNG1. Thus, ~10,000 presence points out of 200,000 random (pre/abs) points sampled across SNG1 and ~24,000 presence points out of 100,000 random (pre/abs) points for SNG2.

Figure 18: Land cover classification map of the SNG2 created using the supervised classification technique in ERDAS Imagine software. This is a 1-meter spatial resolution map.

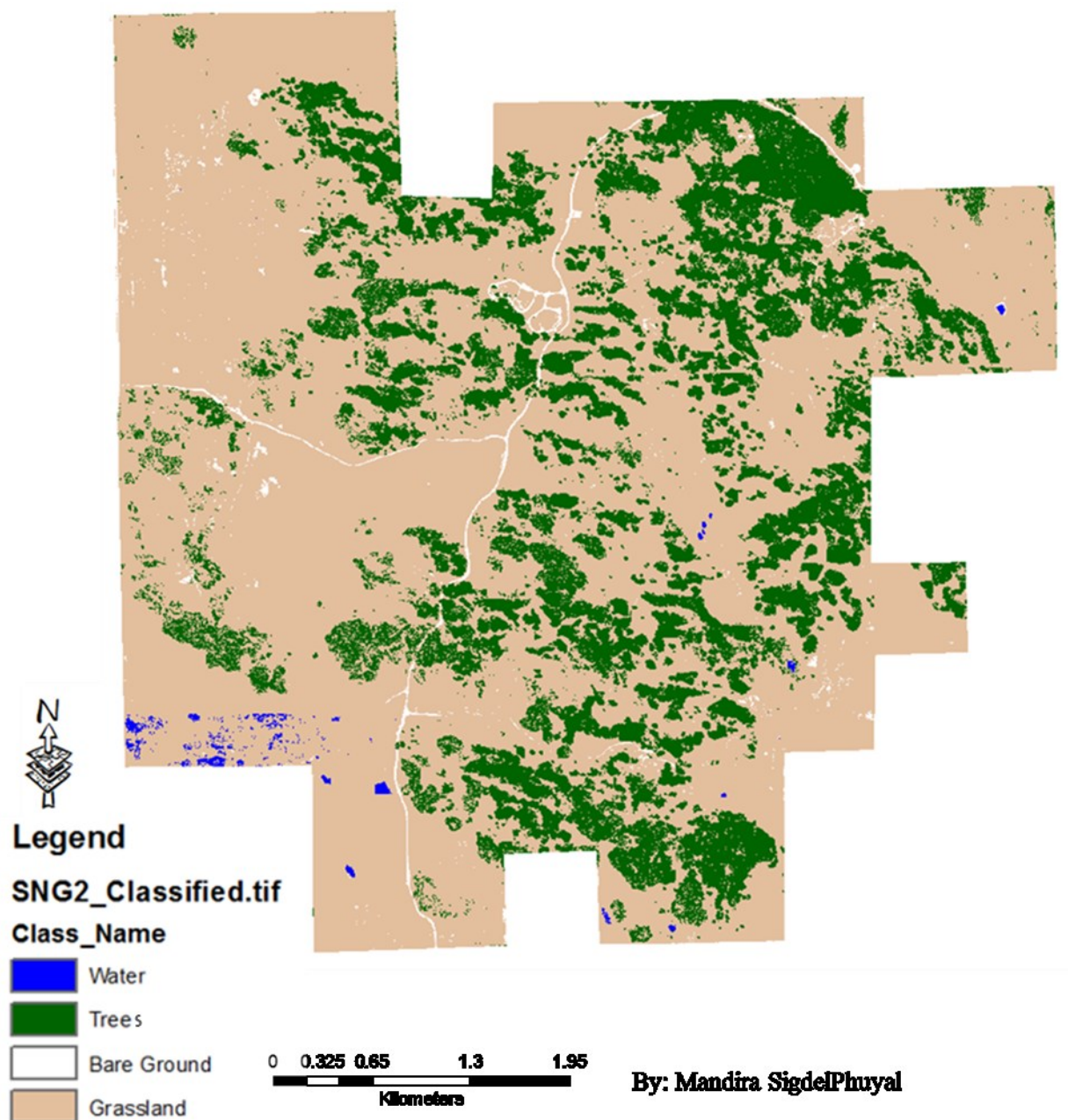


Figure 19: Chart showing the comparative percentage areas covered by five different land covers of the SNG1.

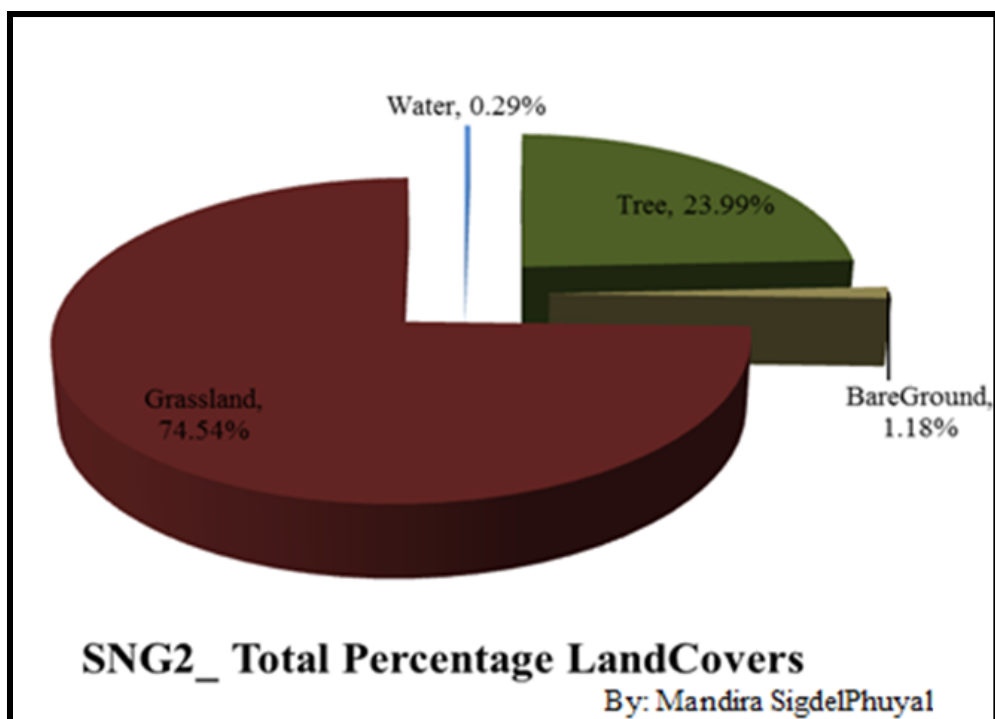


Table 4: Land covers of the SNG2 with their classification accuracy in percentage.

SNG2 Classification Accuracy Statistics				
Land Cover Classes	Area covered %	User Accuracy	Producer Accuracy	Kappa
Tree	23.99%	95.24%	95.24%	0.935
Bare Ground	1.18%	100%	100%	1
Water	0.29%	100%	100%	1
Grassland	74.54%	98.28%	98.28%	0.937
Overall Classification Accuracy = 97.50%				
Overall Kappa Statistics = 0.9383				

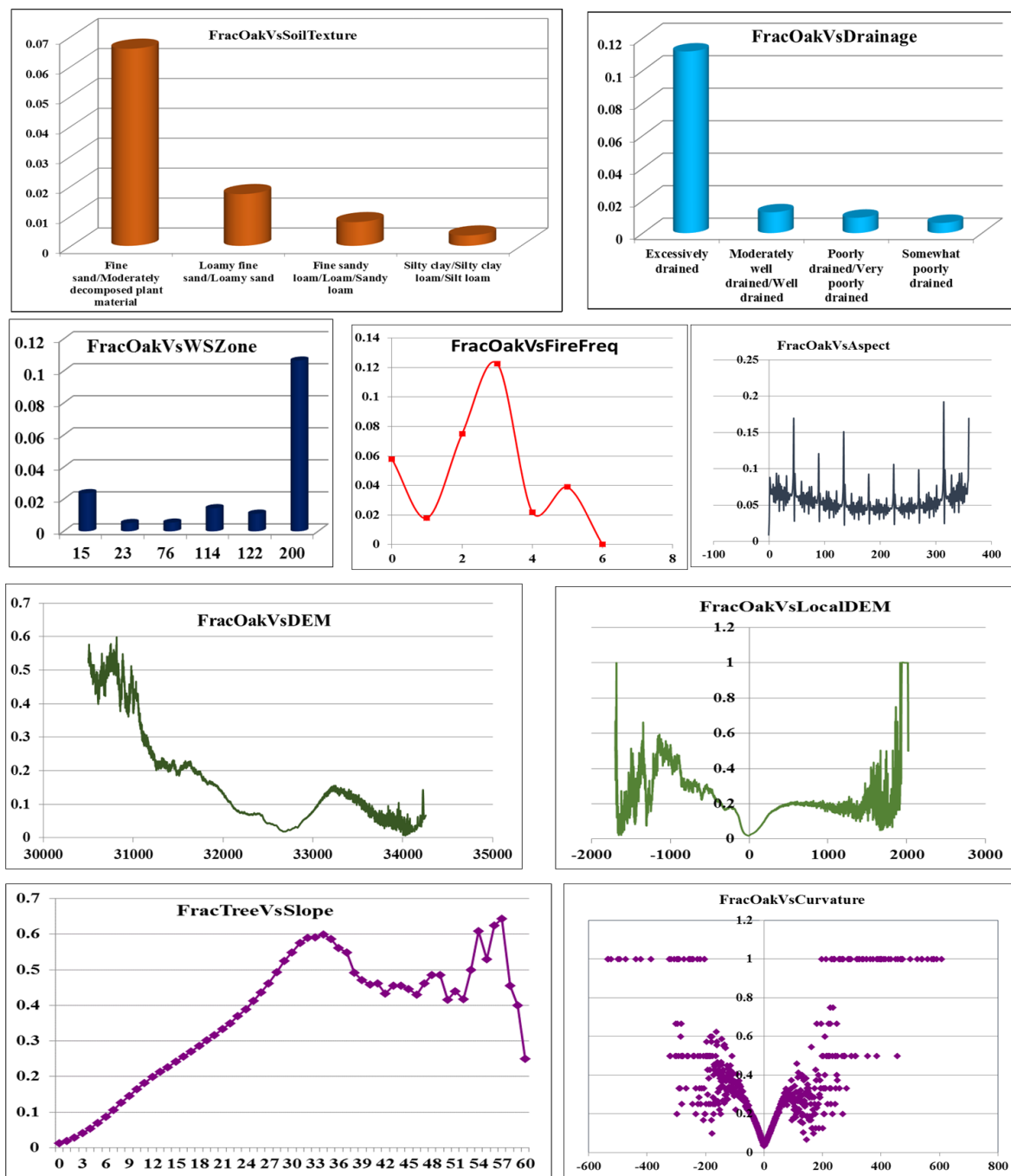
4.2. The empirical variables analysis

The empirical analysis plots of the SNG1 and SNG2 (see Figure 20 and Figure 21) suggest that several environmental variables are correlated with oak distribution within the SNG. Results from both SNG1 and SNG2 showed that higher density of oak trees is found in coarser texture and well-drained soil types. The relationship between oak tree cover and vertical distance to water saturation zone (hereafter, $\text{FracTree(oak)vsWSZone}$) appears bi-modal, perhaps reflecting misclassification of some trees (e.g. willows) growing in lowland areas with high water tables.

The empirical plots (Figure 20 and Figure 21) also provide interesting relationships between oak savanna and different topographic variables. Among the topographical variables, DEM and slope appear to have strong influence for oak savanna distribution on the grassland (in both SNG1 and SNG2), where oaks prefer sloping topography oriented northwards. The overall empirical results indicate that oaks generally dislike flat surface or clayey soil with low water drainage capacity. No clear patterns emerge for the relationship between oak distribution and slope convexity, average elevation or local elevation. Results at both SNG1 and SNG2 suggest that oak trees are most common in areas with moderate fire return frequency. Although the mechanisms for this are not known, it may reflect the combination of negative effects (e.g. seedling mortality in fire) with positive effects (e.g. establishment site availability and suitability following fire).

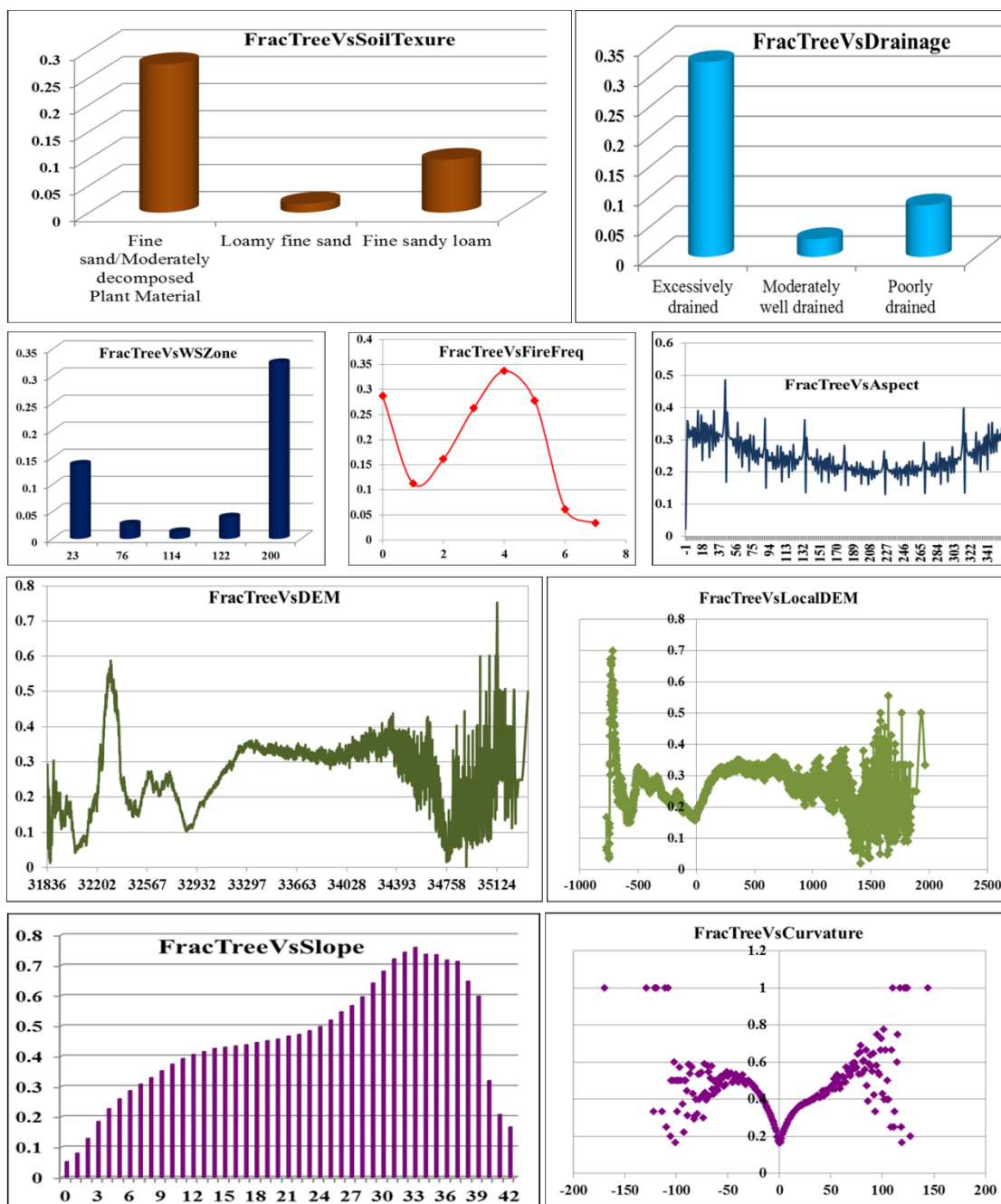
The visual analysis of oak distributions compared to environmental data-layers, provides some insight into likely relationships and potential mechanisms. However, in

multivariate situations, the true correlations may be obscured by correlations among variables, and we lack any predictive capability. The SDM approach will contribute further information on the effects of each variable and interactions among them.

Figure 20: The empirical variables analysis plots of the SNG1.⁵

By: Mandira SigdelPhuyal

⁵ Here, empirical variable plot represents a simple relational plot, where the fraction of oak tree (FracTree) is plotted in Y-axis and environmental variables are plotted in X-axis. The variables units: DEM is in cm, slope and aspect are in Degree, WSZone is in cm (it indicates the depth of ground water aquifer from the surface; details of soil properties names see Table 2.

Figure 21: The Empirical Variables Analysis plots of the SNG2⁶.

By: Mandira SigdelPhuyal

⁶ Here, the empirical variable plot represents a simple relational plot, where the fraction of oak tree (FracTree) is plotted in Y-axis and environmental variables are plotted in X-axis. The variables units: DEM is in cm, slope and aspect are in Degree, WSZone is in cm (it indicates the depth of ground water aquifer from the surface).

4.3. Species Distribution Model Results

4.3.1. *Model Evaluation Statistics*

The different SDM approaches were evaluated based on ROC, TSS, and Kappa statistics scores, all of which have the range of 0-1. If a model has the highest positive evaluation statistics compared to other models, the model is accepted as the best predictive model. Among the different SDMs (Table 5), the Random Forest (RF) model produced consistently higher evaluation statistics for both study areas. The ROC and TSS values suggest excellent predictive ability, although the Kappa scores are relatively low, perhaps, because of the prevalence effect (Liu, White, and Newell 2009). The RF approach allows for non-linear relationships between drivers and response variables, whereas, the linear-based models cannot deal very well with, for example, an optimum response at the center of a variable range.

Further, based on ROC scores, the MAXENT model appeared as the second best predictive model for both SNG1 and SNG2 study areas. However, while the ROC scores are high, the lower scores of other evaluation statistics suggest the model is less effective than the RF.

Table 5: Representation of different SDMs and their evaluation statistics result for the SNG1 and SNG2 (bold fonts show the best model across with its three major evaluation statistics).

Models	The SNG1 Full Model Statistics			The SNG2 Full Model Statistics		
	ROC	TSS	Kappa	ROC	TSS	Kappa
MAXENT	0.864	0.586	0.584	0.865	0.591	0.251
RF	0.99	0.978	0.683	0.999	0.976	0.457
GBM	0.859	0.576	0.453	0.858	0.58	0.244
CTA	0.663	0.318	0.444	0.791	0.497	0.11

(By: Mandira SigdelPhuyal)

4.3.2. *Variable Importance*

The variable importance (hereafter VarImp) is the percentage contribution of each independent variable in explaining oak distributions during the SDM modelling. Table 6 shows VarImp of individual variable for the four SDM models.

For the SNG1, the variable importance scores (Table 6) shows the higher contribution of topographic variables including slope, DEM, and local-DEM among variables. Among the other variables, the VarImp scores indicated that fire frequency and soil drainage properties are influential, although results are inconsistent among models and these variables are generally less influential than expected based on empirical relationships, perhaps, because the variable interactions are better considered when all environmental variables are considered together (i.e. in the SDM).

Similar results occurred in SNG2, where, again ‘Slope’ is ranked as the most influential variable for all the models and the RF model appeared more effective compared to other models (Table 6). As at the SNG1, both DEM and LocalDEM are influential at SNG2, but less so than slope. Fire frequency and drainage class are also influential as at the SNG1.

Table 6: The variable importance scores of the SNG1 and SNG2 study areas generated by four SDM approaches (Maximum Entropy, MAXENT; Random Forest, RF; Generalized Boosted Model, GBM; and Classification Tree Analysis, CTA) in **BIOMOD2** (bold fonts across the columns represented the superiority of the models with the higher individual variable contribution).

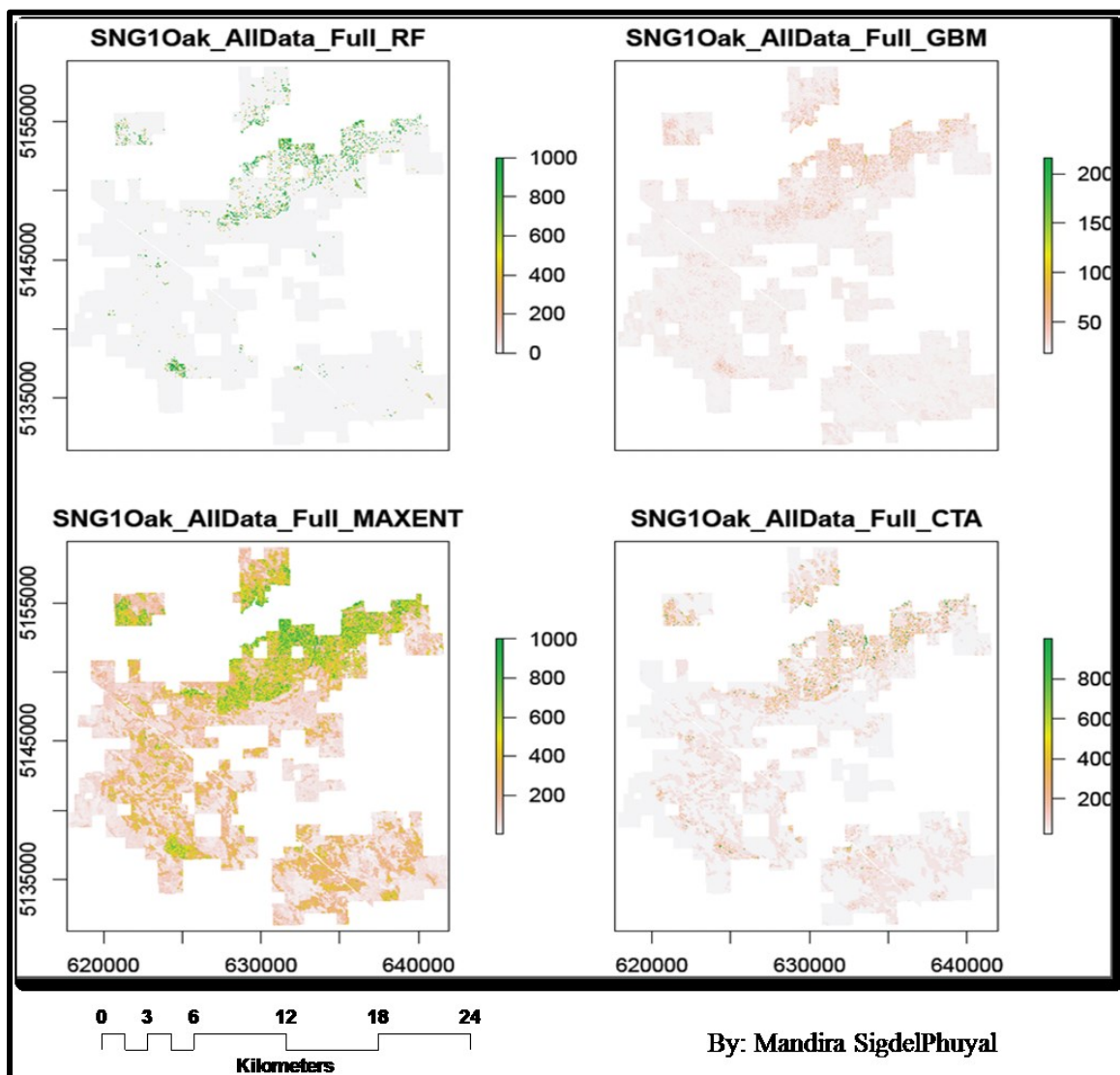
Variables	Variables Importance for the SNG1 (Full Model)				Variables Importance for the SNG2 (Full Model)			
	Models				Models			
	MAXENT	RF	GBM	CTA	MAXENT	RF	GBM	CTA
SNG1_Aspect	0.015	0.283	0.032	0.146	0.042	0.331	0.075	0.136
SNG1_Curvature	0.008	0.209	0	0	0.015	0.208	0.01	0
SNG1_Slope	0.146	0.532	0.494	0.965	0.244	0.508	0.361	0.388
SNG1_DEM	0.046	0.492	0.144	0.397	0.089	0.428	0.207	0.282
SNG1_LocalDEM	0.052	0.47	0.159	0	0.044	0.354	0.021	0.019
SNG1_WSZone	0.278	0.197	0.185	0	0.002	0.033	0.003	0
SNG1_Drainage	0.024	0.252	0.114	0	0.278	0.075	0.326	0.541
SNG1_SoilTexture	0.014	0.09	0	0	0.072	0.004	0	0
SNG1_FireFrequency	0.081	0.302	0.163	0	0.068	0.341	0.12	0.168

4.3.3. *Projection Plots*

Projection plots are the maps of projected species distribution in a defined geographic location as an alternative way of interpreting the performance of the SDMs. Here, the model projection plots (Figure 22 and Figure 23) represent model-based predictions of the probability of finding oak trees in the SNG1 and SNG2 respectively. These plots help to determine quality of model's performance by visual comparison with the original (classified) oak canopy maps (Figure 8 or 15 and 9 or 18 for the SNG1 and SNG2 respectively).

The projected RF model (Figures 22-23) provided similar predictions of oak spatial distributions as the original estimates derived from the high-resolution aerial photographs. Relative to the RF model, the predictive maps using the other SDM models tend to over- and under-predict. This comparison again confirms the superiority of the RF method for prediction of oak presence and absence throughout the geographical space of the SNG.

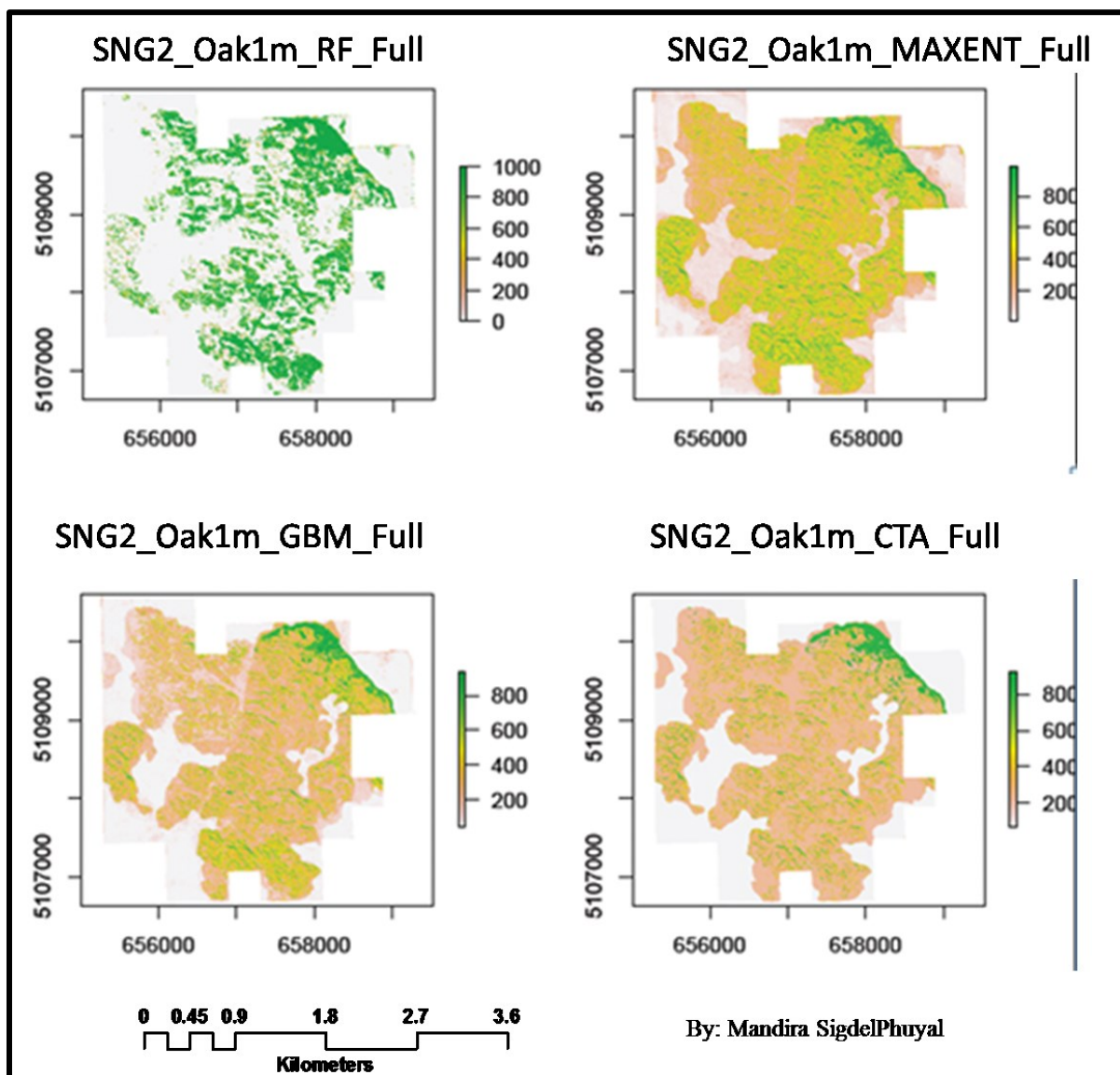
Figure 22: **BIOMOD₂** model predictive probability of oak savanna presence⁷ in the SNG1. Prediction color bars scale probabilities of 0-1 over the range 0-1000 (i.e. color scale = 1000 is equivalent to a tree probability of 1).



⁷ In the Figure 22, for each map, the map title “_AllData_Full” represents the calibration of models using the full model. Projection unit ranges from 0 to 1000 ratio, this is similar to general 0 to 1 (or -1 to 1) probability ratio (unit). The values in X-axis and Y-axis represent the latitude and longitude of the location, and unit is in decimal degree.

Figure 23: *BIOMOD*₂ model predictive probability of oak savanna presence⁸ in SNG2.

Prediction color bars scale probabilities of 0-1 over the range 0-1000 (i.e. color scale = 1000 is equivalent to a tree probability of 1).



⁸ In the Figure 21, for each map, the map title “_AllData_Full” represents the calibration of models using the full model. Projection unit ranges from 0 to 1000 ratio, this is similar to general 0 to 1 (or -1 to 1) probability ratio (unit). Values in X-axis and Y-axis represent the latitude and longitude of the location, and unit is in decimal degree.

4.3.4. *Predictive Response Curve Analysis*

Predictive response curves (Figure 24 and Figure 25) are the 2D response plots of oak savanna verses different environmental variables, estimated using the four SDMs approaches. These response curves are used to understand the form (positive or negative, linear/non-linear, and so forth) of relationships between tree presence and the major drivers.

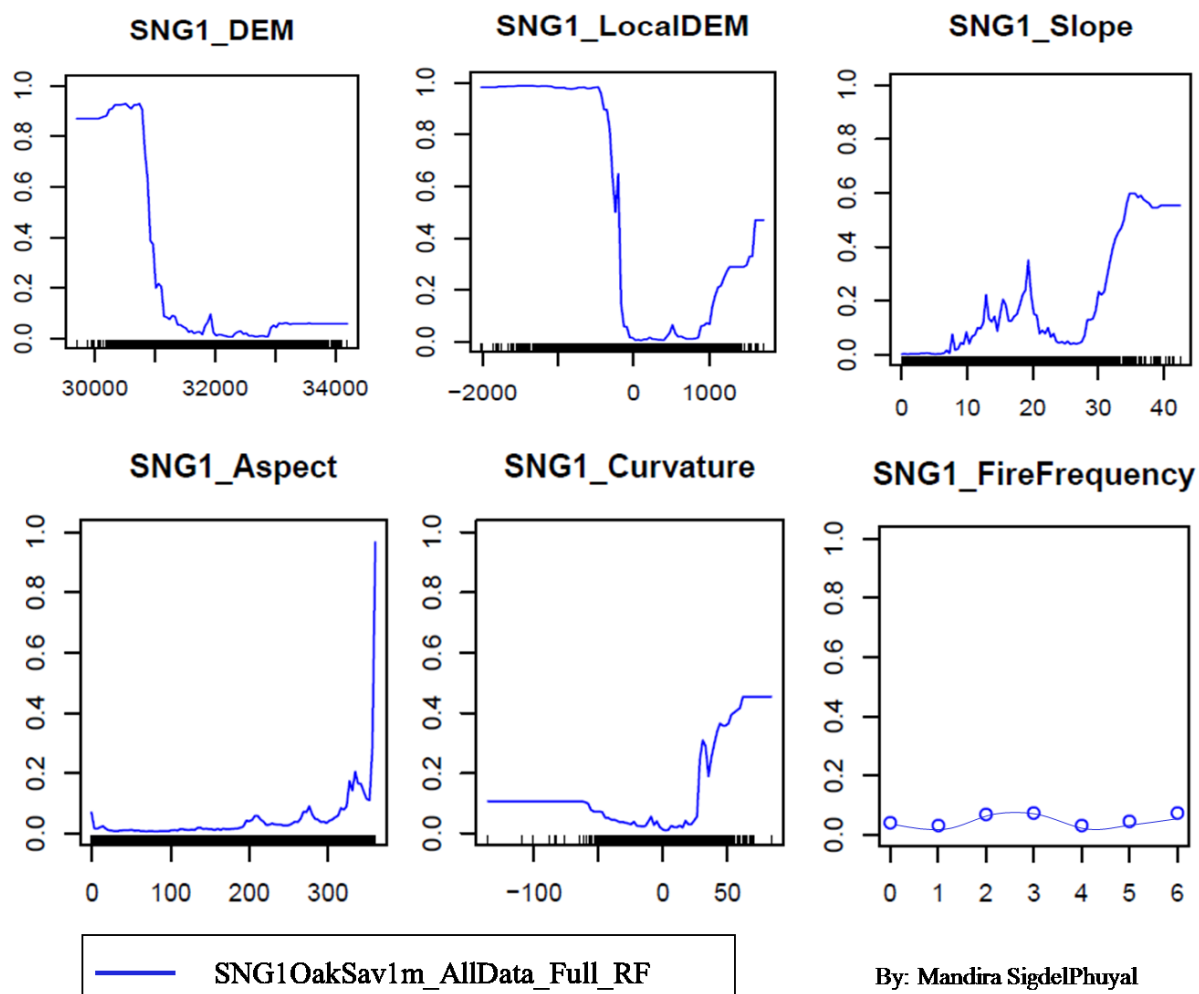
In this research, initially, I plotted the model predicted oak verses environmental variables probability distribution (for both, SNG1 and SNG2) representing oak presence in Y-axis and the environmental variables in X-axis. In such plots (see Appendix VI and VII), I introduced all initially selected environmental variables and all the selected (four SDMs) models. Later, given the superiority of the RF approach for predicting oak occurrence, I concentrated primarily on the RF response plots (Figure 24 and Figure 25) and the top-ranked variables shown in Table 6 and contrast results obtained using other SDMs.

For the SNG1, the response plot (see Figure 24) between oaks presence versus slope indicated the high correlation between oak distribution and steeper slope locations. Other variables such as DEM and localDEM (i.e. continuous variable) again consistently exert an influence but comparatively lower than Slope. DEM has a negative linear relationship with oak, suggesting that oak trees favor low elevation. However, DEM does not account for the local topographic position of a location relative to the top, mid-slope and bottom of local topographic features. Thus, in order to represent this local landscape variability, 'LocalDEM' was created. LocalDEM takes into account the average height of

each selected block so that pixels can be represented as above or below mean height for the block. Despite our expectation that LocalDEM would provide additional explanatory power, the results for this variable were inconsistent. Further, as with the empirical analysis (above), the Aspect variable suggests that oaks slightly favor North-facing slopes (e.g. $45^\circ > \text{aspect} > 315^\circ$).

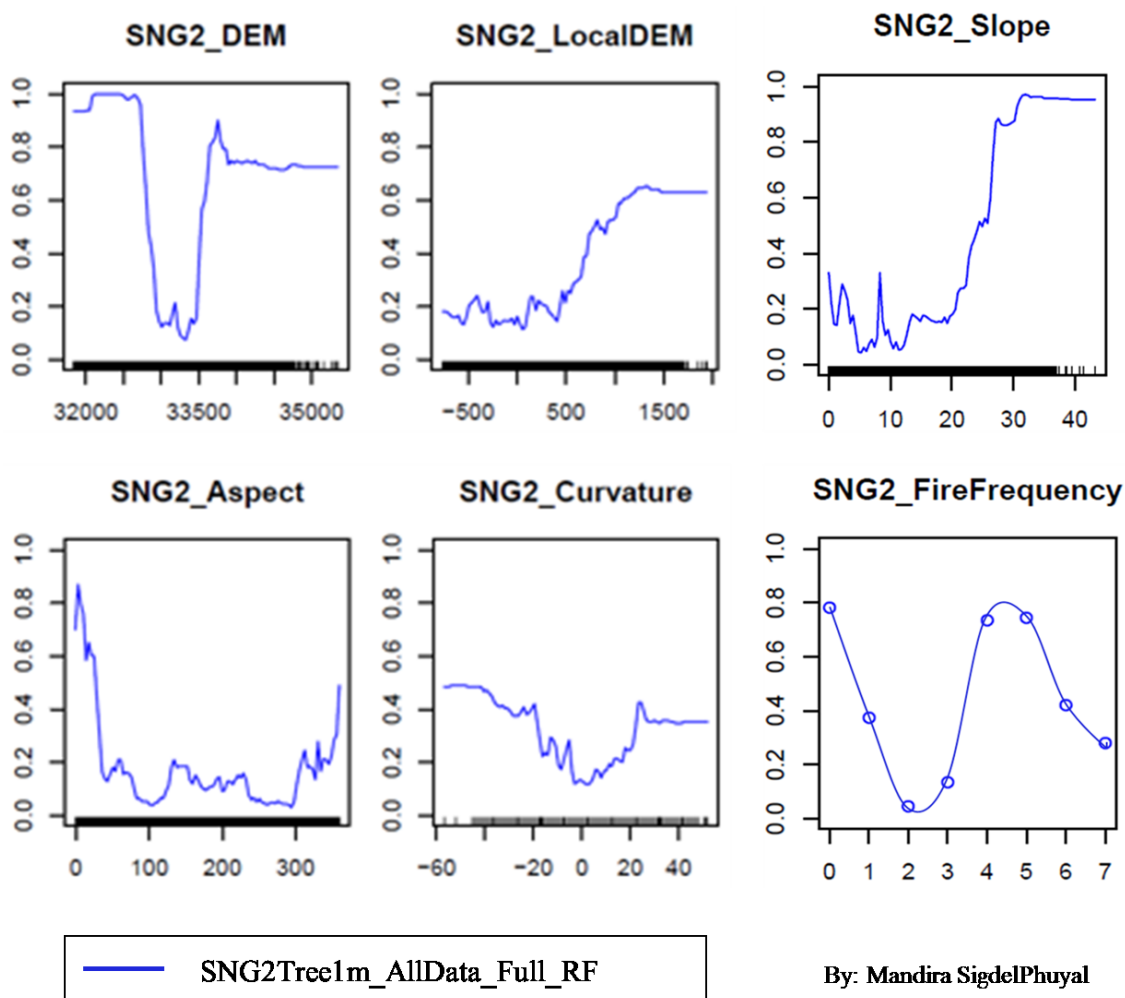
In the SNG2, response plots are little different from results at SNG1. Here, increasing slope and elevation (particularly, LocalDEM) both appeared to favor oak distribution (i.e. oaks favor upland); however, slope appeared as the major driver, with some but hard-to-interpret roles for DEM and curvature. The fitted relationships indicated that the Aspect effect is also similar as that of the SNG1, where oak distribution is mostly on the North-facing slopes (e.g. $45^\circ > \text{aspect} > 315^\circ$). In addition, the fitted relationship between fire and oak for both SNG1 and SNG2 showed the influence of moderate fire frequency for oak establishment, where intermediate fire frequency may favor seedling establishment through removal of litter and reduction in direct competition, without the frequent mortality of seedlings expected in regular fires.

Figure 24⁹: The SNG1 probability response plots of oak versus different (influential) environmental variables inferred using Random Forest (RF) (see. All comparative SDMs response plots for the SNG1 are in Appendix VI).



⁹ Note: in the Figure 24, DEM derivatives and fire are the continuous data. Here, the unit of DEM values are in cm, Slope and Aspect are in Degree, Local DEM is in meter, and fire frequency is the number of fires during the 15-years data. In the legend, “_Full_RF” represents the calibration of model using full model method.

Figure 25¹⁰: The SNG2 probability response plots of oak versus different (highly influential) environmental variables inferred using Random Forest (RF) model (see. All SDMs response plots for the SNG2 are in Appendix-VII).



¹⁰ Note: in the Figure 25, DEM derivatives and fire are the continuous data, whereas, the other remaining are categorical datasets, and the value in the numbers are representing the different categorical classes (detail, see Figure 21 for each categorical dataset). Here, DEM values are in cm, Slope and Aspect are in degree, Local DEM is in meter, and fire frequency is in numbers of repetition of fire. In the legend, “_Full_RF” represents the calibration of model using full model method.

CHAPTER 5: DISCUSSION

5.1. Data Selection and its effect on decision-making

In this research, the 1-m spatial resolution aerial photos were well-suited (relative to satellite data) for classification of land cover and detection of trees. The aerial photographic approach to this work enabled classification of individual tree canopies with high classification accuracy. Differences in the classification accuracies between the two areas of the Sheyenne (SNG1 and SNG2) may have been caused by differences in image quality (atmospheric effects). In addition, spectral similarity among land cover types can lead to misclassification, particularly, among different vegetation with similar reflectance properties. It might have more affected the SNG1 classification accuracy than SNG2 because of its relatively larger geographic extent with more diverse land cover types.

To improve land cover classification, several factors including the image spatial resolution, required spectral band, quality, and the geographical extent need to be considered together. Thus, a combination of a clear image with the higher spatial resolution and required spectral bands (where, the band combination provide better visual contrast to identify each object on the ground) provide the greatest opportunities for an excellent land cover classification.

In addition to the data used for land cover classification, other environmental datasets such as the 1-m spatial resolution digital elevation dataset, used to create various detailed topographic derivative datasets, contributed to understanding oak distribution in

both study areas. Additional data sources (e.g. fire and soil data-layers) were less informative. This is described in more detail in the subsequent sections.

5.2. Empirical vs Predicted result comparison

5.2.1. *Role of Soil types in oak distribution*

Empirical variable relationship plots (Figure 20 and Figure 21) suggested that oak presence is highly correlated with the soil texture, particularly in areas with coarse soils and high water drainage capacity. However, the predicted variable response curves (Figure 24 and Figure 25) showed little or no impact of soil variables on oak establishment. The differences might be because of the available soil data, which was at a spatial scale (vector form) compared to other selected environmental data, where the level of detail may not be appropriate for small-scale analysis using the models. Thus, it might affect the performance of models when they implement together with the other continuous datasets.

5.2.2. *SDM offered contrasting results at landscape level prediction and between two homogeneous environments*

The species distribution modeling approaches were able to predict where oak is more probable in the landscape, but the relationships with soil and topographic variables were not the same between the two regions of the SNG (i.e. SNG1 and SNG2). The performance of various SDMs differed greatly among the selected data and models. It was also different for each variable and each applied model. Thus, the species distribution models have predicted the contrasting result of oak savanna biogeography as

compared to empirical analysis. The differences between the SNG1 and SNG2 results might be that the effect of selected sample size and model instability (Hernandez et al. 2006; Wisz et al. 2008). Although the SNG1 is large relative to the SNG2, the much lower tree cover (5% at SNG1 versus 25% at SNG2) meant that randomly sampled presence locations for the SNG1 were fewer than obtained for SNG2, which may not be sufficient data for efficient model prediction as compared to the data used for empirical analysis.

5.2.3. *Superiority of RF Approach*

The overall model prediction statistics indicated the superiority of Random Forest approach to inform the oak-environment relationship among the selected four different SDMs: MAXENT, GBM, RF, and CTA. The comparative analysis of models prediction statistics, models performance, and empirical study confirmed the RF model as the most successful approach for landscape level prediction. Alternatively, MAXENT might be an acceptable choice, However, MAXENT only uses the limited presence dataset (i.e. it does not use absence data), which is generally considered to be a disadvantage, and the MAXENT model was unstable (would often fail to converge or crash without warning) if we selected large data sample.

Relatively low tree cover, particularly in the SNG1, ultimately affected the number of ‘presence’ points in the random sample. It further appeared to affect the Random Forest model, including the other model’s predictions, and that might have contributed to generating some contrasting results. Thus, the predictions of RF (and may be other SDMs in general) might be improved using a non-random sampling to increase

the proportion of true presence data, rather than working with datasets that are skewed towards absence data.

5.2.4. Slope: The major driver of oak savanna existence

The overall models including RF and other analysis predicted that, in the SNG oak savannas distributions are strongly influenced by the landscape topography. Based on my result, oaks particularly favor an open area where small hills with an intermediate slope exist. This result indicated that distribution of oak might also correlate with the well-drained, coarser soil texture, and north-facing aspects locations. These soil and topography variables suggest that oak seedling establishment is favored in areas that have minimal chance of waterlogged soils, away from valley bottoms.

Beside slope and soil variables, other topographic variables (LocalDEM and DEM) were found to be influential but inconsistent and contrasting between SNG1 and SNG2. LocalDEM, particularly for the SNG2, provided an important insight about oak savanna that they favor upland compared to the SNG1. This may be because the SNG1 covers a large geographic extent compared to the SNG2, such that local effects in SNG1 may have been overwhelmed by coarser-scale differences in soil or other variables.

5.2.5. Disturbances effect

In my research, fire appeared as an important factor for oak establishment; however, it also provided contrasting results between locations and among analysis methods. The empirical analysis predicted that higher abundance of oak occur in locations of intermediate fire frequency (i.e. 3-4 fires in 15 years) and also in fire absent

areas. The RF analysis for SNG1 and SNG2 broadly confirm the empirical analysis, with unimodal relationships indicating a correlation between intermediate fire frequency and increases oak canopy cover. The apparent increase in tree covers at very low (or absent) fire frequency might correspond with closed-canopy formations, where, herbaceous fuel load is insufficient to carry a fire. However, at both SNG1 and SNG2, the analysis confirmed that oaks do not tolerate high and repeated fires, presumably because of seedling and sapling mortality.

5.3. Field Visit: A validation tool

On October 10 2014, I conducted a field visit in the Sheyenne National Grassland as an alternate way of validating my results and for better understanding the results of my analysis. During the field visit, by a visual inspection of oak distribution (see Appendix VIII and Appendix IX) in the SNG1 and SNG2, I found a similar but still a contrasting relationship of oak with environmental variables such as topography, particularly DEM, aspect, and slope. However, this field visit helped better situate the oak species while selecting the sample pixels to perform supervised classification analysis. It helped to generate a precise woody canopy cover map.

In addition, the field visit provided me an opportunity to compare oak trees distributions and physical environmental factors. In the Sheyenne, oaks occurred on sand hills, but the hills might have lower than the average surrounding elevation (particularly, in the northeastern part of the SNG1); however, they are in good concave to convex pattern and well-defined slope profile. Further, the SNG2 is a small area with various small hills, where trees (oak) are widely distributed from the bottom of the hills to the

upper slopes. Thus, the predicted result comparison between slope and DEM (particularly, for the SNG2) is consistent with the heavy distribution of oak trees at small to average slope angles, which either have a low elevation or high compare to surrounding average elevation.

In addition, in the field, I noticed that oak trees are often found in mixed stands with other tree species, especially Aspen, and occasionally willow and cottonwood. The potential misclassification of oak trees resulting from these mixed species might have led to the DEM (and LocalDEM) relationships suggesting that oak are found at low and moderately higher elevations, but less frequently at intermediate elevations.

CHAPTER 6: CONCLUSION

6.1. Summary of the research and results

6.1.1. Influence of data quality in classification accuracy

The high spatial resolution aerial photography and multiple environmental datasets that I selected to map and explore oak distribution in the Sheyenne National Grassland provided the main input for my analysis of oak-environment relationships.

The land cover classification applying a supervised classification technique provided 80 % and 95 % overall classification accuracies, with high user and producer classification agreements for SNG1 and SNG2 study areas, respectively. The Sheyenne National Grassland has an overall oak canopy cover of ~6 %, but with marked (individual) differences between the larger northern block (5% oak cover) and the smaller southern block (25% oak cover).

The land cover classification, however, indicated that accuracy in land cover classification using the remote sensing techniques is highly variable by the extent of study area selected, types of data, the data selection technique, and statistical approach. The quality of the remote sensing data also had a large effect on our ability to classify trees accurately at the SNG1 (relative to the SNG2) even though the aerial survey data were acquired on the same date. The overall analysis suggested that oaks savannas are a complicated habitat to map separating from other woody vegetation with similar reflectance properties.

6.1.1. SDMs offered important understanding of grassland oaks biogeography

Using a variety of SDMs, *BIOMOD₂* helped me to predict, compare, and analyze the effect of different environmental variables on oak savanna distribution at landscape level. Here, the Random Forest (RF) model proved to be flexible and superior in predicting non-linear species-environment relationships at a landscape level among the candidate Species Distribution Models (MAXENT, GBM, RF, and CTA).

Analysis of classified high spatial resolution aerial photographs from the Sheyenne National Grasslands, using statistical SDM approaches allows us to conclude that the oak savanna ecosystem is strongly associated with landscape topography and soil types, with oak trees favoring sloping locations, sandy and well-drained soils, northerly aspect and infrequent (but occasional) fire.

Controlled fires are one of the major practices for maintaining oak savanna and reducing the cover and density of invasive species. In general, it would appear that Bur Oak benefit from a low-to-moderate frequency of fires; however, the extent or frequency of fire should be considered carefully, since too frequent fires appears to inhibit oak canopy cover, presumably through reduced seedling establishment (possibly through increased adult mortality).

6.2. Limitations in research and results

Working with the high spatial resolution geospatial datasets with large numbers of occurrence data and using the multiple models together in *BIOMOD₂*, several challenges need to be considered. The main limitations that I faced during my analysis included

memory space problems in computer, and the length of time it takes SDMs to run. In addition, the performance of SDMs depended on the data size (Hernandez et al. 2006) or data resolution that created similar but conflicting results between the two study areas (SNG1 and SNG2). Because of the limitation in model's data holding capacity, I needed to sub-sample a small fraction of the SNG areas to run the model.

Another main problem that I realized when I started studying about the Sheyenne National Grassland was the available literature is that limited research has been conducted about the Sheyenne National Grassland, with little on the general history and physical geography including the woody cover analysis.

6.3. Research Implications

The techniques, analysis, and results offered by this research might be a good source of information for a person who is interested in working with the SNG or with similar species-environment relationship for other areas. The land cover map and the woody cover classification might be the most accurate and precise data available on the SNG oak distributions. Thus, SNG managers and others interested in the SNG research could use this woody cover dataset to guide SNG management or to inform complementary studies.

In addition, this research provided important insight about general biogeography of remnant oak savanna of the woodland-prairie ecoregion at a landscape to local level. It would help grassland managers and conservationists to better identify oak habitat and to implement conservation and restoration policies.

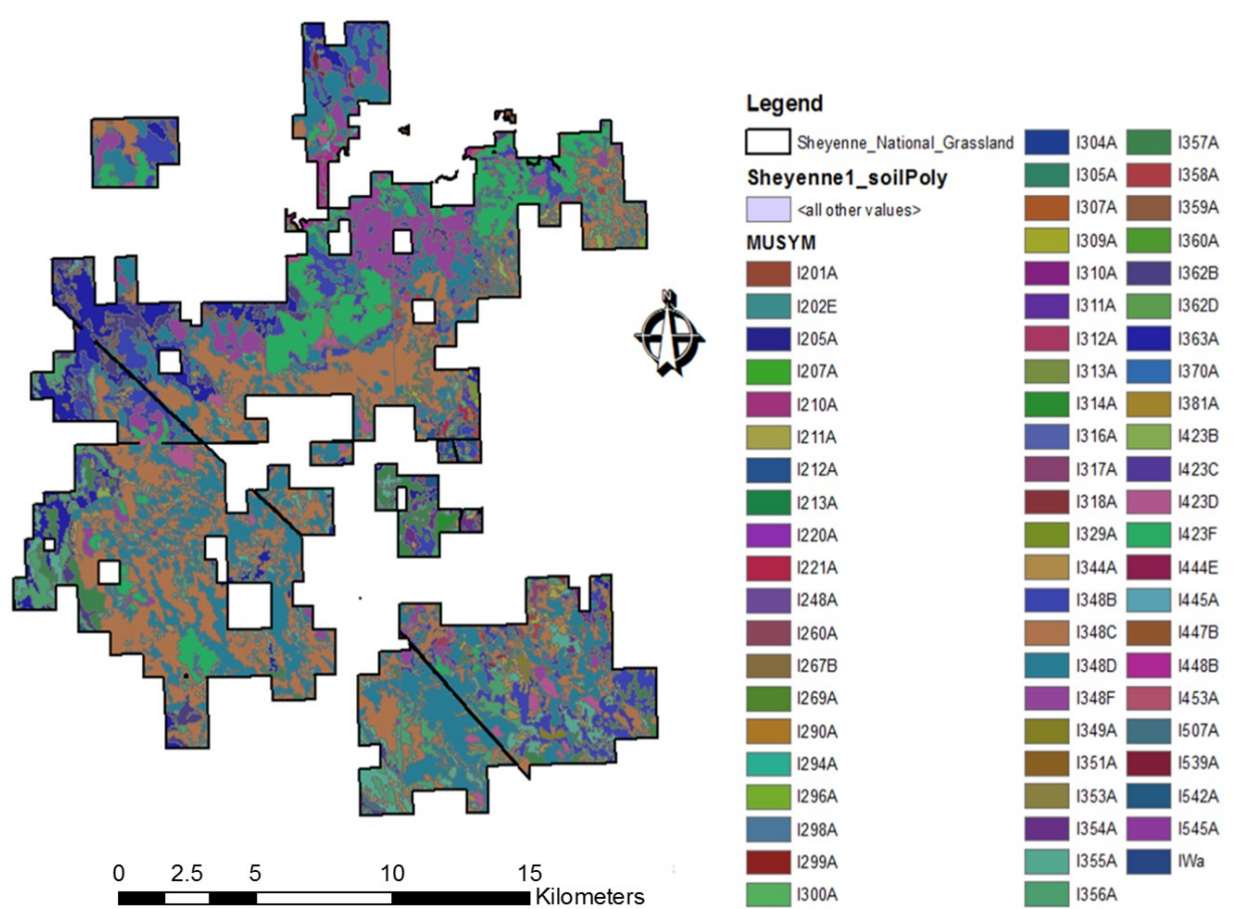
Further, the overall comparative model evaluation statistics suggest that scientists and managers could potentially use species distribution models (particularly the RF approach) to analyze species distributions in the Sheyenne or a different geographic location. This study would better help them determine the data type and appropriate statistical approach according to individual interest for their particular research.

6.4. Future Work

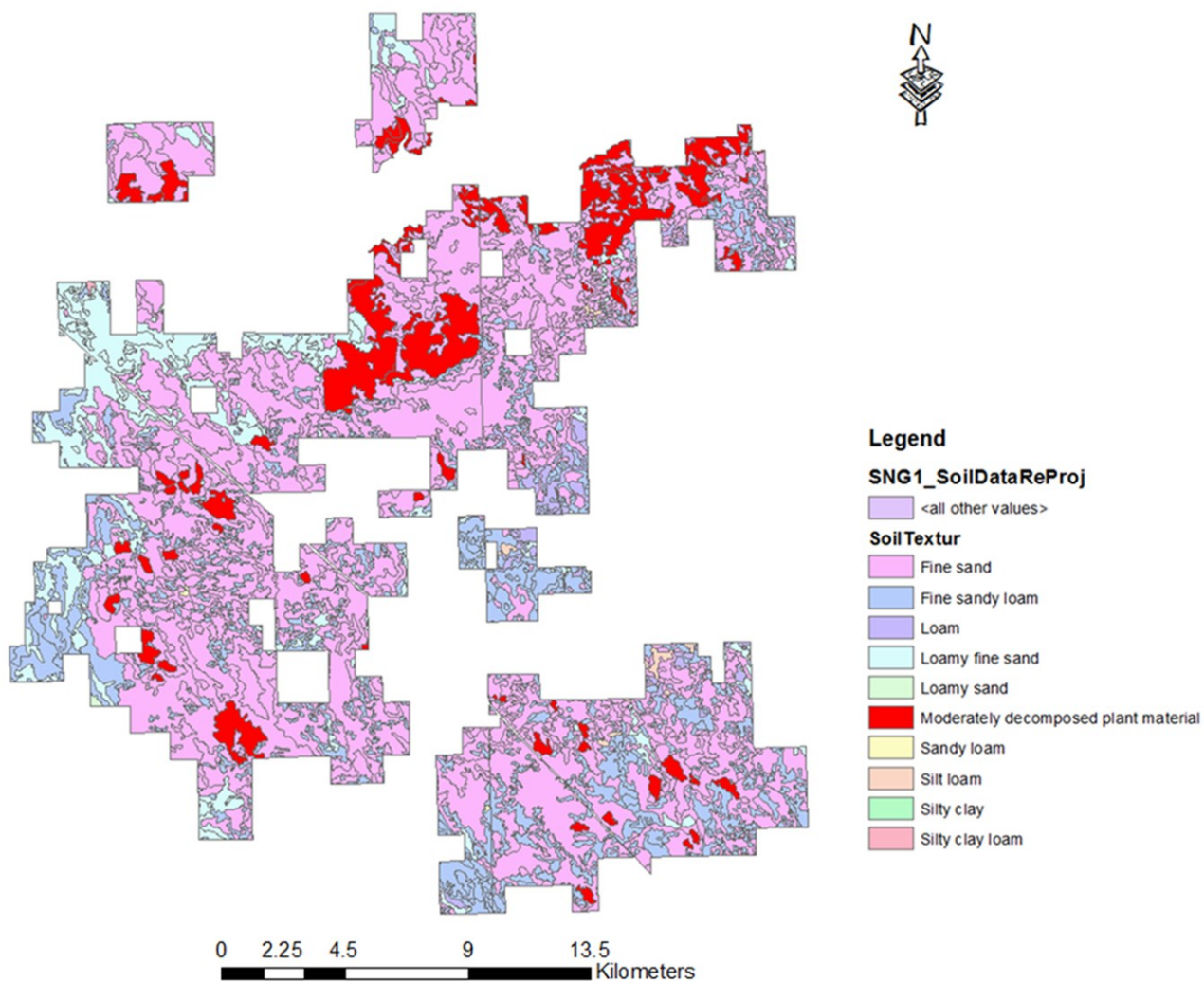
The analysis and prediction results included in this research provided important insight into oak savanna distribution in the protected grassland-savanna of the SNG. However, because of possible limitation and biases in the applied data quality and data sufficiency, the results may not fully explain oak biogeography in North American Grassland like the Sheyenne. Thus, addition of other appropriate data such as long-term fire data, grazing history data, and more accurate soil data, possibly, from field measurements, may provide a more detailed understanding of bur oak biogeography in the SNG.

APPENDICES

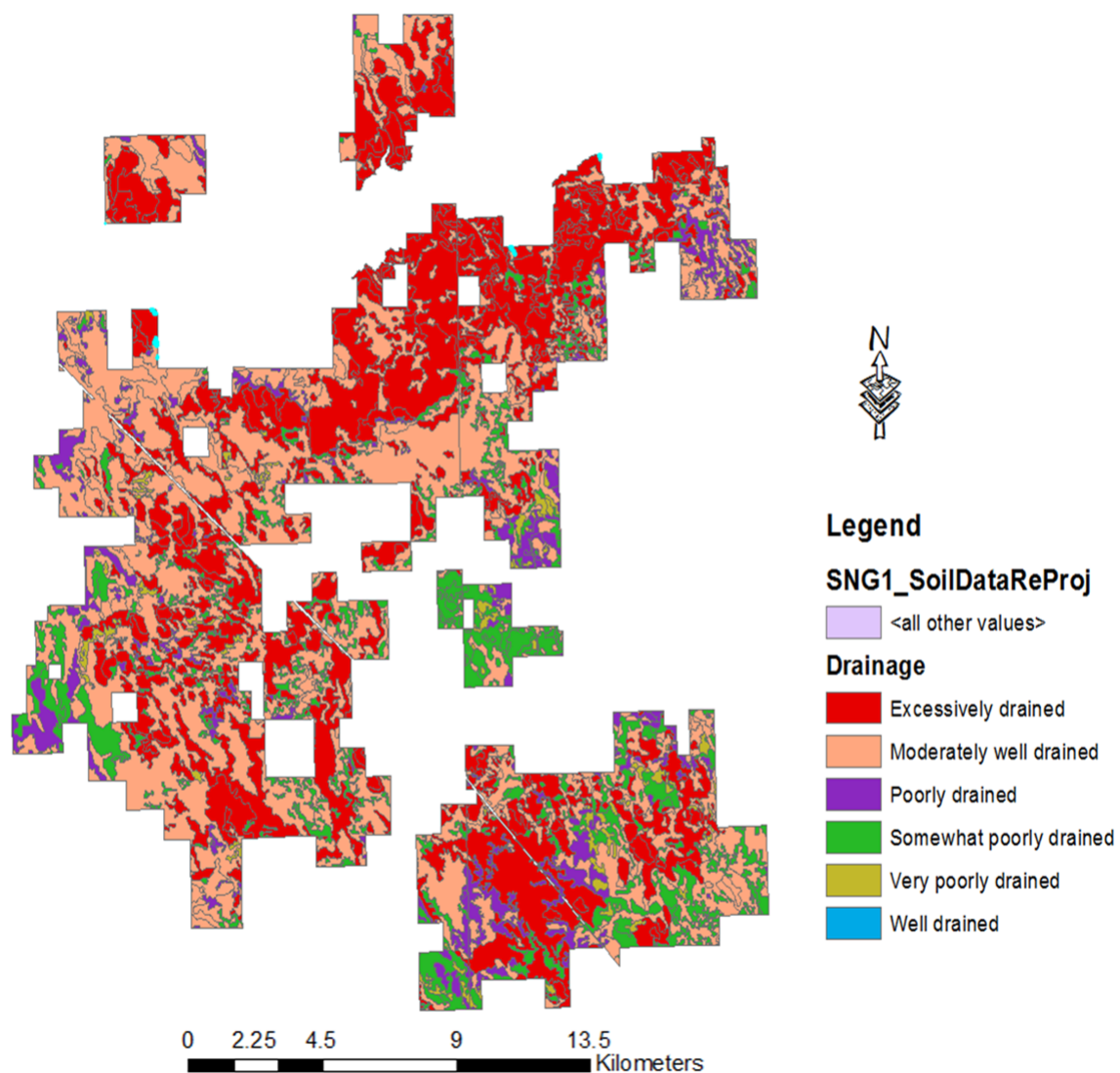
Appendix I: The original soil polygon of the SNG1 representing different soil physical properties provided by the NRCS in their Map Unit System (MUSYM).



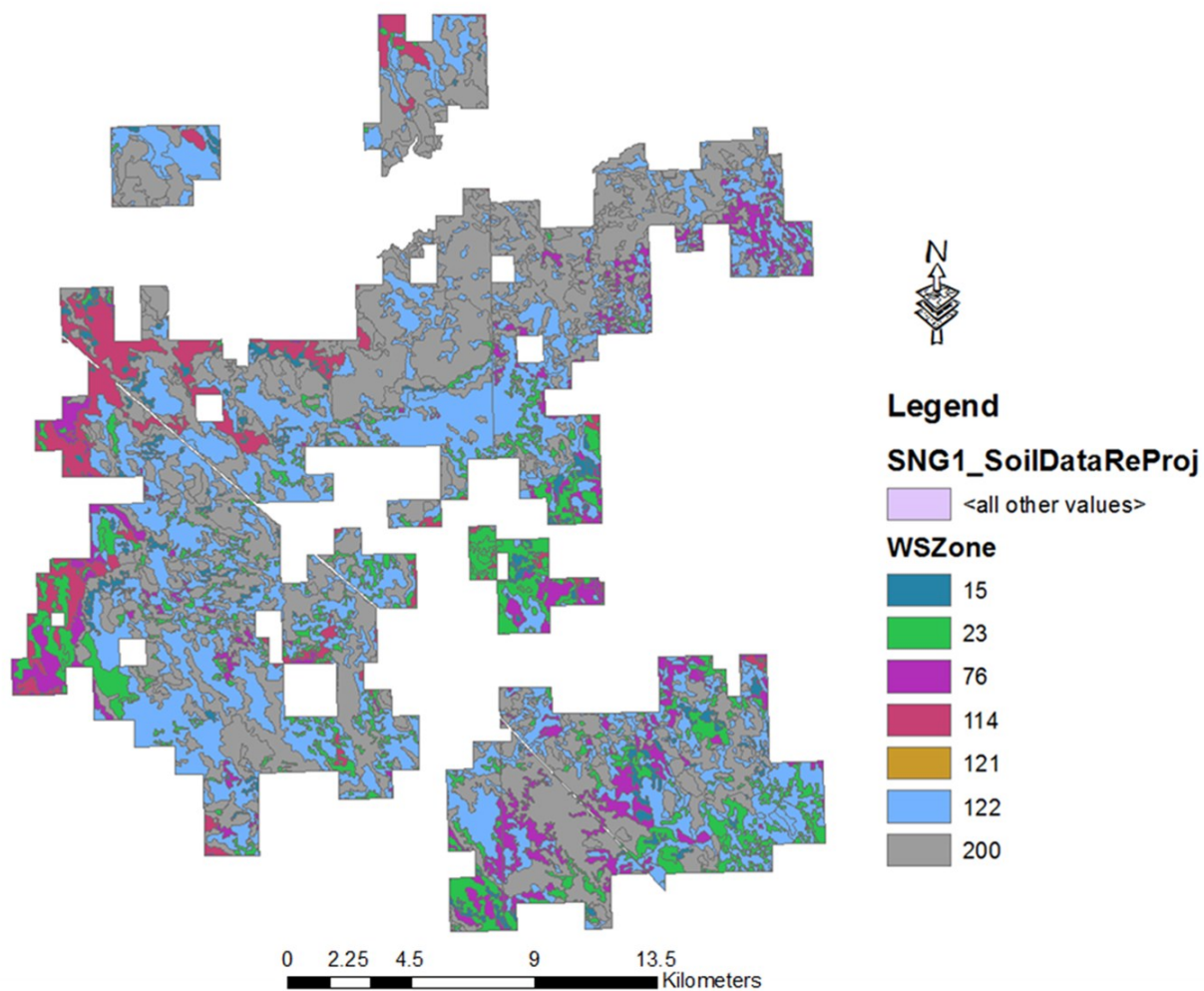
Appendix II: The NRCS soil texture map of the SNG1 created in ArcMap (by: Mandira SigdelPhuyal).



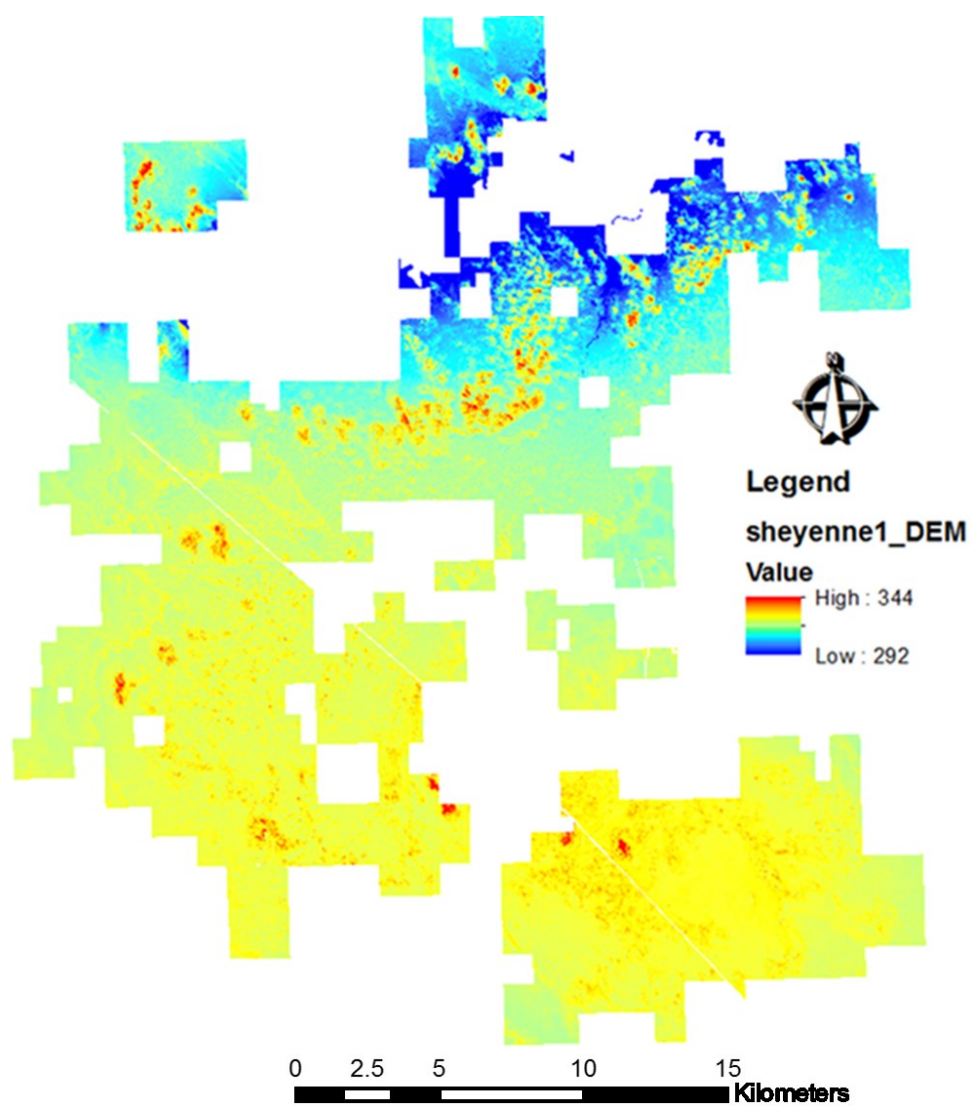
Appendix III: The NRCS soil drainage classes map of the SNG1 created in ArcMap (by: Mandira SigdelPhuyal).



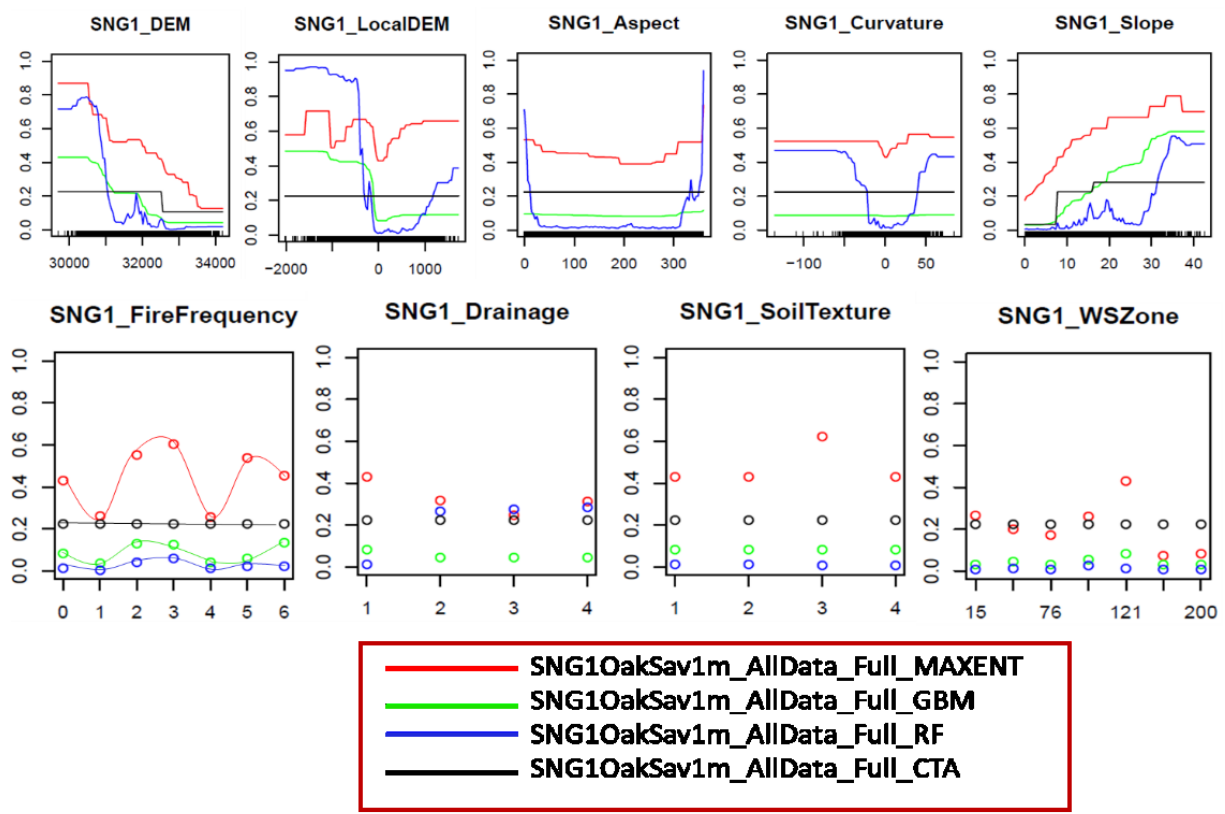
Appendix IV: The NRCS soil water saturation zone (WSZone) map of the SNG1 created in ArcMap. (Unit = cm) (by: Mandira SigdelPhuyal).



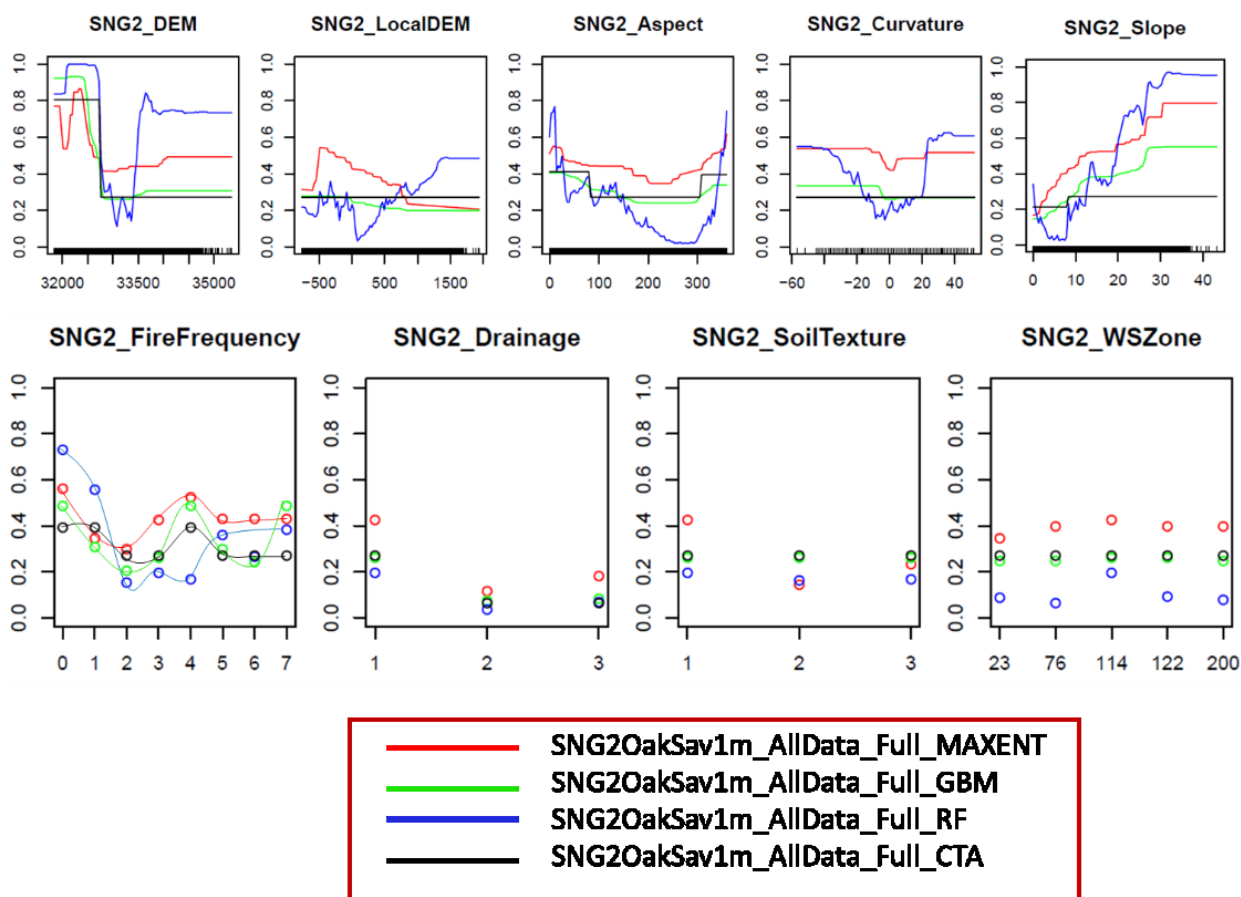
Appendix V: The 1-m spatial resolution original DEM map of the SNG1 created in ArcMap (by: Mandira SigdelPhuyal).



Appendix VI: Oak versus environmental variables response plot for the SNG1 predicted by all four SDMs.



Appendix VII: Oak versus environmental variables response plot for the SNG2 predicted by all four SDMs (by: Mandira SigdelPhuyal).



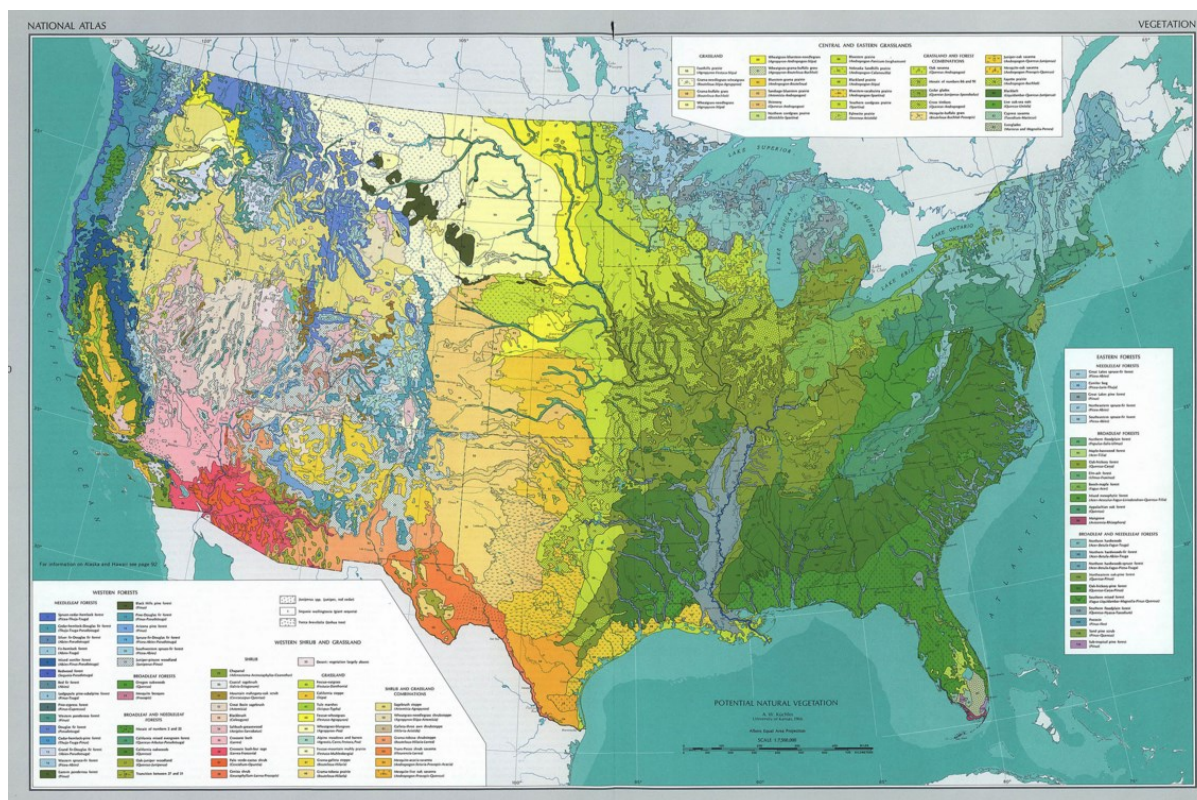
Appendix VIII: Oak savanna distribution in the SNG (Photo taken: 10 October 2014).



Appendix IX: Oak savanna distribution in the SNG (Photo taken: October 10, 2014).



Appendix X: Kuchler's Potential Natural Vegetation map (original) of the United States.



REFERENCES

- Abrams, M. D., W. McShea, and W. Healy. 2002. The postglacial history of oak forests in eastern North America. In *Oak forest ecosystems: ecology and management for wildlife*, eds. William J. McShea and William M. Healy, 34-45. Baltimore: Johns Hopkins University Press. Available from: SDSU Library Catalog, EBSCOhost (accessed 13 December 2014).
- Anderson, R.P., M. Dudík, S. Ferrier, A. Guisan, R. J Hijmans, F. Huettmann, J. R Leathwick, A. Lehmann, J. Li, and L. G Lohmann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29 (2):129-151.
- Apfelbaum, S. I., and A. Haney. 1987. Structure and dynamics of Midwest oak savannas. *Unpublished report to Illinois Non-game Fund. Applied Ecological Services*. <http://www.appliedeco.com/Projects/StructureDynamicsSavannas>. Pdf.
- . 1991. Management of degraded oak savanna remnants in the upper Midwest: preliminary results from three years of study. Paper read at Proceedings of the Oak Woods Management Workshop. Eastern Illinois University, Charleston.
- ArcGIS Pro. *ArcGIS for Desktop*. Available in <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/block-statistics.htm> (accessed 20 November 2015).
- Bader, B. J. 2001. Developing a species list for oak savanna/oak woodland restoration at the University of Wisconsin-Madison Arboretum. *Ecological Restoration* 19 (4):242-250.
- Bailey, R. 2004. Identifying Ecoregion Boundaries. *Environmental management* 34 (1):S14-S26.
- . 2002. *Ecoregion-based design for sustainability*. Springer Science & Business Media.
- . 1976. Ecoregions of the United States. *US Forest Service, Ogden, Utah*. Map scale 1:7,500,000.
- Benayas, J. M. R., A. C. Newton, A. Diaz, and J. M. Bullock. 2009. Enhancement of biodiversity and ecosystem services by ecological restoration: a meta-analysis. *Science* 325 (5944):1121-1124.
- Blumler, M. A., A. Cole, J. Flenley, and U. Schickoff. 2011. History of biogeographical thought. *Handbook of Biogeography*: 23-42.
- Bock, C. E., and J. H. Bock. 1997. Shrub densities in relation to fire, livestock grazing, and precipitation in an Arizona desert grassland. *The Southwestern Naturalist*: 188-193.
- Breiman, L. 2001. Random forests. *Machine learning* 45 (1):5-32.

- Briggs, J. M., A. K. Knapp, J. M. Blair, J. L. Heisler, G. A. Hoch, M. S. Lett, and J. K. McCarron. 2005. An ecosystem in transition: causes and consequences of the conversion of mesic grassland to shrubland. *BioScience* 55 (3):243-254.
- Brudvig, L. A., and H. Asbjornsen. 2005. Oak Regeneration Before and After Initial Restoration Efforts in a Tall Grass Oak Savanna Conference Paper. *The American midland naturalist* 153 (1):180-186.
- Bryan Stotts. 10 October 2014 (personal conversation). United States Forest Service, Shyenenne Ranger District Office, Lisbon, North Dakota.
- Bryce, S.A., Omernik, J.M., Pater, D.A., Ulmer, M., Schaar, J., Freeouf, J., Johnson, R., Kuck, P., and Azevedo, S.H. 1996. *Ecoregions of North Dakota and South Dakota, (color poster with map, descriptive text, summary tables, and photographs)*. Reston, Virginia, U.S. Geological Survey (map scale 1:1,500,000).
- Bucini, G., and N. P. Hanan. 2007. A continental-scale analysis of tree cover in African savannas. *Global Ecology and Biogeography* 16 (5):593-605.
- Camill, P., C. E. Umbanhowar, R. Teed, C. E. Geiss, J. Aldinger, L. Dvorak, J. Kenning, J. Limmer, and K. Walkup. 2003. Late-glacial and Holocene climatic effects on fire and vegetation dynamics at the prairie–forest ecotone in south-central Minnesota. *Journal of Ecology* 91 (5):822-836.
- Coblentz, D. D., and Kurt. H. Riitters. 2004. Topographic controls on the regional-scale biodiversity of the south-western USA. *Journal of Biogeography* 31 (7):1125-1138.
- Considine, C. D., J. W. Groninger, C. M. Ruffner, M. D. Therrell, and S. G. Baer. 2013. Fire history and stand structure of high quality black oak (*Quercus velutina*) sand savannas. *Natural Areas Journal* 33 (1):10-20.
- Coop, J. D., and T. J. Givnish. 2007. Spatial and temporal patterns of recent forest encroachment in montane grasslands of the Valles Caldera, New Mexico, USA. *Journal of Biogeography* 34 (5):914-927.
- Cunningham, M., D. H. Johnson, and D. N. Svingen. 2006. Estimates of breeding bird populations in the Shyenenne National Grassland, North Dakota. *Prairie Naturalist* 38 (1):39-56.
- Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. *Ecology* 88 (11):2783-2792.
- Demharter, T., U. Michel, M. Ehlers, and P. Reinartz. 2011. Object-based change detection: dimension of damage in residential areas of Abu Suruj, Sudan. Paper read at SPIE Remote Sensing.
- Diniz Filho, J. A. F., V. G. Ferro, T. Santos, J. C. Nabout, and R. Dobrovolski. 2010. The three phases of the ensemble forecasting of niche models: geographic range and

- shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae). *Revista Brasileira de Entomologia* 54 (3):339-349.
- Domek, Tom. 1998. *Last call for tallgrass in North Dakota*. North Dakota Outdoors 60(10):14-19. Jamestown, ND: Northern Prairie Wildlife Research Center Online. <http://www.npwrc.usgs.gov/resource/plants/tallgras/index.htm> (Version 02OCT1998; accessed 25 October 2014).
- Eiserhardt, W. L., J.-C. Svenning, W. D. Kissling, and H. Balslev. 2011. Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Annals of botany* 108 (8):1391-1416.
- Elith, J., and C. H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32 (1):66-77.
- Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40 (1):677-697.
- Ffolliott, P., W. McShea, and W. Healy. 2002. Ecology and management of evergreen oak woodlands in Arizona and New Mexico. In *Oak forest ecosystems: ecology and management for wildlife*, eds. William J. McShea and William M. Healy, 304-316. Baltimore: Johns Hopkins University Press. Available from: SDSU Library Catalog, EBSCOhost (accessed 13 December 2014).
- Franklin, J. 2010. *Mapping species distributions: spatial inference and prediction*: Cambridge University Press.
- Friedman, J., T. Hastie, and R. Tibshirani. 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* 28 (2):337-407.
- Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*: 1189-1232.
- Garzon, M. B., R. Blazek, M. Neteler, R. S. De Dios, H. S. Ollero, and C. Furlanello. 2006. Predicting habitat suitability with machine learning models: the potential area of *Pinus sylvestris* L. in the Iberian Peninsula. *Ecological Modelling* 197 (3):383-393.
- Gastón, A., and J. I. García-Viñas. 2011. Modelling species distributions with penalised logistic regressions: A comparison with maximum entropy models. *Ecological Modelling* 222 (13):2037-2041.
- Damien Georges & Wilfried Thuiller. 2013. An example of species distribution modeling with biomod2. *biomod2: getting started*: 1-21.

- George, M. R., and M. F. Alonso. 2008. Oak woodland vegetation dynamics: a state and transition approach. Paper read at Merenlender A, McCreary D, Purcell K (tech. eds): Proceedings of the sixth California oak symposium: today's challenges, tomorrow's opportunities. US Department of Agriculture, Forest Service, Pacific Southwest Research Station, CA, Gen. Tech. Rep. PSW-GTR-217.
- Gibbens, R., R. McNeely, K. Havstad, R. Beck, and B. Nolen. 2005. Vegetation changes in the Jornada Basin from 1858 to 1998. *Journal of Arid Environments* 61 (4):651-668.
- Gill, R., and V. Beardall. 2001. The impact of deer on woodlands: the effects of browsing and seed dispersal on vegetation structure and composition. *Forestry* 74 (3):209-218.
- Gleason, H. A. 1922. The vegetational history of the Middle West. *Annals of the Association of American Geographers* 12 (1):39-85.
- Grundel, R., N. B. Pavlovic, and E. Bollinger. 2007. Distinctiveness, use, and value of Midwestern oak savannas and woodlands as avian habitats. *The Auk* 124 (3):969-985.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8 (9):993-1009.
- Guo, Q., and Y. Liu. 2010. ModEco: an integrated software package for ecological niche modeling. *Ecography* 33 (4): 637-642.
- Hanan, N. P., and C. E. Lehmann. 2011. Tree-grass interactions in savannas: paradigms, contradictions and conceptual models. In *Ecosystem Function in Savannas: Measurement and Modeling at Landscape to Global Scales*, eds. Michael J. Hill and Niall P Hanan, Ch-2: 39-57. Boca Raton: CRC Press.
- Hanspach, Jan, Ingolf Kühn, Sven Pompe, and Stefan Klotz. 2010. Predictive performance of plant species distribution models depends on species traits. *Perspectives in Plant Ecology, Evolution and Systematics* 12 (3): 219-225.
- Harrington, J. A., and E. Kathol. 2009. Responses of shrub midstory and herbaceous layers to managed grazing and fire in a North American savanna (oak woodland) and prairie landscape. *Restoration ecology* 17 (2):234-244.
- Hayes, G. F., and K. D. Holl. 2003. Cattle grazing impacts on annual forbs and vegetation composition of mesic grasslands in California. *Conservation Biology* 17 (6):1694-1702.
- Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29 (5):773-785.
- Herold, M., P. Mayaux, C. Woodcock, A. Baccini, and C. Schmullius. 2008. Some challenges in global land cover mapping: An assessment of agreement and

- accuracy in existing 1 km datasets. *Remote Sensing of Environment* 112 (5): 2538-2556.
- Hill, M. J., M. O. Román & C. B. Schaaf. 2011. Biogeography and Dynamics of Global Tropical and Subtropical Savannas: A Spatiotemporal View. In *Ecosystem Function in Savannas: Measurement and Modeling at Landscape to Global Scales*, eds. Michael J. Hill and Niall P Hanan, 1-37. Boca Raton: CRC Press.
- Hortal, J., N. Roura-Pascual, N. J. Sanders, and C. Rahbek. 2010. Understanding (insect) species distributions across spatial scales. *Ecography* 33 (1):51-53.
- Hovanes, M. E., J. R. G. Deal, and A. H. Rowberg. 1999. Seamless multiresolution display of portable wavelet-compressed images. *Journal of Digital Imaging* 12 (1):109-111.
- Hu, S., and L. Wang. 2013. Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing* 34 (3):790-803.
- Hufkens, K., P. Scheunders, and R. Ceulemans. 2009. Ecotones in vegetation ecology: methodologies and definitions revisited. *Ecological research* 24 (5):977-986.
- Kellman, M., and K. Sanmugadas. 1985. Nutrient retention by savanna ecosystems: I. Retention in the absence of fire. *The Journal of Ecology*: 935-951.
- Kelly, M., Q. Guo, D. Liu, and D. Shaari. 2007. Modeling the risk for a new invasive forest disease in the United States: An evaluation of five environmental niche models. *Computers, Environment and Urban Systems* 31 (6): 689-710.
- Knapp, A. K., J. M. Blair, J. M. Briggs, S. L. Collins, D. C. Hartnett, L. C. Johnson, and E. G. Towne. 1999. The keystone role of bison in North American tallgrass prairie. *BioScience* 49 (1): 39-50.
- Küchler, A. W. 1964. *Potential natural vegetation of the conterminous United States*. American Geographical Society. Vol. 36. New York.
- Lazo, J. D. 2013. *Transferability of species distribution models. A case study of the fungus *Phytophthora cinnamomi* in Andalusia and Southwest Australia*. (Master's thesis, University of Twente). 1-88. Available from https://www.researchgate.net/profile/Joaquin_Duque_Lazo/publication (accessed date 10 November 2014).
- Leathwick, J., J. Elith, M. Francis, T. Hastie, and P. Taylor. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* 321:267-281.
- Lehmann, C. E., T. M. Anderson, M. Sankaran, S. I. Higgins, S. Archibald, W. A. Hoffmann, N. P. Hanan, R. J. Williams, R. J. Fensham, and J. Felfili. 2014. Savanna vegetation-fire-climate relationships differ among continents. *Science* 343 (6170):548-552.

- Leitner, L. A., C. P. Dunn, G. R. Guntenspergen, F. Stearns, and D. M. Sharpe. 1991. Effects of site, landscape features, and fire regime on vegetation patterns in presettlement southern Wisconsin. *Landscape Ecology* 5 (4):203-217.
- Lewis, Roger J. 2000. An introduction to classification and regression tree (CART) analysis. Paper presented in *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*, pp. 1-14.
- Liu, C., M. White, and G. Newell. 2009. Measuring the accuracy of species distribution models: a review. Paper read at Proceedings 18th World IMACs/MODSIM Congress. Cairns, Australia.
- Long III, W., and S. Srihann. 2004. Land cover classification of SSC image: unsupervised and supervised classification using ERDAS Imagine. Paper read at Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International.
- Loveland, T. R., and J. M. Merchant. 2004. Ecoregions and ecoregionalization: geographical and ecological perspectives. *Environmental management* 34 (1):S1-S13.
- Luoto, M., and R. Heikkinen. 2003. Recent developments in spatial methods and data in biogeographical distribution modelling—advantages and pitfalls. *Fennia-International Journal of Geography* 181 (1):35-48.
- Maestre, F. T., M. A. Bowker, M. D. Puche, M. Belén Hinojosa, I. Martínez, P. García-Palacios, A. P. Castillo, S. Soliveres, A. L. Luzuriaga, and A. M. Sánchez. 2009. Shrub encroachment can reverse desertification in semi-arid Mediterranean grasslands. *Ecology Letters* 12 (9):930-941.
- Malanson, George P. 2011. Ecosystems. In *The Sage Handbook of Geographical Knowledge*, eds. John A. Agnew and David N. Livingston. Los Angeles, 452-464. Sage. Available from: SDSU Library Catalog, EBSCOhost (accessed February 22, 2015).
- Manske, L. L., and W. T. Barker. 1988. Habitat usage by prairie grouse on the Sheyenne National Grasslands. *US Forest Service General Technical Report., RM-159:8-20*.
- Matlock, M. D., and R. A. Morgan. 2011. *Ecological engineering design: Restoring and conserving ecosystem services*: John Wiley & Sons.
- McShea, W., and W. M. Healy. 2002. Oaks and acorns as a foundation for ecosystem management. In *Oak forest ecosystems: ecology and management for wildlife*, eds. William J. McShea and William M. Healy, 1-9. Baltimore: Johns Hopkins University Press. Available from: SDSU Library Catalog, EBSCOhost (accessed 13 December 2014).
- McWilliams, W., R. O'Brien, G. Reese, K. Waddell, W. McShea, and W. Healy. 2002. Distribution and abundance of oaks in North America. In *Oak forest ecosystems*:

ecology and management for wildlife, eds. William J. McShea and William M. Healy, 13-33. Baltimore: Johns Hopkins University Press. Available from: SDSU Library Catalog, EBSCOhost (accessed 13 December 2014).

- Millington, A., M. Blumler, and U. Schickhoff. 2011. *The SAGE handbook of biogeography*: Sage.
- _____. 2011. Situating Contemporary Biogeography. In *The SAGE Handbook of Biogeography*, eds. Andrew C. Millington, Mark A. Blumler, and Udo Schickhoff, 1-18. Sage.
- Mistry, J. 2011. The ecosystem dynamics of tropical savannas. In *The SAGE Handbook of Biogeography*, eds. Andrew C. Millington, Mark A. Blumler, and Udo Schickhoff, 281-302. Sage.
- Moisen, G. G., E. A. Freeman, J. A. Blackard, T. S. Frescino, N. E. Zimmermann, and T. C. Edwards. 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling* 199 (2):176-187.
- Marmion, M., M. Parviainen, M. Luoto, R. K. Heikkinen & W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15, 59-69.
- Muller, S., D. Walker, F. Nelson, N. Auerback, J. Bockheim, S. Guyer, and D. Sherba. 1998. Accuracy assessment of a land-cover map of the Kuparuk river basin, Alaska: considerations for remote regions. *Photogrammetric Engineering and Remote Sensing* 64 (6):619-628.
- Mutke, J. 2011. Biodiversity gradients. In *The SAGE Handbook of Biogeography*, eds. Andrew C. Millington, Mark A. Blumler, and Udo Schickhoff, 170-190. Sage.
- NAIP Imagery - USDA Farm Service Agency. *GIS Data, North Dakota*. Gisdata.nd.gov. Available in <http://gisdata.nd.gov/NAIP/> (accessed July 5, 2014).
- Newell, J. A., J. E. Toepfer, and M. A. Rumble. 1988. Summer brood-rearing ecology of the greater prairie chicken on the Sheyenne National Grasslands. *US Forest Service General Technical Report., RM-159:24-31*.
- Nielsen, S., C. Kirschbaum, and A. Haney. 2003. Restoration of Midwest oak barrens: Structural manipulation or process-only? *Conservation Ecology* 7 (2):10.
- Nuzzo, V. A. 1986. Extent and status of Midwest oak savanna: presettlement and 1985. *Natural Areas Journal* 6 (2):6-36.
- Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, and J. C. Morrison. 2001. *Terrestrial Ecoregions of the World: A New Map of Life on Earth* A new global

map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51 (11):933-938.

- Omerik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77 (1):118-125.
- . 1995. Ecoregions: a spatial framework for environmental management. In *Biological assessment and criteria: tools for water resource planning and decision making*, eds. Wayne S. Davis and Thomas P. Simon, Ch-5: 49-62. Lewis Publishers: Boca Raton. New York.
- . 2004. Perspectives on the Nature and Definition of Ecological Regions. *Environmental management* 34 (1):S27-S38.
- Parker, K. C., and J. Bendix. 1996. Landscape-scale geomorphic influences on vegetation patterns in four environments. *Physical Geography* 17 (2):113-141.
- Pennisi, E. 2005. What determines species diversity? *Science* 309 (5731):90.
- Peterson, D. W., P. B. Reich, K. J. Wrage, and J. Franklin. 2007. Plant functional group responses to fire frequency and tree canopy cover gradients in oak savannas and woodlands. *Journal of Vegetation Science* 18 (1):3-12.
- Pearson, Richard G., Wilfried Thuiller, Miguel B. Araújo, Enrique Martinez-Meyer, Lluís Brotons, Colin McClean, Lera Miles, Pedro Segurado, Terence P. Dawson, and David C. Lees. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33(10): 1704-1711.
- Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Paper read at Proceedings of the twenty-first international conference on Machine learning.
- Ransom County North Dakota. *History of the Sheyenne National Grasslands*. <http://www.ransomcountynd.com/index.asp?SEC=21CD6716-D6C3-4963-B29F-9554423B0CE8&Type=NONE> (accessed 3 December 2014).
- . *Sheyenne National Grasslands*. Available in <http://www.ransomcountynd.com/index.asp?Type=NONE&SEC=%7b05B67B83-EC26-4A5B-855F-43D5D61CE444> (accessed 6 December 2014).
- Red River Basin Mapping Initiative (RRBMI). *IWI LiDAR Download Portal*. <http://gis.rrbdin.org/lidarapps.htm> (accessed 2 August 2013).
- Ritchie, J., and C. Haynes. 1987. Holocene vegetation zonation in the eastern Sahara. *Nature* 330 (6149):645-647.
- Ritchie, M. E., D. Tilman, and J. M. Knops. 1998. Herbivore effects on plant and nitrogen dynamics in oak savanna. *Ecology* 79 (1):165-177.

- Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67:93-104.
- Roy, D., L. Boschetti, L. Giglio, M. Hill, and N. Hanan. 2011. Remote sensing of global savanna fire occurrence, extent and properties. *Ecosystem Function in Savannas: Measurement and Modeling at Landscape to Global Scales*.
- Sankaran, M., N. P. Hanan, R. J. Scholes, J. Ratnam, D. J. Augustine, B. S. Cade, J. Gignoux, S. I. Higgins, X. Le Roux, and F. Ludwig. 2005. Determinants of woody cover in African savannas. *Nature* 438 (7069):846-849.
- Sankaran, M., J. Ratnam, and N. Hanan. 2008. Woody cover in African savannas: the role of resources, fire and herbivory. *Global Ecology and Biogeography* 17 (2):236-245.
- Sankaran, M., J. Ratnam, and N. P. Hanan. 2004. Tree–grass coexistence in savannas revisited—insights from an examination of assumptions and mechanisms invoked in existing models. *Ecology Letters* 7 (6):480-490.
- Santana, F. S., and A. M. Saraiva. 2010. Challenges in Ecological Niche Modelling. Conference paper.
- Schetter, T. A., and K. V. Root. 2011. Assessing an imperiled oak savanna landscape in northwestern Ohio using Landsat data. *Natural Areas Journal* 31 (2):118-130.
- Schickhoff, U. 2011. Biogeographical distributions: The role of past environments, physical factors and biotic interactions. In *The SAGE Handbook of Biogeography*, eds. Andrew C. Millington, Mark A. Blumler, and Udo Schickhoff, 141-169. Sage.
- Scheiman, D. M., E. K. Bollinger, and D. H. Johnson. 2003. Effects of leafy spurge infestation on grassland birds. *The Journal of wildlife management*:115-121.
- Schlesinger, W. H., and A. M. Pilmanis. 1998. Plant-soil interactions in deserts. *Biogeochemistry* 42 (1-2):169-187.
- Schröder, B., and R. Seppelt. 2006. Analysis of pattern–process interactions based on landscape models—overview, general concepts, and methodological issues. *Ecological Modelling* 199 (4):505-516.
- Séré, C., A. Ayantunde, A. Duncan, A. Freeman, M. Herrero, S. Tarawali, and I. Wright. 2008. Livestock production and poverty alleviation-challenges and opportunities in arid and semi-arid tropical rangeland based systems. Paper read at *Multifunctional grasslands in a changing world, Volume 1: XXI International Grassland Congress and VIII International Rangeland Congress*, Hohhot, China, 29 June-5 July 2008.

- Sieg, C. H., and A. J. Bjugstad. 1994. Five years of following the western prairie fringed orchid (*Platanthera praeclara*) on the Sheyenne National Grassland, North Dakota. In *Proceedings of the Thirteenth North American Prairie Conference*. Windsor, ON, Canada: Department of Parks and Recreation. p, vol. 141, p. 146.
- Stallins, J. A., and A. J. Parker. 2003. The influence of complex systems interactions on barrier island dune vegetation pattern and process. *Annals of the Association of American Geographers* 93 (1):13-29.
- Stroh, R. K. 2002. *Native woodlands of the Sheyenne National Grassland, North Dakota*. n.p.: [S.l. : Rodney Kenneth Stroh], 2002. Available from: SDSU Library Catalog, EBSCOhost (accessed 25 December 2014).
- Svedarsky, W. D., and G. L. Van Amburg. 1996. *Integrated management of the greater prairie chicken and livestock on the Sheyenne National Grassland*: US Geological Survey, Northern Prairie Wildlife Research Center.
- Tamme, R., I. Hiiesalu, L. Laanisto, R. Szava-Kovats, and M. Pärtel. 2010. Environmental heterogeneity, species diversity and co-existence at different spatial scales. *Journal of Vegetation Science* 21 (4):796-801.
- Taylor, C. M., and R. J. Ellis. 2006. Satellite detection of soil moisture impacts on convection at the mesoscale. *Geophysical Research Letters* 33 (3).
- Terwilliger, V. J., and M. Zeroni. 1994. Gas exchange of a desert shrub (*Zygophyllum dumosum* Boiss.) under different soil moisture regimes during summer drought. *Vegetatio* 115 (2):133-144.
- Thuiller, W. 2003. BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9 (10):1353-1362.
- . 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology* 10 (12):2020-2027.
- Thuiller, W., M. B. Araújo, S. Lavorel, and N. Kenkel. 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14 (5):669-680.
- Thuiller, W., G. F. Midgley, M. Rougeti, and R. M. Cowling. 2006. Predicting patterns of plant species richness in megadiverse South Africa. *Ecography* 29 (5):733-744.
- Thuiller, W., D. Georges, and R. Engler. 2013. Biomod2: Ensemble platform for species distribution modeling. *R package version 2* (7):r560.
- . 2012. *BIOMOD: Tutorial*. 1-88.
- . 2014. *Package 'biomod2'*. 1-93. Available in <https://cran.r-project.org/web/packages/biomod2/biomod2.pdf> (accessed July 2014).

- Georges, Damien and Wilfried Thuiller. 2013. *Multi-species distribution modeling with biomod2*. 1-11. Available in https://r-forge-project.org/scm/viewvc.php/*checkout*/pkg/biomod2/inst/doc/Multi_species_computation.pdf?root=biomod.
- Turner, M. G. 2005. Landscape ecology: what is the state of the science? *Annual Review of Ecology, Evolution, and Systematics*: 319-344.
- USDA Forest Service. *Fire, Fuel and Smoke Science Program, Rocky Mountain Research Station*. Available in <http://www.firelab.org/document/potential-natural-vegetation-groups-v2000> (page modified on 11 March 2014, accessed 20 September 2015).
- USDA Forest Service. 2008. *Description of the Ecoregions of the United States*. (US Ecoregions Maps) Available in <http://www.fs.fed.us/land/ecosysmgmt/index.html>. Maps compiled by Robert G. Bailey, March 1995 (page modified 22 August 2008).
- _____. *Fire, Fuel and Smoke Science Program, Rocky Mountain Research Station*. Available in http://www.fs.fed.us/rm/pubs/rmrs_gtr087/rmrs_gtr087_pg34.jpg (page last modified on 11 March 2014, accessed 10 August 2015).
- _____. *FSGeodata Clearinghouse*. Available in <http://fsgeodata.fs.fed.us/vector/lrsr.php> (accessed date 27 August 2014).
- USDA Natural Resources Conservation Service (NRCS). *Web Soil Survey (WSS)*. Available in <http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx> (page last modified 16 December 2013, accessed 5 January 2015).
- US Environmental Protection Agency. *US EPA Level III and IV Ecoregions*. ArcGIS data. Available in http://geodata.epa.gov/arcgis/rest/services/ORD/USEPA_Ecoregions_Level_III_and_IV/MapServer. Map Service (page last modified 27 August 2015, accessed date 17 October 2015).
- US Environmental Protection Agency, Western Ecology Division. *Ecoregions of North Dakota and South Dakota*. Available in http://archive.epa.gov/wed/ecoregions/web/html/ndsd_eco.html (page last updated 10 June 2015, accessed date 15 December 2015).
- Van Gils, H., F. Conti, G. Ciaschetti, and E. Westinga. 2012. Fine resolution distribution modelling of endemics in Majella National Park, Central Italy. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* 146 (sup1):276-287.
- Veblen T. Thomas. 1989. Biogeography. In *Geography in America*, eds. Gary L. Gaile and Cort J. Willmott, 28-46. Columbus: Merrill Publishing Company.

- Vinod, D., A. R. Bhat, V. Prakash, M. ShivaPrakash, and S. Kannan. 2007. Intensity based image mosaicing. In *Proceedings of the 7th Conference on 7th WSEAS International Conference on Applied Computer Science*, vol. 7, pp. 111-116.
- Wisn, M. S., R. Hijmans, J. Li, A. T. Peterson, C. Graham, and A. Guisan. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14 (5):763-773.
- Young et.al. 2003. Biogeography. In *Geography in America at the dawn of 21st Century*, eds. Gaile and Cort J. Willmott, 17-31. Oxford University Press.
- Zaniewski, A. E., A. Lehmann, and J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157 (2):261-280.