UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INVESTIGATING THE IMPACTS OF ANTHROPOGENIC AND CLIMATIC CHANGES ON THE STEPPE ECOSYSTEM IN CHINA'S LOESS PLATEAU AND THE MIXED-GRASS PRAIRIE REGION OF SUTHWEST OKLAHOMA,USA

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

DONG YAN
Norman, Oklahoma
2014

INVESTIGATING THE IMPACTS OF ANTHROPOGENIC AND CLIMATIC CHANGES ON THE STEPPE ECOSYSTEM IN CHINA'S LOESS PLATEAU AND THE MIXED-GRASS PRAIRIE REGION IN SOUTHWEST OKLAHOMA, USA

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY

____
Dr. Kirsten de Beurs, Chair

____
Dr. Lara Souza

____
Dr. Bruce Hoagland

____
Dr. Jason Julian

____
Dr. Xiangming Xiao

Acknowledgements

I actually found that the acknowledgement section was the most difficult part of the dissertation for me to complete. Because it is difficult for me to show my gratitude, by words, to all the people who supported me through my Ph.D. study.

First, I would like to thank my committee chair, Dr. Kirsten de Beurs, from whom I not only learned how to conduct scientific research, but also how to deal with problems in real life. Without the guidance and suggestions by Dr.de Beurs, I would never been able to complete my PhD study successfully.

Second, I would like to thank my committee members Dr. Bruce Hoagland, Dr. Jason Julian, Dr. Lara Souza and Dr. Xiangming Xiao who provided me crucial comments and suggestions on my course work and dissertation research.

Third, I would like to thank Ms. Martin Shirley, Ms. Deborah Marsh, Dr. Fred Shelley, Dr. Laura Smith and Dr. Darren Purcell who help me with a variety of things through my graduate study at OU.

Last but not the least, I would like to thank my family and all my friends in Norman for their invaluable support during my PhD study.

Table of Contents

List of Tables

List of Figures

Abstract

Grassland ecosystems occupy approximately 40% of the earth's terrestrial area and represent one of most important ecosystems on Earth in terms of its impacts on global food supply, carbon sequestration and maintaining biodiversity. Grassland ecosystems are very sensitive to disturbances caused by either climatic or anthropogenic changes such as changes in precipitation regimes or management practices. The objective of this dissertation is to investigate the impacts imposed by grassland restoration activities and changes in precipitation anomalies on the steppe in China's Loess Plateau and the mixed-grass prairie in southwest Oklahoma. In chapter two, I analyzed how large-scale vegetation conservation programs affected the grassland dynamics in China's Loess Plateau by combining remotely sensed data with socio-economic statistics. The results of this study showed that the impact of vegetation conservation programs on vegetation change in the Loess Plateau is twofold. On the one hand, vegetation conservation programs target marginal lands. Thus, significant vegetation increases due to cropland conversion and afforestation can be found in these regions. On the other hand, intensified agricultural production can be found in croplands with suitable topography and well-established irrigation systems which were not enrolled in conservation programs to offset the agricultural production loss caused by vegetation conservation programs elsewhere. In chapter three, I demonstrated a new methodology on mapping the historical distribution of grassland species in southwest Oklahoma based on the Random Forest classification

algorithm. In this study, elevation, soil pH and soil clay content were found to be significant variables for predicting the distribution of C_3 and C_4 grassland species. With the mapped distribution of grassland species between 1981 and 2010, in chapter four, I examined the relationship between changes in precipitation anomalies and the dynamics of relative abundance of C_3 and C_4 grassland species in southwest Oklahoma. In this study, significant decreases of C_3/C_4 ratio were identified in pasture/hay fields due to the increases in C_4 abundance resulting from the decreases of sparsely vegetated area between 2005 and 2010. I suspect that the increase in C_4 abundance was a drought adaptation strategy adopted by ranchers. Because C_4 species are more tolerant of drought conditions and thus can help to maintain stable forage/hay production when negative precipitation anomalies prevailed during the growing season of C_3 species.

Chapter 1 INTRODUCTION

Grassland ecosystems occupy approximately 40% of the earth's terrestrial area (McSherry & Ritchie, 2013) and about 70% of the world's agricultural area is made up of permanent meadow and pasture (O'Mara, 2012). With such a widespread distribution, grasslands represent one of most important ecosystems on Earth in terms of its impacts on global food supply, carbon sequestration and maintaining biodiversity (Samson & Knopf, 1994; O'Mara, 2012). Grassland ecosystems are very sensitive to disturbances caused by both climatic and anthropogenic changes (Goodin & Henebry, 1997) and there were diverse responses by grassland ecosystems to climatic or anthropogenic changes documented by previous studies.

Overgrazing is the primary form of anthropogenic disturbance that occurred in the Inner Mongolia steppe located in the northern Loess Plateau of China (Gong Li *et al.*, 2000; Hilker *et al.*, 2014). Desertification caused by overgrazing in the Inner Mongolia steppe was found to be able to alter a series of land surface properties and environmental processes such as: albedo and wind regimes (Gong Li et al., 2000). In some cases, anthropogenic disturbances to grassland ecosystems will be followed by changes devoted to restore grassland ecosystems. For instance, to mitigate degraded vegetation condition caused by overgrazing and other poor land management practices such as conversion of grassland to croplands (Qi *et al.*, 2012b), the Chinese government launched one of the largest conservation programs in the world: the Grain to Green Program (GTGP)(Liu *et al.*, 2008). The primary goal is to mitigate soil erosion by increasing vegetation cover through converting steep slope agriculture (croplands on slopes ≥ 15º in northwestern China and ≥ 25º in areas other than northwestern China) to forest or grassland and to carry out afforestation in barren areas(Liu et al., 2008).

The primary effect that climate change and elevated CO_2 were expected to exert on global grassland ecosystems was to change the relative abundance of C_3 and C_4 species (i.e., the C_3/C_4 ratio) (Winslow *et al.*, 2003; Lattanzi, 2010). Plant species can be classified into three functional forms: C_3, C_4 and Cassulacean Acid Metabolism (CAM), based on how carbon dioxide is treated before or in the process of photosynthesis. Carbon dioxide is converted into a three-carbon compound by the C_3 photosynthetic pathway, which is the most typical way of carbon fixation and adopted by approximately 95% of the plant species on the Earth (Bianchi & Canuel, 2011). C_4 plants fix carbon by incorporating carbon dioxide into a four-carbon compound. CAM plants fix carbon dioxide during the night and store it in an acid before carbon dioxide can be used in photosynthesis during daylight hours (Bianchi & Canuel, 2011). C_3 plants tend to occur in environments with moderate temperature and adequate water availability while C_4 and CAM plants are more adapted to high temperature and water stress conditions due to their better water use efficiency (Bianchi & Canuel, 2011; Wang *et al.*, 2013).

The relative abundance of C_3 and C_4 (C_3/C_4) species in grassland ecosystems is found to be an important indicator of climatic disturbances and an indispensable parameter in terms of modeling carbon cycling in grassland ecosystems at global scale (Goodin & Henebry, 1997; Winslow *et al.*, 2003). For instance, elevated atmospheric CO_2 is found to possibly result in a higher C_3/C_4 ratio due to the saturated productivity of C_4 species at the current CO_2 level (Wang *et al.*, 2013). Increased temperature and fire frequency has the potential to lead to a lower C_3/C_4 ratio by favoring the growth of C_4 species (Goodin & Henebry, 1997).

The dynamics of C_3/C_4 ratio over time can facilitate the understanding of those processes that may affect the dynamics such as changes in disturbance regime (Hanberry *et al.*, 2012). An effective way to examine the long term C_3/C_4 ratio dynamics is through remote sensing. A commonly adopted method to discriminate C_3 and C_4 grassland cover types on remotely sensed images is to capture the seasonal greenness asynchrony exhibited by C_3 and C_4 grassland cover types during the growing season(Goodin & Henebry, 1997; Tieszen *et al.*, 1997; Foody & Dash, 2007; Wang *et al.*, 2013). C_3 species begin growth in early spring and reach peak growth in late spring. These species may become senescent or semi-dormant in the summer to avoid hot temperatures and water stress conditions. Some C_3 species may resume growth in the fall (Wang *et al.*, 2013). Growth of C_4 species starts in late spring, reaches the peak during summer. Previous studies on detecting the seasonal greenness asynchrony primarily relied on examining the differences in the shapes of temporal trajectories of photosynthetic greenness using discriminant or clustering analyses (Goodin & Henebry, 1997; Tieszen *et al.*, 1997; Foody & Dash, 2007). In order to effectively detect differences in the greenness trajectory shape, temporally-rich canopy greenness time series were indispensable in these studies. As a result, field canopy greenness measurements (Goodin & Henebry, 1997), observations acquired by Advanced Very High Resolution Radiometer(AVHRR) (Tieszen *et al.*, 1997), and Medium Resolution Imaging Spectrometer (MERIS) (Foody & Dash, 2007) were used in previous studies. C_3 and C_4 grassland cover types in the Great Plains tend to occur in a mixed composition manner (Wang *et al.*, 2013). The utilization of C_3/C_4 ratio retrieved at the scale of AVHRR (1km) or MERIS (1200m) is ill-suitable to analyze the long term dynamics of C_3/C_4 ratio at fine scale due to the presence of mixed pixels. Observations acquired by the Landsat series of satellites, with a spatial resolution ranging from 30m to 60m in the visible to near infrared spectrum, are an ideal data source to examine long term C_3/C_4 ratio dynamics at fine scale. However, the revisit interval of the Landsat series of satellites is 16 days and can be longer if severe cloud contamination is present, which tends to result in a temporally sparse canopy greenness time series that often fails to meet the input requirements by discriminant or clustering analyses. Thus, there is a need for a methodology that will enable the analysis of long term C_3/C_4 ratio dynamics using fine scale observations obtained from satellites such as Landsat.

The primary control on the spatial stratification of C_3/C_4 ratio at global scale was previously believed to be temperature (Ehleringer & Björkman, 1977; Cavagnaro, 1988; Cabido *et al.*, 1997; Ehleringer *et al.*, 1997). For example, in the studies on the changes of distribution of C_3 and C_4 grassland species at different altitudes in central Argentina, balanced abundance of C_3 and C_4 grassland species was found at about 1500m and increases in C_3 and C_4 dominance were observed at above and below 1500m, respectively (Cavagnaro, 1988; Cabido *et al.*, 1997). In these studies, temperature was found to exert significantly higher control on the changes of distribution of C_3 and C_4 species along altitudinal gradients than precipitation (Cavagnaro, 1988; Cabido *et al.*, 1997). In the prairie region of North America, C_3 species were found to dominate the northern Great Plains with C_4 species occupying its southern counterpart (Ehleringer *et al.*, 1997; Epstein *et al.*, 1997). The crossover latitude of C_3 and C_4 species in the Great Plains was found to be between 43-45ºN (Epstein *et al.*, 1997; Winslow *et al.*, 2003). However, in a more recent study on the control of global distribution of relative C_3 and C_4 biomass, the differences in growing season moisture availability and plant water use efficiency were found to be the dominating factors instead of temperature (Winslow *et al.*, 2003). Intensive studies have been conducted to examine the climate control on changes of C_3/C_4 ratio across space, but few studies were devoted to understand how climate could affect the changes of C_3/C_4 ratio over time (Ricotta *et al.*, 2003; Wang *et al.*, 2013).

1.1 Research objectives

The overall research objective of this dissertation was to investigate the impacts of anthropogenic and climatic changes on the dynamics of grassland ecosystems in China's Loess Plateau and the mixed-grass prairie of southwest Oklahoma, USA. There were three specific research objectives associated with the second, the third and the fourth chapter of this dissertation, respectively:

1. Chapter two: To use a multiple lines of evidence approach to investigate the impacts of large scale conservation programs on grassland vegetation dynamics in China's Loess Plateau between 2000 and 2009.
2. Chapter three: To demonstrate a new methodology for distinguishing C_3 and C_4 grassland cover types both in the present and in the recent past based on the Random Forest classification algorithm.
3. Chapter four: To examine the effects of precipitation anomalies on the changes of relative abundance of C_3 and C_4 grassland species in southwest Oklahoma between 1981 and 2010.

In this dissertation, China's Loess Plateau and the mixed-grass prairie in southwest Oklahoma were selected as case studies areas because both the grassland ecosystems in the two study areas are of high conservation significance. Severe soil erosion affects about 70% of China's Loess Plateau(Liu & Diamond, 2005). Severe soil erosion results in significant loss of fertile land resources and imposes food security threats to this region (Schnitzer *et al.*, 2013). Grassland conservation activities were of high importance to mitigate the degraded vegetation condition in this region. Significant declines in mixed grass prairies (30%-99%) and short grass prairie (20%-85%) were also observed within the Great Plains. Transformation of the endangered prairie grasslands can have profound consequences affecting biodiversity conservation and human society (Samson & Knopf, 1994). Conversion of prairie grasslands to row crop agriculture left highly fragmented prairie areas which are more vulnerable to climate driven vegetation changes in any other parts of the United States(Guo, 2000). Prairie grasslands host unique habitats for a variety of endangered species and declines of prairie grasslands may threaten the conservation of keystone species (Samson & Knopf, 1994).

1.2 Organization of dissertation

The first chapter provides an overview of the potential impacts of anthropogenic and climatic changes on grassland ecosystems in China's Loess Plateau and the mixed-grass prairie region of southwest Oklahoma, USA. The research objectives of the dissertation were also stated in the first chapter. In chapter two, I investigated the impacts of large scale conservation programs on grassland vegetation dynamics in China's Loess Plateau between 2000 and 2009 using a multiple lines of evidence approach. Chapter two has been published in the special issue of *Landscape Perspectives on Environmental Conservation* associated with journal *Land* in September, 2013. In chapter three, I demonstrated a new methodology for distinguishing C_3 and C_4 grassland cover types both in the present and in the recent past based on the Random Forest classification algorithm.

Significant variables for predicting the distribution of C_3 and C_4 grassland cover types were also identified. Chapter three will be submitted to the journal *Remote Sensing of Environment*. In chapter four, I examined the effects of precipitation anomalies on the changes of relative abundance of C_3 and C_4 grassland species in southwest Oklahoma between 1981 and 2010. Chapter four will be submitted to *Journal of Biogeography*. Chapter five summarizes the conclusions obtained in the previous chapters and provides suggestions for future researches.

Chapter 2 THE IMPACTS OF WEATHER AND CONSERVATION PROGRAMS ON VEGETATION DYNAMICS IN CHINA'S LOESS PLATEAU

This chapter has been published in the special issue of *Landscape Perspectives on Environmental Conservation* associated with journal *Land* in September, 2013.

Abstract

In this chapter, I present an analysis of the impacts of weather change and large scale vegetation conservation programs on the vegetation dynamics in China's Loess Plateau from 2000 through 2009. I employed a multiple lines of evidence approach in which multi-scale data were used. I employed Normalized Difference Vegetation Index (NDVI) data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) at 500m to identify significant vegetation increases in the Loess Plateau since 2000. I found increases in NDVI for 48% of the Loess Plateau between 2000 and 2009. I were able to attribute up to 37.5% of the observed vegetation increases to weather change, vegetation conservation activities and crop yield increases. I demonstrate that the impact of vegetation conservation programs on vegetation change in the Loess Plateau is twofold. On the one hand, vegetation conservation programs target marginal lands. Thus, significant vegetation increases due to cropland conversion and afforestation can be found in these regions. On the other hand, intensified agricultural production can be found in croplands with suitable topography and well-established irrigation systems which were not enrolled

in conservation programs to offset the agricultural production loss caused by vegetation conservation programs elsewhere.

2.1 Introduction

Forest transitions refer to the change from a net loss in forest cover to a net gain of forested areas within a specific region (Mather & Needle, 1998; Kauppi *et al.*, 2006). The occurrence of forest transitions can improve ecosystem services by increases in forest cover or woody biomass (Rudel *et al.*, 2005). For instance, expanded forest cover can improve water quality by reducing sediment discharges resulting from soil erosion(Rudel *et al.*, 2005). Woody biomass increases can lead to increased carbon sequestration(Rudel *et al.*, 2005). There are two pathways identified to facilitate the development of forest transitions: the 'economic development path' and the 'forest scarcity path' (Rudel *et al.*, 2005). The 'economic development path' describes forest regrowth on abandoned agricultural lands as a result of a decreasing agricultural population, which migrates to urban areas during the process of industrialization and urbanization (Rudel *et al.*, 2005; Mather, 2007). The 'forest scarcity path' is a result of elevated prices of forest products due to the forest scarcity caused by previous agricultural expansion (Rudel *et al.*, 2005; Mather, 2007).

Previous studies on forest transitions in China reveal distinct characteristics in terms of the driving forces and the impacts of forest transitions. Rather than the loss of agricultural population and rising prices of forest products, large scale conservation programs, launched in response to the deteriorated natural environment, were found to be the most significant driving force of Chinese forest transitions (Mather, 2007). Historically, China has suffered from deforestation and widespread soil erosion(Liu & Diamond, 2005). Severe soil erosion associated with degraded vegetation conditions was recognized as the major factor in causing the massive floods in China during the summer of 1998 (Qu, 1999; Uchida *et al.*, 2005). In order to mitigate the degraded environment by protecting natural forests and reducing soil erosion, the Chinese government launched two of the largest conservation programs in the world: the Natural Forest Conservation Program (NFCP) and Grain to Green Program (GTGP)(Liu *et al.*, 2008). NFCP was initiated in 1998 and designed to be implemented in multiple stages between 1998 and 2050(Liu *et al.*, 2008). There are three key goals associated with NFCP: 1) restoring natural forest by issuing logging bans and mountain closure, 2) shifting the major source of timber production from natural forests to plantation forests, and 3) reducing soil and water loss by afforestation(Zhang *et al.*, 2000; Liu *et al.*, 2008). The NFCP has been established in 18 provinces and autonomous regions in China since 2000 with a total investment of about 61 billion Yuan (~9.7 billion dollars) from 1998 to 2005(Zhang *et al.*, 2000; Liu *et al.*, 2008). GTGP was launched in 1999 in three pilot study provinces – Gansu, Shanxi and Sichuan(Liu *et al.*, 2008) – and expanded to 25 provinces and autonomous regions by 2002(Liu *et al.*, 2008). The goal of GTGP is twofold. The primary goal is to mitigate soil erosion by increasing vegetation cover through converting steep slope agriculture (croplands on slopes ≥ 15º in northwestern China and ≥ 25º in areas other than northwestern China) to forest or grassland and to carry out afforestation in barren areas(Liu *et al.*, 2008). GTGP also aims to fight poverty and improve the socioeconomic well-being of participating farmers(Liu *et al.*, 2008).

GTGP received an investment of more than 90 billion Yuan (~14.3 billion dollars) by 2005 and aimed to achieved an increase in vegetation cover of 320,000 km^2 by 2010(Liu *et al.*, 2008).

In this study, I focused on the Loess Plateau which is an area covered by thick and highly erodible loess in central north China(Zheng *et al.*, 2013) (Figure 1). Severe soil erosion affects about 70% of China's Loess Plateau(Liu & Diamond, 2005). Intensive cultivation on marginal land (e.g. land with steep slopes) without effective soil loss protection was identified as one of the major factors contributing to the high erosion rates within this region(Zheng *et al.*, 2013). Severe soil erosion results in significant loss of fertile land resources and imposes food security threats to this region (Schnitzer *et al.*, 2013). The Loess Plateau was a first priority area for NFCP and two of the three pilot study provinces of GTGP (Gansu and Shanxi)(Zhang *et al.*, 2000; Liu *et al.*, 2008) were located here. The Loess Plateau is frequently included in studies which investigate the effects of NFCP and GTGP on vegetation cover change (Cao *et al.*, 2009; Zhou *et al.*, 2009), desertification control(Qi *et al.*, 2012a) and cost-effectiveness of the programs. Typically these studies are only conducted in a small portion of the plateau (e.g. a few counties) and there has been no evaluation of the impacts of NFCP and GTGP on vegetation development in the entire area. In addition, there are debates over the effectiveness of NFCP and GTGP in improving vegetation conditions in the Loess Plateau. For example, Zhou et al. (2009) and Cao et al. (2009) attributed the increase in vegetation cover in Shaanxi Province to GTGP. Afforestation failure by GTGP, however, was also found in some counties of Shaanxi Province due to inappropriately selected tree species(Cao *et al.*, 2009).

In this paper, I employed a multiple lines of evidence approach to investigate the effects of large scale vegetation conservation programs on the vegetation development in China's entire Loess Plateau. I carried out five different analyses to determine the effect of these large scale vegetation conservation programs on the vegetated land surface:

1) I employed satellite derived vegetation indices with a spatial resolution of 500m to identify significant vegetation changes since 2000.

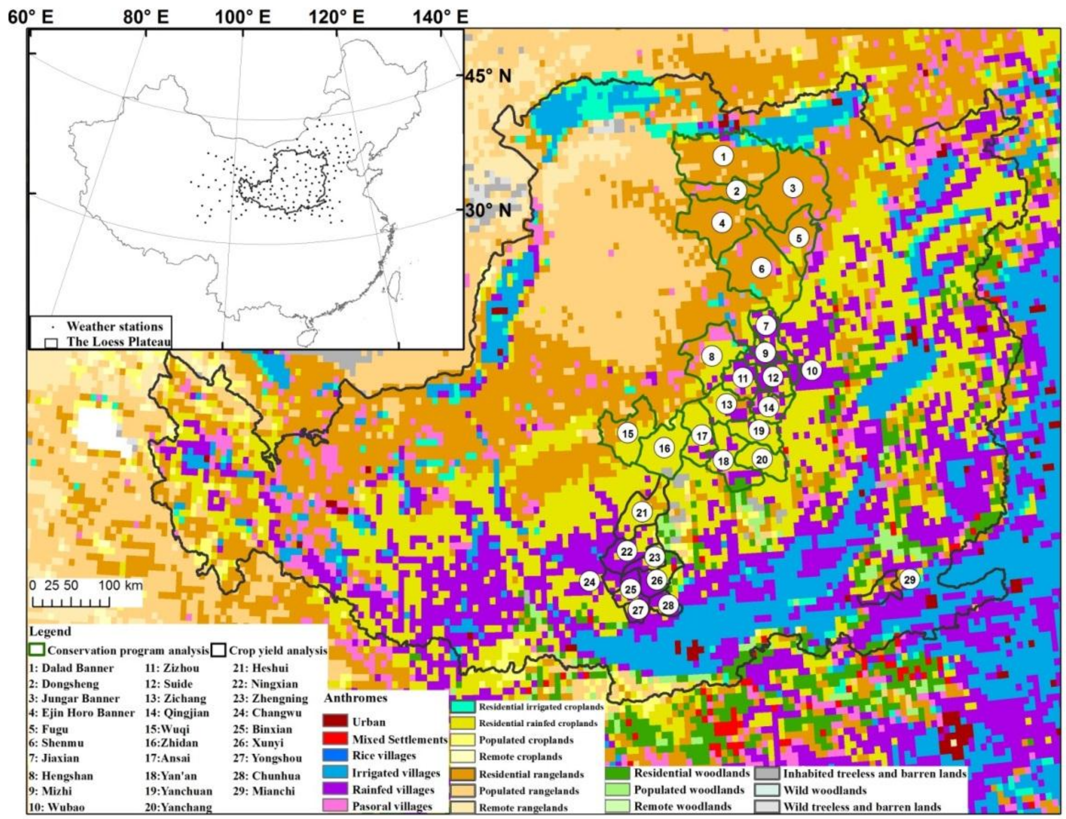
2) I examined monthly temperature and precipitation data to determine if significant weather changes were contributing to the observed vegetation increases.

3) I employed satellite derived land cover data and vegetation conservation data reported by statistical yearbooks to investigate if the observed vegetation increases can be attributed to land cover change caused by vegetation conservation activities.

4) I used anthropogenic biomes to evaluate how large scale vegetation conservation programs choose target regions.

5) I investigated crop yield data to investigate the effects of vegetation conservation programs on croplands in areas with relatively shallow slopes.

Figure 2.1 Upper Left inset: Location of the Loess Plateau and the 138 weather stations in China. Main map frame: Location of key case study counties within the Loess Plateau. Case studies counties selected for conservation program analyses are outlined in green (County #1 to #20). Case studies counties selected for crop yield analyses are outlined in black (County #21 to #29). The spatial distribution of anthropogenic biomes is based on the anthropogenic biomes in 2000 as described in Ellis et al. (2010).



2.2 Study region

2.2.1 Loess Plateau

The Loess Plateau, with an area of about 626,000km^2, is located in the middle reaches of the Yellow River. Elevation of the Loess Plateau ranges from 1000m to 1600m above sea level(Chen *et al.*, 2007). The Loess Plateau extends from 101ºE to 114ºE and 34ºN to 41ºN, covering the entire area of Ningxia Hui Autonomous Region and Shanxi Province and parts of Shaanxi, Gansu, Henan and Qinghai Provinces and Inner Mongolia Autonomous Region (Figure 2.1). The Loess Plateau hosts a population of about 86 million mostly rural residents(Wei *et al.*, 2006). The provincial capitals located in the Loess Plateau include Hohhot, Lanzhou, Taiyuan, Xi'an, Xining and Yinchuan. The Loess Plateau features a northwest-southeast climate gradient. The mean annual precipitation increases from 300mm in the northwest to 700mm in the southeast(He *et al.*, 2006). The mean annual temperature ranges from about 4°C in the northwest to 14°C in the southeast(He *et al.*, 2006).

2.2.2 Study regions for conservation program analysis and crop yield analysis

Besides studying the entire Loess Plateau, I also selected 20 counties in the northern Loess Plateau to analyze the effects of vegetation conservation programs on vegetation dynamics in more detail. Of these 20 counties, four were in the region of Ordos, Inner Mongolia, nine counties were in the region of Yulin, Shaanxi Province and the remaining seven counties were located in the region of Yan'an, Shaanxi Province (Figure 2.1). In addition, I selected nine counties as case study areas to investigate the effects of GTGP on the crop yield of croplands that were not directly participating in GTGP (Figure 2.1).

Detailed explanations of the selection criteria for the case study counties for the vegetation conservation and crop yield analyses are available in section 2.3.4 and 2.3.6, respectively.

2.3 Data and Methods

Determining the effect of large conservation programs over large areas is not straightforward. There are several different factors that play a role in the location and the success of these programs. In this study I took a multiple lines of evidence approach to determine the effect of the large scale conservation programs on the vegetated land cover in the Loess Plateau. First, I determined where vegetation is changing according to remotely sensed vegetation index data (section 2.3.1). Then I went through multiple steps to attribute the vegetation increases to one of the following potential causes: weather (2.3.2), land cover change (2.3.3), vegetation conservation activities reported in statistical yearbooks (2.3.4), the effect of GTGP and NFCP on elevated areas and areas with steep slopes (2.3.5), and finally I investigate how crop yield has changed as an indirect result of conservation programs (2.3.6). An overview of the data used in the following seven sections can be found in table 2.1.

Table 2.1 Overview of data sources

Dataset	Region	Spatial Resolution	Temporal Resolution and Time	Reference
MODIS Nadir BRDF-Adjusted Reflectance (MCD43A4)	Loess Plateau	500m	16-day 2000 - 2009	Schaaf et al. (2002)
MODIS global 500m land cover product (MCD12Q1)	Loess Plateau	500m	Annual 2003 - 2008	Friedl et al. (2010)
Temperature and Precipitation	Loess Plateau	Stations	Monthly 1999 - 2009	Chinese National Meteorological Information Center (2008)
Agricultural Productivity	Five counties in Xianyang (Shaanxi) and three counties in Qingyang (Gansu) and one county in Sanmenxia (Henan)	County	Annual 2000 – 2008*	Regional statistical yearbooks
ASTER Digital Elevation Map	Loess Plateau	1 arc-second (~30m)	2011	Tachikawa et al. (2011)
Anthropogenic biomes	Loess Plateau	0.083º	2000	Ellis and Ramankutty (2008)
Vegetation Conservation Statistics	Seven counties in Yan'an (Shaanxi) and nine counties in Yulin (Shaanxi) and four counties in Ordos (Inner Mongolia)	County	2000 to 2008**	Regional statistical yearbooks

* Except in 2002 and 2005 for Sanmenxia and Xianyang, respectively.

** 2000 to 2008 for Yan'an; 2000 to 2007 for Yulin and Ordos

I employed a stepwise approach rather than a conventional multiple regression procedure to attribute observed vegetation increases to different factors because data availability was variable in space and time and thus a unified multiple regression approach could not be used. For example, for any location we were interested in, a multiple regression procedure would require us to have a complete set of response and predictor variables. However, the crop yield and vegetation conservation data only covered a limited portion of the area with observed vegetation increases.

2.3.1 Vegetation changes measured by MODIS Nadir BRDF-Adjusted Reflectance data and NDVI maximum value composites

The Moderate Resolution Imaging Spectroradiometer (MODIS) Nadir BRDF-Adjusted Reflectance (NBAR) data(Schaaf *et al.*, 2002) at 500m spatial resolution, with temporal resolution of 16 days (MCD43A4) and temporal coverage from Feb 24, 2000 through Dec 26, 2009, were employed to derive the Normalized Difference Vegetation Index (NDVI) as:

$$NDVI = (NIR - RED) / (NIR + RED) \quad (2.1)$$

Where, NIR is MODIS band 2 (841-876 nm), RED is MODIS band 1 (620-670 nm).

The MODIS reflectance dataset is available as 10°×10°latitude/longitude tiles and the Loess Plateau is covered by four tiles (h26v4, h25v5, h26v5, h27v5). The MODIS reflectance dataset is originally produced every 8 days based on 16 days of acquisition. After calculating NDVI for each original reflectance composite, I generated 16-day NDVI composites by averaging the original consecutive 8-day rolling NDVI composites. With a temporal resolution of 16 days there are a total of 227 NDVI composites for the time period from 2000 to 2009 (20 NDVI composites for year 2000 and 23 composites for each year from 2001 to 2009).

I applied the seasonal Mann-Kendall (SMK) trend test to estimate vegetation change from the 272 NDVI composites between 2000 and 2009. The SMK test statistic was calculated based on the rank of a season (in this case, a 16-day period) within the time series of the same season across all years(de Beurs & Henebry, 2004). Two steps were carried out to calculate the SMK test statistic: 1) the Mann-Kendall(MK) test statistic is the sum of the number of times each season has a higher rank than the same season in any previous year; 2) The SMK test statistic for the entire time series is calculated by summing the MK test statistic for each season(de Beurs & Henebry, 2004). SMK is then corrected for autocorrelation among seasons. The advantage of the seasonal Mann-Kendall trend test over simple linear regression in terms of estimating changes in NDVI time series was previously discussed by de Beurs and Henebry(de Beurs & Henebry, 2004). The SMK trend test was applied to each 500m pixel in the MODIS NDVI dataset. I interpreted the area having a positive SMK test statistic with a p-value lower than or equal to 0.01 and annual missing data rates lower than or equal to 40% as highly significant positive vegetation change. The area having a negative SMK test statistic with a p-value lower than or equal to 0.01 and missing data rates lower than or equal to 40% was interpreted as highly significant negative vegetation change.

Besides applying SMK change analysis, I generated monthly NDVI maximum value composites to represent the overall vegetation condition for each month. I also calculated annual accumulated NDVI on a pixel basis by summing NDVI for the entire year using the 16-day NDVI composites.

2.3.2 Weather Data

Weather data used in this study included monthly temperature and precipitation records from 138 weather stations, spread over the Loess Plateau (Figure 1). The weather data were collected and compiled by the Chinese National Meteorological Information Center (http://cdc.cma.gov.cn). The data were of high accuracy after eliminating errors by extreme value testing, temporal consistency testing and manual error checks(Chinese National Meteorological Information Center., 2008). I applied kriging interpolation to generate monthly raster layers at a spatial resolution of 500m for temperature and precipitation from 1999 to 2009. I then applied the SMK trend test to the time series of monthly temperature and precipitation data from 1999 to 2009 on a pixel by pixel base in order to examine whether there was significant weather change that could be related to the vegetation dynamics. I took the area with a *p*-value lower than or equal to 0.1 as a region showing significant weather change since 1999.

In order to determine whether there were significant impacts on vegetation change imposed by weather changes, I employed the Spearman's rank correlation coefficient to calculate the lagged correlation between the monthly NDVI maximum value composites and the precipitation in the previous month. I also computed the correlation coefficient between temperature and NDVI with the same monthly lag.

In addition, I employed the *p*-values associated with NDVI and weather trend analyses to assess the probability of error in terms of attributing NDVI increases to significant changes in temperature or precipitation. I calculated the error by estimating the probability that both changes in NDVI and weather were correctly estimated.

2.3.3 Land Cover Data

The MODIS global 500m land cover product (MCD12Q1) was employed to derive the land cover information over the study period. As the quality of the data was higher for the years starting in 2003 (M. Friedl, personal communication, October 18, 2010), MODIS land cover data from 2003 through 2008 were selected for this study. The MODIS land cover data provides multiple land cover classification schemes including: IGBP global vegetation, University of Maryland land cover classification, MODIS-derived LAI/fPAR and the Plant Functional Type. I picked the IGBP classification scheme which was accompanied by a separate data layer that provides the assessment of relative classification quality for each 500m pixel. The classification quality ranges from 0 to 100 with a higher number representing a higher probability that the classification was correct. The IGBP classification scheme allowed an analysis of fourteen land cover categories including twelve vegetation classes and four classes of non-vegetated land.

Year-to-year land cover change inferred from direct comparison of the MODIS land cover data across years could contain substantial spurious changes due to inaccurate classifications resulting from potential mixed pixels at 500m spatial resolution, the limited separability of certain land cover classes in the MODIS spectral-temporal space, phenological changes and disturbances (e.g. fires) (Friedl *et al.*, 2010). In order to avoid characterizing the land cover change during the study period by direct comparison of the MODIS land cover data, I divided the six-year MODIS land cover datasets into two groups: one from 2003 to 2005 and the other from 2006 to 2008. For each group, I derived a three-year-summarized land cover map. Pixels that were classified as the same land cover type for at least two consecutive years were considered as stable. The land cover type of these pixels was assigned as the land cover type with the highest frequency during the three years. Other pixels were assigned to a new class labeled as random classification. In order to evaluate the efficiency of this grouping strategy, I compared the land cover assessment scores of the pixels that were considered as stable with those considered as random classifications. I employed a two-sample t test to determine if the mean land cover assessment scores of stable pixels were significantly higher than the scores of the random classification pixels.

I explored the land cover change for three broad land cover categories: natural vegetation, cropland, and barren or sparsely vegetated to identify vegetation dynamics that could be caused by large scale vegetation conservation programs. To create the natural vegetation class we combined the ten natural vegetation classes under the IGBP classification scheme.

I employed the relative classification quality scores to assess the potential errors in the land cover change analysis. I first calculated the average classification quality scores for each of the three land cover categories in the two time periods: from 2003 to 2005 and from 2006 to 2008. The error rate for any type of land cover change was calculated to estimate the probability that the land cover types in both time periods were correctly classified.

Besides investigating land cover changes according to MODIS data, I also explored anthropogenic biomes (Anthromes). Anthromes provide a way of understanding the terrestrial biosphere in terms of form and intensity of human influence on biomes (Ellis & Ramankutty, 2008; Ellis *et al.*, 2010). Anthromes are organized according to population density with village anthromes supporting more than 100 people / km^2 and residential anthromes supporting between 10 and 100 people / km^2. Populated (1-10) and remote anthromes (< 1) support fewer people. Anthromes from the year 2000 were employed in this study to identify the preference of large scale vegetation conservation programs in choosing target regions in the Loess Plateau based on population density and agricultural productivity. Anthromes are mapped at a spatial resolution of 0.083º and are available at http://ecotope.org/anthromes/v2/data/. We examined the distributions of land cover changes that may be caused by vegetation conservation programs (e.g. croplands and barren lands converted to natural vegetation) within different anthromes. We compared the distribution of different types of land cover change (e.g. croplands converted to natural vegetation vs. croplands left unchanged) to understand how large scale vegetation conservation programs choose target regions.

2.3.4 Vegetation conservation data from statistical yearbooks

In order to identify land cover change that occurred at scales finer than 500m and thus may not have been identified in the land cover change analysis using MODIS derived land cover maps, I employed vegetation conservation data documented in regional statistical yearbooks.

Forestry and agricultural statistics documented in a regional statistical yearbook are published by the regional bureau of statistics using a methodology that integrates household survey and a bottom-up reporting system(Gale, 2002; Statistic Bureau of Xianyang., 2008). The bottom-up reporting system works in such a way that the collection of the target statistics begins with the statistics reported by officials at the bottom level (e.g. the head official of a village)(Gale, 2002). The reported statistics are aggregated at each upper level before being integrated with data acquired using probability based survey methods and published by the regional bureau of statistics(Gale, 2002; Statistic Bureau of Xianyang., 2008).

Inaccurate national agricultural statistics such as inflated crop yield resulting from underreported sown or cultivated area had been reported by previous studies using these data(Hansen *et al.*, 2003). The quality of the official statistics has been improved after probability based survey methods were integrated with the bottom-up reporting system. Numerous peer-reviewed articles on the economic development of China that employ the official statistics have been published in professional journals(Chow, 2006).

I examined the effects of the vegetation conservation programs on vegetation increases in 20 counties located in northern Shaanxi Province and western Inner Mongolia. The 20 counties were chosen based on: 1) the availability of vegetation conservation data; 2) observed NDVI increases not explained by land cover change in natural vegetation accounted for at least 80% of the total natural vegetation within the county.

I identified 20 counties in Shaanxi Province and western Inner Mongolia (Figure 2.1). SMK derived significant increases in precipitation coincided with more than 40% of the unexplained vegetation change in eight of the 20 counties. No more than 15% coincidence between SMK derived increases in precipitation and NDVI were found in the remaining 12 counties. Significant increases in temperature were not identified in any of the selected 20 counties.

Vegetation conservation data are available for the regions of Yan'an and Yulin in Shaanxi Province (2000 to 2008 for Yan'an; 2000 to 2007 for Yulin) and for the region of Ordos in Inner Mongolia from 2000 to 2007(Statistic Bureau of Yulin., 1999-2007; Statistic Bureau of Yan'an., 1999-2008; Statistic Bureau of Ordos., 2000-2008). These data document the annual area of vegetation conservation activity including artificial planting of forest and grassland, aerial seeding, cropland converted to forest or grassland, and mountain closures (Table 2.2).

Table 2.2 Vegetation conservation activities in Ordos, Yulin and Yan'an

Region	Artificial planting	Aerial seeding	Cropland converted to forest and grassland	Mountain closure	Dominant anthromes
Ordos, Inner Mongolia	Available	Available	N/A	N/A	Residential and populated rangelands
Yulin, Shaanxi	Available	Available	Available	N/A	Residential rangelands
Yan'an, Shaanxi	Available	Available	Available	Available	Residential rainfed croplands

I employed spearman rank partial correlation analysis to understand if vegetation conservation activities documented in statistical yearbooks in combination with weather variables (precipitation and temperature) can be used to explain the observed positive vegetation change. The dependent variable was defined as the spatially averaged annual maximum NDVI of natural vegetation pixels with 80% observed vegetation increases not explained by land cover change. Three independent variables were used in the analysis:1) accumulated vegetation conservation activities from 2000 to the previous year (e.g. accumulated conservation activities for year 2004 were the sum of conservation activities conducted from 2000 through 2003); 2) spatially averaged annual precipitation for each year; 3) spatially averaged mean annual temperature for each year. The partial correlation analysis was carried out separately for each of the three independent variables while the other two independent variables were controlled. A two-tailed Student's t test was carried out to test if the partial correlations were significant. I examined the partial correlation between these variables from 2003 through 2009. I did not include the years between 2000 and 2002 in the analysis because newly afforested/mountain closure/cropland conversion areas may not exhibit increases in NDVI immediately. I compared the results for the three independent variables to see which combination could explain most of the observed vegetation change.

2.3.5 Elevation and slope

The goals of GTGP and NFCP have led us to hypothesize that the observed land cover change including conversion from cropland or sparsely vegetated area to natural vegetation occurring on steep slopes ($\geq 15°$ in the Loess Plateau(Liu *et al.*, 2008)) and in relatively high elevation regions (above 1200 m) could be the result of the implementation of one of the two vegetation conservation programs. In order to determine whether the observed vegetation changes were indeed related to GTGP and NFCP, we compared the spatially averaged slope and elevation (derived in ArcGIS) on which the land cover change occurred to that of the cropland and sparsely vegetated area experiencing no dynamics. We retrieved the elevation data from the ASTER Global Digital Elevation Map (GDEM) which has a spatial resolution of 1 arc-second (~30m) and is comprised of 22,600 $1° \times 1°$lat/lon tiles with a geographic projection(Tachikawa *et al.*). We generated a mosaic of 89 tiles to cover the Loess Plateau. We re-projected and resampled the ASTER DEM mosaic to

Albers Equal Area Projection with 500m spatial resolution to match the projection and spatial resolution of other datasets used in this study.

2.3.6 Crop yield data

I investigated the relationship between observed crop yield and observed NDVI changes for croplands that experienced no land cover change during the study period. I only investigated counties with available crop yield data and in which the area of croplands showing significant increases in NDVI accounted for more than 60% of the total cropland area of the county. I found nine qualified counties in the Loess Plateau. Five of these counties were in Shaanxi Province, three in Gansu Province and one in Henan Province. Significant increases in temperature and precipitation detected by the SMK analysis were not found in any of the nine counties. Dominant anthromes within the selected nine counties are rainfed villages and residential rainfed croplands according to the anthromes in 2000 (Figure 2.1).

Agricultural productivity data were obtained from officially published regional statistical yearbooks(Statistic Bureau of Yan'an., 1999-2008; Statistic Bureau of Sanmenxia., 2000-2001, 2003-2008; Statistic Bureau of Xianyang., 2008). Agricultural productivity data are available for the region of Xianyang in Shaanxi Province, the region of Qingyang in Gansu province and for the region of Sanmenxia in Henan Province from 2000 to 2008 (except in 2002and 2005 for Sanmenxia and Xianyang, respectively). Data for annual sown area, crop production and crop yield are available in the agricultural productivity dataset for major crops including: wheat, corn, bean, potato, rapeseed, tobacco, vegetables and apple.

I calculated the sown area weighted crop yield for each county as the sum of crop yields of major crops, which were weighted by the sown area ratio. The sown area ratio of a specific crop was computed by dividing the sown area of the crop by the total sown area of all crops in a county.

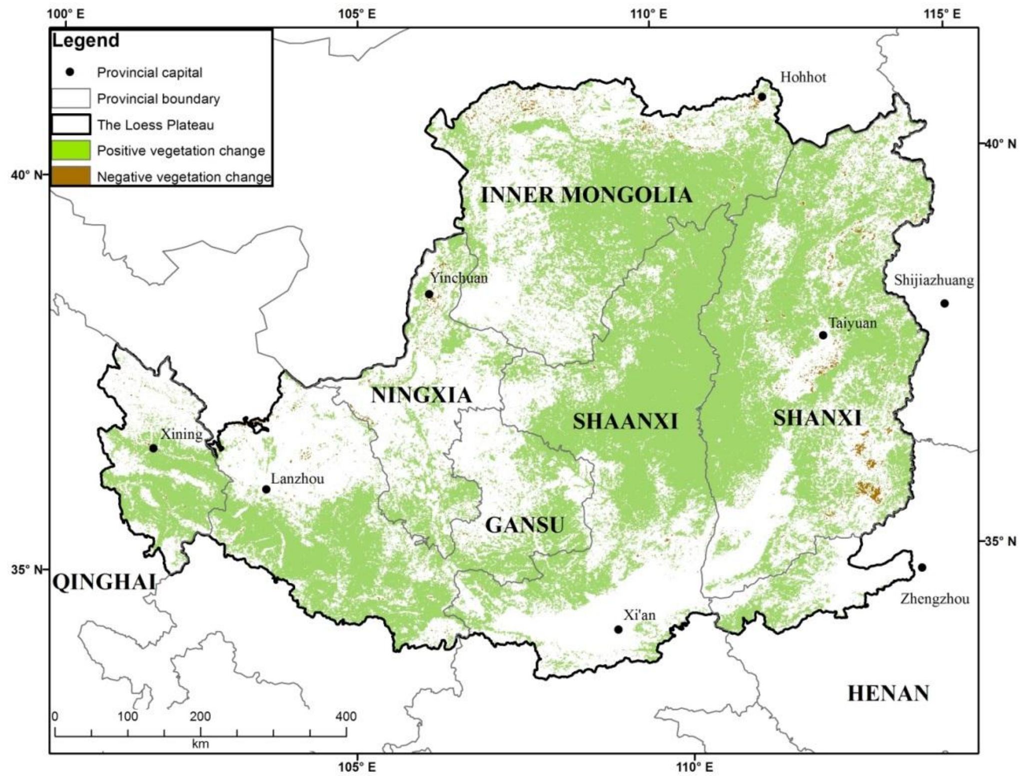
I employed spearman rank partial correlation analysis to understand if crop yield change or weather change or the combination of the two factors were responsible for the observed positive vegetation changes that were not attributable to land cover change. I used spatially averaged annual accumulated NDVI of cropland pixels in a specific county as the dependent variable. Three independent variables were used in the analysis:1) sown area weighted crop yield (kg/ha); 2) spatially averaged annual precipitation (mm); 3) spatially averaged mean annual temperature (°C). The partial correlation analysis was carried out separately for each of the three independent variables while the other two independent variables were controlled. A two-tailed Student's t test was carried to test if the partial correlations were significant. I compared the analysis results for the three independent variables to see which combination can explain the most of the observed vegetation change.

2.4 Results and Discussion

2.4.1 Vegetation change since 2000

I found significant vegetation increases as measured by NDVI in 48.2% ($3.02\times10^5 km^2$) of the Loess Plateau (Figure 2.2). The area with positive vegetation change spreads across the Loess Plateau from northeast to southwest with the major part in the north of Shaanxi province and south of Gansu and Qinghai provinces. I found significant negative vegetation change in 0.8% (4.9×10^3 km^2) of the Loess Plateau (Figure 2.2) and no significant vegetation change in 51% of the study area ($3.20\times10^5 km^2$). I excluded the region showing no significant vegetation change from the remaining analyses.

Figure 2.2 Significant vegetation change between 2000 and 2009 in the Loess Plateau. The significant vegetation change was identified by applying seasonal Mann-Kendall trend test to 500m MODIS NDVI time series acquired between 2000 and 2009. Green pixels represent significant positive vegetation change, which accounted for 48.2% of the Loess Plateau. Brown pixels indicate significant negative vegetation change, which accounted for 0.8% of the Loess Plateau. Pixels in white represent either areas with no significant vegetation change or areas with high missing data rates due to frequent cloud cover.

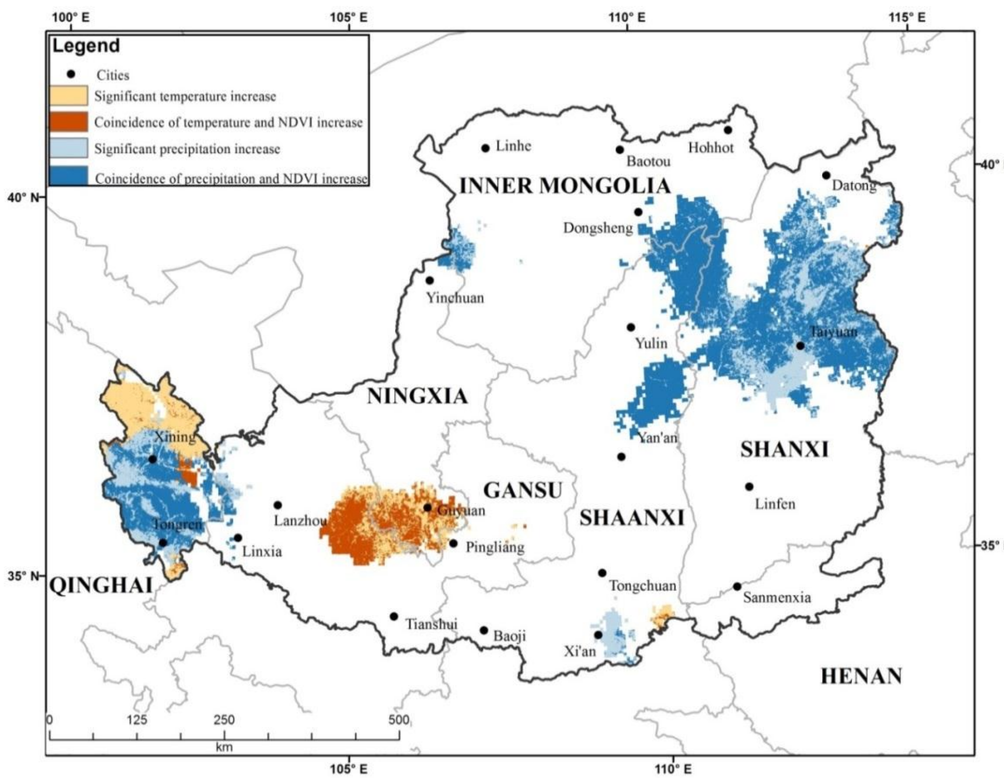


I investigated the distribution of positive vegetation change in terms of annual maximum NDVI in 2000. I found that the majority of the positive vegetation change occurred in the area that had a maximum NDVI ranging from 0.2–0.4 in 2000, and the pixels with NDVI values lower than 0.6 in 2000 accounted for 75.1% of the area with positive change. The distribution of positive vegetation change indicated that the majority of the positive vegetation change occurred in the region with relatively low vegetation cover at the beginning of the study period.

2.4.2 Impact of weather change on vegetation change in the Loess Plateau

I found significant positive correlation between precipitation and lagged NDVI in 82.6% of the Loess Plateau and between NDVI and temperature in 95.2% of the Loess Plateau. Precipitation is the main factor constraining vegetation growth in the far north (mean annual precipitation: 300mm(He *et al.*, 2006)) but not in the far south, which receives more precipitation (mean annual precipitation: 700mm(He *et al.*, 2006)). The central area, with moderate precipitation, revealed significant correlation between NDVI and precipitation. Despite the positive correlation between weather and NDVI, the NDVI increases cannot be entirely explained by weather changes. Significantly increasing precipitation ($p \leq 0.1$) was only found in the northeastern part and the far southwestern part of the Loess Plateau and accounted for just 22.2% of the total area of positive vegetation change (Figure 3). I found significant warming ($p \leq 0.1$) in the southwest corner of the Loess Plateau but this warming coincides with positive vegetation change in just 3.8% of the pixels (Figure 2.3). Based on the error assessments using *p*-values, the NDVI increases that can be attributed to significant increases in precipitation and temperature were 19.8% to 22.2% and 3.4% to 3.8%, respectively.

Figure 2.3 Significant temperature and precipitation increases from 1999 to 2009 in the Loess Plateau. The significant temperature and precipitation increases were identified by applying seasonal Mann-Kendall trend test to monthly 500m temperature and precipitation time series interpolated from weather station based monthly observations. Blue pixels represent significant precipitation increases, which accounted for 16.8% of the Loess Plateau. Dark blue pixels indicate the coincidence of significant NDVI and precipitation increases, which accounted for 10.7% and 22.2% of the entire Loess Plateau and positive vegetation change area, respectively. Brown pixels indicate significant temperature increases, which accounted for 4.7% of the Loess Plateau. Dark brown pixels indicate the coincidence of significant NDVI and precipitation increases, which accounted for 1.8% and 3.8% of the entire Loess Plateau and positive vegetation change area, respectively.



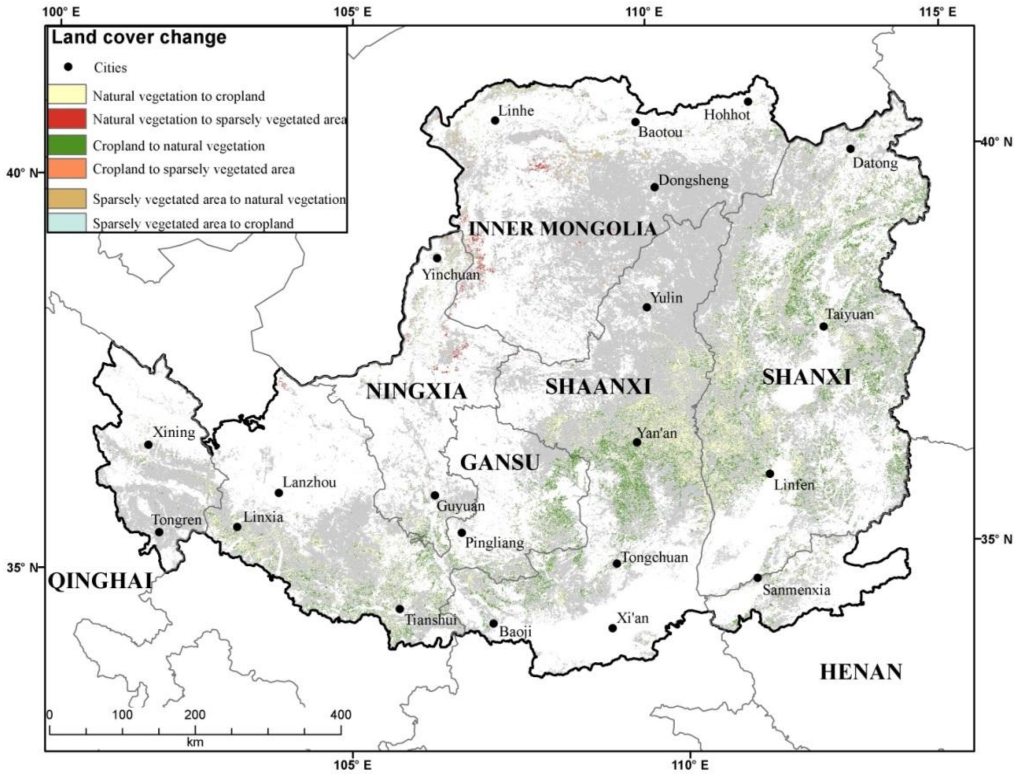
2.4.3 Land cover change revealed by MODIS

I compared the land cover classification quality between stable and random classification pixels derived from our three-year grouping strategy during 2003 to

2005 and 2006 to 2008. Results of the two-sample t test show that the mean land cover classification quality for stable pixels was significantly higher than that of random classification pixels during both periods of 2003-2005 (70.9 vs 61.7, p = 0.01) and 2006-2008 (72.0 vs 63.1, p = 0.01). The higher land cover classification quality associated with the stable pixels suggests that the three-year average strategy was able to identify high quality classification results and thus can ensure reliable land cover change detection.

A little more than 13% ($3.97 \times 10^4 km^2$) of the area with positive vegetation change in the MODIS time series reveals a change in terms of the three broad land cover categories of natural vegetation, cropland, and barren or sparsely vegetated area (Figure 2.4). Three types of land cover changes can be used to explain NDVI increases: from cropland or sparsely vegetated area to natural vegetation or from sparsely vegetated area to cropland. These three types of land cover changes accounted for 7.6% of the area with significant increases in NDVI. The land cover classification errors for these three types of land cover changes were 53% (from cropland to natural vegetation), 45% (from sparsely vegetated area to natural vegetation) and 50% (from sparsely vegetated area to cropland), respectively. Based on the probability of land cover change analysis errors, the percentage of NDVI increases that were attributable to land cover changes ranged from 3.6% to 7.6%.

Figure 2.4 Land cover change within the positive vegetation change region of the Loess Plateau between 2003-2005 and 2006-2008. The land cover change analysis was conducted among three broad land cover categories: natural vegetation, cropland and barren or sparsely vegetated area. Between 2003-2005 and 2006-2008, there was 4.1% and 6.2% increase in natural vegetation and sparsely vegetated area within the positive vegetation change region, respectively. A 10.6% decrease in cropland area was also identified.



The remainder of the area with positive vegetation change reveals no land cover change observable with 500m satellite data. We found that within the area with positive vegetation change based on the MODIS time series analysis, croplands are decreasing while natural vegetation and sparsely vegetated areas are increasing (Table 2.3). The area identified as random classification accounts for less than 3% of the total area with positive vegetation change. Therefore it is reasonable to assume that the random classification would not cause significant bias in the land cover change analysis.

Table 2.3 Land cover change shown by three-year-summarized MODIS land cover data.

	Natural Vegetation (km^2)	Cropland (km^2)	Barren or Sparsely Vegetated (km^2)
2003-2005	2.09×10^5	7.13×10^4	6.40×10^3
2006-2008	2.17×10^5	6.38×10^4	6.80×10^3
Change (%)	4.13%	-10.62%	6.22%

The conversion between natural vegetation and cropland accounted for the majority of land cover changes detected by the MODIS land cover dataset. This conversion mainly occurred in central Shaanxi and northern Shanxi Province. Although there was a 6.2% increase in the area of barren land, increases in barren land only accounted for about 1% of the area showing land cover change. We found that the distribution of different types of land cover change were population dependent (Table 2.4). For example, the majority of pixels (58.6% = 1.7% + 49.3% + 7.6%) identified as croplands which were converted to natural vegetation occurred in anthromes classified as residential anthromes while 37.2% (4.1% + 29.7% + 3.4%) of conversion of croplands to natural vegetation were found in anthromes with higher population density. In contrast, more unchanged croplands can be found in village anthromes (60.5%) compared to residential anthromes (37.1%). The conversion of barren or sparsely vegetated area to natural vegetation that occurred in village, residential and populated anthromes accounted for 11.7%, 47.1% and 41.2% of the total conversion, respectively.

Table 2.4 Distribution of land cover changes in major anthromes within the area of positive vegetation change.

Anthromes	***Land cover change according to MODIS 12 Land Cover data (%) 2003 – 2005 compared to 2006 - 2008***					***Population density (persons/km^2)***
	Cropland to natural vegetation	**Barren to natural vegetation**	**Unchanged natural vegetation**	**Unchanged croplands**	**Unchanged barren area**	
Irrigated villages	4.1	11.7	3.2	11.8	1.2	≥100
Rainfed villages	29.7	N/A	12.7	47.5	N/A	≥100
Pastoral villages	3.4	N/A	4.9	1.2	N/A	≥100
Residential irrigated croplands	1.7	5.9	1.5	1.4	2.4	10-100
Residential rainfed croplands	49.3	5.9	28.5	31.3	N/A	10-100
Residential rangelands	7.6	35.3	31.9	4.4	15.7	10-100
Populated rangelands	N/A	41.2	10.5	N/A	47.0	1-10
Remote rangelands	N/A	N/A	N/A	N/A	27.7	0-1
All	95.8	100	93.2	97.6	94	

To summarize, anthromes with higher population densities revealed less cropland conversion than anthromes with lower population densities. The difference in the distribution of cropland conversion and unchanged cropland among anthromes can be explained by the concern for food security by policymakers of GTGP(Xu *et al.*, 2006). Declining land productivity resulting from severe soil erosion in the Loess Plateau endangers food security in this region(Chen *et al.*, 2007).The implementation of GTGP was predicted to result in a 10% to 14% loss in grain production with higher losses predicted for some provinces within the Loess Plateau (e.g. 14% to 17% for Shaanxi Province)(Feng *et al.*, 2005). In order to achieve the dual goal of vegetation conservation and minimizing grain production loss, GTGP was designed to focus more on marginal croplands with low productivity in remote and mountainous regions hosting lower population density (e.g. croplands in residential anthromes)(Uchida *et al.*, 2005; de Beurs *et al.*, 2013; Kelly & Huo, 2013). By examining the characteristics of 3397 parcels of land in Shaanxi Province, Kelley and Huo (Kelly & Huo, 2013) reported that land parcels with steeper slope, lower yield and greater distance from farmers' houses had a significantly higher probability of being enrolled in GTGP. Xu et al.(Xu *et al.*, 2006) reported that the yield of cropland plots which were transformed under GTGP was about 30% to 50% of the yield by non-GTGP cropland plots based on data acquired by a national survey of GTGP. The targeting strategies of GTGP reported by previous studies agree with what I found by examining the distribution of land cover changes over anthromes: GTGP left croplands with higher productivity in areas with higher population density (e.g. croplands in village anthromes) unchanged to offset the decrease in grain production due to loss of croplands elsewhere.

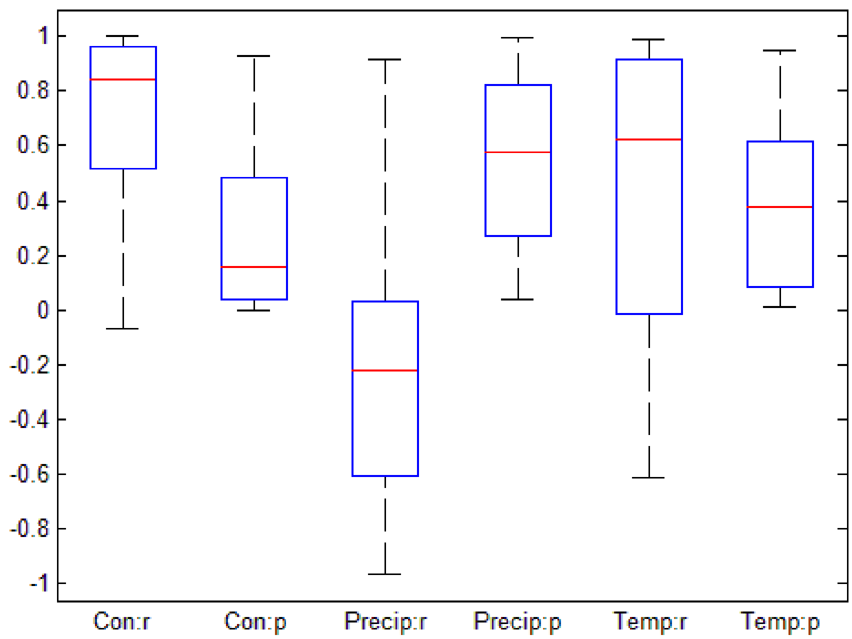
Conversion of sparsely vegetated areas occurred mainly in anthromes with higher population densities, e.g. residential and populated rangelands instead of remote rangelands. I suspect that since afforestation requires manual labor, a certain population density is necessary, and thus more re-vegetated barren lands were found in anthromes with higher population densities(de Beurs *et al.*, 2013).

2.4.4 Vegetation conversion Reported in Statistical Yearbooks

I found that for 68.7% of the area with significant vegetation changes, we could not explain the changes directly by MODIS land cover changes or weather changes. In order to identify conservation activities that may be missing in the MODIS land cover change analysis, I examined the statistical yearbook data for 20 counties located in northern Shaanxi Province and western Inner Mongolia Positive vegetation change within the 20 counties accounts for approximately 21% of the total positive vegetation changes within the Loess Plateau.

I found that accumulated vegetation conservation activities, mean annual temperature and annual precipitation were significantly correlated with maximum annual NDVI ($P < 0.05$) within five, four and two of the 20 selected counties, respectively. The distribution of the partial correlation coefficients and p-values for the three independent variables are shown in Figure 2.5.

Figure 2.5 Partial correlation analysis results using vegetation conservation data acquired from regional statistical yearbook. ‘Con:r’ and ’Con:p’ refer to the correlation coefficients and p values of the partial correlation between accumulated vegetation conservation and NDVI, with precipitation and temperature as control variables. ‘Precip:r’ and ’ Precip:p’ refer to the correlation coefficients and p values of the partial correlation between NDVI and precipitation, with conservation and temperature as control variables. ‘Temp:r’ and ’ Temp:p’ represent the correlation coefficients and p values of the partial correlation between NDVI and temperature, with annual precipitation and accumulated conservation as control variables



There was no significant weather change derived from the SMK trend analysis in 12 of the 20 selected counties. By employing the partial correlation analysis, however, I identified three counties (Yan’an, Yanchuan and Zhidan) where conservation activities, annual precipitation and mean annual temperature were all found to be significantly correlated with maximum annual NDVI ($P < 0.05$), when the other two variables were controlled. This demonstrated the importance of using multi-variable analysis in change attribution.

In the remaining eight counties where significant increases in precipitation was detected by SMK trend test, conservation activities were found to be significantly correlated with annual maximum NDVI (P < 0.05) in Shenmu and Zichang. In Zichang county, besides conservation activities, mean annual temperature was also identified to be significantly correlated with annual maximum NDVI.

Conservation program related vegetation increases within the selected counties have been identified by using 30m Landsat images or field measured data acquired at very fine scale in previous studies. Zhou et al.(Zhou *et al.*, 2012) investigated the effects of GTGP on land cover changes in Ansai county, Yan'an between 1995 and 2010 using multi-temporal Landsat TM/ETM+ images. They identified a 21.4% increase in newly forested area by 2010(Zhou *et al.*, 2012). The increases in newly forested area were attributed to conversions of cropland and shrub-grassland to forests by GTGP(Zhou *et al.*, 2012). Cao et al.(Cao *et al.*, 2009) investigated the effects of afforestation, cropland conversion and grazing prohibition by GTGP on vegetation dynamics in five counties within the regions of Yan'an and Yulin by sampling vegetation changes in 150 0.5ha-sized plots. They reported a 12.5% increase in the average vegetation cover of the five counties between 1998 and 2005 due to GTGP. The unexplained positive NDVI increases in natural vegetation area of the 21 selected counties account for 19.9% of the total area with NDVI increases in the Loess Plateau. I suspect that the vegetation increases within the selected 21 counties were due to vegetation conservation activities that occurred at scales finer than 500m, thus land cover changes resulting from these conservation activities were missed in the MODIS land cover change analysis.

2.4.5 The effect of GTGP on areas with steep slopes and sparsely vegetated areas

The goal of GTGP is to convert croplands on steep slopes to forest and grassland and afforest barren and sparsely vegetated areas. Steep slopes are defined as slopes greater or equal to 15° in northwestern China(Liu *et al.*, 2008) where the major part of the Loess Plateau is located. The land cover change analyses based on MODIS have confirmed a decrease in the area of cropland and an increase in the area of natural vegetation over the study period (Table 3). I found that 41.6% (8977.50 km^2) of the area that changed from cropland to natural vegetation was located in areas with slopes steeper than 15°. The spatially averaged slope on which croplands were converted to natural vegetation is 14.3° (Table 2.5), which is significantly steeper than the average slope of unchanged croplands (10.9°)($P<0.01$). This may indicate the possible effect of GTGP. The conversion from sparsely vegetated area to natural vegetation was mainly detected in the region of Ordos, which lies between the Hobq Desert and the Mu Us Desert. I speculate that the vegetation increases in this region were due to vegetation conservation activities by GTGP.

Table 2.5 Slope and elevation results

Land cover change type	*Slope (°)*	*STD*[a]	*Elevation (m)*	*STD*[b]
Cropland to Natural Vegetation	14.3†	8.7	1348.4*	408.7
Unchanged Cropland	10.9‡	8.3	1101.1*	380.8
Sparsely Vegetated to Natural Vegetation	3.5*	3.9	1231.7*	189.2
Unchanged Sparely Vegetated	3.6*	3.9	1202.2*	174.4

a: standard deviation of spatially averaged slope.

b: standard deviation of spatially averaged elevation.

†,‡: The mean values were significantly different from each other (P<0.01).

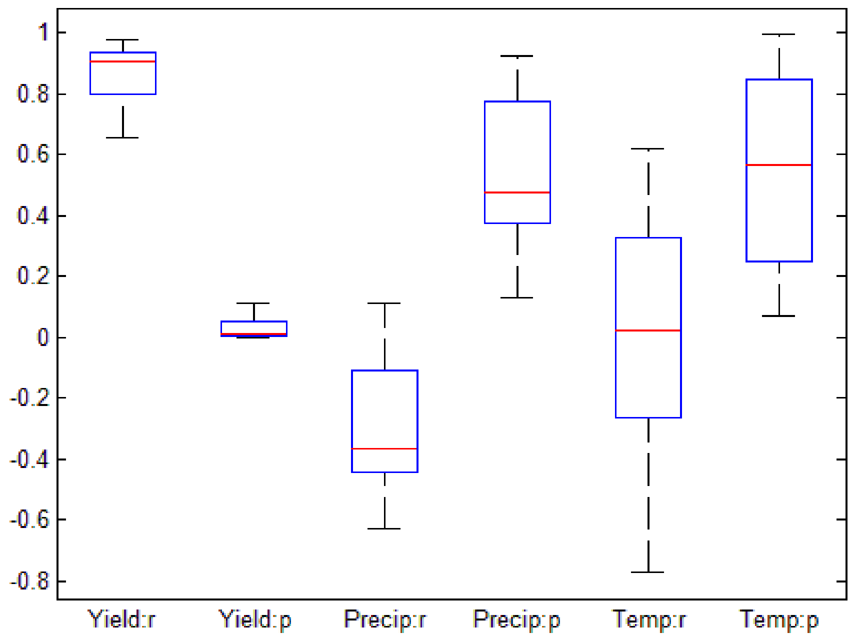
*: No significant differences were identified between the mean values.

2.4.6 The effect of conservation programs on unchanged croplands

The pixels identified as unchanged croplands that did not experience significant weather change but did reveal significant NDVI increases accounts for 12.2% of the total area with vegetation increases. I found significant positive relationships between sown area weighted crop yield and annual accumulated NDVI for these areas. Thus, it appears that crop yields are increasing in unchanged croplands in the Loess Plateau. Due to limited availability of crop yield data, I have not obtained crop yield data covering all the unchanged croplands area.

The results in figure 2.6 reveal a strong correlation between increased crop yields and increases in accumulated NDVI in seven of the selected nine counties ($P < 0.05$), when weather variables were controlled. Croplands with observed NDVI increases in the seven counties accounted for 2.1% of the positive vegetation change area. I did not identify any county where annual precipitation or mean temperature were significantly correlated with accumulated NDVI.

Figure 2.6 Partial correlation analysis using crop yield data acquired from regional statistical yearbook. 'Yield:r' and 'Yield:p' refer to the correlation coefficients and p values of the partial correlation between sown area weighted crop yield and accumulated NDVI, with precipitation and temperature as control variables. 'Precip:r' and 'Precip:p' refer to the correlation coefficients and p values of the partial correlation between NDVI and precipitation, with crop yield and temperature as control variables. 'Temp:r' and 'Temp:p' represent the correlation coefficients and p values of the partial correlation between NDVI and temperature, with annual precipitation and sown area weighted crop yield as control variables



As shown in tables 2.4 and 2.5, the spatially averaged slope in unchanged croplands was significantly lower than the slope of croplands converted to natural vegetation, and the majority of unchanged croplands (60.5%) are found in anthromes with high population densities (e.g. irrigated and rainfed villages). I suggest that these croplands were not participating in GTGP. They were more profitable and located in areas with suitable topography. Positive vegetation change in these croplands may be caused by significant increases in crop yield due to intensified agricultural production to offset economic loss caused by GTGP. Increases in crop yield were achieved through intensified agricultural production on non-GTGP croplands by means such as utilizing improved seeds and increasing crop diversity(Xu *et al.*, 2006). For example, Xu et al.(Xu *et al.*, 2006) found increases in crop yield in Delong county of Ningxia Hui Autonomous Region located in the western Loess Plateau from about 2516 kg/ha to about 4312 kg/ha after GTGP was initiated in this county. I did not find any articles or other relevant data about the contribution of fertilizer and technology to the increased crop yield in this region. However, it seems reasonable to speculate that increased utilization of fertilizer and advances in agricultural production technology were also potential causes of crop yield increases.

2.5 Conclusion

Change attribution is one of the most challenging parts of any change analysis based on satellite data. The observed NDVI increases that I was able to explain accounted for about 34.7% to 37.5% of the total changes observed in the NDVI time series analysis. I was able to attribute about 23.2% to 26.0% of the changes to precipitation and temperature increases. I was able to attribute 3.6% to 7.6% of the NDVI increases to large scale land cover changes caused by GTGP and NFCP programs. I also attributed 5.8% (1.7% + 1.8% + 2.3%) of the observed NDVI increases to the interacting effects of changes in weather variables and conservation programs. Based on the available county level crop yield data, I was able to identified 2.1% the vegetation increase was a result of crop production intensifications in unchanged cropland areas (Table 2.6).

Table 2.6 Change attribution table.

Total percentage of the Loess Plateau with positive vegetation change:	***48.2%***
Attribution:	
Significant increase in precipitation	19.8% - 22.2%
Significant increase in temperature	3.4% - 3.8%
Land cover change caused by GTGP or NFCP shown by MODIS	3.6% - 7.6%
GTGP and NFCP activities combined with changes in annual precipitation and mean annual temperature	1.7%
GTGP and NFCP activities combined with changes in mean annual temperature	1.8%
GTGP and NFCP activities combined with significant increase in precipitation	2.3%
Crop yield increases	2.1%
Total percentage of vegetation change area explained:	**34.7 - 37.5%***

* Overlapping effects of different causes were removed

Chapter 3 IDENTIFICATION OF C3 and C4 GRASSLAND COVER TYPES IN SOUTHWEST OKLAHOMA USING RANDOM FOREST CLASSIFICATION

Abstract

The objective of this chapter is to demonstrate a new methodology to differentiate C_3 and C_4 grassland cover types both at present and in the recent past using the 'Random Forest' classification algorithm. A total of 18 predictor variables were employed by the 'Random Forest' classification algorithm to map the occurrence of C_3 and C_4 grassland cover types in southwest Oklahoma in 1988, 2005, 2010 and 2013. The 18 predictor variables fell into three groups: spectral predictor variables derived from Landsat surface reflectance, topographic predictor variables derived from ASTER DEM, and soil edaphic variables extracted from the Oklahoma gSSURGO dataset. Training data for random forest classification in 2013 were obtained by field survey and manual delineation on aerial photos or Landsat imageries based on field knowledge. Training data for land cover classifications in 1988, 2005 and 2010 were acquired by identifying the 'core areas' obtained from the 2013 random forest classifications using FRAGSTATS. The Random Forest classification algorithm generated highly accurate classification results for C_3/C_4 cover discrimination in 2013 with an overall classification error of 0.8%. The idea of using patch 'core areas' obtained in present time to train classifier for historical land cover classification was proved to be effective with the overall classification error for 1988, 2005 and 2010 being 10.35%, 7.94% and 8.43%, respectively. Elevation,

soil pH and soil clay content were found to contribute more to C_3/C_4 cover discrimination than variables describing seasonal greenness development did in Southwest Oklahoma.

3.1 Introduction

Plant photosynthesis is the process of converting light energy to chemical energy that can be stored in carbohydrate molecules. Carbohydrate generated in plant photosynthesis is synthesized from carbon dioxide and water. Plant species can be classified into three functional forms: C_3, C_4 and Cassulacean Acid Metabolism (CAM), based on how carbon dioxide is treated before or in the process of photosynthesis. Carbon dioxide is converted into a three-carbon compound by the C_3 photosynthetic pathway, which is the most typical way of carbon fixation and adopted by approximately 95% of the plant species on the Earth (Bianchi & Canuel, 2011). C_4 plants fix carbon by incorporating carbon dioxide into a four-carbon compound. CAM plants fix carbon dioxide during the night and store it in an acid before carbon dioxide can be used in photosynthesis during daylight hours (Bianchi & Canuel, 2011). C_3 plants tend to occur in environments with moderate temperature and adequate water availability while C_4 and CAM plants are more adapted to high temperature and water stress conditions due to their better water use efficiency (Bianchi & Canuel, 2011; Wang *et al.*, 2013).

The relative abundance of C_3 and C_4 (C_3/C_4) species in grassland ecosystems is found to be an important indicator of natural or anthropogenic disturbances and an indispensable parameter in terms of modeling carbon cycling in grassland ecosystems at global scale (Goodin & Henebry, 1997; Winslow *et al.*, 2003). For instance, elevated atmospheric CO_2 is found to possibly result in a higher C_3/C_4 ratio due to the saturated productivity of C_4 species at the current CO_2 level (Wang *et al.*, 2013). Increased temperature and fire frequency has the potential to lead to a lower C_3/C_4 ratio by favoring the growth of C_4 species (Goodin & Henebry, 1997).

The dynamics of C_3/C_4 ratio over time can facilitate the understanding of those processes that may affect the dynamics such as changes in disturbance regime (Hanberry *et al.*, 2012). An effective way to examine the long term C_3/C_4 ratio dynamics is through remote sensing. A commonly adopted method to discriminate C_3 and C_4 grassland cover types on remotely sensed images is to capture the seasonal greenness asynchrony exhibited by C_3 and C_4 grassland cover types during the growing season(Goodin & Henebry, 1997; Tieszen *et al.*, 1997; Foody & Dash, 2007; Wang *et al.*, 2013). C_3 species begin growth in early spring and reach peak growth in late spring. These species may become senescent or semi-dormant in the summer to avoid hot temperatures and water stress conditions. Some C_3 species may resume growth in the fall (Wang *et al.*, 2013). Growth of C_4 species starts in late spring, reaches the peak during summer. Previous studies on detecting the seasonal greenness asynchrony primarily relied on examining the differences in the shapes of temporal trajectories of photosynthetic greenness using discriminant or clustering analyses (Goodin & Henebry, 1997; Tieszen *et al.*, 1997; Foody & Dash, 2007). In order to effectively detect differences in the greenness trajectory shape, temporally-rich canopy greenness time series were indispensable in these studies. As a result, field canopy greenness measurements (Goodin & Henebry, 1997), observations acquired by Advanced Very High Resolution Radiometer(AVHRR) (Tieszen *et al.*, 1997), and Medium Resolution Imaging Spectrometer (MERIS) (Foody & Dash,

2007) were used in previous studies. C_3 and C_4 grassland cover types in the Great Plains tend to occur in a mixed composition manner (Wang *et al.*, 2013). The utilization of C_3/C_4 ratio retrieved at the scale of AVHRR (1km) or MERIS (1200m) is ill-suitable to analyze the long term dynamics of C_3/C_4 ratio at fine scale due to the presence of mixed pixels. Observations acquired by the Landsat series of satellites, with a spatial resolution ranging from 30m to 60m in the visible to near infrared spectrum, are an ideal data source to examine long term C_3/C_4 ratio dynamics at fine scale. However, the revisit interval of the Landsat series of satellites is 16 days and can be longer if severe cloud contamination is present, which tends to result in a temporally sparse canopy greenness time series that often fails to meet the input requirements by discriminant or clustering analyses. Thus, there is a need for a methodology that will enable the analysis of long term C_3/C_4 ratio dynamics using fine scale observations obtained from satellites such as Landsat.

In order to compensate for the low temporal resolution of Landsat observations, in this chapter, I use a tree-based classification model and incorporate topographical and soil edaphic variables to aid in the discrimination of C_3 and C_4 grassland cover types both in the present time and recent past. There are two advantages of employing a tree-based prediction model in this study: 1) tree-based classification models do not require a temporally rich canopy greenness time series; 2) tree-based classification models are nonparametric and thus are valid even when predictor variables reveal different statistical distributions. A list of commonly encountered C_3 and C_4 species in the study area is shown in table 3.1. Descriptions of the species characteristics were retrieved from the Oklahoma Vascular Plant Database at: http://www.oklahomaplantdatabase.org/

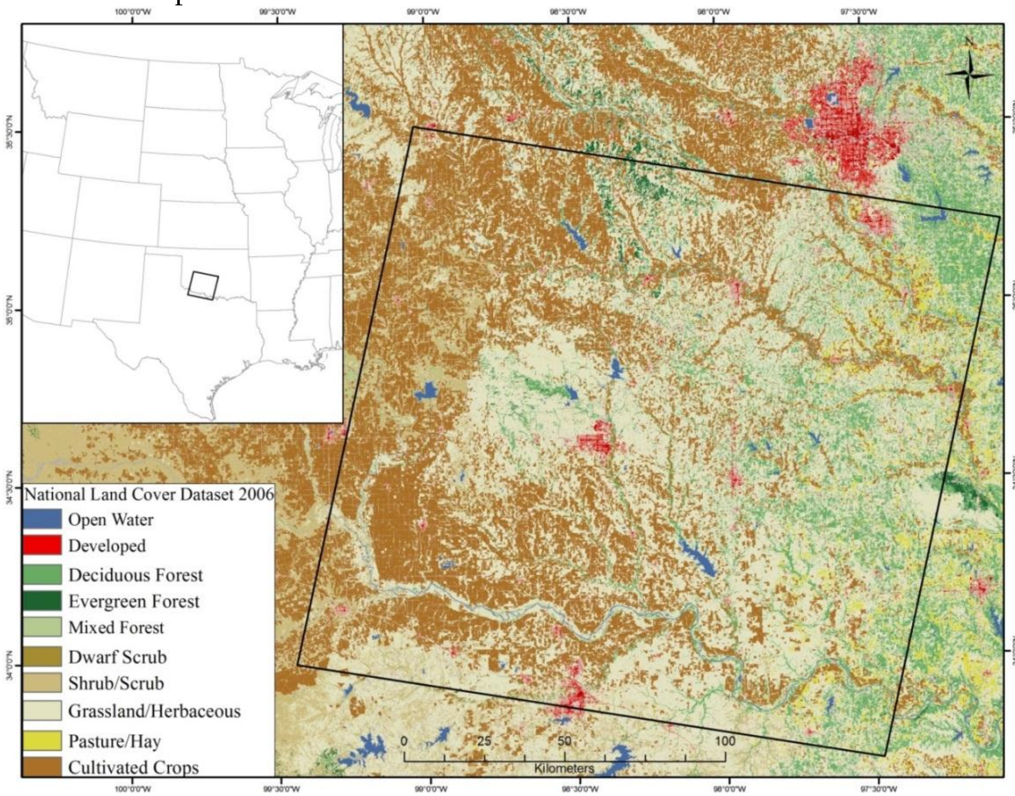
Table 3.1 Commonly encountered C_3 and C_4 species in Southwest Oklahoma

Species	Plant Functional Form	Duration	Origin	Growth habitat
Aegilops cylindrica Host (Jointed goatgrass)	C_3	Annual	Introduced	Graminoid
Andropogon gerardii Vitman (Big bluestem)	C_4	Perennial	Native	Graminoid
Bouteloua curtipendula (Michx.) Torr. (Sideoats grama)	C_4	Perennial	Native	Graminoid
Bouteloua dactyloides (Nutt.) J.T. Columbus. (Buffalograss)	C_4	Perennial	Native	Graminoid
Bromus arvensis L. (Japanese brome)	C_3	Annual	Introduced	Graminoid
Bromus secalinus L. (Rye brome)	C_3	Annual	Introduced	Graminoid
Bromus tectorum L. (Downy brome)	C_3	Annual	Introduced	Graminoid
Elymus virginicus L. (Virginia wildrye)	C_3	Annual	Native	Graminoid
Hordeum pusillum Nutt. (Little barley)	C_3	Annual	Native	Graminoid
Nassella leucotricha (Trin. & Rupr.) Pohl (Texas needlegrass)	C_3	Annual	Native	Graminoid
Schizachyrium scoparium (Michx.) Nash (Little bluestem)	C_4	Perennial	Native	Graminoid
Sporobolus compositus (Poir.) Merr. (Tall Dropseed)	C_4	Perennial	Native	Graminoid

3.2 Study area

In this chapter, I focused on the area covered by the Landsat scene of path 28 and row 36 in southwest Oklahoma (Figure 3.1). The dominant geomorphic province in the study area is red-bed plains, which is characterized by flat plains and gently rolling hills formed by Permian red shales and sandstones (Curtis *et al.*, 2008). Other geomorphic provinces such as sandstone and limestone hills, granite mountains and sand-dune belts are embedded within the landscape (Curtis *et al.*, 2008). Annual precipitation in the study area is primarily received in spring and autumn months and features a decreasing gradient from east to west (Johnson, 2008). Mean annual precipitation varies from approximately 1000mm in the east to about 800mm in the west (Johnson, 2008). Major types of potential natural vegetation in the study area are post oak-blackjack forest, mixedgrass and tallgrass prairie (Duck & Fletcher, 1945; Hoagland, 2008). Mean annual temperature decreases from 17°C in the south to about 15°C in the north (Johnson, 2008). Mixedgrass prairie mainly occurs in the western portion of the study area as influenced by the decreasing average annual precipitation from east to west. Much of the mixedgrass prairie has been converted to cultivated crops of wheat and cotton as shown by the 2006 National Land Cover Database in figure 3.1(Hoagland, 2008; Xian *et al.*, 2009).

Figure 3.1 An overview of land cover types in the study area. Land cover types shown in the map are based on the 2006 National Land Cover Database.



3.3 Data

3.3.1 National Land Cover Database

In this study, the classification was only conducted in areas with the land cover of shrublands, grasslands or pasture/hay fields according to land cover types identified by the National Land Cover Database (NLCD). The NLCD datasets were generated across the conterminous United States with a spatial resolution of 30m based on Landsat Thematic Mapper (TM) or Landsat Enhanced Thematic Mapper Plus (ETM+) images. Shrublands, grasslands and pasture/hay fields are defined as follows within the National Land Cover Database(Xian *et al.*, 2009):

- Shrublands: "*Areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.*"
- Grasslands: "*Areas dominated by gramanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.*"
- Pasture/Hay: "*Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation.*"

Land cover classes other than these classes were excluded from this analysis using a land cover mask derived from NLCD for each targeted year. The 1992 NLCD dataset was used to generate the land cover mask for 1988. The land cover mask derived from the 2006 NLCD dataset was applied to the targeted years of 2005, 2010 and 2013.

3.3.2 Landsat surface reflectance data

Landsat surface reflectance data were acquired from two different sources. Surface reflectance for the new Landsat 8 Operational Land Imager (OLI) (30m) were retrieved by applying the atmospheric correction algorithm of Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) in ENVI to Landsat OLI images. Surface reflectance for Landsat Thematic Mapper (TM) (30m) and

Enhanced Thematic Mapper Plus (ETM+) (30m) were acquired from the Landsat Surface Reflectance Climate Data Records (CDRs), which are atmospherically corrected by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS)(Masek *et al.*, 2006). The Landsat Surface Reflectance CDRs were downloaded from the USGS EarthExplorer at http://earthexplorer.usgs.gov/. For each targeted year, Landsat images acquired during early spring, late spring and early summer were downloaded. This is to capture the phenological asynchronicity between croplands, and C_3 and C_4 grassland cover types. The actual images included in the analysis were carefully chosen to avoid images with high cloud cover. A complete list of acquisition dates and sensors for Landsat images included in the analysis is shown in table 3.2.

Table 3.2 A list of Landsat images used in Random Forest classifications

Year	Early spring	Late spring	Early summer
1988	March 26th / TM	May 13th / TM	June 30th / TM
2005	March 17th / ETM+	April 26th / TM	June 21st / ETM+
2010	March 31st / ETM+	May 2nd / ETM+	June 19th / ETM+
2013	March 7th / ETM+	May 18th / OLI	June 27th / ETM+

Two sets of spectral predictor variables were used in the detection algorithm for C_3 and C_4 cover types. Typically spectral reflectance data from different wavelengths reveal very high correlations. The principal component analysis is able to generate a new feature space, in which the correlations between different wavelengths are eliminated and the variance of the original data is partitioned along a number of mutually orthogonal axes (Richards & Jia, 2006). The larger variance a principal component carries, the more it will contribute to increase the separability between different target classes (Richards & Jia, 2006). Thus, the principal component analysis can help to reduce noises in predictor variables. To reduce the correlation between the different bands and to reduce the number of predictors, the first set of spectral variables was generated by applying the principal component transformation to the surface reflectance acquired by the six multi-spectral bands of Landsat images.

The first two principal components typically accounted for more than 90% of the total variance of the original six bands. Thus, I chose to only use these first two principal components for each date in the Random Forest classification.

The second set of spectral predictor variables was used to capture the differences in the phenological development of different land cover classes from early spring to early summer. An effective way to track the phenological development of plant canopies is by examining the changes in photosynthetic greenness at canopy level. Spectral vegetation indices are commonly used tools to quantify the level of photosynthetic greenness of plant canopies (Tucker, 1979). Soil backgrounds were found to be a significant source of noise in measuring photosynthetic greenness of plant canopies using vegetation indices(Huete, 1988). Since the effects of soil backgrounds are more pronounced in grassland areas, I employed the Soil Adjusted Vegetation Index (SAVI) to quantify the photosynthetic greenness of grassland canopy. SAVI can be calculated as (NIR - RED)*(1 + L) / (NIR + RED + L), where NIR is the surface reflectance of the near infrared band

(760 – 900 nm), RED is the surface reflectance of the red band (630 – 690 nm) and L is a correction factor with a recommended value of 0.5(Huete, 1988). In this study, four phenological metrics were calculated as follows:

- dSAVI1: $SAVI_{Early\ spring}$ - $SAVI_{Late\ spring}$
- dSAVI2: $SAVI_{Late\ spring}$– $SAVI_{Early\ summer}$
- dSAVI3: $SAVI_{Early\ spring}$ - $SAVI_{Early\ summer}$
- $SAVI_{sum}$: The sum of SAVI over the three dates

3.3.3 Digital elevation data

A previous study, which was situated in the upland prairie of eastern Nebraska, investigated if grassland community patterns were significantly affected by topographical positions (Schacht et al., 2000). They found that aspect was a significant variable in terms of predicting the occurrence of C_3 and C_4 grassland cover types (Schacht *et al.*, 2000). In this study, elevation, slope and aspect were used as topographical predictors in the classification.

Elevation information were acquired from the global Digital Elevation Model (DEM) generated using observations by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) with a spatial resolution of 1 arc-second (approximately 30m) (Tachikawa *et al.*, 2011). The ASTER Global DEM is available as 1º by 1º tiles. A total of nine ASTER DEM tiles were downloaded for the study area. Downloaded ASTER DEM tiles were first mosaicked and re-projected to the UTM projection to be matched with Landsat images. Slope and aspect were calculated based on the ASTER DEM in ENVI. Slope was calculated as percent grade in ENVI with the formula of 100 × Rise / Run. Aspect was calculated as the clockwise increase of angle from the north.

3.3.4 Soil edaphic quality data

Previous studies on the impacts imposed by soil edaphic factors on cool season grass seed germinations and herbicide efficacy show that soil properties such as soil texture and soil organic matter content can affect the growth of C_3 species significantly (Wicks *et al.*, 1971; Blackshaw *et al.*, 1994). For example, given adequate moisture, seed germination rates were found to be higher in coarse-textured soils than that in fine-textured soils (Wicks *et al.*, 1971). The efficacy of herbicides used to control the growth of C_3 species was found to be negatively correlated with soil organic matter content (Blackshaw *et al.*, 1994).

To improve the detection of the occurrence of C_3 and C_4 grassland cover types, soil edaphic variables were extracted from the Gridded Soil Survey Geographic (gSSURGO) Database for Oklahoma (Soil Survey Staff, 2012) and the National Value Added Look Up (valu1) Table database (Soil Survey Staff, 2012). The gSSURGO dataset is derived from the standard Soil Survey Database created by United States Department of Agriculture, Natural Resources Conservation Service. The Oklahoma gSSURGO dataset is available as a mosaicked ESRI GRID file covering the entire state of Oklahoma with a spatial resolution of 10m. I used a total of five soil edaphic variables as predictor variables: total sand, total clay, total silt, soil organic matter content and soil pH. The soil edaphic predictor variables were extracted based on the relationship between the gSSURGO dataset and the National

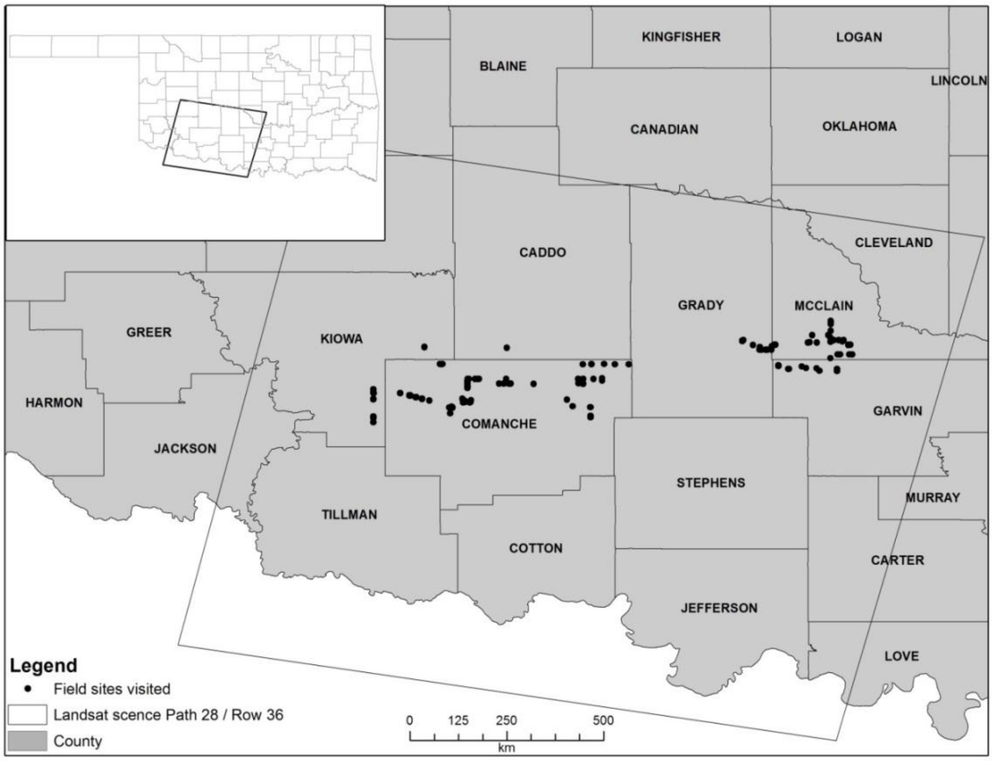
Value Added Look Up Table database. Two steps were taken to extract soil edaphic predictor variables. First, for a soil edaphic predictor variable ***m*** in a soil map unit with the ID of ***k*** in the National Value Look Up Table database, the average value of ***m*** within ***k*** was obtained by averaging the value of ***m*** over all the sub-components within ***k***. Second, I assigned the average value of variable ***m*** to the gSSURGO grid cells with the same map unit ID as that of ***k***. I repeated the above two steps for the other soil predictor variables.

3.4 Methods

3.4.1 Training data collection

Field surveys were carried out in May and June 2012 and from April to September in 2013 to collect geo-referenced occurrence data for C_3 cover types. Occurrence data for C_3 cover types were mainly made up of C_3 species commonly seen in the study area such as: Downy/Japanese/Rye brome, Jointed goatgrass and Texas needle grass. Field sites visited during the 2012 and 2013 field surveys are shown in Figure 3.2. C_3 cover type occurrence data were collected on both publicly accessible areas and private properties following different procedures. Within publicly accessible areas, C_3 patches were delineated by walking along C_3 patch edges with a Garmin eTrex Venture® HC GPS unit, which has a positioning accuracy varying between 3 to 5 meters, depending on the availability of satellites. C_3 cover occurrence data on private properties were collected in two steps: 1) Determine if C_3 cover type occupied a dominant canopy cover within the property by visual estimation from outside of the property. Record the coordinates of private properties where C_3 cover type occupied a dominant canopy cover using the Garmin eTrex Venture® HC GPS unit. 2) Within the marked properties, delineate a pixel as covered by C_3 cover type if the value of dSAVI2 for this pixel is greater than the median of the dSAVI2 values exhibited by the C_3 patches collected in public accessible areas. A total of 1236 Landsat pixels were collected as training data for C_3 cover type within public accessible and private properties.

Figure 3.2 The field sites visited in 2012 and 2013 field surveys.



Training data for C_4 cover types were collected based on the computed phenological metrics. A Landsat pixel will be delineated as a C_4 cover types if: 1) the pixel is classified as a land cover type of shrublands, grasslands or pasture/hay field by NLCD; 2) the pixel exhibited a negative dSAVI1 and dSAVI2. In other words, the pixel exhibited continuous increases in photosynthetic greenness from early spring to early summer. A total of 2050 Landsat pixels were delineated as covered by the C_4 cover types.

Although the NLCD land cover mask had been applied to exclude barren areas, forests and croplands, some areas with the land cover of winter wheat, rock outcrop and sparse vegetation cover were misclassified as areas with herbaceous cover. Training data for rock outcrop and sparsely vegetated area were collected by manual delineation on aerial photographs acquired by the National Agriculture Imagery Program (NAIP) in May, 2010. The delineated area of rock outcrop and sparsely vegetated area equaled to 1368 Landsat pixels. Most of the remaining croplands in the study area were winter wheat fields. Based on the typical winter wheat production calendar in the study area, winter wheat is expected to experience a growth stage change from possibly the stage of jointing or first node of stem visible on the early spring image to the stage of maturation on the late spring image(Paulsen, 1997). The growth stage transition corresponds to a positive dSAVI1, which indicates a decrease in photosynthetic greenness from early spring to late spring. In this study, a Landsat pixel was delineated as a cropland pixel with winter wheat cover if: the pixel exhibited a positive dSAVI1 and the pixel was on the edge of a cropland patch identified by the NLCD dataset. A total of 1733 Landsat pixels was collected as training pixels for croplands.

3.4.2 Random Forest Classification

I used the Random Forest classification algorithm implemented in the 'randomForest' package in R (Liaw & Wiener, 2002) to discriminate C_3 and C_4 grassland cover types at both present time and in the recent past. A total of 18 predictor variables were generated for each targeted year (Table 3.3). The predictor variables fell into three groups: spectral predictor variables derived from Landsat surface reflectance, topographic predictor variables derived from ASTER DEM, and soil edaphic variables extracted from the Oklahoma gSSURGO dataset. The targeted land cover classes are C_3 cover type, C_4 cover type, croplands and rock outcrops and sparsely vegetated area (RSV).

Table 3.3 A list of predictor variables used in Random Forest classifications

	Variable name	**Unit**	**Spatial resolution**
Spectral variables	The first two principal components of the early spring Landsat image	N/A	30m
	The first two principal components of the late spring Landsat image	N/A	30m
	The first two principal components of the early summer Landsat image	N/A	30m
	dSAVI1	N/A	30m
	dSAVI2	N/A	30m
	dSAVI3	N/A	30m
	$SAVI_{sum}$	N/A	30m
Topographical variables	Elevation	Meter	~30m
	Slope	Percentage	~30m
	Aspect	Degree	~30m
Soil edaphic properties	Total sand	Percentage	10m
	Total silt	Percentage	10m
	Total clay	Percentage	10m
	Organic matter	Percentage	10m
	pH	N/A	10m

Bias and variance represent the two most important error sources of a predictive model. Tree-based classification models are low-bias but high-variance procedures (Hastie *et al.*, 2009). **B**ootstrap **ag**gregation or **Bag**ging is an ensemble learning approach, developed to reduce the variance in statistical classification or regression by averaging the outputs from an ensemble of classification or regression models(Breiman, 1996). The bagging of classification trees can be implemented in three steps (Breiman, 1996; Liaw & Wiener, 2002; Hastie *et al.*, 2009):

- Create *B* bootstrap replicates of the original training data. (Suppose there are ***N*** cases in the original training data, a bootstrap replicate of the original training data can be created by sampling with replacements for ***N*** times from the training data. Only one case will be drawn from the original data each time. In other words, a bootstrap replicate is of the same size as the original data but the bootstrap replicate can have duplicate cases within it)
- Construct a committee of *B* classification trees with each one grown as an unpruned classification tree from a bootstrap replicate. Each tree generated in this process is identically distributed. Each node split in any of the *B* classification trees is determined by selecting the best splitting variable among all the predictors associated with the training data.
- Classify each new sample as a class that has the majority of votes from the tree committee grown in the previous step.

Since each tree is identically distributed, the bias of the bagged tree committee equals to the bias of any individual tree within the committee (Hastie *et* *al.*, 2009). The variance of the bagged tree committee is, however, significantly reduced by averaging the outputs from all the individual trees(Hastie *et al.*, 2009).

The idea of Random Forest (RF) classification is essentially the same as that of bagging. The only difference between RF classification and classifications generated by bagged classification trees is that: when making the tree node split, instead of making the split decision based on the best variable selected from all the available predictors, make the split based on the best variable selected from a random subset of the available predictors (Breiman, 2001; Liaw & Wiener, 2002).

To effectively reduce the variance of the Random Forest classifier, 2000 classifications were used in the Random Forest classification in each targeted year. The default values were used for all the other parameters of a Random Forest classifier.

3.4.3 Random Forest classification accuracy assessment

The Random Forest classification algorithm is designed in such a way that a multiple-fold cross validation can be conducted with the Random Forest Classifier being trained at the same time, and there is no need to run a separate validation (Breiman, 2001; Liaw & Wiener, 2002). Accuracy assessment for Random Forest classification can be obtained by examining the so called 'Out-of-bag' (OOB) predictions. OOB predictions are Random Forest predictions for OOB samples. Suppose there are a total of *B* classification trees grown from *B* bootstrap replicates of the training data in a 'Random Forest'. On average, a sample ***S*** in the training dataset will not be present in approximately 36% of the *B* bootstrap replicates(Liaw

& Wiener, 2002). In other words, 36% of the *B* classification trees in the forest were grown independently from ***S***. The average prediction, made by this subset of classification trees, for ***S*** is called an OOB prediction.

In this study, the training dataset is made up of four targeted land cover classes with X_i ($i=C_3$, C_4, *Croplands*, *RSV*) samples for each class. Take the class of C_3 cover type as an example, the OOB error rate can be calculated as OOB_{C3} / X_{C3}, where OOB_{C3} is the total number of times that a sample from the class of C_3 was not correctly classified as C_3 based on OOB predictions. The 'OOB estimate of error rate' (*i.e.* the overall classification error for all the land cover classes) can be obtained by repeating the previous process for the remaining targeted land cover class and average the individual OOB error rate(Liaw & Wiener, 2002). Provided that enough classification trees had been grown, the 'OOB estimate of error rate' is an unbiased measurement of the overall classification accuracy(Breiman, 2001; Liaw & Wiener, 2002).

3.4.4 Evaluation of predictor importance

For any predictor variable (e.g. aspect, slope)**,** its importance in Random Forest classification can be evaluated using an index named 'Decrease in classification accuracy' (Breiman, 2001; Liaw & Wiener, 2002). The importance index can be calculated as the decrease in the classification accuracy if the variable is removed from the classification. The larger the decrease, the more important the variable was in the classification. The importance index for each of the 18 predictor variables was calculated automatically by the R 'randomForest' package.

3.4.5 Discriminating C_3 and C_4 grassland cover types in the recent past

Discriminating C_3 and C_4 grassland cover types in the recent past was based on the classification results obtained in 2013. The training dataset for historical image classification was created by identifying these land cover patches whose land cover types were of high probability to be stable over time. Two assumptions were made to identify stable land cover patches

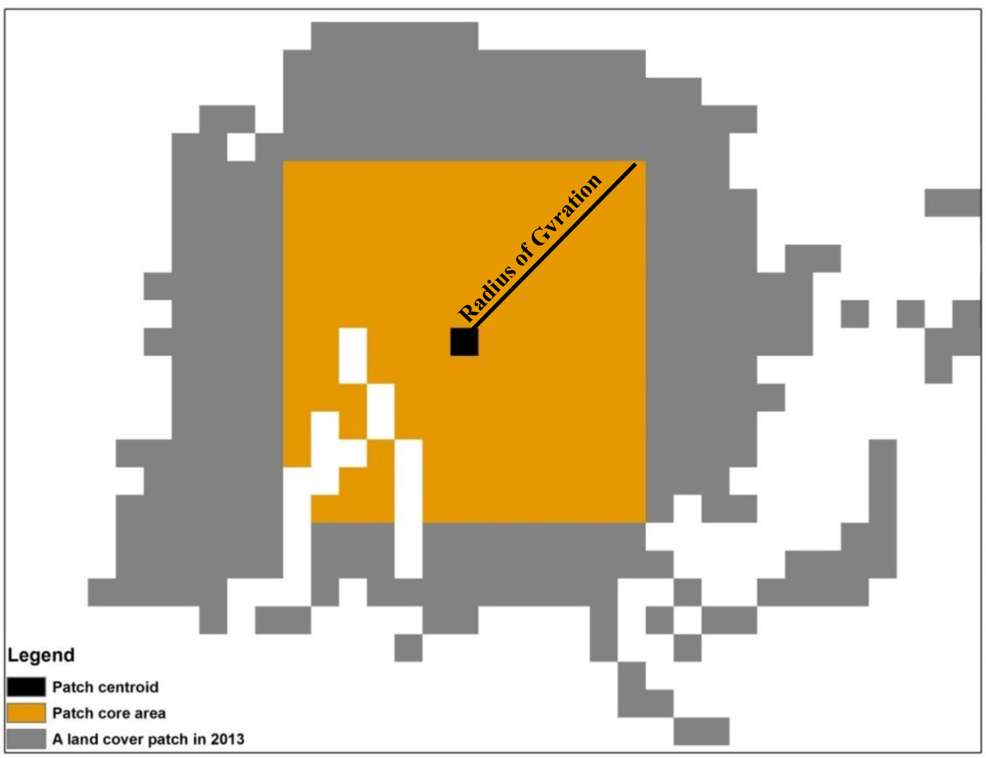
- The land cover types of patches with large sizes are of higher probability to be stable over time
- The areas that are closer to the center of a land cover patch is of higher probability to be stable over time

In this study, the stable area was referred to as the 'core area' of a patch generated in the 2013 Random Forest classification. Two steps were taken to identify core areas for each land cover type obtained from the 2013 classification:

- First, statistics of patch size, patch radius of gyration and the patch centroids were obtained for each patch in each land cover class of the 2013 classification using FRAGSTATS 4.2. FRAGSTATS is a software package for spatial pattern analysis on maps showing categorical or continuous phenomenon(McGarigal, 2013). A patch centroid is defined as the average location of all the cells in the patch. Patch radius of gyration is defined as the mean distance between the centroids of each cell in the patch and the patch centroid (McGarigal, 2013).

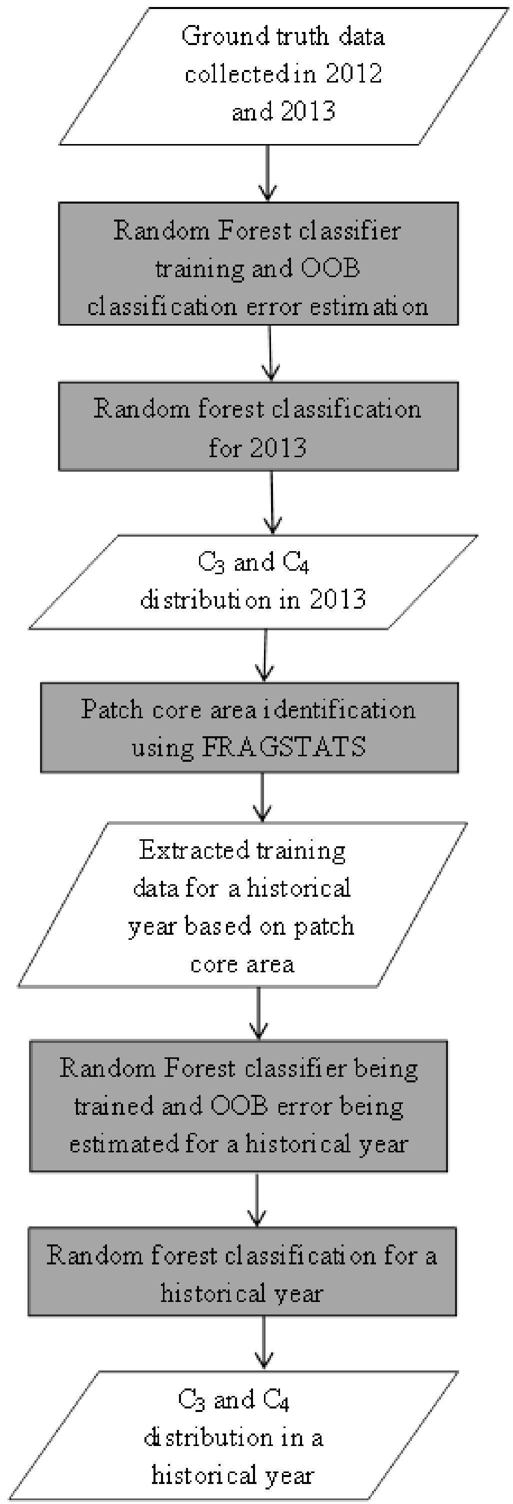
- Second, the cells whose centroids were within a distance that was less than or equal to the radius of gyration from the patch centroids were defined as the core area of a patch. The core area of a patch was delineated as a square area whose center is identical to the patch centroid and the length of a diagonal of the square area is two times as long as the radius of gyration of the patch (Figure 3.3). The core area was delineated as a rectangle instead of a circle because it is easier to implement automatic delineation of rectangular areas that of circular areas.

Figure 3.3 An example of identification of patch core areas based on patch radius of gyrations



For each land cover type classified in 2013, the core area in the patches with a patch size in the top one percentile of all the patches of the same land cover type were used to extract training data for 1988, 2005 and 2010. To examine if using the core areas from smaller patches to extract training data will result in a higher classification error rate, using the year of 2010 as an example, I compared the classification error rates generated by the training data extracted from the top 1% and the top 20% of all the patches obtained in the 2013 classification. A schematic flowchart demonstrating the steps used to identify present and historical C_3 and C_4 grassland cover types is shown in Figure 3.4

Figure 3.4 A flowchart demonstrating the steps took to discriminate C_3 and C_4 grassland cover types in the recent past



3.5 Results

3.5.1 Random Forest classification for the year of 2013

Based on the training data collected for 2013, the OOB estimate of error rate for 2013 is 0.64%. The confusion matrix for 2013 Random Forest classification is shown in table 3.4. The trained Random Forest classifier was then applied to the entire study area and the classification result for 2013 is shown in figure 3.5. The classification results show that, in 2013, the occurrence of C_3 cover type was mainly in the counties located in western portion of the study area such as: Caddo, Comanche, Kiowa and Tillman. In contrast, most of the C_4 cover type occurrence was found in counties located in the eastern portion of the study area such as: Carter, Garvin, Jefferson and Stephens. Rock outcrop and sparsely vegetated area were mainly found in the Blue Canyon Wind Farm and the Wichita Mountains Wildlife Refuge in the northwestern corner of Comanche County. Most of the remaining croplands were found in the southern half of Comanche County, the east half of Tillman County and the west half of Cotton County. The remaining croplands were often found on the edge of fields that were identified as croplands by the NLCD dataset. The variable importance as indicated by the mean decrease in classification accuracy is shown in Figure 3.6. Removing dSAVI1, dSAVI2 and dSAVI3 from the Random Forest classification in 2013 resulted in the highest decreases in classification accuracy. Thus, the three phenological predictors are the three most important predictor variables. In contrast, the exclusion of the predictor variables such as the soil organic matter content, slope and aspect caused relatively low decreases in classification accuracy, which means these three predictor variables were the three least important predictor variables in the 2013 Random Forest classification.

Table 3.4 The confusion matrix for 2013 Random Forest classification

		Random Forest Prediction				
		C_3	Croplands	C_4	RSV	OOB estimate of error
Ground observations	C_3	1222	6	0	8	1.13%
	Croplands	1	2049	0	0	0.04%
	C_4	0	0	1733	0	0.00%
	RSV	36	0	0	1332	2.63%
	Overall					0.80%

Figure 3.5 Discrimination of C_3 and C_4 grassland cover types using Random Forest classification in 2013.

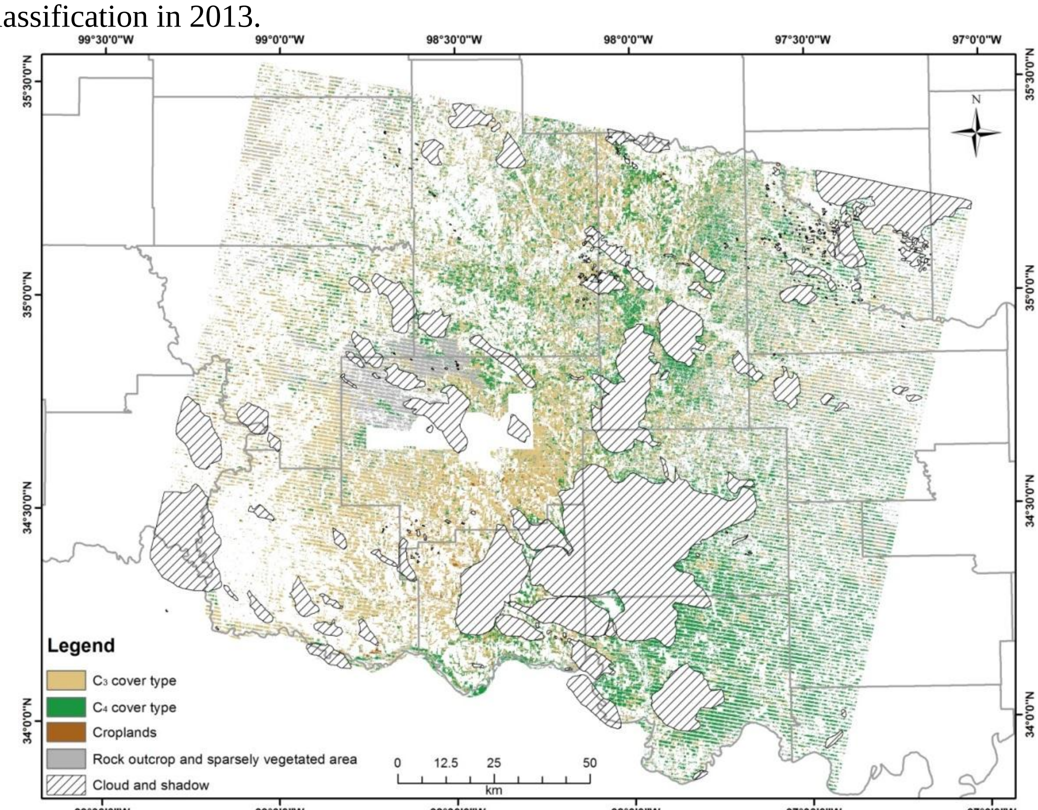
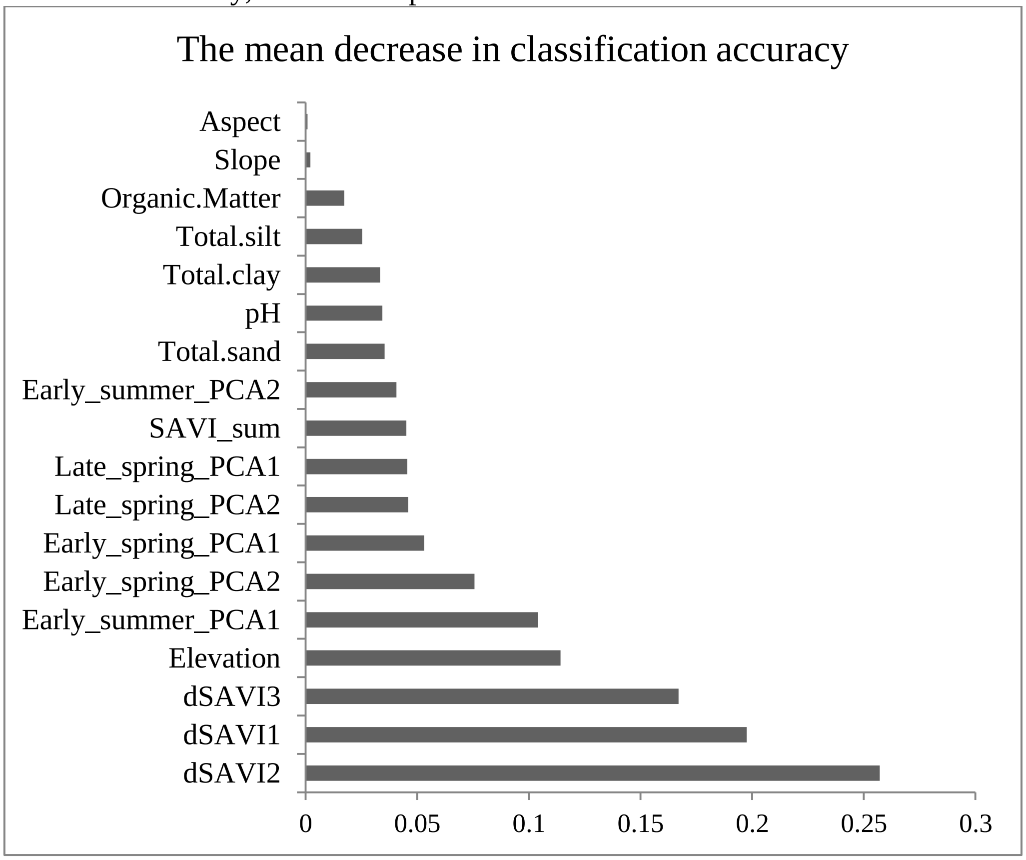


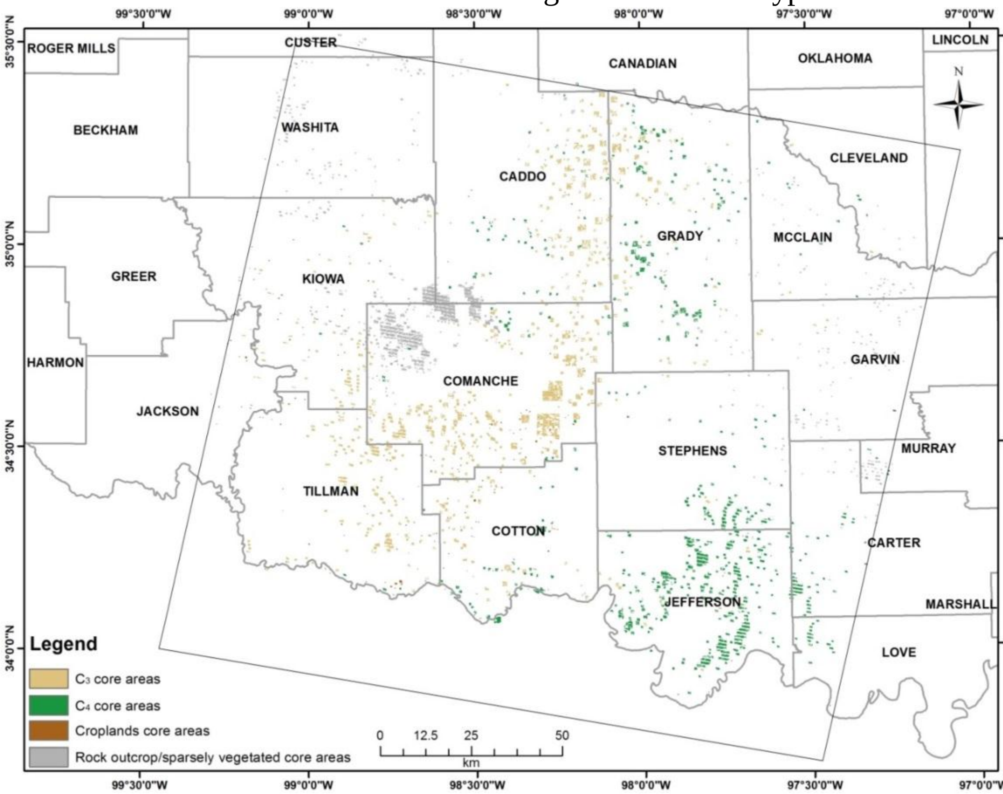
Figure 3.6 Evaluation of variable importance in 2013 Random Forest classification based on the mean decrease in classification accuracy. The larger the decrease in classification accuracy, the more important the variable will be in the classification.



3.5.2 Discrimination of C_3 and C_4 cover types in 1988, 2005 and 2010.

The ‘core areas’ extracted from the top 1% land cover patches obtained from the 2013 Random Forest classification are shown in Figure 3.7. The ‘core areas’ for the C_3 cover type and the rock outcrop and sparsely vegetated area were mainly located in Comanche County, Kiowa County and Tillman county. The ‘core areas’ for the C_4 cover type clustered in the southern Stephens County and across the Jefferson County. ‘Core areas’ for croplands spread across the southern Comanche County, the western Tillman County and the eastern Cotton County.

Figure 3.7 The 'core areas' extracted from the top 1% of the patches obtained in 2013 random forest classification for each targeted land cover type



The confusion matrices for 1988, 2005 and 2010 Random Forest classifications, derived from the top 1% land cover patches of 2013 are shown in table 3.5, 3.6 and 3.7, respectively. The overall OOB error rate ranged from 7.94% in 2005 to 10.35% in 1988. The C_3 cover type had the highest OOB error rate in all the three years. The OOB error rate of the C_3 cover type varied between 10.75% in 2005 and 13.43% in 1988. Croplands had the lowest OOB error rate of 4.49% and 5.15% in 2005 and 2010, respectively. The rock outcrop/sparsely vegetated area had the lowest OOB error rate of 7.39% in 1988.

Table 3.5 The confusion matrix for 1988 Random Forest classification based on training data extracted from the top 1% land cover patches obtained in 2013

		Random Forest Prediction				
		C_3	Croplands	C_4	RSV	OOB estimate of error
Ground observations	C_3	7513	449	628	89	13.43%
	Croplands	393	5084	45	10	8.10%
	C_4	801	83	7618	118	11.62%
	RSV	301	12	321	7950	7.39%
	Overall					10.35%

Table 3.6 The confusion matrix for 2005 Random Forest classification based on training data extracted from the top 1% land cover patches obtained in 2013

		Random Forest Prediction				
		C_3	Croplands	C_4	RSV	OOB estimate of error
Ground observations	C_3	11820	437	847	139	10.75%
	Croplands	363	8099	22	5	4.59%
	C_4	980	100	11782	160	9.52%
	RSV	355	19	346	12074	5.63%
	Overall					7.94%

Table 3.7 The confusion matrix for 2010 Random Forest classification based on training data extracted from the top 1% land cover patches obtained in 2013

		Random Forest Prediction				
		C_3	Croplands	C_4	RSV	OOB estimate of error
Ground observations	C_3	9021	344	771	78	11.68%
	Croplands	280	6226	54	4	5.15%
	C_4	830	74	8970	83	9.91%
	RSV	258	8	284	9099	5.70%
	Overall					8.43%

The confusion matrix for 2010 Random Forest classification, derived from the top 20% land cover patches of 2013 is shown in table 3.8. The overall OOB error rate was 25.24%. The C_4 cover type had the highest OOB error rate of 35.61% with croplands having the lowest OOB error rate of 10.07%.

Table 3.8 The confusion matrix for 2010 Random Forest classification based on training data extracted from the top 20% land cover patches obtained in 2013

		Random Forest Prediction				
		C_3	Croplands	C_4	RSV	OOB estimate of error
Ground observations	C_3	17148	1879	4337	1204	30.20%
	Croplands	1505	17873	432	64	10.07%
	C_4	5144	935	15656	2581	35.61%
	RSV	2128	335	2797	18475	22.16%
	Overall					25.24%

Random Forest classification results for 1988, 2005 and 2010 are shown in figure 3.8, 3.9 and 3.10, respectively. Random forest classifiers used in the land cover classifications of the three years were all trained based on the top 1% land cover patches of 2013. The spatial distribution of different land cover types in 1988, 2005 and 2010 were generally similar to that in 2013.

Results of the variable importance evaluations for 1988, 2005 and 2010 are shown in figure 3.11, 3.12 and 3.13, respectively. Aspect and slope were the least important predictor variables in Random Forest classifications in all the three years of 1988, 2005 and 2010. Elevation, soil pH and the total clay content were the three most important predictor variables in 1988 and 2005. Elevation, dSAVI3 and the total clay content were the three most important predictor variables in 2010.

Figure 3.8 Discrimination of C_3 and C_4 grassland cover types using Random Forest classification in 1988.

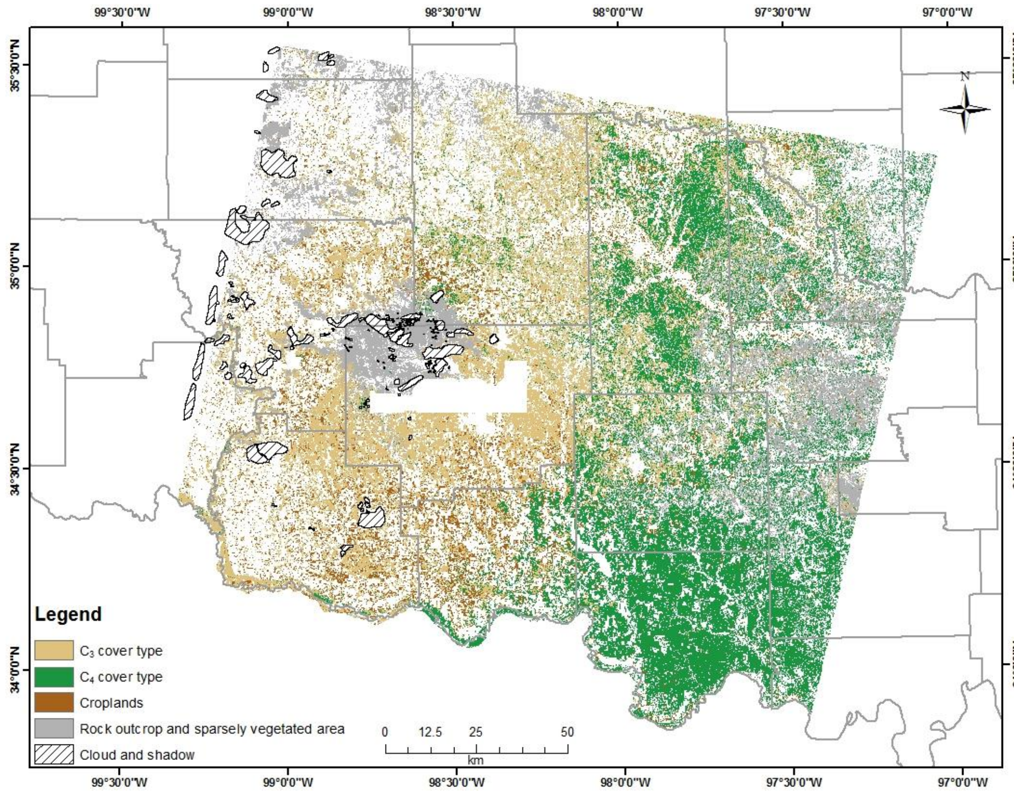


Figure 3.9 Discrimination of C_3 and C_4 grassland cover types using Random Forest classification in 2005.

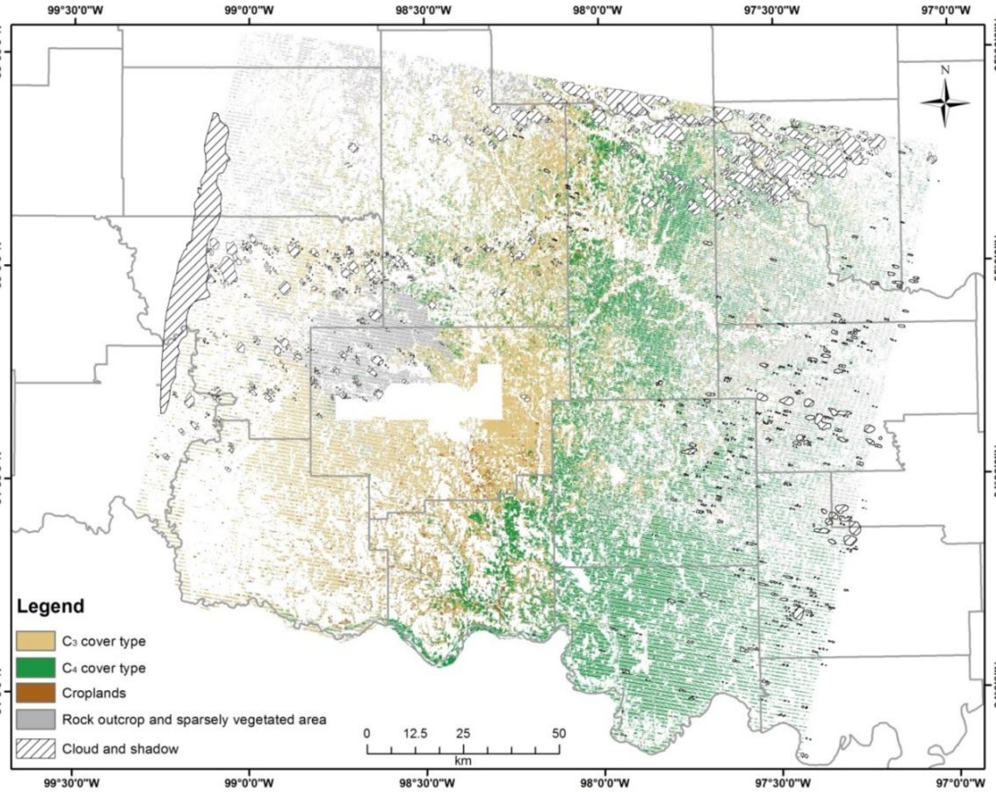


Figure 3.10 Discrimination of C_3 and C_4 grassland cover types using Random Forest classification in 2010.

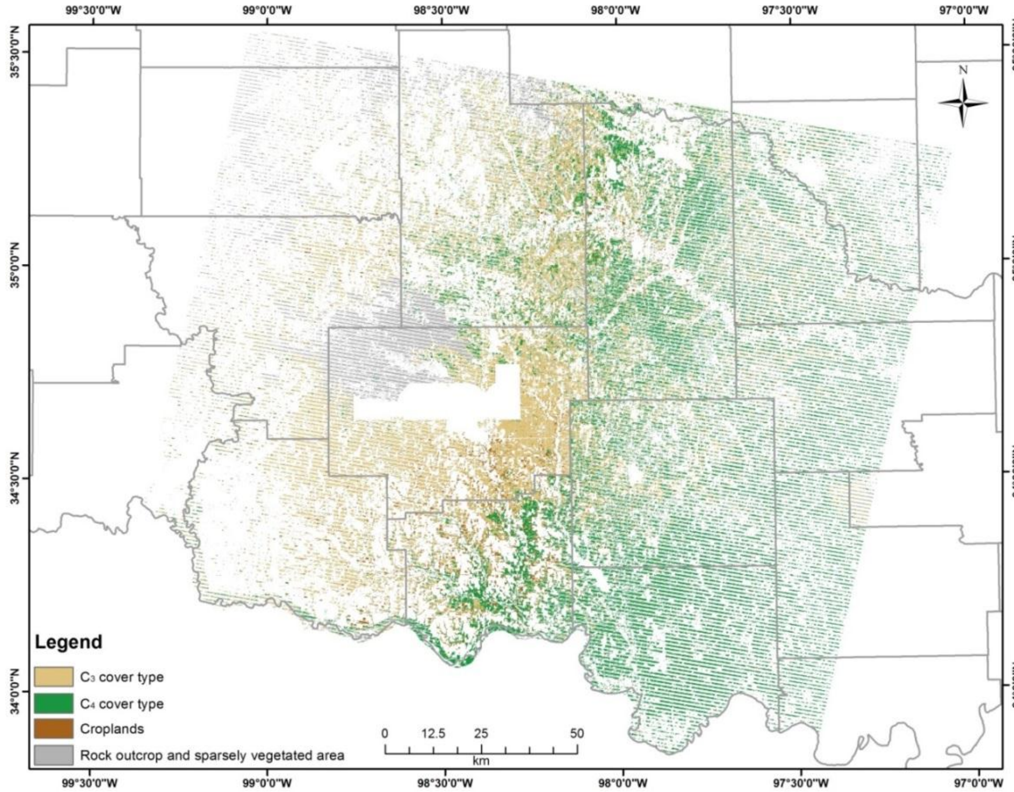


Figure 3.11 Evaluation of variable importance in 1988 Random Forest classification based on the mean decrease in classification accuracy. The larger the decrease in classification accuracy, the more important the variable will be in the classification.

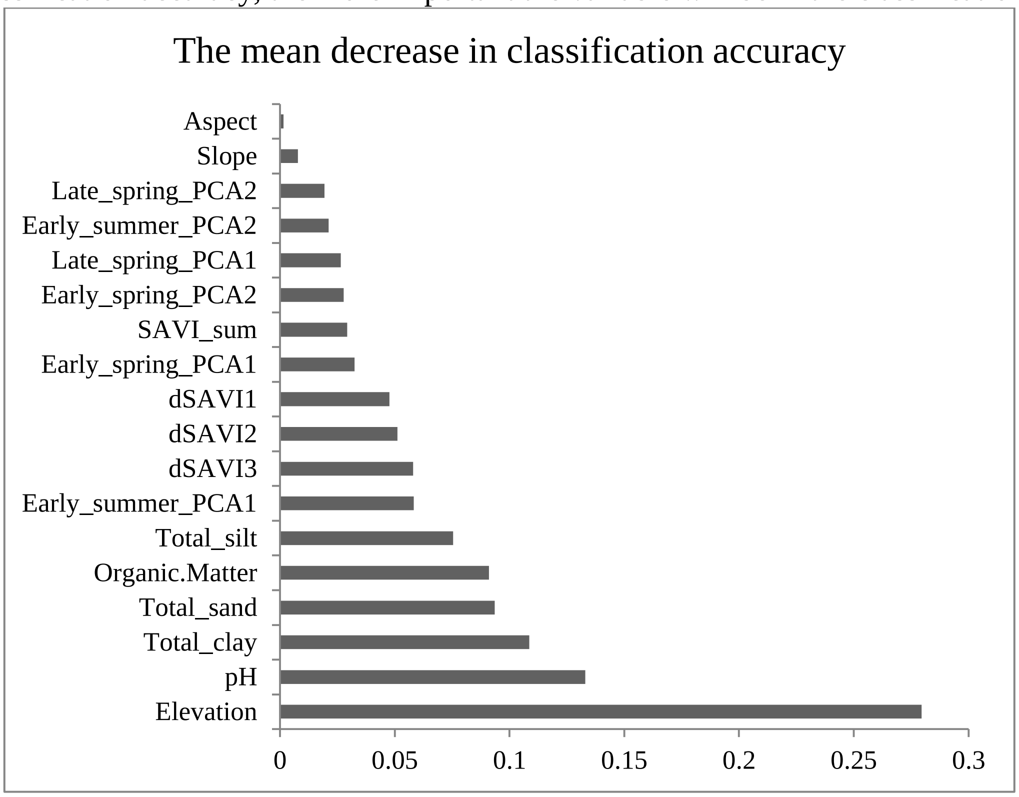


Figure 3.12 Evaluation of variable importance in 2005 Random Forest classification based on the mean decrease in classification accuracy. The larger the decrease in classification accuracy, the more important the variable will be in the classification.

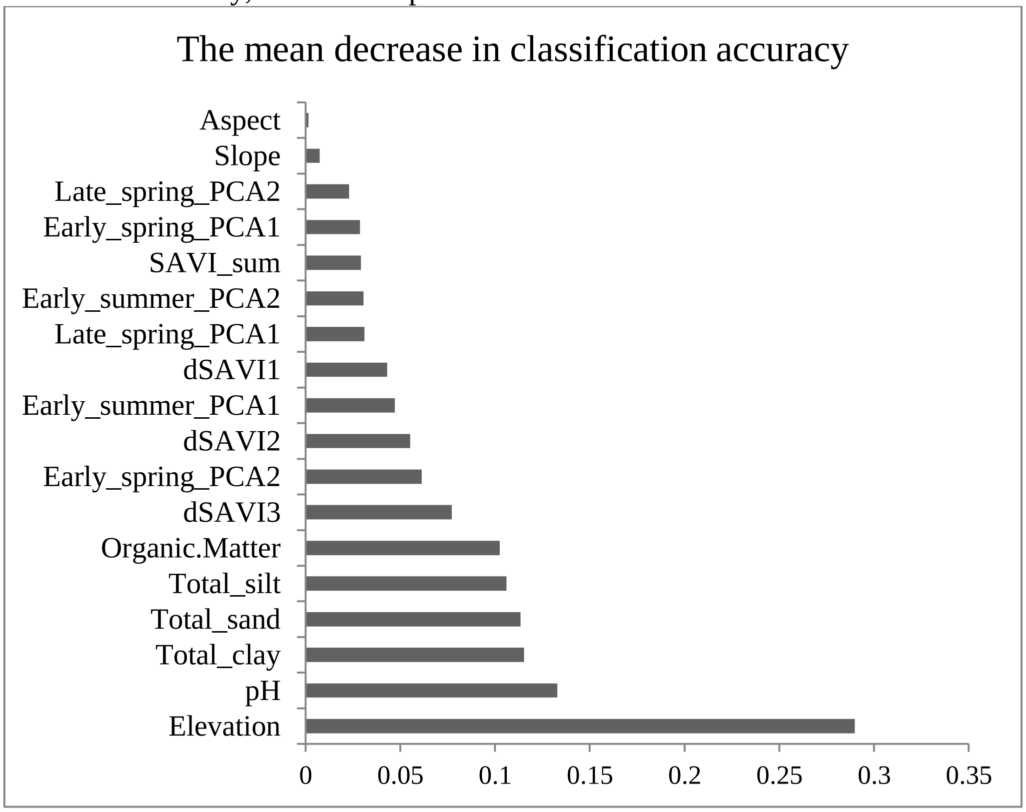
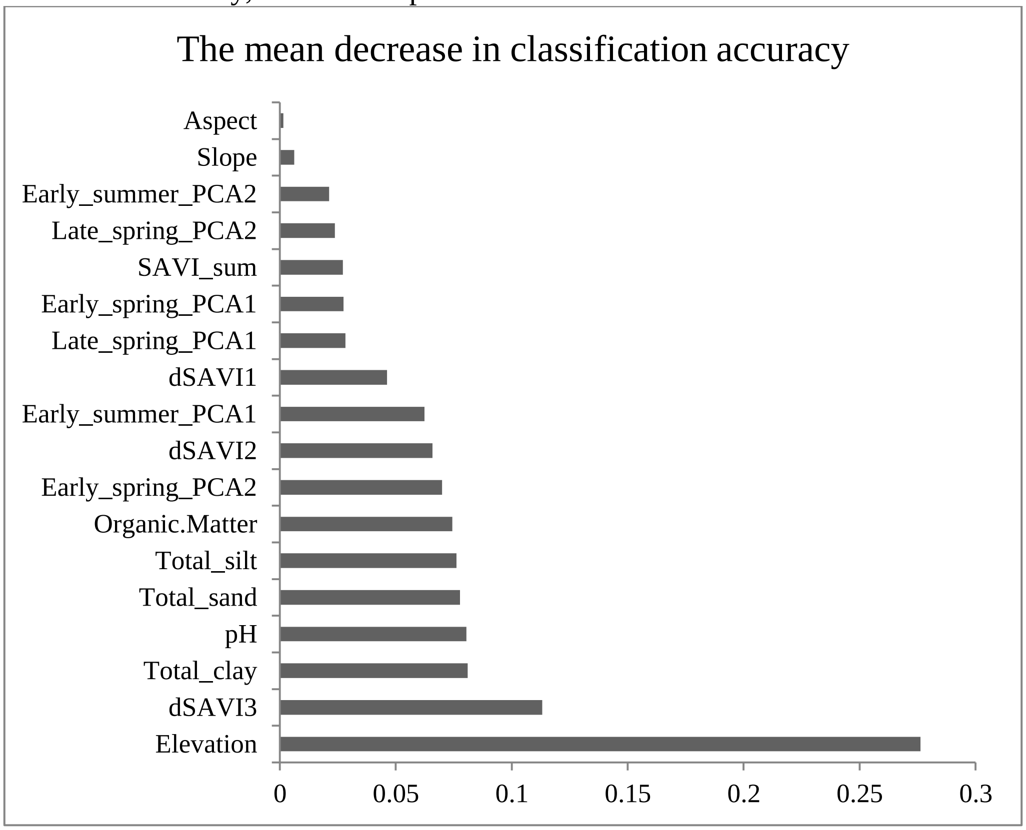


Figure 3.13 Evaluation of variable importance in 2010 Random Forest classification based on the mean decrease in classification accuracy. The larger the decrease in classification accuracy, the more important the variable will be in the classification.



3.6 Discussion

3.6.1 The effects of training data selection on historical species occurrences mapping

The availability of reliable ground truth data is crucial to map historical species occurrences. Historical aerial photography and land survey records were found to be reliable surrogates of ground truth data to map historical tree species occurrences (Hanberry *et al.*, 2012; Kellner *et al.*, 2012). The availability of fine scale ground truth data for grassland species, however, was sparse in my study area. Since there can be significant inter-annual variations in in the relative abundance of C_3 and C_4 cover types due to changes in weather conditions each year (Wang *et al.*, 2013), it is crucial to identify these areas which land cover types are stable over time and thus can be used as training data for historical grassland species mapping.

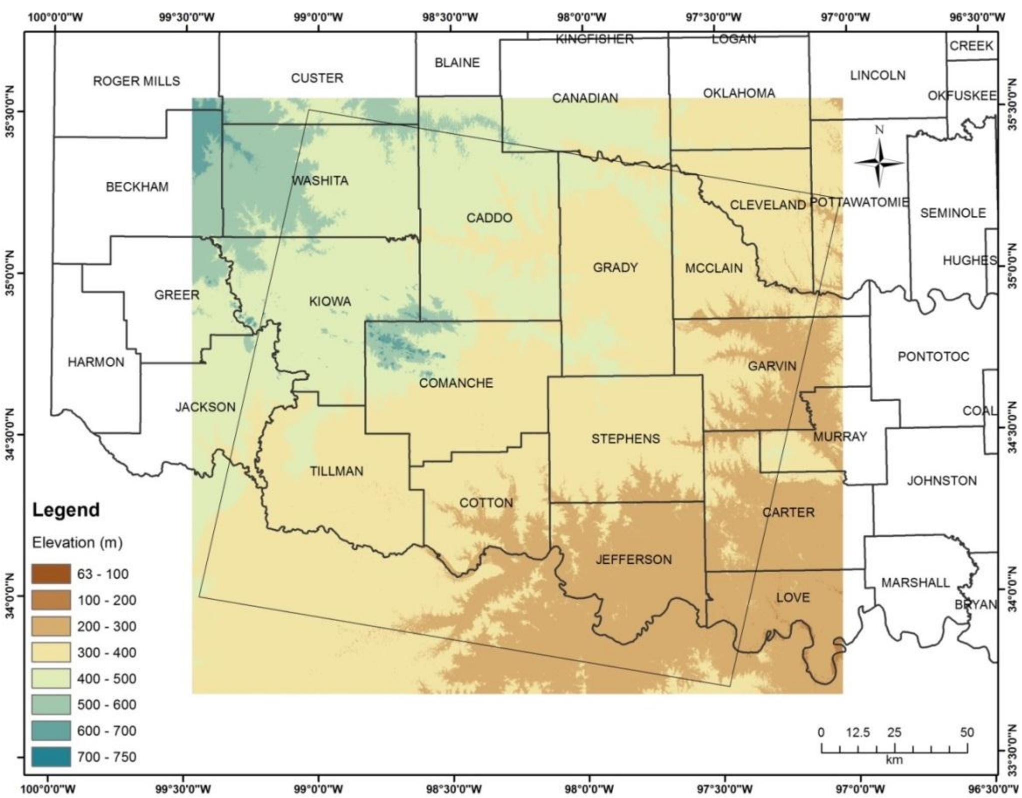
In this study, the utilization of patch 'core areas' of 2013 as ground truth data for historical grassland species mapping was proven to be effective. The OOB error rate for the C_3 cover type was the highest among the four land cover types in all the three targeted historical years. Using the 'core areas' extracted from the top 1% patches with C_3 cover type in 2013, as ground truth data, resulted in a maximum classification error of 13.43% (Table 3.5). Since an OOB error estimate is essentially the same as an error estimate that can be obtained by cross validation (Breiman, 2001; Liaw & Wiener, 2002), a maximum error rate of 13.43% in 1988 indicated that at least 86.57% of the extracted 'core areas' in each land cover categories maintained the same land cover type between 1988 and 2013.

Comparing the OOB error estimates for 2010, obtained using 'core areas' extracted from patches with different sizes, in table 3.7 and table 3.8 can shed light on the effects of training data selection on historical grassland species mapping. There were significant increases in the OOB error estimates for all the land cover types in table 3.8 as compared to the OOB error estimates shown in table 3.7. I believe that the increases in OOB error estimates as shown in table 3.8 were due to collecting training data patches whose land cover types were not stable over time. Since the OOB error estimates in table 3.8 were obtained by using the top 20% of the land cover patches in 2013, smaller patches were included as ground truth data in the estimating the OOB error. These smaller patches experienced land cover changes from 2010 to 2013 and resulted in the increases in OOB error estimates.

3.6.2 Determine the most important predictor variable in the discrimination of C_3 and C_4 grassland cover types in southwest Oklahoma

There were dramatic changes in the ranks of predictor variable importance between 2013 and the three historical years. In 2013, the top three most important variables were all the predictors computed to capture the seasonal greenness asynchrony exhibited by different land cover types (Figure 3.6). In contrast, for 1988, 2005 and 2010, Elevation, pH and the total clay content were the predictor variables most frequently evaluated as the top three most important predictors. Elevation in the study area increases from about 63m in the southeast to about 700m in the northwest as shown in figure 3.14. The 'core areas' extracted for different land cover categories fell into almost three distinctive elevation zones with core areas for rock outcrop/sparsely vegetated areas in the zone with the highest elevation, the 'core areas' for C_3 cover type and croplands being in the middle zone and the majority of 'core areas' for C_4 cover type lying in the zone with the lowest elevation (Figure 3.7 and 3.14). The differences in elevation for the 'core areas' of different land cover categories may be more pronounced than the differences in seasonal greenness development, which made elevation the most important predictor variable in the Random Forest classifications for 1988, 2005 and 2010. Elevation being an important variable in predicting the occurrences of C_3 and C_4 cover types, as found in this study, is consistent with the results obtained by previous studies. C_4 grassland cover types tend to occupy regions with low elevation and latitude across the globe (Bremond *et al.*, 2012). The decreases in C_4 species richness and biomass with increasing elevation and latitude were found to be caused by the decreases in temperature(Bremond *et al.*, 2012).

Figure 3.14 The elevation gradient in the study area. Elevation in the study area increases from the southeast to the northwest.



Unlike elevation, I did not find distinctive spatial gradients for soil pH and soil clay content in the study area. There were no previous studies documenting the effects of soil pH on the occurrences of C_3 and C_4 cover types. Soil pH being an important predictor in discriminating C_3 and C_4 cover types in the study area might just be a special case and thus might not be applicable to other regions.

I found that soil clay content was an important variable in predicting C_3/C_4 cover type occurrence. I suspect that this variables was important because its controls soil moistures. Soil texture and structure exert primary control on soil moisture properties such as plant available water (PAW). Increases in the amounts of clay soils typically result in elevated PAW (O'Geen, 2012). In addition to temperature, PAW was found to be a secondary factor that controls the distribution of C_3 and C_4 cover types (Lattanzi, 2010; Wang *et al.*, 2013). In temperate grasslands, such as the study area, the C_4 species were found to gain increases in dominance with increasing aridity due to their higher plant water use efficiency (Lattanzi, 2010).

Elevation and soil clay content were not identified as significant variables in predicting C_3/C_4 cover type occurrence in 2013. I suspect that these variables were not identified as significant in the Random Forest classification as a result of the way that the training data were collected. In delineating training data for C_3/C_4 cover types for 2013, I only focused on the asynchronous canopy greenness development between these two cover types, which could result in collecting training data that are more separable by spectral properties instead of by elevation and soil clay content. In contrast, the training data for random forest classifications in 1988, 2005 and 2010 were collected based on core areas extracted from very large C_3/C_4 patches obtained in 2013. These large C_3/C_4 patches were more evenly distributed across the study area, thus can reveal the inherent differences in the habitats of C_3/C_4 cover types. When using training data derived from these large patches, predictor variables that are able to represent these inherent differences such as elevation and soil clay content began to emerge.

3.7 Conclusion

In this study, the Random Forest classification algorithm was employed to discriminate the C_3 grassland cover type from the C_4 cover type both in the present time and in the recent past. The Random Forest classification algorithm generated highly accurate classification results for C_3/C_4 cover discrimination in 2013 with an overall classification error of 0.8%. The idea of using patch 'core areas' obtained in present time to train classifier for historical land cover classification was proved to be effective with the overall classification error for 1988, 2005 and 2010 being 10.35%, 7.94% and 8.43%, respectively. Elevation, soil pH and soil clay content were found to contribute more to C_3/C_4 cover discrimination than variables describing seasonal greenness development did in Southwest Oklahoma.

Chapter 4 LANDSCAPE DYNAMICS OF C_3 AND C_4 GRASSLAND SPECIES BETWEEN 1981 AND 2010 IN SOUTHWEST OKLAHOMA

Abstract

In this chapter, I aimed to examine the effects of precipitation anomalies on the changes of relative abundance of C_3 and C_4 grassland species in southwest Oklahoma between 1981 and 2010. Precipitation anomalies were computed based on the 1981-2010 precipitation normal and monthly precipitation summaries observed at 35 GHCN weather stations in the study area. Abundance of C_3 and C_4 in 1985, 1988, 2005 and 2010 were obtained by training a Random Forest classifier with the 'patch core areas' identified by the 2013 Random Forest classification . I examined the changes of C_3/C_4 ratio over different land cover types, along an elevation gradient and across areas with different levels of soil clay content in southwest Oklahoma between three time periods: 1985-1988, 2005-2010 and 1985-2010. The 1992 and 2006 NLCD datasets were employed to provide land cover masks to exclude areas with land cover other than shrublands, grasslands and pasture/hay from the analysis. The distributions of elevation and soil clay content in the study area were obtained from the ASTER DEM and the Oklahoma gSSURGO dataset, respectively. The difference between precipitation anomalies in the growing seasons of C_3 and C_4 species fluctuated through the years. The time periods between 1983 and 1988 and between 2002 and 2009 were found to be relatively favorable for the growth of C_3 and C_4 species, respectively. Significant decreases of C_3/C_4 ratio were identified in

pasture/hay fields due to the increases in C_4 abundance, resulting from the decreases of sparsely vegetated area between 2005 and 2010. I suspect that the increases in C_4 abundance was a drought adaptation strategy adopted by local ranchers. Because C_4 species are more tolerant of drought conditions and these species can help to maintain stable forage/hay production when negative precipitation anomalies prevailed during the growing season of C_3 species.

4.1 Introduction

Grassland ecosystems occupy approximately 40% of the earth's terrestrial area (McSherry & Ritchie, 2013) and about 70% of the world's agricultural area is made up of permanent meadow and pasture (O'Mara, 2012). With such a widespread distribution, grasslands represent one of most important ecosystems on Earth in terms of its impacts on global food supply, carbon sequestration and maintaining biodiversity (Samson & Knopf, 1994; O'Mara, 2012). The primary effect that climate change and elevated CO_2 were expected to exert on global grassland ecosystems was to change the relative abundance of C_3 and C_4 species (i.e., the C_3/C_4 ratio) (Winslow *et al.*, 2003; Lattanzi, 2010). The primary control on the spatial stratification of C_3/C_4 ratio at global scale was previously believed to be temperature (Ehleringer & Björkman, 1977; Cavagnaro, 1988; Cabido *et al.*, 1997; Ehleringer *et al.*, 1997). For example, in the studies on the changes of distribution of C_3 and C_4 grassland species at different altitudes in central Argentina, balanced abundance of C_3 and C_4 grassland species was found at about 1500m and increases in C_3 and C_4 dominance were observed at above and below 1500m, respectively (Cavagnaro, 1988; Cabido *et al.*, 1997). In these studies, temperature was found to exert significantly higher control on the changes of distribution of C_3 and C_4 species along altitudinal gradients than precipitation (Cavagnaro, 1988; Cabido *et al.*, 1997). In the prairie region of North America, C_3 species were found to dominate the northern Great Plains with C_4 species occupying its southern counterpart (Ehleringer *et al.*, 1997; Epstein *et al.*, 1997). The crossover latitude of C_3 and C_4 species in the Great Plains was found to be between 43-45ºN (Epstein *et al.*, 1997; Winslow *et al.*, 2003). However, in a more recent study on the control of global distribution of relative C_3 and C_4 biomass, the differences in growing season moisture availability and plant water use efficiency were found to be the dominating factors instead of temperature (Winslow *et al.*, 2003). Intensive studies have been conducted to examine the climate control on changes of C_3/C_4 ratio across space, but few studies were devoted to understand how climate could affect the changes of C_3/C_4 ratio over time (Ricotta *et al.*, 2003; Wang *et al.*, 2013). In a study on the climate dependency of C_3/C_4 ratio at four sites within the Great Plains, C_3 and C_4 were derived from NDVI time series acquired by MODIS (Moderate Resolution Imaging Spectroradiometer) during 2000-2009. Seasonal changes in temperature and precipitation were found to exert significant control on changes of C_3/C_4 ratio at different times of a year (Wang *et al.*, 2013).

In this chapter, I aimed to detect how changes of C_3/C_4 ratio were affected by differences in seasonal precipitation anomalies in southwest Oklahoma between 1981 and 2010. I hypothesized that the C_3/C_4 ratio would increase during a time period when moisture availability during the growing seasons of C_3 species was higher than that of the growing seasons of C_4 species during the targeted time period. Moisture availability in this study was measured by the difference between the observed growing season precipitation and the 30-year growing season precipitation normal.

4.2 Study area

In this chapter, I focused on the area covered by the Landsat scene of path 28 and row 36 in southwest Oklahoma (Figure 3.1). The dominant geomorphic province in the study area is red-bed plains, which is characterized by flat plains and gently rolling hills formed by Permian red shales and sandstones (Curtis *et al.*, 2008). Other geomorphic provinces such as sandstone and limestone hills, granite mountains and sand-dune belts are embedded within the landscape (Curtis *et al.*, 2008). Annual precipitation in the study area is primarily received in spring and autumn months and features a decreasing gradient from east to west (Johnson, 2008). Mean annual precipitation varies from approximately 1000mm in the east to about 800mm in the west (Johnson, 2008). Major types of potential natural vegetation in the study area are post oak-blackjack forest, mixedgrass and tallgrass prairie (Duck & Fletcher, 1945; Hoagland, 2008). Mean annual temperature decreases from 17°C in the south to about 15°C in the north (Johnson, 2008). Mixedgrass prairie mainly occurs in the western portion of the study area as influenced by the decreasing average annual precipitation from east to west. Much of the mixedgrass prairie has been converted to cultivated crops of wheat and cotton as shown by the 2006 National Land Cover Database in figure 3.1 (Hoagland, 2008; Xian *et al.*, 2009).

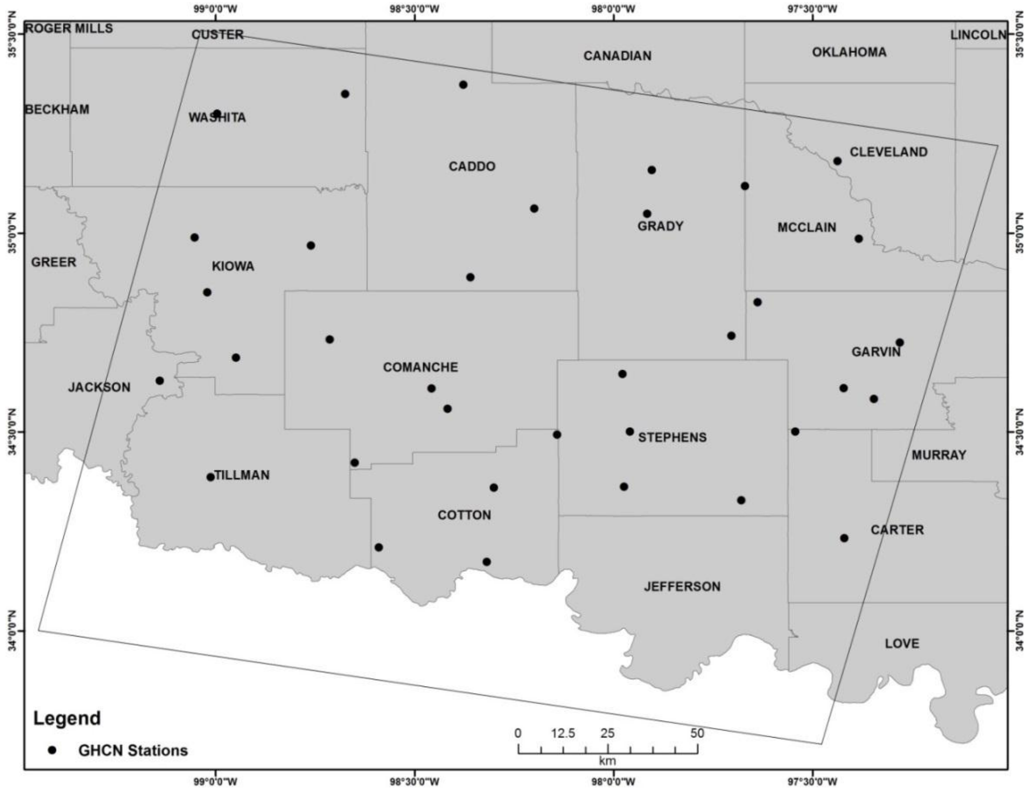
A list of commonly encountered C_3 and C_4 species in the study area is shown in table 3.1. Descriptions of the species characteristics were retrieved from the Oklahoma Vascular Plant Database at: http://www.oklahomaplantdatabase.org/

4.3 Data

4.3.1 1981-2010 Monthly precipitation normals and monthly precipitation summaries

The 1981 to 2010 Monthly Precipitation Normals (MPN) for 35 Global Historical Climatology Network (GHCN) weather stations within the study area were downloaded from the NOAA National Climate Data Center at: http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals. Locations of the 35 weather stations are shown in Figure 4.1. The GHCN collects precipitation data from approximately 7500 land surface stations all over the globe (Peterson *et al.*, 1998). The 1981-2010 monthly precipitation normals are made up of two components: monthly rainfall normals and monthly snowfall normals. A number of quantities are available for each of the two components such as: average monthly totals, quartiles and frequencies of occurrence (Durre *et al.*, 2013).

Figure 4.1 Locations of GHCN weather stations within the area covered by the Landsat scene Path 28 / Row 36



Monthly rainfall and snowfall totals between 1981 and 2010 were downloaded for the same 35 GHCN weather stations from the NOAA National Climate Data Center at: http://www.ncdc.noaa.gov/ghcnm/.

4.3.2 National Land Cover Database

In this study, the dynamics of C_3/C_4 ratio were only analyzed within the area with the land cover of shrublands, grasslands or pasture/hay fields according to land cover types identified by the National Land Cover Database (NLCD). Land cover classes other than these classes were excluded from this analysis using a land cover mask derived from NLCD for each targeted year. The 1992 NLCD dataset was used to generate the land cover mask and used as a reference for targeted years before

1992. The land cover mask derived from the 2006 NLCD dataset was applied to the targeted years after 2005.

4.3.3 Digital elevation data

In chapter three, elevation was identified as one of the most important variables in terms of predicting the occurrences of C_3 and C_4 grassland species. In this study, elevation data were used to examine if there were significant changes of C_3/C_4 ratio along the altitudinal gradient within the study area. Elevation information were acquired from the global Digital Elevation Model (DEM) generated using observations by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) with a spatial resolution of 1 arc-second (approximately 30m) (Tachikawa *et al.*, 2011). The ASTER Global DEM is available as 1º by 1º tiles. A total of nine ASTER DEM tiles were downloaded for the study area.

4.3.4 Soil clay content

In chapter three, soil clay content was identified as one of the most important variables in terms of predicting the occurrences of C_3 and C_4 grassland species. Soil moisture available for plant to use is controlled by soil clay content. Increases in the amounts of clay soils typically result in elevated plant available water (O'Geen, 2012). In temperate grasslands, such as the mixed-grass prairie of southwest Oklahoma, C_4 species were found to gain increased dominance in areas with low soil clay content due to their higher plant water use efficiency (Lattanzi, 2010). In this chapter, the distribution of soil clay content within the study area was used as a reference to examine if there were significant changes of C_3/C_4 ratio in the areas with different levels of soil clay content. Soil clay content was extracted using the Gridded Soil Survey Geographic (gSSURGO) Database for Oklahoma and the National Value Added Look Up (valu1) Table database (Soil Survey Staff, 2012). Two steps were taken to extract soil clay content. First, for a soil map unit with the ID of ***k*** in the National Value Look Up Table database, the average value of soil clay content within ***k*** was obtained by averaging the soil clay content over all the sub-components within ***k***. Second, I assigned the average soil clay content to the gSSURGO grid cells with the same map unit ID as that of ***k***.

4.3.5 Landsat surface reflectance data

Landsat surface reflectance for Landsat Thematic Mapper (TM) (30m) and Enhanced Thematic Mapper Plus (ETM+) (30m) were acquired from the Landsat Surface Reflectance Climate Data Records (CDRs), which are atmospherically corrected by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek *et al.*, 2006). The Landsat Surface Reflectance CDRs were downloaded from the USGS EarthExplorer at http://earthexplorer.usgs.gov/. For each targeted year, Landsat images acquired during early spring, late spring and early summer were downloaded. In addition to the surface reflectance datasets listed in table 3.2, Landsat surface reflectance CDR for April 3rd, May 5th and Aug 9th of 1985 were also used in this study.

4.4 Methodology

4.4.1 Conversion of snowfall to equivalent amount of liquid water

A snow-to-liquid-equivalent ratio (SLR) of 10 has regularly been adopted by previous studies to convert the amount of snowfall to equivalent amount of liquid water (Roebber *et al.*, 2003; Baxter *et al.*, 2005). In this study, the 1981-2010 monthly snowfall normals and monthly snowfall totals observed at the 35 GHCN stations were all divided by 11.6 to be converted to equivalent amount of liquid water. The SLR value of 11.6 is the long term average SLR reported for Norman, OK (Baxter *et al.*, 2005).

4.4.2 Identifying C_3 or C_4 favorable years based on normalized precipitation anomaly during C_3 and C_4 growing seasons.

C_3 and C_4 species exhibit active growth during different times of a year. Germination of C_3 annuals can start as early as the beginning of fall (Tyrl *et al.*, 2008). The active growth of C_3 species begin growth in early spring and reach peak growth in late spring. C_3 species may become senescent or semi-dormant in the summer to avoid hot temperatures and water stress conditions (Winslow *et al.*, 2003). Some C_3 species may resume growth in the fall (Wang *et al.*, 2013). Growth of C_4 species starts in late spring, reaches the peak during summer. In this study, the growing season for C_3 grassland species was defined as from September to May in the next year (Tyrl *et al.*, 2008). For C_4 grassland species, the growing season was defined as from March to August(Tyrl *et al.*, 2008).

To determine which years were more favorable for the growing of C_3 species and which were more favorable for the growth of C_4 species, I investigated the precipitation data from 35 GHCN weather stations. For each of the 35 GHCN weather stations, monthly precipitation was computed as the sum of monthly rainfall and the equivalent amount of liquid water converted from monthly snowfall. The monthly precipitation normal was computed as the sum of the monthly rainfall normal and the monthly normal of liquid water converted from monthly snowfall normal.

The normalized precipitation anomaly (*NPA*) during the growing season of C_3/ C_4 species at a particular GHCN weather station $\boldsymbol{i}$ ($\boldsymbol{i}$ = 1, 2, ..., 35) in a year between 1981 and 2010 was computed as:

$$NPA_{i-C3} = \frac{\sum_{SEP}^{MAY} P_j - \sum_{SEP}^{MAY} PN_j}{\sum_{SEP}^{MAY} PN_j} \tag{4.1}$$

$$NPA_{i-C4} = \frac{\sum_{MAR}^{AUG} P_j - \sum_{MAR}^{AUG} PN_j}{\sum_{MAR}^{AUG} PN_j} \tag{4.2}$$

where P_j is the monthly precipitation in month j and PN_j is the precipitation normal for month j.

For each year, the difference between the *NPA* for C_3 and C_4 species at station $\boldsymbol{i}$: $dNPA_i$ ($\boldsymbol{i}$ = 1, 2, ..., 35) was computed by subtracting $NPA_{i\text{-}C4}$ from $NPA_{i\text{-}C3}$. The *dNPA* was then averaged over the 35 GHCN stations to derive the *dNPA-k* ($\boldsymbol{k}$ = 1981, 1982, ..., 2010), which represents the average difference between the *NPA* for the growing seasons of C_3 and C_4 species in a particular year $\boldsymbol{k}$. A positive *dNPA-k* indicates that the moisture condition in year ***k*** was relatively favorable for the growth of C_3 species and a negative *dNPA-k* was relatively favorable for C_4 species growth.

4.4.3 Retrieving historical grassland species abundance

In this study, the historical abundance of C_3 and C_4 species was retrieved from the land cover classification results obtained in Chapter three. The abundance of C_3 or C_4 species was calculated by dividing the number of pixels classified as C_3 or C_4 cover type by the number of pixels from all land cover types (i.e., C_3 cover type, C_4 cover type, Croplands and rock outcrop and sparsely vegetated area). The C_3/C_4 ratio was then calculated by dividing C_3 abundance by the abundance of C_4 cover types.

4.4.4 Characterizing the dynamics of C_3/C_4 ratio between 1981 and 2010

To investigate if there were significant impacts exerted by the differences in seasonal moisture availability on C_3/C_4 ratio dynamics, changes in C_3/C_4 ratio were examined over two time periods: one with a continuously positive *dNPA* and the other one with a continuously negative *dNPA*. Changes in C_3/C_4 ratio were determined by comparing the C_3/C_4 ratio retrieved during the two years that were at the beginning and the end of a time period. To avoid a biased comparison of C_3/C_4 ratio which could be caused by changes in vegetation phenology, the beginning and ending years were selected in a way that anniversary Landsat CRDs must be available at each of three dates as required by Random Forest classifications in these two years. The C_3/C_4 ratios between 1985 and 1988, and between 2005 and 2010 were compared for the time periods with a continuously positive *dNPA* and a continuously negative *dNPA*, respectively. A third comparison was made between C_3/C_4 ratios in 1985 and 2010 to examine the changes in C_3/C_4 ratio over the entire study period.

For each of three time periods, changes in C_3/C_4 ratio were examined within different types of land cover (shrublands, grasslands and pasture/hay fields) and along the gradient of elevation and soil clay content. The gradient of elevation and soil clay content were generated by classifying elevation/soil clay content into four groups with different ranges as:

- Group 1: between the minimum and the first quartile;
- Group2: between the first quartile and the median;
- Group3: between the median and the third quartile;
- Group4: between the third quartile and the maximum;

The pixels with a value of NAN (i.e., no data), as a result of Landsat ETM+ scan line corrector failure, being covered by cloud or cloud shadow, or located within in the area of NLCD land cover masks, in any of the two years under comparison were excluded from the analysis.

4.4.5 Homogeneity test using contingency table and Pearson's Chi-squared test

To determine if the changes in C_3/C_4 ratio between two years were statistically significant, a contingency table and Pearson's Chi-squared test were employed to examine if the abundance distribution among the four land cover types (*i.e.*, C_3 cover type, C_4 cover type, Croplands and rock outcrop and sparsely vegetated area) was homogeneous between the two years. Changes in C_3/C_4 ratio obtained from the two years with homogeneous abundance distribution among four land cover types can be viewed as random results, which are unreliable. The Chi-squared test statistic (χ^2) can be derived by:

- Filling in a contingency table with the abundance distribution of the four land cover types during the two years. An example contingency table is shown in Table 4.1.
- Calculate the expected abundance of a land cover ***i*** in year ***j*** ($\boldsymbol{E_{ij}}$) as:

$$E_{ij} = \frac{(i^{th}\ row\ total)(j^{th}\ column\ total)}{grand\ total} \qquad (4.3)$$

- For each cell in the contingency table, the departure of the observed abundance of land cover ***i*** in year ***j*** ($\boldsymbol{O_{ij}}$) from the expected abundance ($\boldsymbol{E_{ij}}$) can be computed as:

$$Abundance\ departure = \frac{(O_{ij}-E_{ij})^2}{E_{ij}} \qquad (4.4)$$

- The χ^2 can be obtained by summing up the abundance departures in all the cells.

Table 4.1 An example of the contingency table used in this study. *P*x_ A and *P*x_ B are the abundance of land cover X in year A and B, respectively.

Land cover	Year A	Year B	Row Total
C_3 cover type	Pc_{3}_ A	Pc_{3}_B	Pc_{3}_ A + Pc_{3}_B
C_4 cover type	Pc_{4}_ A	Pc_{4}_ B	Pc_{4}_ A + Pc_{4}_B
Croplands	P_{crop}_A	P_{crop}_B	P_{crop}_ A + P_{crop}_B
RSV	P_{rsv}_A	P_{rsv}_B	P_{rsv}_ A + P_{rsv}_B
Column Total	100	100	Grand total: 200

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

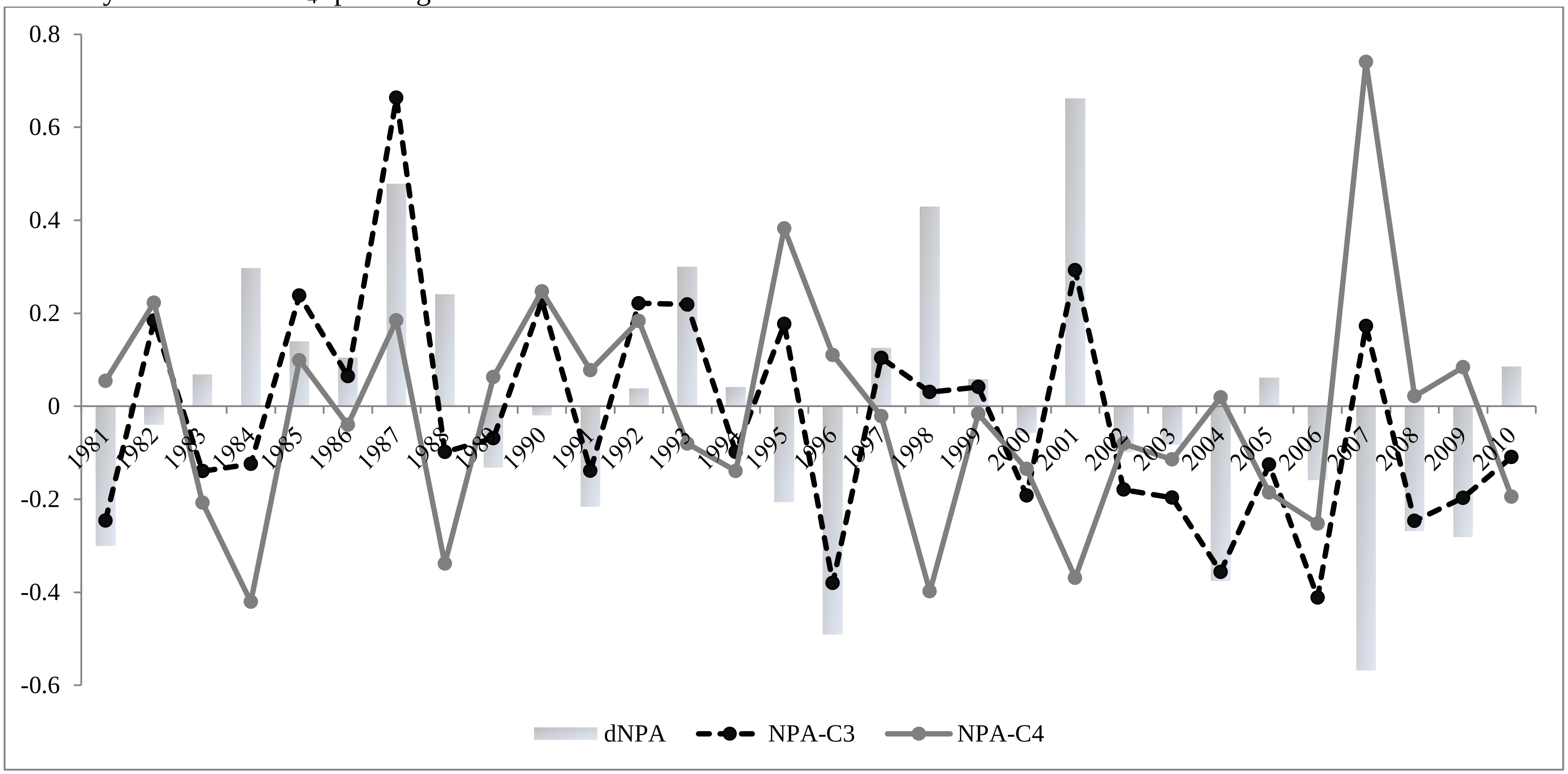
In this study, χ^2 obtained by a contingency table was then used to perform a Pearson's Chi-squared test with three Degree-of-Freedom (df) in R. The number of Degree-of-Freedom was determined based on the number of rows and columns in a contingency table as: df = (Number-of-Rows -1) (Number-of-Columns -1). The Null hypothesis for the Chi-squared test was that the abundance distribution among the four land cover categories was homogenous between the two targeted years. The alternative hypothesis was that the abundance distributions of the two years were different from each other because there were significant changes in the abundance between the two years for at least one of the four land cover categories.

4.5 Results

4.5.1 The inter-annual variation of dNPA between 1981 and 2010

The inter-annual variation of NPA in the growing season of C_3 and C_4 species, and the annual *dNPA* between 1981 and 2010 is shown in Figure 4.2. The annual *dNPA* had been fluctuating through the years between 1981 and 2010, except for the periods from 1983 to 1988 and from 2003 to 2009. A *dNPA* of 0.66 occurred in 2001, which represented the highest *dNPA* between 1981 and 2010. The *dNPA* of -0.56 in 2007 was the lowest *dNPA* between 1981 and 2010. There were constantly positive *dNPA*s between 1983 and 1988 and almost continuously negative *dNPA*s were seen between 2002 and 2009.

Figure 4.2 Inter-annual variations of the difference between the normalized departure from precipitation normal (dNPA) during the C_3 and C_4 growing seasons from 1981 to 2010. A year with a positive dNPA is the year with a moisture condition that was relatively favorable for C_3 species growth. The years with negative dNPAs were the time periods when moisture conditions were relatively favorable for C_4 species growth.



4.5.2 Distribution of C_3 and C_4 grassland species between 1981 and 2010

The distribution of C_3 and C_4 grassland species was retrieved for 1985, 1988, 2005 and 2010 using the methodology demonstrated in Chapter three. The Random Forest classification results for 1988, 2005 and 2010 are shown in Figure 3.8, 3.9 and 2.10. Distribution of C_3 and C_4 grassland species in 1985 is shown in Figure 4.3. The accuracy assessment of 1985 Random Forest classification based on OOB error estimate is shown in Table 4.2. Distribution of C_3 and C_4 grassland species in 1985 was similar to that of the other three years with C_3 species mainly found in the west portion of the study area and C_4 species occupying the east.

Figure 4.3 Random Forest classification result for 1985. The classification result was obtained using the methodology demonstrated in chapter three

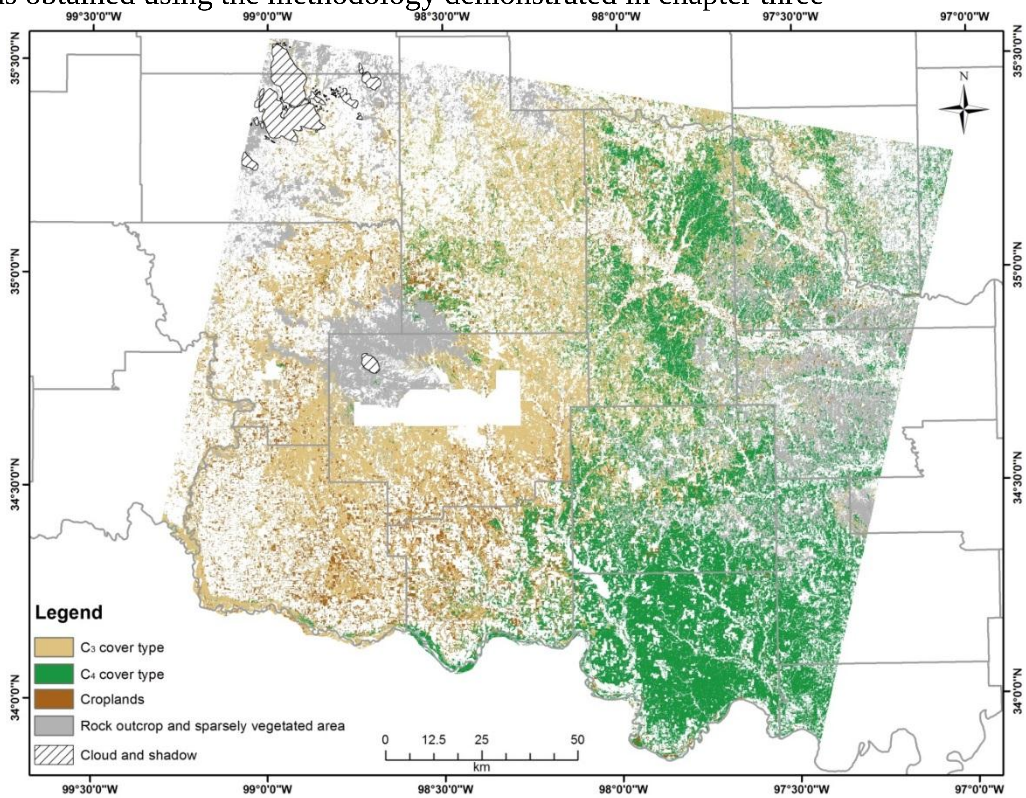


Table 4.2 Accuracy assessment for Random Forest classification of 1985

	Random Forest Prediction					
		C_3	Croplands	C_4	RSV	OOB estimate of error
Ground observations	C_3	7625	486	543	94	12.83%
	Croplands	424	5134	40	7	8.40%
	C_4	756	84	7794	81	10.57%
	RSV	298	12	311	8052	7.16%
	Overall					9.88%

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

4.5.3 Changes in C_3/C_4 ratio stratified by NLCD land cover

The changes in C_3/C_4 ratio in the areas with the land cover of shrublands, grasslands and pasture/hay are shown in table 4.3, 4.4 and 4.5, respectively. There was a decrease in C_3/C_4 ratio in all of the three targeted time periods. However, statistically significant decreases in C_3/C_4 ratio were only seen between 2005 and 2010 and between 1985 and 2010 in the areas covered by pasture/hay fields (P < 0.01) (Table 4.4 and Table 4.5). The significant decreases of C_3/C_4 ratio were all caused by the increases in C_4 abundance (Table 4.4 and Table 4.5).

Table 4.3 A comparison of the abundance distribution among the four land cover categories between 1985 and 1988. The pixels with a NAN value in any of the two years were removed from the statistics.

	Shrublands†				
Year	C_3	C_4	Cropland	RSV	C_3/ C_4
1985	44.18%	21.40%	4.24%	30.19%	2.06
1988	43.27%	22.24%	3.97%	30.52%	1.95
	Grasslands†				
Year	C_3	C_4	Cropland	RSV	C_3/ C_4
1985	33.29%	40.01%	5.43%	21.26%	0.83
1988	33.09%	40.79%	5.50%	20.61%	0.81
	Pasture/Hay†				
Year	C_3	C_4	Cropland	RSV	C_3/ C_4
1985	36.62%	28.19%	24.80%	10.38%	1.30
1988	35.32%	28.17%	26.58%	9.94%	1.25

†: The abundance distributions of the two targeted years were homogeneous with a $P > 0.1$.

Note:

1**.** 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

2. The land cover types of shrublands, grasslands and pasture/hay were identified by the National Land Cover Database. The land cover types of C_3, C_4, cropland and Rock outcrop and Sparsely Vegetate area were obtained using the Random Forest classification algorithm.

Table 4.4 A comparison of the abundance distribution among the four land cover categories between 2005 and 2010. The pixels with a NAN value in any of the two years were removed from the statistics.

	Shrublands†				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	65.08%	6.01%	12.54%	16.37%	10.83
2010	63.86%	8.32%	10.98%	16.84%	7.68
	Grasslands†				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	45.65%	34.88%	8.11%	11.36%	1.31
2010	44.31%	39.27%	7.48%	8.94%	1.13
	Pasture/Hay***				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	21.54%	54.44%	9.14%	14.88%	0.40
2010	21.13%	70.14%	7.33%	1.39%	0.30

†: The abundance distributions of the two targeted years were homogeneous with a $P > 0.1$.

***: The abundance distributions of the two targeted years were heterogeneous. There was significant change in the abundance of at least one land cover category between the two targeted years with a $P < 0.01$.

Note:

1**.** 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

2. The land cover types of shrublands, grasslands and pasture/hay were identified by the National Land Cover Database. The land cover types of C_3, C_4, cropland and Rock outcrop and Sparsely Vegetate area were obtained using the Random Forest classification algorithm.

Table 4.5 A comparison of the abundance distribution among the four land cover categories between 1985 and 2010. The pixels with a NAN value in any of the two years were removed from the statistics.

	Shrublands†				
Year	C_3	C_4	Cropland	RSV	C_3/ C_4
1985	67.44%	3.86%	9.93%	18.77%	17.48
2010	62.70%	8.46%	6.62%	22.21%	7.41
	Grasslands†				
Year	C_3	C_4	Cropland	RSV	C_3/ C_4
1985	42.03%	35.42%	6.62%	15.93%	1.17
2010	39.24%	44.94%	4.56%	11.27%	0.87
	Pasture/Hay***				
Year	C_3	C_4	Cropland	RSV	C_3/ C_4
1985	16.29%	57.51%	5.59%	20.61%	0.28
2010	15.80%	78.98%	3.76%	1.46%	0.20

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

***: The abundance distributions of the two targeted years were heterogeneous. There was significant change in the abundance of at least one land cover category between the two targeted years with a $P < 0.01$.

Note:

1**.** ‘RSV’ stands for Rock outcrop and Sparsely Vegetate area.

2. The land cover types of shrublands, grasslands and pasture/hay were identified by the National Land Cover Database. The land cover types of C_3, C_4, cropland and Rock outcrop and Sparsely Vegetate area were obtained using the Random Forest classification algorithm.

4.5.4 Changes in C_3/C_4 ratio along elevation gradient

Distribution of the altitudinal gradient in southwest Oklahoma is shown in figure 4.4. Elevation in the study area increases from 63m in the southeast to 756m in the northwest. The changes in C_3/C_4 ratio along the altitudinal gradient in the study area are shown in table 4.6, 4.7 and 4.8, respectively. Significant changes in C_3/C_4 ratio were only found between 1985 and 2010 in the southeast portion of the study area, where elevation ranges from the minimum (63m) to the median (355m) (Table 4.8). Significant changes C_3/C_4 ratio in the low elevation area between 1985 and 2010 were represented by decreases in C_3/C_4 ratio due to significant increases in C_4 abundance.

Figure 4.4 The Southeast-Northwest elevation gradient in southwest Oklahoma. Elevation increases from 63m in the southeast to 756m in the northwest.

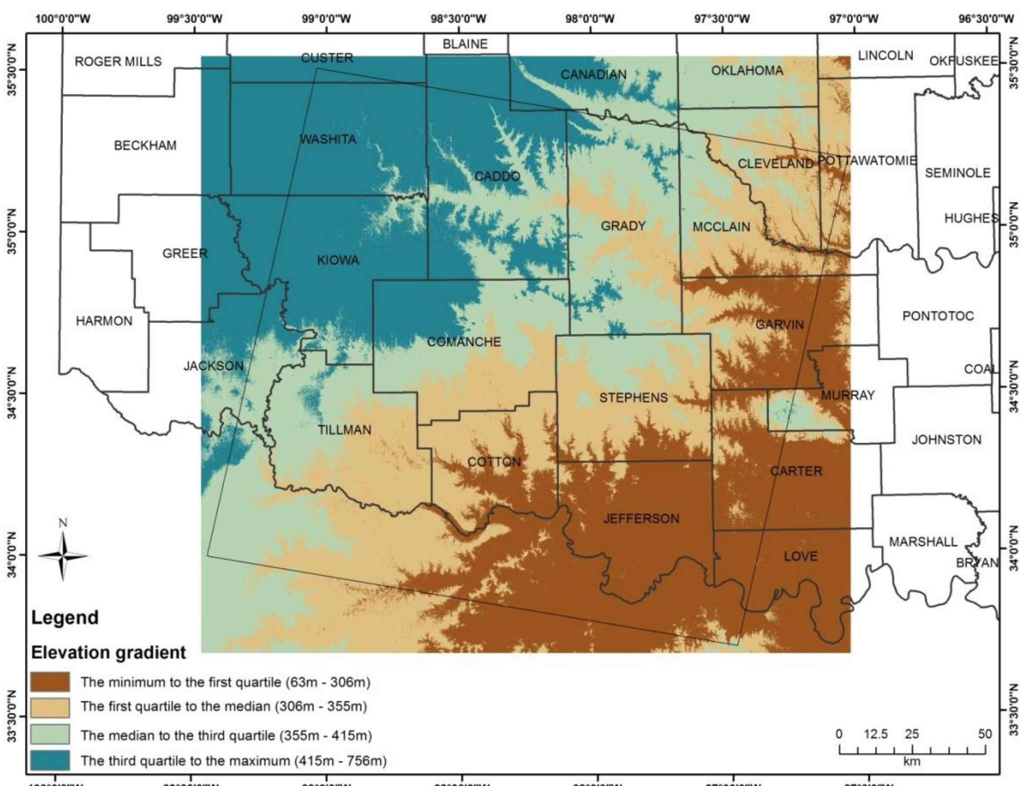


Table 4.6 The changes in C_3/C_4 ratio along the elevation gradient between 1985 and 1988 in southwest Oklahoma. The pixels with a NAN value in any of the two years were removed from the statistics.

	Group1: 63m – 306m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	2.78%	80.40%	4.33%	12.49%	0.03
1988	2.83%	80.18%	5.05%	11.94%	0.04
	Group2: 306m – 355m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	32.02%	30.13%	17.29%	20.56%	1.06
1988	30.74%	31.46%	17.37%	20.44%	0.98
	Group3: 355m – 415m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	55.06%	27.98%	7.09%	9.87%	1.97
1988	54.44%	29.20%	7.29%	9.07%	1.86
	Group4: 415m – 756m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	45.13%	7.12%	8.73%	39.02%	6.34
1988	45.29%	6.65%	9.72%	38.34%	6.81

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

Table 4.7 The changes in C_3/C_4 ratio along the elevation gradient between 2005 and 2010 in southwest Oklahoma. The pixels with a NAN value in any of the two years were removed from the statistics.

	Group1: 63m – 306m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	6.99%	77.55%	9.29%	6.17%	0.09
2010	5.01%	85.87%	7.69%	1.43%	0.06
	Group2: 306m – 355m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	45.63%	31.30%	16.20%	6.87%	1.46
2010	47.49%	34.42%	15.97%	2.12%	1.38
	Group3: 355m – 415m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	59.12%	31.99%	3.45%	5.44%	1.85
2010	57.26%	36.57%	2.81%	3.36%	1.57
	Group4: 415m – 756m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	50.24%	12.25%	3.56%	33.96%	4.10
2010	45.37%	16.13%	2.75%	35.75%	2.81

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

Table 4.8 The changes in C_3/C_4 ratio along the elevation gradient between 1985 and 2010 in southwest Oklahoma. The pixels with a NAN value in any of the two years were removed from the statistics.

	Group1: 63m – 306m **				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	5.39%	80.78%	5.04%	8.79%	0.07
2010	3.13%	92.82%	3.10%	0.95%	0.04
	Group2: 306m – 355m **				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	41.46%	30.18%	14.99%	13.37%	1.38
2010	44.63%	41.70%	10.38%	3.29%	1.08
	Group3: 355m – 415m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	59.95%	29.89%	2.53%	7.64%	2.01
2010	54.32%	39.88%	1.77%	4.03%	1.36
	Group4: 415m – 756m †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	47.87%	9.52%	3.01%	39.59%	5.03
2010	40.59%	13.85%	2.20%	43.36%	2.93

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

**: The abundance distributions of the two targeted years were heterogeneous. There was significant change in the abundance of at least one land cover category between the two targeted years with a $P < 0.05$.

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

4.5.5 Changes in C_3/C_4 ratio stratified by soil clay content

The distribution of soil clay content in southwest Oklahoma is shown in figure 4.5. Areas with low soil clay content were mainly found in Caddo, Grady and Stephens County while areas with high soil clay content concentrated in the southwest portion of the study area such as Kiowa, Tillman, Cotton and southwest Comanche County. Changes in C_3/C_4 ratio in the areas with varying soil clay content between 1985 and 1988, between 2005 and 2010 and between 1985 and 2010 are shown in table 4.9, 4.10 and 4.11, respectively. No significant changes in C_3/C_4 ratio were found in the areas with varying soil clay content during the three targeted time periods

Figure 4.5 The spatial distribution of total soil clay content in southwest Oklahoma. Total soil clay content ranges from 0.50% to 50.00% in the study area.

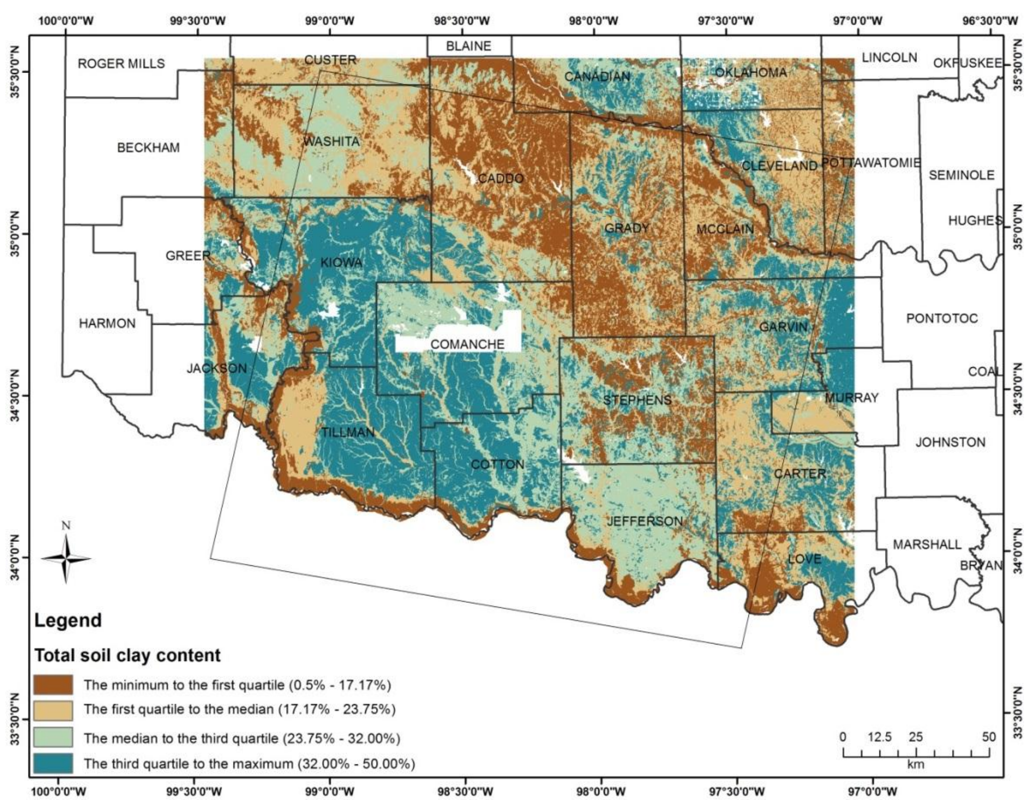


Table 4.9 The changes in C_3/C_4 ratio between 1985 and 1988 stratified by the distribution of soil clay content in southwest Oklahoma. The pixels with a NAN value in any of the two years were removed from the statistics.

	Group1: 0.50% - 17.17% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	38.66%	41.43%	5.04%	14.87%	0.93
1988	38.46%	41.39%	4.72%	15.43%	0.93
	Group2: 17.17% - 23.75% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	30.14%	34.74%	5.41%	29.70%	0.87
1988	28.69%	38.22%	5.71%	27.37%	0.75
	Group3: 23.75% - 32.00% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	25.90%	49.49%	7.15%	17.46%	0.52
1988	25.23%	49.66%	8.21%	16.90%	0.51
	Group4: 32.00% - 50.00% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	45.10%	15.65%	21.76%	17.49%	2.88
1988	45.26%	14.86%	22.42%	17.46%	3.05

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

Note: ‘RSV’ stands for Rock outcrop and Sparsely Vegetate area.

Table 4.10 The changes in C_3/C_4 ratio between 2005 and 2010 stratified by the distribution of soil clay content in southwest Oklahoma. The pixels with a NAN value in any of the two years were removed from the statistics.

	Group1: 0.50% - 17.17% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	46.86%	37.92%	2.62%	12.60%	1.24
2010	43.25%	44.21%	2.36%	10.17%	0.98
	Group2: 17.17% - 23.75% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	42.41%	35.32%	5.43%	16.84%	1.20
2010	40.51%	41.32%	4.77%	13.40%	0.98
	Group3: 23.75% - 32.00% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	36.61%	44.92%	9.01%	9.46%	0.81
2010	37.46%	45.54%	8.59%	8.41%	0.82
	Group4: 32.00% - 50.00% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
2005	58.71%	15.71%	17.19%	8.39%	3.74
2010	58.16%	21.72%	15.57%	4.55%	2.68

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

Table 4.11 The changes in C_3/C_4 ratio between 1985 and 2010 stratified by the distribution of soil clay content in southwest Oklahoma. The pixels with a NAN value in any of the two years were removed from the statistics.

	Group1: 0.50% - 17.17% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	45.68%	39.62%	2.73%	11.98%	1.15
2010	38.00%	49.89%	1.49%	10.62%	0.76
	Group2: 17.17% - 23.75% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	37.78%	33.14%	2.98%	26.10%	1.14
2010	33.23%	46.81%	2.47%	17.49%	0.71
	Group3: 23.75% - 32.00% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	31.89%	46.91%	6.40%	14.80%	0.68
2010	31.37%	52.52%	4.76%	11.35%	0.60
	Group4: 32.00% - 50.00% †				
Year	C_3	C_4	Cropland	RSV	C_3/C_4
1985	55.34%	15.17%	15.11%	14.38%	3.65
2010	57.00%	25.69%	10.00%	7.31%	2.22

†: The abundance distributions of the two targeted years were homogeneous, with a $P > 0.1$.

Note: 'RSV' stands for Rock outcrop and Sparsely Vegetate area.

4.6 Discussion

4.6.1 A comparison of the inter-annual variation of dNPA during 1983 – 1988 and 2002-2009

dNPA measures the difference between the moisture anomaly during the growing seasons of C_3 and C_4 species. *dNPA* maintained a relatively constant state during both 1983-1988 and 2002-2009. The constant states, however, were caused by quite different changes of precipitation anomaly during C_3 and C_4 growing seasons in the two targeted time periods. During 1983-1988, the NPA of C_3 and C_4 growing seasons were highly synchronous. A positive or negative NPA in a C_3 growing season were always accompanied by a positive or negative NPA in a C_4 growing season, except for the year of 1986 (Figure 4.2). In other words, there were either above-normal or drought conditions in the growing seasons of both C_3 and C_4 species. The occurrence of continuously positive *dNPAs* was only due to a more above-normal moisture condition or a less intense drought condition in C_3 growing seasons. In contrast, there were asynchronous changes of NPA between C_3 and C_4 growing seasons during four of the eight years from 2002 to 2009 (i.e., 2004, 2007, 2008 and 2009) (Figure 4.2). The occurrences of negative *dNPAs* in these four years were caused by an above-normal moisture condition in C_4 growing seasons being accompanied by a drought condition in C_3 growing seasons.

No significant changes of C_3/C_4 ratio between 1985 and 1988 were found, independent of the different land cover types, elevation gradients, soil clay contents.

I suspect this may be caused by the synchronous changes in NPA, which exerted similar impacts on the growth of C_3 and C_4 species.

4.6.2 The relationship between the inter-annual variations of dNPA and land cover changes between 1981 and 2010

By examining C_3/C_4 ratio changes in areas with different land cover types, statistically significant changes were only found in areas covered by pasture/hay during 2005-2010 and 1985-2010 (Table 4.4, 4.5). By examining C_3/C_4 ratio changes along the southeast-northwest elevation gradient, significant changes were only found in areas with an elevation below the median elevation in the study area between 1985 and 2010 (Table 4.8). Significant C_3/C_4 ratio changes identified in this study were all related to the decrease of C_3/C_4 ratio, due to the significant increase of C_4 abundance, during the targeted time period. By further examining the changes in abundance distribution among the four land cover categories, increases of C_4 abundance can always be attributed to the substantial decreases of rock outcrop and sparsely vegetated area (Table 4.4, 4.5, 4.8). This is consistent with the land cover changes inferred from the Random Forest classification results for 1985, 1988, 2005 and 2010 (Figure 4.3, 3.8, 3.9 and 3.10). Substantial decreases of sparsely vegetated area did not occur until 2010. Sparsely vegetated areas that were constantly present in 1985, 1988 and 2005 in southern McClain County and western Garvin County were replaced by C_4 cover type in 2010. I suspect the replacement of sparsely vegetated areas by C_4 cover type in pasture/hay areas between 2005 and 2010 was a drought adaptation strategy adopted by local ranchers.

Based on the definition by the NLCD dataset, only pasture/hay fields would have been subject to intensive management practices among the three land cover types (i.e., Shrublands, Grasslands and Pasture/Hay) examined in this study. Based on the inter-annual variations of NPA shown in figure 4.2, beginning with 2002, NPA in C_4 species growing seasons were constantly higher than that in the growing seasons of C_3 species, except for 2005. The higher NPA in C_4 species growing seasons were either related to drought conditions that were less intense than that in the growing seasons of C_3 species (e.g., 2002, 2003 and 2006) or above-normal moisture availability conditions that were not present in C_3 species growing seasons (e.g., 2004, 2007, 2008 and 2009).

In Oklahoma, alfalfa and small grain hay crops such as wheat, rye and oat represent the most important C_3 hays (Edwards *et al.*, 2014). The changes of sown area and production of alfalfa and small grain hay in McClain County and Garvin County between 2007 and 2012 are shown in table 4.12. The sown area and hay production statistics were based on data published in the 2012 USDA Census of Agriculture (USDA, 2012). Except for the sown area of small grains in Garvin County, there were decreases in all the other available statistics for major C_3 hay crops and some decreases were as high as 48% (e.g., the decrease in alfalfa production of Garvin County between 2007 and 2012).

Table 4.12 Changes in the sown area and production of alfalfa and small grains in Garvin and McClain County between 2007 and 2012.

	Garvin County			
	Sown area (acres)		Production (tons)	
	Alfalfa	Small grains	Alfalfa	Small grains
2007	17535	8618	68099	N/A
2012	12951	9278	35311	16125
	McClain County			
	Sown area (acres)		Production (tons)	
	Alfalfa	Small grains	Alfalfa	Small grains
2007	10022	11069	34842	N/A
2012	5745	8548	15361	12946

Note: the statistics for the production of small grains in 2007 were not available.

Since I was not able to find the data regarding the changes of C_4 hay sown area and production between 2007 and 2012 in Garvin and McClain County, I speculated that the increases in C_4 abundance in southern McClain County and western Garvin County might be a result of rancher's preferences to use C_4 species to maintain steady forage or hay production when dramatic decreases of C_3 hays were encountered. The advantage of planting C_4 species lies in that during drought prevalence, such as in the year of 2002, 2003 and 2006, C_4 species could have a higher chance to survive because C_4 species are more tolerant of drought conditions due to their higher water use efficiency (Lattanzi, 2010; Wang *et al.*, 2013).

Significant changes of C_3/C_4 ratio in low elevation areas can also be explained by the land cover changes in pasture/hay fields, because almost all the pasture/hay fields were located in the eastern portion of the study area, where elevation is the below the median.

4.7 Conclusion

In this study, the effects of precipitation anomalies on the relative abundance of C_3 and C_4 grassland species were examined in southwest Oklahoma between 1981 and 2010. The difference between precipitation anomalies in the growing seasons of C_3 and C_4 species fluctuated through the years between 1981 and 2010. The time periods between 1983 and 1988 and between 2002 and 2009 were found to be relatively favorable for the growth of C_3 and C_4 species, respectively. No significant changes of C_3/C_4 ratio were identified during 1983-1988, which was caused by the synchronous changes in NPA during the growing seasons of C_3 and C_4 species. Significant decreases of C_3/C_4 ratio were identified in pasture/hay fields due to the increases in C_4 abundance resulting from the decreases of sparsely vegetated area between 2005 and 2010. I suspect that the increase in C_4 abundance was a drought adaptation strategy adopted by ranchers. Because C_4 species are more tolerant of drought conditions and thus can help to maintain stable forage/hay production when negative precipitation anomalies prevailed during the growing season of C_3 species.

Chapter 5 CONCLUSION

5.1 Summary of conclusions and suggestions for future researches

Before this dissertation, there were no studies attempting to conduct an overall evaluation of the effects of large scale conservation programs on the vegetation dynamics of China's Loess Plateau or to investigate the dynamics of the relative abundance of C_3 and C_4 grassland species at a spatial resolution of 30m. The multiple lines of evidence approach employed in chapter two was proven to be effective in terms of identifying the complex effects exerted by conservation programs on vegetation development of the Loess Plateau. On the one hand, vegetation conservation programs target marginal lands. Thus, significant vegetation increases due to cropland conversion and afforestation can be found in these regions. On the other hand, intensified agricultural production can be found in croplands with suitable topography and well-established irrigation systems which were not enrolled in conservation programs to offset the agricultural production loss caused by vegetation conservation programs elsewhere. In the study shown in chapter two, vegetation changes identified with remotely sensed data were only validated in an indirect manner with archived conservation statistic data. For future researches that have to deal with the validation of changes that might be caused by conservation activities across broad spatial areas, I would suggest the stratified validation by field surveys or based on aerial photos.

In chapter three, the combination of 'patch core areas' with the Random Forest classification algorithm were proved to be of high accuracy to infer historical grassland species in the past. The 'patch core areas' were also found to be able to help identify the important variables in terms of C_3 and C_4 grassland species, which were ignored in previous studies such as elevation and soil clay content.

In chapter four, significant decreases of C_3/C_4 ratio were identified in pasture/hay fields due to the increases in C_4 abundance, resulting from the decreases of sparsely vegetated area between 2005 and 2010. I suspect that the increase in C_4 abundance was a drought adaptation strategy adopted by local ranchers. Because C_4 species are more tolerant of drought conditions and these species can help to maintain stable forage/hay production when negative precipitation anomalies prevailed during the growing season of C_3 species. It is very important to note, however, that the comparison of C_3/C_4 were ratios were made between C_3/C_4 ratios retrieved at two individual years. A biased comparison can occur if significant disturbance events happened in one or both of the two years such as fire or intense drought. To avoid a biased comparison of C_3/C_4 ratios retrieved from two time periods, it would be better if the C_3/C_4 ratio was obtained from a C_3/C_4 classification result that had been aggregated over several years within the targeted time period. The aggregated C_3/C_4 classification result can be generated in a way such as that a pixel will be labelled as being covered by C_3 or C_4 species only if the pixel was classified as C_3 or C_4 cover at multiple years within the targeted time period. The aggregated C_3/C_4 classification result will be able to effectively remove the noise due to random disturbance events such as fire or intense drought.

References

Baxter, M.A., Graves, C.E. & Moore, J.T. (2005) A Climatology of Snow-to-Liquid Ratio for the Contiguous United States. *Weather & Forecasting*, **20**, 729-744.

Bianchi, T.S. & Canuel, E.A. (2011) *Chemical Biomarkers in Aquatic Ecosystems*. Princeton University Press, Princeton, NJ.

Blackshaw, R.E., Moyer, J.R. & Kozub, G.C. (1994) EFFICACY OF DOWNY BROME HERBICIDES AS INFLUENCED BY SOIL PROPERTIES. *Canadian Journal of Plant Science*, **74**, 177-183.

Breiman, L. (1996) Bagging predictors. *Machine Learning*, **24**, 123-140.

Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.

Bremond, L., Boom, A. & Favier, C. (2012) Neotropical C3/C4 grass distributions – present, past and future. *Global Change Biology*, **18**, 2324-2334.

Cabido, M., Ateca, N., Astegiano, M. & Anton, A. (1997) Distribution of C_3 and C_4 grasses along an altitudinal gradient in Central Argentina. *Journal of Biogeography*, **24**, 197-204.

Cao, S., Chen, L. & Yu, X. (2009) Impact of China's Grain for Green Project on the landscape of vulnerable arid and semi-arid agricultural regions: a case study in northern Shaanxi Province. *J. Appl.Ecol.*, **46**, 536-543.

Cavagnaro, J.B. (1988) Distribution of C_3 and C_4 grasses at different altitudes in a temperate arid region of Argentina. *Oecologia*, **76**, 273-277.

Chen, L., Wei, W., Fu, B. & Lu, Y. (2007) Soil and water conservation on the Loess Plateau in China: review and perspective. *Progress in Physical Geography*, **31**, 389-403.

Chinese National Meteorological Information Center. (2008) Monthly surface meteorological dataset of China. In. China Meteorological Data Sharing Service System, Beijing, China.

Chow, G. (2006) Are Chinese Official Statistics Reliable? *CESifo Economic Studies*, **52**, 19.

Curtis, N.M., Ham, W.E. & Johnson, K.S. (2008) Geomorphic Provinces of Oklahoma. *Earth Sciences and mineral resources of Oklahoma* (ed. by K.S. Johnson and K.V. Luza). Oklahoma Geological Survey, Norman, OK.

de Beurs, K.M. & Henebry, G.M. (2004) Trend Analysis of the Pathfinder AVHRR Land (PAL) NDVI Data for the Deserts of Central Asia. *IEEE Geosci. Remote Sens. Lett.*, **1**, 282-286.

de Beurs, K.M., Yan, D. & Karnieli, A. (2013) The effect of large scale conservation programs on the vegetative development of China's Loess Plateau *Dryland East Asia: land dynamics amid social and climate change* (ed. by J. Chen, S. Wan, G. Henebry, J. Qi, G. Gutman, G. Sun and M. Kappas). De Gruyter, Berlin, Germany and Boston, USA.

Duck, L.G. & Fletcher, J.B. (1945) The Game Types of Oklahoma. *A survey of the game and furbearing animals of Oklahoma*. Oklahoma Game and Fish Commission, Division of Wildlife Restoration and Research, Oklahoma City, OK.

Durre, I., Squires, M.F., Vose, R.S., Yin, X., Arguez, A. & Applequist, S. (2013) NOAA’s 1981–2010 U.S. Climate Normals: Monthly Precipitation, Snowfall, and Snow Depth. *Journal of Applied Meteorology and Climatology*, **52**, 2377-2395.

Edwards, J., Warren, J. & Redfearn, D. (2014) PSS-2071: Sod-seeding winter wheat in bermudagrass pastures. In. Oklahoma State University, Still water, OK.

Ehleringer, J. & Björkman, O. (1977) Quantum Yields for CO_2 Uptake in C_3 and C_4 Plants. *Plant Physiology*, **59**, 86-90.

Ehleringer, J.R., Cerling, T.E. & Helliker, B.R. (1997) C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*, **112**, 285-299.

Ellis, E.C. & Ramankutty, N. (2008) Putting people in the map: anthropogenic biomes of the world. *Front. Ecol. Environ.*, **6**, 439-447.

Ellis, E.C., Klein Goldewijk, K., Siebert, S., Lightman, D. & Ramankutty, N. (2010) Anthropogenic transformation of the biomes, 1700 to 2000. *Glob. Ecol. Biogeogr.*, **19**, 589-606.

Epstein, H.E., Lauenroth, W.K., Burke, I.C. & Coffin, D.P. (1997) Productivity Patterns of C3 and C4 Functional Types in the U.S. Great Plains. *Ecology*, **78**, 722-731.

Feng, Z.M., Yang, Y.Z., Zhang, Y.Q., Zhang, P.T. & Li, Y.Q. (2005) Grain-for-green policy and its impacts on grain supply in west China. *Land Use Policy*, **22**, 301-312.

Foody, G.M. & Dash, J. (2007) Discriminating and mapping the C3 and C4 composition of grasslands in the northern Great Plains, USA. *Ecological Informatics*, **2**, 89-93.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. & Huang, X.M. (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.*, **114**, 168-182.

Gale, F. (2002) China's Statistics: Are they reliable? *USDA ERS Agric. Inf. Bull.*, **775**, 50-53.

Gong Li, S., Harazono, Y., Oikawa, T., Zhao, H.L., Ying He, Z. & Chang, X.L. (2000) Grassland desertification by grazing and the resulting micrometeorological changes in Inner Mongolia. *Agricultural and Forest Meteorology*, **102**, 125-137.

Goodin, D.G. & Henebry, G.M. (1997) A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model. *Remote Sensing of Environment*, **61**, 270-278.

Guo, Q.F. (2000) Climate change and biodiversity conservation in Great Plains agroecosystems. *Global Environmental Change-Human and Policy Dimensions*, **10**, 289-298.

Hanberry, B.B., He, H.S. & Palik, B.J. (2012) Comparing Predicted Historical Distributions of Tree Species Using Two Tree-based Ensemble Classification Methods. *The American Midland Naturalist*, **168**, 443-455.

Hansen, J., Fuller, F.H. & Hsu, H. (2003) Sources of Discontinuity and Uncertainty in Chinese Agricultural Data. In: *Staff General Research Papers*. Department of Economics,Iowa State University, Ames, USA.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning:Data Mining, Inference, and Prediction, Second Edition*. Springer, New York City, NY.

He, X.B., Zhou, J., Zhang, X.B. & Tang, K.L. (2006) Soil erosion response to climatic change and human activity during the Quaternary on the Loess Plateau, China. *Reg. Environ. Chang.*, **6**, 62-70.

Hilker, T., Natsagdorj, E., Waring, R.H., Lyapustin, A. & Wang, Y. (2014) Satellite observed widespread decline in Mongolian grasslands largely due to overgrazing. *Global Change Biology*, **20**, 418-428.

Hoagland, B.W. (2008) Vegetation of Oklahoma. *Earth Sicences and Mineral Resources of Oklahoma* (ed. by K.S. Johnson and K.V. Luza). Oklahoma Geological Survey, Norman, OK.

Huete, A.R. (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, **25**, 295-309.

Johnson, H.L. (2008) Climate of Oklahoma. *Earth Sciences and mineral resources of Oklahoma* (ed. by K.S. Johnson and K.V. Luza). OKlahoma Geological Survey, Norman, OK.

Kauppi, P.E., Ausubel, J.H., Fang, J., Mather, A.S., Sedjo, R.A. & Waggoner, P.E. (2006) Returning forests analyzed with the forest identity. *Proceedings of the National Academy of Sciences*, **103**, 17574-17579.

Kellner, J.R., Asner, G.P., Cordell, S., Thaxton, J.M., Kinney, K.M., Kennedy-Bowdoin, T., Knapp, D.E., Questad, E.J. & Ambagis, S. (2012) Historical Land-Cover Classification for Conservation and Management in Hawaiian Subalpine Drylands. *Pacific Science*, **66**, 457-466.

Kelly, P. & Huo, X. (2013) Do farmers or governments make better land conservation choices? Evidence from China's sloping land conversion program. *Journal of Forest Economics*, **19**, 32-60.

Lattanzi, F.A. (2010) C3/C4 grasslands and climate change. *Grassland Science in Europe*, **15**, 3-13.

Liaw, A. & Wiener, M. (2002) Classification and Regression by randomForest. *R news*, **2**, 18-22.

Liu, J., Li, S., Ouyang, Z., Tam, C. & Chen, X. (2008) Ecological and socioeconomic effects of China's policies for ecosystem services. *Proc. Natl. Acad. Sci. USA*, **105**, 9477-9482.

Liu, J.G. & Diamond, J. (2005) China's environment in a globalizing world. *Nature*, **435**, 1179-1186.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Feng, G., Kutler, J. & Teng-Kui, L. (2006) A Landsat surface reflectance dataset for North America, 1990-2000. *Geoscience and Remote Sensing Letters, IEEE*, **3**, 68-72.

Mather, A.S. (2007) Recent asian forest transitions in relation to forest-transition theory. *Int. For. Rev.*, **9**, 491-502.

Mather, A.S. & Needle, C.L. (1998) The forest transition: A theoretical basis. *Area*, **30**, 117-124.

McGarigal, K. (2013) *FRAGSTATS 4.2 Help*. University of Massachusetts, Amherst, MA.

McSherry, M.E. & Ritchie, M.E. (2013) Effects of grazing on grassland soil carbon: a global review. *Global Change Biology*, **19**, 1347-1357.

O'Geen, A.T. (2012) Soil Water Dynamics. *Nature Education Knowledge* **3**, 1.

O'Mara, F.P. (2012) The role of grasslands in food security and climate change. *Annals of Botany*, **110**, 1263-1270.

Paulsen, G.M. (1997) Growth and Development. *Wheat Production Handbook*. Kansas State University Agricultural Experiment Station and Cooperative Extension Service, Manhattan, KS.

Peterson, T.C., Vose, R., Schmoyer, R. & Razuvaëv, V. (1998) Global historical climatology network (GHCN) quality control of monthly temperature data. *International Journal of Climatology*, **18**, 1169-1179.

Qi, Y., Chang, Q., Jia, K., Liu, M., Liu, J. & Chen, T. (2012a) Temporal-spatial variability of desertification in an agro-pastoral transitional zone of northern Shaanxi Province, China. *Catena*, **88**, 37-45.

Qi, Y., Dong, Y., Peng, Q., Xiao, S., He, Y., Liu, X., Sun, L., Jia, J. & Yang, Z. (2012b) Effects of a conversion from grassland to cropland on the different soil organic carbon fractions in Inner Mongolia, China. *Journal of Geographical Sciences*, **22**, 315-328.

Qu, G. (1999) Environmental Protection Knowledge. In. China's Red Flag Publishing House, Beijing.

Richards, J.A. & Jia, X. (2006) *Remote Sensing Digital Image Analysis: An introduction*, 4th edn. Springer, New York City, NY.

Ricotta, C., Reed, B.C. & Tieszen, L.T. (2003) The role of C3 and C4 grasses to interannual variability in remotely sensed ecosystem performance over the US Great Plains. *International Journal of Remote Sensing*, **24**, 4421-4431.

Roebber, P.J., Bruening, S.L., Schultz, D.M. & Cortinas, J.V. (2003) Improving Snowfall Forecasting by Diagnosing Snow Density. *Weather and Forecasting*, **18**, 264-287.

Rudel, T.K., Coomes, O.T., Moran, E., Achard, F., Angelsen, A., Xu, J. & Lambin, E. (2005) Forest transitions: towards a global understanding of land use change. *Global Environmental Change*, **15**, 23-31.

Samson, F. & Knopf, F. (1994) Prairie conservation in North-America. *Bioscience*, **44**, 418-421.

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X.W., Tsang, T., Strugnell, N.C., Zhang, X.Y., Jin, Y.F., Muller, J.P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R.P., Hu, B.X., Liang, S.L., Privette, J.L. & Roy, D. (2002) First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.*, **83**, 135-148.

Schacht, W.H., Volesky, J.D., Bauer, D., Smart, A.J. & Mousel, E.M. (2000) Plant community patterns on upland prairie in the eastern Nebraska Sandhills. *Prairie Naturalist*, **32**, 43-58.

Schnitzer, S., Seitz, F., Eicker, A., Güntner, A., Wattenbach, M. & Menzel, A. (2013) Estimation of soil loss by water erosion in the Chinese Loess Plateau using Universal Soil Loss Equation and GRACE. *Geophys. J. Int.*, **193**, 1283-1290.

Statistic Bureau of Ordos. (2000-2008) *Statistical Yearbook of Ordos*. China Statistics Press, Beijing, China.

Statistic Bureau of Sanmenxia. (2000-2001, 2003-2008) *Statistical Yearbook of Sanmenxia*. China Statistics Press, Beijing, China.

Statistic Bureau of Xianyang. (2008) *Statistical Yearbook of Xianyang*. China Statistics Press, Beijing, China.

Statistic Bureau of Yan'an. (1999-2008) *Statistical Yearbook of Yan'an*. China Statistics Press, Beijing,China.

Statistic Bureau of Yulin. (1999-2007) *Statistical Yearbook of Yulin*. China Statistics Press, Beijing, China.

Tachikawa, T., Hato, M., Kaku, M. & Iwasaki, A. Characteristics of ASTER GDEM version 2. *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* (ed by, pp. 3657-3660.

Tachikawa, T., Hato, M., Kaku, M. & Iwasaki, A. (2011) Characteristics of ASTER GDEM version 2. *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* (ed by, pp. 3657-3660.

Tieszen, L.L., Reed, B.C., Bliss, N.B., Wylie, B.K. & DeJong, D.D. (1997) NDVI, C_3 and C_4 Production, and Distributions in Great Plains Grassland Land Cover Classes. *Ecological Applications*, **7**, 59-78.

Tucker, C.J. (1979) RED AND PHOTOGRAPHIC INFRARED LINEAR COMBINATIONS FOR MONITORING VEGETATION. *Remote Sensing of Environment*, **8**, 127-150.

Tyrl, R.J., Bidwell, T.G., Masters, R.E. & Elmore, R.D. (2008) *Field guide to Oklahoma Plants:commonly encountered prairie,shrubland, and forest species.*, Second edn. Oklahoma State University, Stillwater,OK.

Uchida, E., Xu, J.T. & Rozelle, S. (2005) Grain for green: Cost-effectiveness and sustainability of China's conservation set-aside program. *Land Econ.*, **81**, 247-264.

USDA (2012) *2012 Census of Agriculture: Summary and State data*. United States Department of Agriculture, Washington, DC.

Wang, C.Z., Hunt, E.R., Zhang, L. & Guo, H.D. (2013) Phenology-assisted classification of C_3 and C_4 grasses in the US Great Plains and their climate dependency with MODIS time series. *Remote Sensing of Environment*, **138**, 90-101.

Wei, J., Zhou, J., Tian, J., He, X. & Tang, K. (2006) Decoupling soil erosion and human activities on the Chinese Loess Plateau in the 20th Century. *Catena*, **68**, 10-15.

Wicks, G.A., Burnside, O.C. & Fenster, C.R. (1971) Influence of Soil Type and Depth of Planting on Downy Brome Seed. *Weed Science*, **19**, 82-86.

Winslow, J.C., Hunt Jr, E.R. & Piper, S.C. (2003) The influence of seasonal water availability on global C_3 versus C_4 grassland biomass and its implications for climate change research. *Ecological Modelling*, **163**, 153-173.

Xian, G., Homer, C. & Fry, J. (2009) Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. *Remote Sensing of Environment*, **113**, 1133-1147.

Xu, Z.G., Xu, J.T., Deng, X.Z., Huang, J.K., Uchida, E. & Rozelle, S. (2006) Grain for green versus grain: Conflict between food security and conservation set-aside in China. *World Development*, **34**, 130-148.

Zhang, P., Shao, G., Zhao, G., Le Master, D.C., Parker, G.R., Dunning, J.B. & Li, Q. (2000) China's forest policy for the 21st century. *Science*, **288**, 2135-2136.

Zheng, M., Qin, F., Yang, J. & Cai, Q. (2013) The spatio-temporal invariability of sediment concentration and the flow–sediment relationship for hilly areas of the Chinese Loess Plateau. *Catena*, **109**, 164-176.

Zhou, D., Zhao, S. & Zhu, C. (2012) The grain for green project induced land cover change in the Loess Plateau: a case study with Ansai County, Shaanxi Province, China. *Ecological Indicators*, **23**, 88-94.

Zhou, H., Van Rompaey, A. & Wang, J. (2009) Detecting the impact of the "grain for green" program on the mean annual vegetation cover in the Shaanxi Province, China using SPOT-VGT NDVI data. *Land Use Policy*, **26**, 954-960.