A Data-driven Approach for Mapping Grasslands at a Regional Scale.

By

Dana L. Peterson

M.A. University of Kansas, 1999
B.G.S., University of Kansas, 1996

Submitted to the graduate degree program in Geography and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Stephen L. Egbert

Jude H. Kastens

Kelly Kindscher

J. Chris Brown

William C. Johnson

Date Defended: 23 January 2019

The dissertation committee for Dana L. Peterson certifies that this is the approved version of the following dissertation:

A Data-driven Approach for Mapping Grasslands at a Regional Scale

Chair: Stephen L. Egbert

Date Approved: 23 January 2019

Abstract

The goal of this research was to use a data-driven approach to develop a regional scale grassland mapping protocol with the following objectives. First, identify and characterize the spatial distribution of grassland types and land use across Kansas as well as the static or dynamic nature of grasslands over time using multi-year U.S. Department of Agriculture (USDA) Farm Service Agency (FSA) 578 data. Second, evaluate the spectral separability of four hierarchies of grassland types and land use using FSA 578 data, multi-seasonal Landsat 8 spectral bands, Landsat 8 Normalized Difference Vegetation Index (NDVI) data, and Moderate Resolution Imaging Spectrometer (MODIS) NDVI time series. Third, determine the optimal data combination, and the appropriate thematic resolution, for mapping grassland type by evaluating the modeling performance of the Random Forest (RF) classifier.

A county-level analysis of the multi-year FSA 578 data found that the data were not all-inclusive of total grasslands across Kansas, but were sufficient to illustrate regional trends in grassland type, land use, and field size. Eastern Kansas was found to be more diverse in grassland type, more variable in land use, and contained a high number of smaller fields. Conversely, western Kansas consisted of larger fields that were primarily grazed native grasslands and land enrolled in the Conservation Reserve Program (CRP). These results indicate a more complex grassland landscape to map in eastern Kansas, while also providing guidance for training sample distributions for image classification.

Jeffries-Matusita (JM) distance statistics were calculated for three-date multispectral Landsat 8, three-date Landsat 8 NDVI, and 23-period, 16-day composite Terra MODIS NDVI time series. The results indicate that combining the three datasets maximized the spectral separability of grassland types across all four grassland-type hierarchies. A comparison of the

three datasets showed that multispectral Landsat 8 data had the highest JM distance statistics (which indicates the most separability). JM distance statistics calculated by-band and by-period consistently showed that information from spring and fall was more important than summer for separating grassland types. The results showed lower separability for land-use classes within a grassland type versus between grassland types. The spectral separability of pairwise comparisons incorporating land use between grassland types varied, indicating that land use does affect spectral separability in some instances. On the other hand, JM distance statistics did not substantially drop when more refined grassland types were aggregated to coarser grassland type classes (e.g. Level-1: cool- and warm-season), indicating that land use does not negatively affect the spectral separability of functional grassland types. The results indicate low spectral separability between brome and fescue but moderate to high separability between native and CRP, suggesting the use of a Level-1 or Level-2 thematic classification scheme for the study area.

Finally, random forest models were constructed and evaluated using 2015 FSA 578 data and four datasets of remotely sensed data in two adjacent Landsat scenes (path/rows). Models were created for each of the four grassland hierarchies. The results showed that out-of-bag (OOB) error increased with grassland hierarchy complexity (the number of thematic classes) and OOB error was lowest for the combined remotely sensed dataset. Mapping CRP as a separate grassland type resulted in low producer's accuracy levels, with CRP largely mapped as warm-season grasslands, suggesting the Level-1 classification scheme was appropriate for regional mapping of grassland types. Path/rows 27/33 and 28/33 had OOB overall accuracy levels of 87% and 92%, respectively. User's and producer's accuracy levels indicate that cool-season grasslands were mapped more accurately in path/row 27/33 where that class is more dominant

than in 28/33. Using test data (withheld verification data) unexpectedly increased overall accuracy levels by 4% and 6% over OOB accuracies, which may have resulted from varying data proportions between OOB and test data, suggesting the need for further evaluation.

Acknowledgments

The research which is the subject of this dissertation has been financed, in part, with federal funds from the Fish and Wildlife Service, a division of the United States Department of Interior, and administered by the Kansas Department of Wildlife, Parks, and Tourism. The contents and opinions, however, do not necessarily reflect the views or policies of the United States Department of Interior or the Kansas Department of Wildlife, Parks, and Tourism. The research was also partially funded by the Kansas Water Office, under contract Number 16-107: "Kansas Land Cover Mapping Initiative: A 10-Year Update". I would also like to thank the Kansas Farm Service Agency, for sharing data under the Memorandum of Understanding (MOU). These data were instrumental in this research for the development of the grassland cover mapping protocols. Lastly, I acknowledge the Biofuels and Climate Change - Farmers' Land Use Decisions (BACC–FLUD) project (supported by the National Science Foundation, Award Number EPS-0903806) for purchasing the 2004-2007 FSA 578 data and CLU boundaries from 2007 from Farm Market iD.

A warm thank you to my doctoral committee members: Dr. Stephen Egbert, Dr. Jude Kastens, Dr. Kelly Kindscher, Dr. J. Chris Brown and Dr. William Johnson. I appreciate the guidance, constructive conversations and recommendations the committee made to improve my research. Dr. Stephen Egbert, I appreciate the countless hours of discussions, holding me accountable and encouraging me to make progress. Your positive attitude and affirmations helped me more than you know. I thank Dr. Jude Kastens for his help in learning MatLab, sharing code, and for the numerous discussions that refined and strengthened my research. I greatly appreciate the

meticulous editing and feedback Dr. Stephen Egbert and Dr. Jude Kastens provided on my dissertation.

I would like to thank the Kansas Applied Remote Sensing Program (KARS) at Kansas Biological Survey (KBS) (Dr. Edward Martinko, Director) for the amazing staff and facilities to conduct research. The interdisciplinary expertise at the KBS is unique and remarkable. During my career at KBS, I have been fortunate to have three mentors, Dr. Stephen Egbert, Dr. Mark Jakubauskas and Jerry Whistler. I would not be where I am today without working for and alongside them in numerous engaging research projects. A special thank you to Dr. Jerry deNoyelles, Jennifer Delisle, Brandy Hildreth, Debbie Baker, Jerry Whistler, Gina Ross, Mike Houts, John Lomas, and Chris Bishop. I grateful for the help Chris Bishop and John Lomas provided in preprocessing imagery used in my research.

Finally, many thanks to my family and friends who have provided love and support throughout this endeavor. To my parents, J.C. and Mary Gumm, thank you for teaching me perseverance and tenacity, foundations of a strong work ethic. And to do the best you can, that is all anyone can ask of you. And I thank my sisters, Tammy and Terri, for the numerous phone conversations after many long days at work and for encouraging me to keep going. Lastly, 25 years ago I married my best friend, Leland. Leland, you have always been by and on my side, cheering me on in life's challenges. I appreciate the sacrifices you've made to allow me to pursue my Ph.D. And to my sons, Charlie and Calvin, my greatest teachers, you teach me about faith, patience, and grace and inspire me to be the best mother and person I can be.

Table of Contents

List of Figures

List of Tables

Introduction

1.1 *Background and Problem Statement*

Grasslands cover 40.5% of the earth's surface, more than either forest or cropland (Gibson, 2009). Grasslands provide habitat to support wildlife, forage for domestic livestock, serve as filters for water quality, provide venues for recreational interests, and serve as a major global carbon sink. While expansive, grasslands are potentially the most threatened biome due to land conversion and intensive land use (Samson *et al.*, 2004). Globally, the conversion of grassland to cropland represents the leading cause of landscape fragmentation and lost grassland extent (Gibson, 2009). In addition, the quality of remaining native grassland has been modified or degraded by invading non-native species, fire suppression and overgrazing by domestic livestock (Weaver, 1954; Gibson, 2009; Risser, 1988).

The tallgrass prairies of the Great Plains in North America, considered one of the more biologically diverse grasslands in the world (Risser, 1988), have the greatest reported loss of grassland area with estimates of only 9.4% - 13% of the original tallgrass prairie remaining (Gibson, 2009; Samson *et al.*, 2004). It has been estimated that the tallgrass prairie once occupied 167 million acres, stretching east into western Ohio, west to the eastern third of Kansas and Nebraska, north into southern Manitoba, Canada and south into portions of Texas (Robertson *et al.*, 1997). Fragmentation of the tallgrass prairie in the eastern Great Plains began in the early 1800s when European settlers converted "the Great American Desert" into cropland and non-native grasslands for domestic livestock grazing (Samson *et al.*, 2004). Tallgrass prairie remnants remain almost exclusively on rocky substrates that are unable to be plowed. Furthermore, most of the Great Plains and eastern tallgrass prairie remnants are privately owned and subjected to a variety of land management practices, including grazing and haying for

domestic livestock (Owensby, 1993). Reports indicate that 18% of tallgrass prairie remains in Kansas, the largest percent of any state. Furthermore Kansas has the largest contiguous tract of tallgrass prairie located in the hilly region in eastern Kansas known as the Flint Hills (Risser, 1988). Meanwhile other states, including Indiana, Illinois, Iowa, Minnesota, Missouri, North Dakota ,and Missouri contain less than a half percent of their original tallgrass prairie (Risser, 1988; Robertson & Schwartz, 1994b). Mapping and monitoring the extent, distribution, and condition of remaining tallgrass prairie are critical to ensure preservation and sustainability of these biologically diverse grasslands.

Accurate and ongoing mapping of the landscape provides tools for understanding the changing landscape, including the environmental and socio-economic drivers, and provides tools for planning and conservation. For decades researchers have used remotely sensed data to map and monitor grasslands, including the tallgrass prairie. Studies have used remote sensing technology to monitor and model biophysical characteristics of grasslands, including functional distributions (i.e. C3 and C4 grasslands), productivity (biomass and cover) and grassland use that can alter grassland biophysical characteristics and quality. For example, several studies have used remotely sensed data to map or predict distributions and abundance of C3 and C4 grasslands. Tieszen *et al.* (1997) used time series AVHRR Normalized Difference Vegetation Index (NDVI) data to characterize the spatial and temporal distribution of C3 and C4 grasslands in the Great Plains over a five-year period. Davidson and Csillag (2003) compared three approaches using AVHRR NDVI to predict the relative abundance of C4 cover in a Canadian mixed-grass prairie. They found a two-date ratio, early season NDVI to late season NDVI, best predicted C4 abundance (Davidson & Csillag, 2003). Meanwhile Foody and Dash (2007) used a 30-week time series of MERIS Terrestrial Chlorophyll Index (MTCI) data to map high, medium

and low C3 cover in South Dakota with an overall accuracy of 77%. In addition, Gu and Wylie (2015) leveraged the spatial resolution of Landsat NDVI and the temporal resolution of MODIS NDVI in a rule-based piecewise regression to produce a 30-m grassland productivity map of the Greater Platte River Basin, Nebraska. Understanding productivity and the abundance of C3 and C4 grasslands is important as the two grassland types respond differently to environmental change due to grazing intensity, fire frequency, nutrient regimes, and climate change (Tieszen *et al.*, 1997).

Other studies have used remotely sensed data to map thematic grassland classes that are represented by either their dominant functional group or as native and non-native grassland types. Using multi-seasonal ASTER NDVI, Wang *et al.* (2010) mapped cool-season (non-native) and warm-season (native) grasslands in western Missouri with an accuracy of 80%. The authors found that spring and summer NDVI provided the highest separability between these two grassland types due to their asynchronous phenology, with maximum productivity reached in May and July for cool- versus warm-season grasslands, respectively. Another study showed discriminant analysis and MODIS NDVI time series spectrally separated native and non-native dry mixed-grass prairie in Alberta, Canada with an overall accuracy of 73% (McInnes *et al.*, 2015). Meanwhile a mapping effort by Peterson *et al.* (2008) found that multi-seasonal Landsat Thematic Mapper (TM) data better separated native (warm-season) and non-native (cool-season) grassland types in the Flint Hills ecoregion than coarser resolution MODIS NDVI time series. Many of these studies and mapping efforts rely on the asynchronous phenology of cool- and warm-season grasslands. However, grasslands are used and managed extensively and intensively. The type, combination, timing and intensity of land management practices within grassland types alter the biophysical properties of grasslands, including vegetation productivity

and composition, and soil structure and chemistry, which in turn results in altered spectral responses that complicate the ability to accurately map grassland types. Several studies have used remotely sensed data to characterize and monitor land management practices and land use intensity occurring within grasslands. For example, Guo *et al.* (2003) and Guo *et al.* (2000) used multi-seasonal field data and Landsat TM imagery to show that biophysical and spectral characteristics were significantly different among three common land management practices in cool-season (non-native) and warm-season (native) grasslands in Douglas County, Kansas. Discriminant analysis showed the two grassland types and the three treatments in the two grassland types could be separated with an accuracy of 90.1% and 70.4%, respectively (Price *et al.*, 2002a). Peterson *et al.* (2002b) obtained similar results when using discriminant analysis to separate grazed cool- and warm-season grasslands in the same county. Another study by Lauver and Whistler (1993) found significant differences in the biophysical characteristics (species diversity, plant cover, and biomass) of high-quality (hayed) and low quality (overgrazed) tallgrass prairie remnants in Anderson County, Kansas that were mapped using single-date Landsat TM data and probability thresholding with moderate success (63% overall accuracy). Another study by Franke *et al.* (2012) found that multi-temporal RapidEye data and a decision tree classifier could map grassland land use intensity in a 500 km^2 grassland area in Germany with accuracies up to 85.7%. A recent study Halabuk *et al.* (2015) used MODIS NDVI and EVI to detect haying events in prairie hay meadows in Slovakia with accuracy levels as high as 85%. While these studies provide examples of successful results for grassland mapping and monitoring, they primarily occur on a relatively small scale.

However, there are land cover datasets that contain grassland information at larger scales. One national mapping effort was coordinated under the Multi-Resolution Land Characteristics

(MRLC) Consortium that began in the 1980s. Initially the MRLC was developed as a means to build and share a national Landsat imagery archive using agreed-upon image processing standards. Given that the cost of a single Landsat image in the mid-1990s was $3,000-$4,000, multi-temporal, regional scale mapping was too costly for federal or state agencies or academic institutions (Wulder *et al.*, 2012). In 2008 a data policy change made Landsat data freely available. The MRLC consortium's focus shifted from creating a national imagery archive to creating and maintaining a series of national land cover datasets produced collaboratively by members in the consortium (Wickham *et al.*, 2014). The MRLC Consortium land cover products include the National Land Cover Dataset (NLCD), Coastal Change Analysis Program (C-CAP), Gap Analysis Program (GAP), and Landscape Fire and Resource Management Planning Tools (LANDFIRE) (Wickham *et al.*, 2014).

The NLCD is a national mapping effort led by USGS. National NLCD databases were produced in 1992, 2001, 2006, 2011 and 2016, switching from a ten-year to five-year update cycle (Wickham *et al.*, 2014). The NLCD maps sixteen land cover classes at a 30m spatial resolution. "Grassland/Herbaceous" and "Pasture/Hay" are the two grassland classes in the NLCD, but the Pasture/Hay class can include native hay meadows or non-native grasslands, as well as alfalfa. The 2011 NLCD was produced using Landsat imagery and ancillary geospatial data and a decision tree classifier (DTC) (Homer *et al.*, 2015) and had reported accuracy level of 82% (Wickham *et al.*, 2017). And previous national NLCD Level II (2001 and 2006) products have reported overall accuracy levels of ~85%; however, accuracy levels vary by region.

In 1999 the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) began producing an annual Cropland Data Layer (CDL) for several states that has now grown into a national mapping product. As the name implies, the effort

focuses on mapping crop types. The spatial and temporal resolution of imagery used has varied over the years, but more recently NASS has used a combination of Landsat 8 and MODIS NDVI and ancillary data (soils, topography) in a decision tree classifier (DTC). The USDA Farm Service Agency's (FSA) annotated Common Land Unit (CLU) database is used for image classification model development and accuracy assessment. Map accuracy levels vary by crop type, and for the 2015 CDL the overall accuracy level was reported at 85%. For non-cropland classes (grassland, woodland, etc.), however, the CDL uses the NLCD map for training and validation, with no accuracy levels reported for these classes. The two grassland classes are "Other Hay/Non-Alfalfa" and "Grassland/Pasture". Another potential issue with using the CDL is that in some years the CDL grouped grassland enrolled in the Conservation Reserve Program with fallow/idle cropland but in other years grouped CRP with the pasture/grassland class. Given that the CDL does not focus on mapping non-cropland classes, NASS refers end-users to the NLCD for those classes. While both the NLCD and CDL contain aggregate grassland classes, the classes primarily represent land use and do not distinguish land cover in terms of their functional group or classification as native, non-native, and CRP.

The land cover from the GAP was initially produced in the mid-to late-1990s with a mapping focus on natural vegetation using NatureServe's Hierarchical Ecological System for classification. The national GAP mapping effort collaborated with states to develop their state-wide map product. As a result, independent supervised classification approaches (which potentially varied by classifier, thematic resolution, minimum mapping unit, training data, etc.) were used, making it difficult to edge-match states. In Kansas, 40 alliance-level natural vegetation classes were mapped with varying success. The Kansas GAP map overall map had classification accuracies of 89%, 66%, and 52% for Anderson Level I, Formation Level, and

Alliance Level mapping, respectively. GAP has been updated by the USGS for several regions in the US, but not for the North Central Region where Kansas is located.

While land cover maps containing grassland classes represent a snapshot in time, they can be used in a time series to help understand landscape changes and respective drivers of landscape change. For example, Drummond (2007) created five land cover maps between 1973 and 2000 in two ecoregions in the Great Plains to create a time series of regional loss and expansion of grasslands. The author found distinct temporal trends in the conversions between cropland and grassland that were attributed to changes in socioeconomics and policy. A somewhat controversial study by Wright and Wimberly (2013) used the CDL from 2006 and 2011 to quantify the large conversions of grassland to cropland that the authors claim to be the result of increased soy and corn production. However, a response (Cooper, 2015) expressed concerns with regard to how the thematic data were aggregated and handled in the change detection analysis.

1.2 *Research Objectives*

While previous studies in eastern Kansas have evaluated the biophysical characteristics of grasslands and have used field and satellite-acquired spectral data to statistically discriminate between grassland types and land management practices, little research has focused on identifying an optimal classification approach using satellite imagery for mapping grassland types at a regional scale. The goal of this research was to determine an optimal classification approach, i.e., which combination of remotely sensed imagery and thematic classification scheme most accurately maps dominant grassland types (warm- and cool-season) across eastern Kansas. To achieve this research goal there were three main objectives.

1. Identify the dominant land use within the two grassland types (warm- and cool-season grasslands) using United States Department of Agriculture (USDA) Farm Service Agency (FSA) data and characterize the static or dynamic nature of land use in grassland types in eastern Kansas.
2. Determine the spectral separability of grassland types and land use using multi-seasonal Landsat 8 spectral bands, Landsat 8 NDVI, and Moderate Resolution Imaging Spectrometer (MODIS) NDVI time series.
3. Determine the optimal combination of data for mapping and the appropriate thematic resolution for mapping grassland type by comparing modeling performance using a Random Forest (RF) modeling approach.

These three objectives represent three research components in the dissertation.

1.2.1 *Research Component 1: Characterizing County-Level Spatial and Temporal Distributions of Grassland Types and Land Use in Kansas*

The first research component identified dominant land use occurring in warm- and cool-season grasslands in eastern Kansas over a six-year period (2004-2007, and 2015). Furthermore, the research evaluated whether land use within grasslands remains static or changes inter-annually. USDA FSA data were analyzed at the county level and at the field level. Both native and non-native grasslands are managed landscapes where a combination of land management practices are utilized to maintain and maximize vegetation productivity. The timing, frequency and intensity of a land management practice and the combination of practices varies by grassland type and by land owner, which in turn complicates mapping grassland types using remotely sensed imagery. While previous research indicated that land management (or land use) complicates the mapping of warm- and cool-season grassland types, there is no documented

information on the prevalence of land use between and within grassland types in Kansas or on the static or dynamic nature of land use over time. Without an understanding of land use trends, it is difficult to fully understand the impact land use has on the ability to map warm- and cool-season grassland types. Furthermore, testing the hypothesis that inter-annual grassland type and land use within grasslands are static can shed light on the efficacy of using out-of-year training data for image classification when within-year training data are unavailable.

1.2.2 *Research Component 2: Exploring the Spectral Characteristics and Separability of Four Grassland Type Hierarchies Using Landsat 8 and MODIS NDVI*

This research component used Jeffries-Matusita (JM) distance statistics and spectral profiles to compare the spectral separability of four hierarchies of grassland types in northeastern Kansas. Three remotely sensed datasets from 2015 (three-date Landsat 8 multispectral, three-date Landsat NDVI, and 23-period, 16-day composite MODIS NDVI time series) and 2015 reference data from the USDA FSA were used in the analyses. The results will be used to determine the optimal dataset(s) for regional scale mapping of grassland types, the hierarchy of grassland types to be mapped, and whether land use affects the spectral separability of grassland types.

1.2.3 *Research Component 3: Evaluating the Utility of Random Forest and Landsat 8 and MODIS NDVI Data for Mapping Grassland Types at a Regional Scale*

The third research component compared data combinations and thematic classification hierarchies for optimal mapping of grassland type across two Landsat 8 scenes (Landsat Worldwide Reference System (WRS) path/rows). Specifically, this objective compares model development and performance of Random Forest (RF) using multispectral Landsat 8, Landsat 8 NDVI, and MODIS NDVI separately and then combined. Out-of-Bag (OOB) errors from the

models were used to identify the optimal thematic grassland hierarchy for mapping grasslands and the optimal data input for this study. Models were applied to the data to produce grassland type maps for both path/rows and both OOB and test data were used to assess accuracy levels for the maps produced. The results of this research will be used to formulate a methodology for mapping functional grassland types at a regional scale.

1.3 *Research Design*

1.3.1 *Study Area:*

The general focus of the study area is on grasslands in the central Great Plains, but the study areas for the individual research components vary on data and topic. For the first research component Kansas was defined as the study area. For the second research component, the study area was defined as the area of Landsat path/row 27/33 that falls within Kansas (Figure 1). For the third research component, the study area consisted of the area within two adjacent Landsat path/rows (27/33 and 28/33) that fall in Kansas.

Kansas exhibits an east-west precipitation gradient and a north-south temperature gradient, with higher precipitation occurring in the east and lower temperatures occurring in the north. Kansas grasslands follow the east-west precipitation gradient with tallgrass prairie in the relatively wet east, mixed prairie in central Kansas and shortgrass prairie in the dryer west.

Haying and grazing are two of the common land uses for both grassland types. However, the timing, intensity and frequency of management practices within each grassland type vary by land owner and by climate conditions in a given year. In addition to grazing, prescribed burning is a commonly used management practice to maintain species diversity and prevent woody encroachment in native warm-season grasslands.

1.3.2 *Data Sources*

FSA Database. The United States Department of Agriculture (USDA) Farm Service Agency (FSA) maintains annual field-level records of acreage, land cover and intended land use for all fields participating in a USDA program. These data are known as FSA 578 data. In Kansas the FSA data are maintained by county FSA field offices where land owners or producers indicate land cover and land use for the upcoming year. In the 1990s, county field offices maintained the data as photocopies of aerial photos with land cover and land use information annotated on the photocopies. Depending on the county, these photocopies were made available to the public by request. In the mid-2000s counties began delineating and maintaining digital geospatial databases of field boundaries, called Common Land Units (CLUs). A common land unit is defined as the smallest land unit that has the same ownership, land cover, and land use. The extents of these units, which are subject to modification by FSA at any time, can be defined based on a change in any of these variables in addition to natural features such as waterways or forests or manmade features such as roads.

The agricultural marketing firm, Farm Market iD, is a proponent of making FSA and CLU data publicly available, and following a legal battle, Farm Maket iD was successful in making the FSA and CLU database publically available in early 2008 until the enactment in May of The Food, Conservation and Energy Act (known as the Farm Bill) of 2008, which then revoked public access to both datasets. During the three month window when data were available, Farm Market iD acquired the FSA and CLU data. Farm Market iD now packages and sells these data along with additional proprietary data about agricultural land owners and producers. The Biofuels and Climate Change - Farmers' Land Use Decisions (BACC–FLUD) project (supported by the National Science Foundation, Award Number EPS-0903806)

purchased the 2003-2007 FSA 578 data and CLU boundaries from 2007 from Farm Market iD. The 2007 Kansas CLU data layer contains more than 1.3 million geospatial features representing field (or management unit) boundaries. The 2003-2007 FSA 578 tabular data contain several key attributes that are relevant to this study. The attribute "Code" contains grassland type information that can be categorized into cool- or warm-season grassland types. The attribute "Intended Land Use Code" identifies the land use that the land owner intends to use the grassland for during that growing season or year. The Intended Land Use Codes for grasslands in Kansas include forage, grazing, and left standing (not grazed or hayed) (Table 1.). In addition to the 2000-2007 FSA 578 and 2007 CLU boundaries, the 2015 CLU and FSA 578 data were recently acquired through a Memorandum of Use (MOU) with the Kansas FSA office as part of a state-funded land cover update. The CLU and FSA 578 data are rich datasets that are ideal for this study by providing the spatial resolution, attribute information, sample size, and a temporal span that would be prohibitive to collect through an independent field campaign.

Satellite Imagery. Factors to consider when selecting appropriate data for mapping include cost, coverage, and resolution (spatial, spectral, radiometric, and temporal). One common theme in the previously described studies is the utility and importance of temporal resolution. Multi-temporal data capture variations in vegetation phenology and disturbances over a growing season. In terms of spatial resolution, several studies contend that higher spatial resolution data (e.g. RapidEye, 5m) are needed for effective grassland mapping and monitoring (Ali *et al.*, 2016; Corbane *et al.*, 2015; Schuster *et al.*, 2015). However, many studies, including several previously discussed, illustrate the utility of moderate resolution Landsat (30m) and MODIS NDVI (250m) in regional mapping applications and studies (Brown *et al.*, 2013;

Peterson *et al.*, 2008; Wardlow *et al.*, 2007). A critical advantage for using Landsat and MODIS are that these data are freely available.

Landsat 8 imagery contains eight Operational Land Imager (OLI) multispectral bands (1-7, 9; 30-m spatial resolution), one OLI panchromatic band (8; 15-m spatial resolution) and two Thermal Infrared Sensor (TIRS) bands (10 and 11; 30-m resampled spatial resolution). Terra MODIS NDVI provides a 250-m time series of 23 16-day NDVI composites per year.

1.3.3 *Data Analysis*

Objective 1. Using intended use codes in the FSA tabular data from 2003-2007 and 2015, summary statistics were calculated by year to identify and characterize dominant land use stratified by county, year, and grassland type. Where field-level data exist for multiple years (i.e. County FIPS, Tract Number, Field Number, Farm Number and Acreage remains the same), the data in the FSA tables were subset to summarize temporal change in grassland type and land use in the study area. County level maps of land use dominance, land use stability (no change over time) and land use change by grassland type were created to identify spatial patterns and dynamics of grassland type and land use in the study area.

Objective 2. Next, for each observation in the FSA 578 data and the CLU boundary shapefile, a unique identifier were created by concatenating the following attributes: State FIPS code, County FIPS code, Tract Number, and Farm Number (SCTF). Boundaries change, and tract and farm numbers can change. This unique identifier allows the CLU boundaries to be attributed with the FSA 578 data. Because there is a many-to-one relationship between tracts and CLU polygons (e.g. multiple crops were planted in a given field/CLU), only FSA and CLU data that have a one-to-one relationship were used for spectral profile analysis and image classification. A total of 18,707 data samples were used in the analysis.

Landsat path/row 27/33 was defined as the study area for determining the spectral separability of four grassland hierarchies using three separate datasets of remotely sensed imagery. The datasets of remotely sensed imagery assembled for the study were multispectral Landsat 8 surface reflectance, Landsat 8 NDVI, and MODIS NDVI time series. Three Landsat 8 surface reflectance images were ordered and acquired using USGS's EarthExplorer (EE) tool https://earthexplorer.usgs.gov/ to represent the spring, summer and fall portions of the growing season. A biweekly time series of 231-meter Terra MODIS 16-day composite NDVI from the 2015 growing season was downloaded from NASA's EarthData online tool, https://earthdata.nasa.gov/. The MODIS time series dataset was reprojected from the native Sinusoidal projection to Albers Equal Area projection and clipped to the Landsat path/row extent. The MODIS NDVI time series dataset was then resampled to 30-meter pixels and snapped to the Landsat 8 footprint. The three datasets were stacked to create a 44-band imagery dataset.

For each dataset separately and then combined, the Jeffries-Matusita (JM) distance statistic was calculated and evaluated for each pair of grassland classes (using grassland type and intended land use) to determine overall and seasonal separability using the full time series and individual time-periods, respectively.

The JM distance measures the separability between two classes by considering the distance between class centers simultaneously with intra-class spread (variability) and has shown utility in remote sensing applications (Davis *et al.*, 1978; Kastens *et al.*, 2017; Masialeti *et al.*, 2010; Richards & Jia, 1999; Swain & King, 1973; Wardlow *et al.*, 2007). Assuming multivariate normal distributions, the JM distance is calculated as:

$$JM_{ij} = 2(1 - e^{-B}) \text{ where}$$

$$B = \frac{1}{8}D^2 + \frac{1}{2}\ln\left(\left|\frac{\Sigma_j + \Sigma_k}{2}\right| \Big/ \sqrt{\left|\Sigma_j\right|\left|\Sigma_k\right|}\right),$$

$$D^2 = \left(\mu_j - \mu_k\right)^T \left(\frac{\Sigma_j + \Sigma_k}{2}\right)^{-1} \left(\mu_j - \mu_k\right), \text{ and where}$$

μ_j and μ_k correspond to class-specific, expected spectral values, and Σ_j and Σ_k are unbiased estimates for the class-specific covariance matrices. The JM distance ranges between zero and two. A JM distance approaching two suggests distinct distributions, or high separability, between two classes, while a JM distance closer to zero suggests overlapping distributions and little separability between two classes. The JM distance for each of the three datasets was plotted and statistics evaluated and described for all pairwise class comparisons.

Objective 3. For the third objective, two adjacent Landsat path/rows, 27/33 and 28/33, were defined as the study area. A 44-band dataset comprised of Landsat 8, Landsat 8 NDVI and MODIS NDVI were constructed for both path/rows. Training data from the FSA 578 data were used to train Random Forest models that were then used to produce grassland maps for the study area. Because image dates used for creating multi-season Landsat imagery varied by path/row, models were developed for each path/row separately and results compared to determine the generality of the mapping approach.

The number of training sites for cool and warm season grasslands were selected in an attempt to represent approximate proportions of the overall landscape within each path/row. A maximum of 10,000 training sites was selected using a random stratified design. For example, if 70% of the grassland types in the path/row were estimated as warm-season, then 70% of the training sample sites were represented by warm-season grasslands. However, actual representations are unknown and estimates were obtained using the 2015 FSA and CLU data.

Next, supervised classifications were run separately for each dataset and then combined using the Random Forest classifier (Breiman, 1994) in MatLab. Ten forests were created for each run. Random Forest implicitly builds a classification ensemble—hundreds or thousands of trees. Each tree is built using a subset of training data known as a "bootstrap sample," with the remaining third of the data providing Out-of-Bag (OOB) samples that were used to produce an unbiased estimate for the predictive error of the random forest model. To develop each component tree, a random subset of predictor variables (the size of which is typically the square root of the total number of predictors, or one half or twice the square root) was used. To apply a RF model, each tree or submodel in the forest or ensemble "votes" on the classification and the majority vote determines the final classification. For this study, one thousand trees were grown with predictor subset size determined by the square root of the total number of predictors. Out-of-bag error were calculated and used to assess forest performance. In addition, a probabilistic independent validation ("test") dataset was used to calculate traditional accuracy statistics in addition to quantity and spatial allocation errors (Pontius Jr & Millones, 2011). The classification maps within the overlap area between the two adjoining Landsat 8 path/rows were compared to assess the level of agreement in the mapped classes to indicate how the two RF models performed with different Landsat image dates and different proportions of grasslands.

1.4 *Significance of Study*

Previous research indicates that human land use complicates mapping grassland types by altering the vegetation phenology through removal of vegetation (haying, mowing, grazing) or the non-removal of biomass from prior growing seasons (standing dead litter). Identifying the dominant land use in cool- and warm-season grasslands in eastern Kansas is necessary for understanding the potential implications land use has on mapping these two grassland types.

Additionally, little research has been done to identify an optimal classification approach in terms of thematic resolution and source data to map grassland types and land use at a regional scale. This study compared the spectral properties of four grassland hierarchies using three remotely sensed datasets: multispectral Landsat 8, Landsat 8 NDVI, and Terra MODIS NDVI. The results of the analysis were used to characterize and quantify the separability of grassland types and land use to help identify the thematic classification scheme and the optimal data for mapping. A random forest modeling approach was used to compare the mapping ability for each of the three datasets separately and combined for each grassland hierarchy for two path/rows. The results will be used to inform regional-scale mapping of grasslands in Kansas.

1.5 *Figures and Tables*

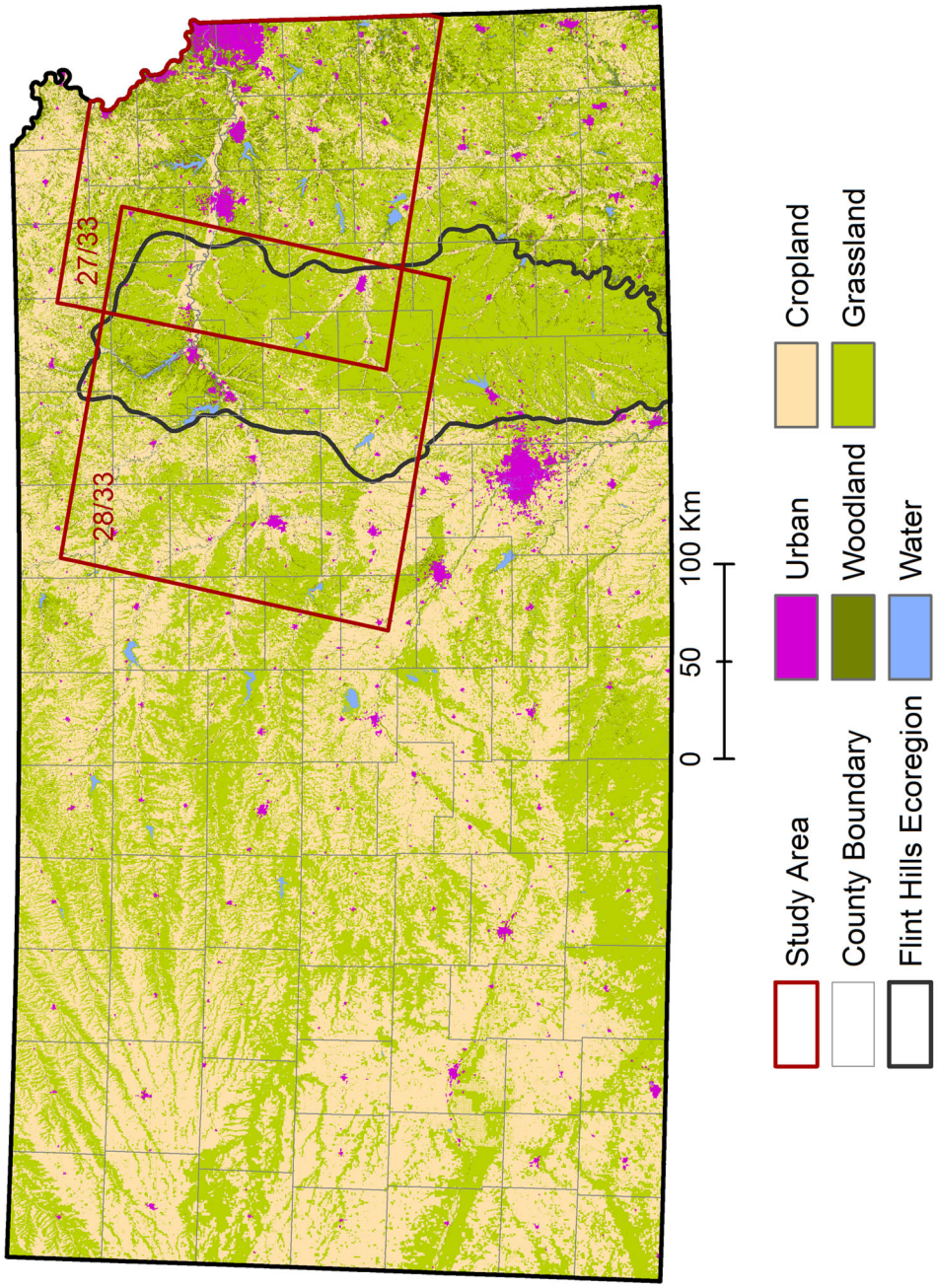


Figure 1.0.1 Kansas with study areas for research components 2 and 3 shown in red and the Flint Hills ecoregion overlaid.

Table 1.0.1. The four grassland hierarchies used in the analyses.

Level-1	Level-2	Level-3	Level-4
Warm-Season (W)	CRP (Crp)	CRP (Crp)	CRP (Crp)
	Native (N)	Native (N)	Native Forage (Nfg)
			Native Grazed (Ngz)
			Native Left Standing (Nls)
Cool-Season (C)	Cool-Season (C)	Fescue (F)	Fescue Forage (Ffg)
			Fescue Grazed (Fgz)
			Fescue Left Standing (Fls)
		Brome (B)	Brome Forage (Bfg)
			Brome Grazed (Bgz)
			Brome Left Standing (Bls)

Table 1.0.2. The three types of intended land use occurring in Kansas grasslands and their associated definitions.

Intended Land Use	Definition
Forage	Intended for harvesting as food for livestock. Does not include crops grown for the intended purpose of grazing by livestock or grown for the intended purpose of grain which may be fed to livestock.
Grazing	Intended solely for pasture for livestock to roam and feed on.
Left Standing	Intended to be left in the field unharvested. Not intended to be mechanically or manually harvested for any purpose, grazed by domesticated livestock, or otherwise harvested in any manner. Typically used for erosion control and nutrient retention.

References

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: from observation to management–a review. *Journal of Plant Ecology*. doi:10.1093/jpe/rtw005

Brown, J. C., Kastens, J. H., Coutinho, A. C., de Castro Victoria, D., & Bishop, C. R. (2013). Classifying multiyear agricultural land use data from Mato Grosso using time series MODIS vegetation index data. *Remote Sensing of Environment, 130*, 39-50.

Cooper, G. (2015). University of Wisconsion Study Based on Shaky Foundation of Faulty Data and Conclusion. Retrieved from http://ethanolrfa.org/2015/04/university-of-wisconsin-study-based-on-shaky-foundation-of-faulty-data-and-conclusions/

Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., . . . Michael, F. (2015). Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation, 37*, 7-16. doi:http://dx.doi.org/10.1016/j.jag.2014.11.005

Davidson, A., & Csillag, F. (2003). A comparison of three approaches for predicting C4 species cover of northern mixed grass prairie. *Remote Sensing of Environment, 86*(1), 70-82. doi:http://dx.doi.org/10.1016/S0034-4257(03)00069-5

Davis, S. M., Landgrebe, D. A., Phillips, T. L., Swain, P. H., Hoffer, R. M., Lindenlaub, J. C., & Silva, L. F. (1978). Remote sensing: the quantitative approach. *New York, McGraw-Hill International Book Co., 1978. 405 p., 1*.

Drummond, M. A. (2007). Regional dynamics of grassland change in the western Great Plains. *Great Plains Research*, 133-144.

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*(2), 89-93. doi:http://dx.doi.org/10.1016/j.ecoinf.2007.03.009

Franke, J., Keuck, V., & Siegert, F. (2012). Assessment of grassland use intensity by remote sensing to support conservation schemes. *Journal for Nature Conservation, 20*(3), 125-134. doi:http://dx.doi.org/10.1016/j.jnc.2012.02.001

Gibson, D. J. (2009). *Grasses and grassland ecology*: Oxford University Press.

Gu, Y., & Wylie, B. K. (2015). Developing a 30-m grassland productivity estimation map for central Nebraska using 250-m MODIS and 30-m Landsat-8 observations. *Remote Sensing of Environment, 171*, 291-298.

Guo, X., Price, K. P., & Stiles, J. (2003). Grasslands Discriminant Analysis Using Landsat TM Single and Multitemporal Data. *Photogrammetric Engineering & Remote Sensing, 69*(11), 1255-1262. doi:10.14358/PERS.69.11.1255

Guo, X., Price, K. P., & Stiles, J. M. (2000). Biophysical and Spectral Characteristics of Cool- and Warm-Season Grasslands under Three Land Management Practices in Eastern Kansas. *Natural Resources Research, 9*(4), 321-331. doi:10.1023/a:1011513527965

Halabuk, A., Mojses, M., Halabuk, M., & David, S. (2015). Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series. *Remote Sensing, 7*(5), 6107.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., . . . Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens, 81*(5), 345-354.

Kastens, J. H., Brown, J. C., Coutinho, A. C., Bishop, C. R., & Esquerdo, J. C. D. (2017). Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PloS one, 12*(4), e0176168.

Lauver, C. L., & Whistler, J. L. (1993). A hierarchical classification of Landsat TM imagery to identify natural grassland areas and rare species habitat. *Photogrammetric Engineering and Remote Sensing, 59*(5), 627-634.

Masialeti, I., Egbert, S., & Wardlow, B. D. (2010). A comparative analysis of phenological curves for major crops in Kansas. *GIScience & Remote Sensing, 47*(2), 241-259.

McInnes, W. S., Smith, B., & McDermid, G. J. (2015). Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(4), 1395-1403. doi:10.1109/JSTARS.2015.2416713

Owensby, C. E. (1993). *Introduction to range management*: Custom Academic Publishing Company.

Peterson, D. L., Price, K. P., & Martinko, E. A. (2002b). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing, 23*(23), 5015-5030.

Peterson, D. L., Whistler, J. L., Egbert, S. L., & Martinko, E. A. (2008). *2005 Kansas Land Cover Patterns: Phase II-Final Report*. Retrieved from Kansas Biological Survey:

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32*(15), 4407-4429.

Price, K. P., Guo, X., & Stiles, J. M. (2002a). Comparison of Landsat TM and ERS-2 SAR data for discriminating among grassland types and treatments in eastern Kansas. *Computers and Electronics in Agriculture, 37*(1), 157-171.

Richards, J., & Jia, X. (1999). Remote sensing digital image analysis: An introduction Springer-Verlag. *Berlin Heidelberg*.

Risser, P. G. (1988). *Diversity in and among grasslands*: National Academy of Sciences Washington, DC.

Robertson, K. R., Anderson, R. C., & Schwartz, M. W. (1997). The tallgrass prairie mosaic *Conservation in highly fragmented landscapes* (pp. 55-87): Springer.

Robertson, K. R., & Schwartz, M. W. (1994b). Prairies. *The changing Illinois environment: critical trends, 3*, 1-32.

Samson, F. B., Knopf, F. L., & Ostlie, W. R. (2004). Great Plains ecosystems: past, present, and future. *Wildlife Society Bulletin, 32*(1), 6-15.

Schuster, C., Schmidt, T., Conrad, C., Kleinschmit, B., & Förster, M. (2015). Grassland habitat mapping by intra-annual time series analysis – Comparison of RapidEye and TerraSAR-X satellite data. *International Journal of Applied Earth Observation and Geoinformation, 34*, 25-34. doi:http://dx.doi.org/10.1016/j.jag.2014.06.004

Swain, P., & King, R. (1973). Two effective feature selection criteria for multispectral remote sensing.

Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K., & DeJong, D. D. (1997). NDVI, C3 and C4 production, and distributions in Great Plains grassland land cover classes. *Ecological applications, 7*(1), 59-78.

Wang, C., Jamison, B. E., & Spicci, A. A. (2010). Trajectory-based warm season grassland mapping in Missouri prairies with multi-temporal ASTER imagery. *Remote Sensing of Environment, 114*(3), 531-539.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment, 108*(3), 290-310.

Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., & Coulston, J. (2014). The multi-resolution land characteristics (MRLC) consortium—20 years of development and integration of USA national land cover data. *Remote Sensing, 6*(8), 7424-7441.

Wickham, J., Stehman, S. V., Gass, L., Dewitz, J. A., Sorenson, D. G., Granneman, B. J., . . . Baer, L. A. (2017). Thematic accuracy assessment of the 2011 national land cover database (NLCD). *Remote Sensing of Environment, 191*, 328-341.

Wright, C. K., & Wimberly, M. C. (2013). Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences, 110*(10), 4134-4139. doi:10.1073/pnas.1215404110

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment, 122*, 2-10.

2 Chapter 2: Characterizing County-Level Spatial and Temporal Distributions of Grassland Types and Land Use in Kansas

2.1 *Abstract*

This study characterizes grassland types and land use across Kansas and evaluates both the dynamic and static nature of grassland type and land use over time to inform a methodology for land cover mapping of grassland types. USDA Farm Service Agency (FSA) data from 2004-2007 and 2015 and Common Land Unit (CLU) geospatial data from 2007 and 2015 were used as the primary data inputs for county summaries of grassland type and land use. The results show regional trends in grassland type in Kansas that are largely driven by climate and soils. During the study period, native warm-season grasslands were most common in the western two-thirds of the state while cool-season fescue and brome grasslands were most common in different portions of the eastern third of Kansas. Fescue were more common in southeastern Kansas where soils are more clay pan while brome were more common in northeastern Kansas and was prevalent in some central counties with well-drained soils. Field size also varied across the state and by grassland type. There were a large number of small fields (< 20 acres) of brome and fescue in eastern and central Kansas while western Kansas was composed of fewer but larger fields of native grassland and land enrolled in the Conservation Reserve Program (CRP). The results also show that the FSA 578 data did not represent all grasslands, especially in the early FSA 578 data (2007 and prior) where most of the Flint Hills area was not included. Meanwhile 2015 FSA 578 data showed an increase across the state, largely in the Flint Hills, suggesting that the establishment and funding of two emergency response FSA programs in the 2014 farm bill increased participation in FSA programs. As anticipated, the change in land use was greater than

the change in grassland type in terms of acres and the number of fields. The results indicate mapping grassland types in eastern Kansas may be more complex than in the central and western parts of the state due to the larger number of grassland types and the more fragmented landscape.

2.2 *Introduction*

The composition, spatial distributions, granularity, and the static and dynamic nature of the landscape are among the factors that influence and potentially impact mapping endeavors using remotely sensed data. This study characterizes grassland types, land use, and field size across Kansas in order to make informed decisions on developing a methodology for image classification to map grassland types in Kansas. Knowledge of the characteristics of grasslands in Kansas will aid in developing a mapping approach, including training sample allocation, thematic resolution of mapped classes, the source imagery used, and the efficacy of using out-of-year training data for image classification.

In Kansas both native and non-native grasslands are managed landscapes where combinations of land management practices are utilized to maintain and maximize vegetation productivity largely for grazing and livestock support. The timing, frequency, and intensity of land management practice, and the combination of practices, varies by grassland type and by land owner, which in turn complicates mapping grassland types using remotely sensed imagery. While previous research has shown that land use and management complicates the mapping of warm- and cool-season grassland types Guo *et al.*, 2000), there is no documented information on the prevalence of grassland types and land use between and within grassland types in Kansas or on the static or dynamic nature of grassland type and land use over time. Without an understanding of these trends, it is difficult to fully understand the impact land management practices have on the ability to map warm- and cool-season grassland types. And without an

understanding of the spatial distributions of grassland types, it is difficult to develop a thematic classification system and proportioned training and validation data allocation for image classification. Furthermore, knowledge of the granularity or size of the features (i.e., parcels) to be mapped can be used to determine the required spatial resolution of the input source imagery as well as how training data will be identified and extracted. Lastly, testing the assumption that both inter-annual grassland type and land use within grasslands are static can shed light on the possible efficacy of using out-of-year training data for image classification when within-year training data are unavailable.

The objectives of this research are three-fold: 1) to identify the dominant grassland types and land use in grasslands and characterize their spatial trends in Kansas; 2) characterize trends in field size within grassland types across the state; and 3) determine the dynamic characteristics of grassland type and land use over time.

2.3 *Methods*

2.3.1 *Study Area*

Situated in the central Great Plains, Kansas exhibits both precipitation and temperature gradients. Precipitation has an west-east gradient where the western third of Kansas's long-term mean (1895-2015) annual precipitation is 531mm, the central third is 660mm and the western third is 945mm (Lin *et al.*, 2017). The west-east gradient of annual precipitation in Kansas largely drives plant productivity (e.g. biomass and canopy height) and thus the west-east grassland gradient from shortgrass to mixed grass to tallgrass prairie, respectively. In addition, higher levels of precipitation in eastern Kansas support more forests and woodlands. Meanwhile temperature affects the distribution of functional grassland types with C3 dominating northern regions and C4 the southern regions. Temperature is the main driving factor influencing the natural distributions

of C3 and C4 grasslands species in the Great Plains. Additionally, C3 and C4 grassland species each have competitive advantages under certain CO2 levels, soil types, frequency and intensity of precipitation events, and type, frequency and intensity of land management practices (including fire). With increasing CO2 levels and climate change, there are anticipated shifts in species distributions which in turn affect ecosystem functions, soil biochemistry and the global carbon cycle.

The regional trends in native grasslands across Kansas largely follow the west-east precipitation gradient. However, temperature, soils, disturbance, and land use are additional factors that affect grassland community distribution, and thus the boundaries for these native grassland types shift inter-annually. Grasslands in the western fourth to third of the state consist of native warm-season shortgrass species such as sideoats grama (*Bouteloua curtipendula*), blue grama (*Bouteloua gracilis*), and buffalograss (*Bouteloua dactyloides*), and moister areas may include *Schizachyrium scoparium*. Several forb and shrub species are supported in these semiarid grasslands including soapweed yucca (*Yucca glauca*) and common ragweed (*Ambrosia psilostachya*). Cropland and land enrolled in the Conservation Reserve Program (CRP) are dominant across Kansas, with CRP representing the majority of grasslands in several western counties. Dominant Kansas agricultural crops include corn, soybeans, grain sorghum, alfalfa, and winter wheat. Central Kansas grasslands consist of a mix of species from the eastern tallgrass prairie and western shortgrass prairie. Dominant species include little bluestem (*Schizachyrium scoparium*), big bluestem (*Andropogon gerardii*), indiangrass (*Sorghastrum nutans*), sideoats grama, blue grama, and buffalograss, as well as native cool-season grass species such as western wheatgrass (*Pascopyrum smithii*).

In the eastern third of Kansas, grasslands consist both of native, warm-season grasslands, and non-native, cool-season grasslands. The warm-season grasslands are either native tallgrass prairie or have been reseeded using a native seed mixture. The tallgrass prairie is dominated by bunchgrasses including big bluestem), little bluestem), and indiangrass. In high-quality prairie remnants, dozens of native forb species can be present including leadplant (*Amorpha canescens*), butterfly milkweed (*Asclepias tuberosa*), and purple coneflower (*Echinacea angustifolia*). The dominant warm-season grassland species fix carbon using C4 photosynthesis. The typical phenology of warm-season grasslands is spring green-up, peak productivity in late spring to early summer when temperatures increase, followed by senescence in fall (Weaver, 1954). The Flint Hills region lies on the western edge of the tallgrass prairie and is the largest remaining tract of native tallgrass prairie in the world. Cool-season grasslands become more prevalent in the eastern part of the state. Cool-season grasslands are planted with non-native herbaceous species such as smooth brome (*Bromus inermis*) and tall fescue (*Festuca arundinacea*) that fix carbon using C3 photosynthesis. The typical phenology of cool-season grassland is early spring green-up, peak productivity in late spring, a mid-summer semi-dormancy and, with sufficient precipitation, a second, smaller growth period in early fall (Weaver, 1954). Cropland, woodland, and forest primarily occupy the river lowlands and riparian areas in eastern Kansas.

Haying (forage) and grazing are two of the common land uses for all grassland types across the Kansas. However, the timing, intensity and frequency of management practices within each grassland type vary by land owner and by climate conditions in a given year. In addition to haying and grazing, prescribed burning is a required management practice for maintaining diversity in native warm-season grasslands and preventing woody encroachment. Herbicide

applications and fertilization are among the other commonly used land management practices to optimize vegetation productivity in non-native, cool-season grasslands.

2.3.2 *Data Sources*

The two primary datasets used for this study are United States Department of Agriculture Farm Service Agency (FSA) tabular data and FSA Common Land Unit (CLU) geospatial data. The FSA tabular and CLU data contain the attribute information, spatial representation, sample size, and temporal span to characterize spatial and temporal trends in grassland type and land use across Kansas.

The FSA maintains annual field-level records of acreage, land cover, and intended land use for all fields enrolled in a USDA program. Kansas FSA 578 data are maintained by county FSA field offices where land owners or producers report land cover and land use by November 15th for the upcoming year. Accurate reporting by land owners ensures program eligibility. The field boundaries, called CLUs, are defined as the smallest land unit that has the same ownership, land cover, and land use. The extents of these units, which are subject to modification by FSA at any time, can be defined based on a change in ownership or land use.

The 2004-2007 FSA tabular and 2007 FSA CLU data were purchased by the Biofuels and Climate Change - Farmers' Land Use Decisions (BACC–FLUD) project (supported by the National Science Foundation, Award Number EPS-0903806) from Farm Market iD, an agricultural database and analytics firm. The 2015 FSA 578 and CLU data were acquired through a Memorandum of Use with the Kansas FSA office as part of a statewide land cover mapping endeavor. The 2007 and 2015 Kansas CLU data layers each contained more than one million geospatial features representing polygonal field boundaries. The FSA 578 data contained several key attributes relevant to this study including reported acreage, grassland type

information that distinguished between cool- and warm-season, and land management practices that land owners intend to use the grassland for during the upcoming growing season or year. The intended land use codes for grasslands in Kansas included forage, grazing, and left standing (Table 1).

2.3.3 *Data Processing*

The tabular data were delivered from Farm Market iD in comma separated value (csv) files. The data tables were imported into Microsoft SQL Server Management Studio 2008 R2 for data processing and analysis. Each entry or row in the tabular data was attributed with a state, county code, tract number, field number and sub-field identifier. These attributes were concatenated to create a unique identifier for each row named "SCTF". This unique identifier was constructed for all of the 2004-2007, 2015 FSA 578 data and the 2007 and 2015 CLU geospatial layers.

To ensure data quality, dominant grassland types (>1,000 samples) were inspected to verify that locations identified as grasslands were actually grasslands. Using ESRI ArcGIS, the SCTF identifier was used to join tabular data to the 2007 CLU polygon data layer. The CLU polygons from 2007 were then overlaid on the closest year of high-resolution FSA NAIP aerial imagery and existing land cover maps. Evaluating the classes identified two grassland types from the 2004 data that did not fall on grassland areas. More than 50% of the fields attributed as Side Oats Grama and Sand Bluestem from 2004 were located on cropland fields and excluded from the analysis. The cause for these errors is unknown, but illustrates rational for data quality checks. Common Bermuda was considered a crop for sod and was also excluded from the analysis.

The 119 grassland types listed in the FSA multi-year data were recoded into seven dominant grassland types that were then used in county and field-level analysis (Table 2). CRP fields were recoded using the conservation practice attribute (Table 2). Conservation practices indicating warm-season grassland were aggregated to a "CRP warm-season" class (e.g. Establishment of Permanent Native Grasses) whereas the conservation practice, Establishment of Introduced grasses and legumes, indicated cool-season grassland (Banks, 2012) and was recoded to a "CRP cool-season" class. Conservation practices that could represent grassland and/or woodland or wetland (e.g. Riparian Buffers) or that did not represent field-sized grasslands (grass waterways) were recoded to "CRP other." The resulting seven classes were CRP warm-season, CRP cool-season, CRP other, Native warm-season, Brome cool-season, Fescue cool-season, and Grass other. While there were a large number of grassland types in the Grass other class, 69 of the 88 grassland types had fewer than 100 fields across all years of FSA 578 data.

As a further check of data quality, a data completeness assessment was performed to evaluate the completeness and representativeness of the FSA 578 data. The FSA reported grassland and CRP acreages were calculated and compared to state and county-level grassland acreages calculated from two supplemental sources. For grassland acreage information, an annual land cover time series (Gao *et al.*, 2017) was used as supplemental data and for CRP acreage information, the county-level USDA FSA cumulative annual reported acreages were used as supplemental data (Gao *et al.*, 2017; USDA, 2017). While all land use/land cover maps contain misclassification errors Peterson *et al.* (2008) and Peterson *et al.* (2017) reported an overall map accuracy of 90% and 93% for 2005 and 2105 data, respectively. County level maps showing proportions of grassland type, land use, and data completeness for each year were created in ESRI ArcMap to identify dominant grassland types and lands use in the study area.

Understanding the distribution and proportions of grasslands is useful for determining training sample allocations for image classification.

Another consideration for land cover mapping in addition to understanding regional variability of grassland type and land use is field size or spatial granularity. Two commonly used sensors for statewide mapping endeavors, Landsat and MODIS have spatial resolutions of 30 meters and 231 meters, respectively. Characterizing the granularity of the field size in the study area informs several components of developing a land cover methodology. Granularity can determine the spatial resolution of the imagery required for mapping, dictate the minimum mapping unit for the mapped classes, and/or the method used for extracting pixel information (pure vs. mixed pixels) used for training the image classification. To characterize the field size of grasslands (including CRP) in Kansas, the nine agricultural statistic districts (ASDs) comprising the state were aggregated into eastern, central and western tiers that roughly correspond to the precipitation gradient (Figure 1). Frequency distributions of the count and acreage of field size were created using 10-acre incremental intervals of the 2015 FSA 578 data.

To characterize the dynamic vs. static trends in grassland types and land use across Kansas, field-level data existing across three or more years where the unique identifier, SCTF, and reported acreage remained constant (within 5%) were extracted for analysis. Table 5 shows that 181,667 fields totaling 3.275 million acres met the criteria for change analysis of grassland type and 193,900 fields totaling 3.4 million acres met the criteria for change analysis of land use.

Next, each change scenario trajectory was evaluated and labeled as "likely" or "unlikely." For example, if the grassland trajectory was brome→native→brome→native, the trajectory was labeled "unlikely," whereas a trajectory of CRP→CRP→native→native was labeled as "likely" since it is plausible that a CRP contract expired and the grassland subsequently labeled as native.

Land use change was independent of grassland type change. There were no "unlikely" land use trajectories since land owners can change land use from one year to the next based on climate, grazing systems, economic conditions, or government policy. Finally, the acres of change and no change were summarized at the county and state level.

2.4 *Results and Discussion*

The completeness assessment of grasslands and CRP was a necessary data quality check since subsequent analysis and county level maps of grassland type and land use utilized these data. The comparison of the statewide grassland acreages from the annual land cover data and annual FSA 578 data showed the 2004-2007 FSA data from Farm Market iD were not inclusive of all grasslands across Kansas (Table 3). The comparison suggests that the FSA data represented on average 31% of the statewide grassland area, ranging between 30% in 2007 to 34% in 2004. While the grassland acres mapped in annual land cover data remained relatively consistent, the 2015 FSA data showed a large increase (75%) in statewide representation of grasslands (Table 3). Figure 2 shows the county level change in acreage in the FSA data between 2007 and 2015. Only Comanche and Mitchell counties showed decreases in FSA acres between 2007 and 2015. Meanwhile there were large increases in FSA acres from 2007 to 2015 in multiple counties in the Flint Hills (e.g. Cowley, Butler, and Marion), Smoky Hills (Trego, Ellis and Russell) and Red Hills (Barber and Clark).

During the 2004-2007 time-frame there were two grassland-related FSA programs available to land owners, operators, or producers. The Conservation Reserve Program (CRP) began in 1989 and offers voluntary enrollment to agricultural producers where "environmentally sensitive farmland" is taken out of production and planted with long-term grass or tree cover to reduce soil erosion, enhance wildlife habitat, and improve water quality. CRP contracts have an

option for a 10- or 15-year enrollment. CRP land can only be grazed or hayed under FSA authorization as emergency relief in response to natural disasters, primarily drought or fire. The Grassland Reserve Program (GRP) began in 1985 but was later repealed by the 2014 Farm Bill. However, GRP contracts established prior to February 2014 remained valid. The voluntary GRP offered 10, 15, or 20-year contracts that provided annual rental payments or conservation easements to land owners and operators to protect rangeland or pastureland from overuse or conversion to farmland or development (USDA, 2009). The FSA data obtained for this study do not indicate lands in the GRP program. It was difficult to discern whether FSA data were incomplete for non-CRP grasslands, or that a large proportion of grasslands were not enrolled in an FSA program in the 2004-2007 time-frame. There were however, three additional FSA disaster assistance programs established between 2007 and 2015 may have resulted in the observed increased grassland acreage in 2015 FSA data.

In 2008 the Livestock Forage Program (LFP), Livestock Indemnity Program (LIP), and the Emergency Assistance for Livestock, Honey Bees, and Farm-Raised Fish Program (ELAP) were established, but were not authorized and funded until the 2014 Farm Bill (Agricultural Act of 2014; P.L. 113-79) (Stubbs, 2018). The LFP compensates land owners or producers who lost grazing opportunity due to natural disasters on native or established non-native grassland that is used specifically for grazing. The LIP compensates land owners or producers who lost considerable livestock from adverse weather, and the ELAP compensates land owners or producers for livestock losses resulting from disease, adverse weather or shortages of water or feed (Stubbs, 2018). These voluntary programs are freely available to land owners given eligibility requirements are met.

While the FSA data underrepresented grassland acres, the comparison of CRP in FSA data and county-level USDA reported CRP acres showed that the 2004-2007 and 2015 FSA data only slightly underrepresented CRP acres, with 2004 having the lowest percent representation of 93.5% (Table 4). And while grassland acres in the FSA data increased from 2007 to 2015, CRP acres decreased in both the FSA data and the county-level USDA data. Expired CRP acres would explain a portion of the observed increase in grassland acres between 2007 and 2015 if land owners chose to use the conservation cover for forage or grazing. The 2010 National Resource Inventory (USDA, 2013a; USDA, 2013b) showed that between 2007 and 2010, CRP acres declined by almost 18%, with 55% of those acres converted to cropland and 41% used as pasture or rangeland.

In addition to providing a completeness assessment, Figures 2 through 7 show annual, county-level proportions of total grassland and CRP acres obtained from the FSA and supplemental data. While small variations could be the result of changing land cover, the variability suggests some limitations of the data. For example, total grassland acres in Meade County (located in southwest Kansas on the southern boundary of the state) increased from 290 thousand acres to 326 thousand acres between 2005 and 2006. These types of inter-annual fluctuations in total grassland acres may be due to errors or misclassification in the annual land cover data and/or expired CRP contracts. Even with some inter-annual variability, the data were sufficient to identify and characterize regional trends in grassland extents, grassland type and land use.

Figures 2 – 7 show the southern Flint Hills and the Red Hills regions of Kansas had the greatest total acres of grasslands (FSA and non-FSA Data), which is primarily due to shallow soils preventing these regions from being plowed for agricultural crop production. Butler,

Greenwood, Cowley, and Chase Counties consistently had the largest grassland acres in the southern Flint Hills and Barber, Clark and Comanche had the largest reported acres in the Red Hills. Meanwhile there were many counties across the state with relatively small grassland acres remaining including in northeastern Kansas (e.g. Doniphan, Atchison, Wyandotte, and Johnson), in southwestern Kansas (e.g. Haskell, Gray, Grant) and in central Kansas (e.g. Rice, Harvey, Sumner, Pratt, and Edwards). According to land cover data and knowledge of the study area, these counties were dominated by cropland or are heavily urbanized (e.g. Wyandotte, Johnson, and Sedgwick) (Figure 1).

As previously mentioned, the lack of warm-season native grassland in Flint Hills represented in the 2004-2007 FSA data was either the result of missing FSA data or more likely, the lack of participation by land owners in FSA programs. The 2015 map in Figure 7 illustrates the changes in the fractions that more closely match the expected quantity of grasslands in the Flint Hills and western Kansas shown in Figure 1. These differences highlight the importance of understanding the study area and assessing data quality for data produced by other entities. Without the completeness assessment, multi-year data, and prior knowledge of the study area, grassland type proportions obtained from the 2004-2007 FSA data would have incorrectly proportioned the sampling sites used for training data in image classification which in turn could potentially increase classification error in a derived land cover product. These results also underscore the importance of scrutinizing data and using multi-year information for data comparisons for quality assurance where possible.

While there was inter-annual variability in total grassland acreage, Figures 8 – 12 show there were consistent regional trends in grassland types across the state. Native warm-season and CRP grasslands were more prevalent in the western two thirds of the Kansas, while cool-season

brome and fescue were more common in the eastern half to third of the state. The prevalence of of cool-season grasslands in eastern Kansas corresponded to the precipitation gradient across Kansas where eastern Kansas has adequate precipitation and soil types to support introduced cool-season grassland types like brome and fescue. The data also indicated finer scale variability in the dominance of brome versus fescue in Kansas driven by regional climate and soils. Southeastern Kansas had distributions of fescue where there is a longer growing season, higher average precipitation and slow draining claypan soils Shoup *et al.*, 2010). Meanwhile, central and eastern Kansas had distributions of brome where there are deeper well-drained soils. The 2004-2007 data showed small fractions of brome as far west as Ellis County in central Kansas and Haskell County in southwest Kansas, where brome is irrigated for livestock grazing (Lamond *et al.*, 1992).

Warm-season CRP was more prevalent over cool-season CRP across all years of data with the largest fractions of warm-season CRP in southwest Kansas. Eastern Kansas had a small number of counties containing cool-season CRP in the 2004-2007 FSA data, but not in 2015. It is possible that cool-season CRP were reported in counties where county extensions supported particular seeding varieties for CRP. Additionally, in 2004 several western counties (Thomas, Gove, Scott and Pratt) contained cool-season CRP as well, though in subsequent years no cool-season CRP acres were reported. It was uncertain if this change was real or noise in the FSA data. Furthermore, the county-level fractions of CRP visually appear to be relatively stable from 2004-2007 with a dramatic statewide decrease in CRP in 2015. While CRP reported acres declined in all counties (USDA, 2017), the reduced fraction of CRP shown in Figure 12 is exaggerated by the additional warm-season native grassland data in the 2015 FSA data.

Regarding land use, most of the grasslands across the state were either grazed or used for forage with very little grassland left standing (Figures 8-12). However, central Kansas had a few counties (Rice, McPherson, Reno, Kingman, Sedgwick, and Sumner) that contained notable left standing native grassland in all years of data. Northwestern Kansas (e.g. Cheyenne, Rawlins, Decatur, Norton) mostly consisted of grazed native grasslands, while southwestern and portions of central Kansas (e.g. Stafford, Reno, and Kingman) consisted of warm-season CRP, followed by grazed native grasslands.

The 2015 FSA data indicate fescue was primarily used for grazing followed by forage, whereas the land use of brome in 2015 was more variable, where some counties showed a dominance for forage and others for grazing. Interestingly, the 2004-2007 data indicated there was more forage production of fescue and brome in Kansas than in 2015. Brome and fescue were introduced to Kansas in the late 1800s as supplemental forage for livestock operations. Native grassland is typically grazed in the spring and summer while brome and fescue provide opportunity to lengthen the spring and fall grazing and forage periods. The Kansas State University (KSU) extension office indicates that brome provides excellent spring and fall pasture as well as excellent forage, producing as much as 3-4 tons of forage per acre (Lamond *et al.*, 1992). And in southeast Kansas, fescue is commonly fertilized in late summer and allowed to grow until late fall to provide winter grazing. Incorporating these grassland types allows land owners and producers to maximize livestock production and profits. Such local and regional practices are largely supported by local entities such as Kansas State Research and Extension and the Kansas Livestock Association. For example, the Southeast Agricultural Research and Extension Center located in southeast Kansas conducts grazing research exclusively on introduced grass species, unlike other extension centers across Kansas that focus heavily on

native grasslands (KSU, 2018). Annual climate conditions likely factor into land owners' decisions on how grasslands will be used in the upcoming year. For example, drought conditions could cause a shift from forage to grazing if summer grazing on native grasslands is less productive and requires supplemental summer grazing.

Figure 13 shows the frequency distributions of field sizes for grassland types among the three tiers of ASDs across Kansas. Moving from west to east there was a decrease in the frequency and acreage of large fields (> 160 acres) (Figure 13), meaning fewer large fields in eastern Kansas. The western tier had 8,800 native grassland and CRP fields greater than 160 acres, totaling 3.47 million acres or 54% of the total FSA grassland acreage. In the central tier there were 6,300 native grassland and CRP fields greater than 160 acres, totaling 2.18 million acres or 34% of the total FSA grassland acreage. Meanwhile the eastern tier had the smallest number (4,856) and total acreage (1.59 million acres) of large fields of grassland, 90% of which were native grassland. There were large fields of CRP in both the west and central tiers whereas brome and fescue were only present in the east tier.

The frequency, acreage and composition for small fields varied among the tiers as well. For example, the western tier had 54,500 native grassland and CRP fields of less than 10 acres; however, the total acreage of the small fields represented only 3% of the grassland acres. Moving eastward there was a substantially higher number of smaller fields. The central tier had 175,106 fields of less than 10 acres and 7.2% of the total grassland acreage, while the eastern tier had 147,930 fields of less than 10 acres, accounting for 8.6% of the acres reported. The composition of small fields varied across tiers. In the west, smaller fields were primarily a mix of grassland and CRP, while moving east there was more variety of grassland types, with brome dominant in the central and eastern tiers and fescue in the eastern tier. The increased frequency and small

acreage of grassland types in the central and eastern tiers are indicative of a more fragmented landscape. Understanding the field-size and level of fragmentation among grassland types across Kansas can be used to help determine the minimum mapping unit and source imagery used for mapping. While MODIS NDVI with its coarser spatial resolution may be suitable for mapping the larger grassland extents in the west, there were many small fields of CRP (e.g. corners of center pivots, which typically are concave in shape and roughly eight acres in size) and native grassland in western Kansas that may not be mapped accurately using a 231m spatial resolution (or approximately 13.2 ac/pixel). Likewise, Landsat data may be more suitable in the east given the frequency of small fields; however, the limited temporal resolution of Landsat may not be adequate for separating the phenological differences among grassland types.

Statewide, a higher percentage of fields and acres remained unchanged for grassland type than land use. Grassland type remained unchanged for 87% of grassland acres and 84% of fields while land use remained unchanged for 60% of grassland acres and 77% of fields (Table 6). Figures 14 and 15 show county-level percent of acres that changed in grassland type and land use, respectively. Generally speaking, counties in central and eastern Kansas exhibited more change than counties in western Kansas. The higher average precipitation levels in central and eastern Kansas support the variety of grassland types (native and non-native) that have been established and provide options to change from one land use to another, whereas in western Kansas, the shortgrass prairie provides yield for grazing but not additional vegetation for forage. Crops such as forage sorghum and alfalfa are typically used as supplemental forage in the semi-arid west.

2.5 *Conclusions*

The assessment and analysis of multiple years of FSA data showed variability in degree of completeness, meaning that the 2004-2007 FSA data were not all-inclusive of total grassland acres in Kansas. Even so, the data were sufficient to identify several regional trends in grassland type, land use, and field size. Eastern Kansas was found to have more grassland types, with the inclusion of non-native brome and fescue, a larger number of small fields, and more variability in land use, which together creates a more fragmented and complex landscape for mapping grasslands in that region. Western Kansas had larger fields that primarily consisted of CRP and grazed native grassland, creating a comparatively simpler landscape for mapping grasslands. The inclusion of 2015 data provided a more complete representation of grassland type and land use in Kansas compared to 2004-2007 data, which possibly was the result of three new FSA programs that were implemented in the interim. These results will be used to inform a grassland mapping approach for Kansas, including training data allocation for image classification.

2.6 *Figures and Tables*

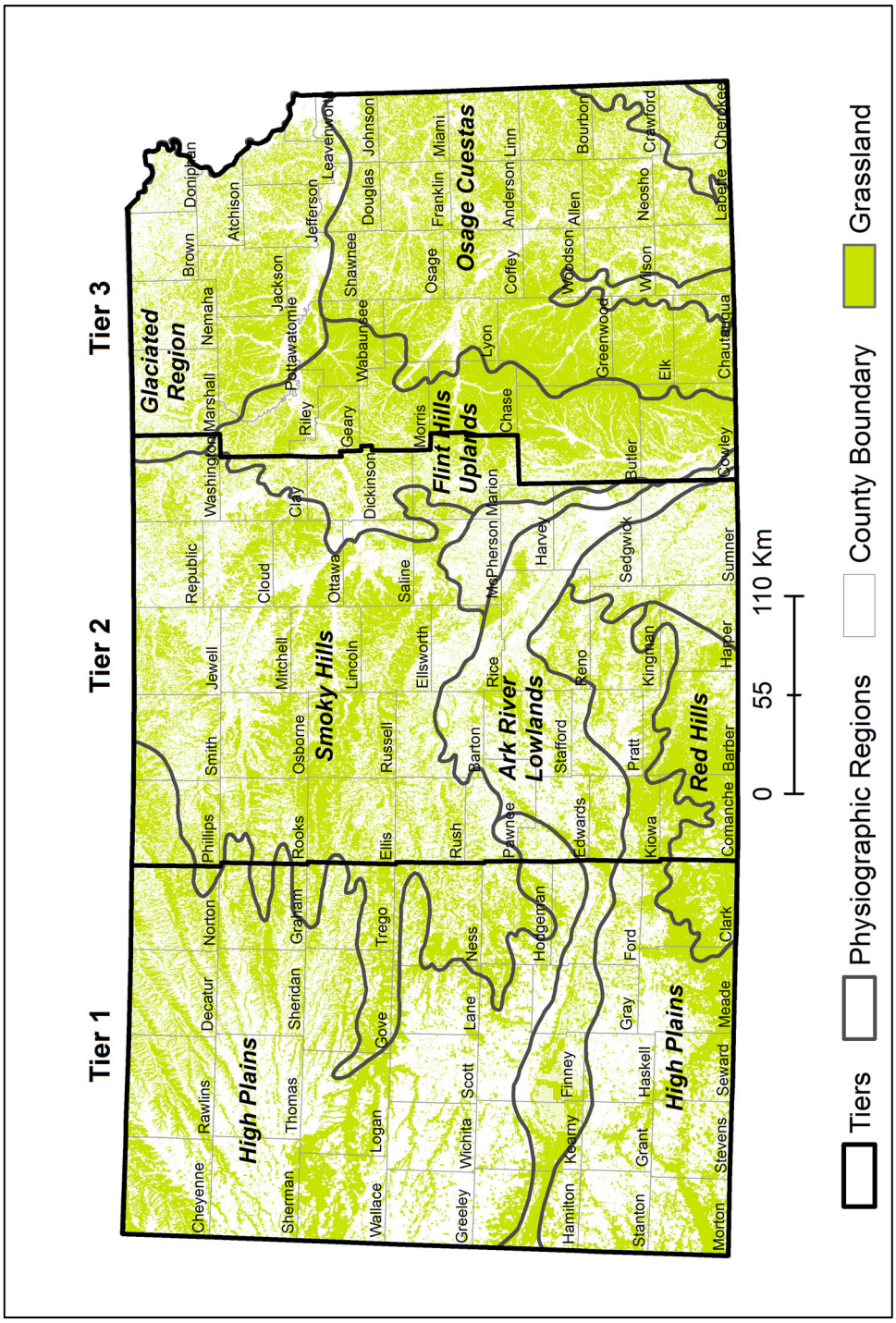


Figure 2.1. The three tiers of Agricultural Statistics Districts (ASDs) used for analyzing field size of grasslands in Kansas overlaid physigraphic provinces of Kansas and 2015 grassland map derived from remotely sensed data.

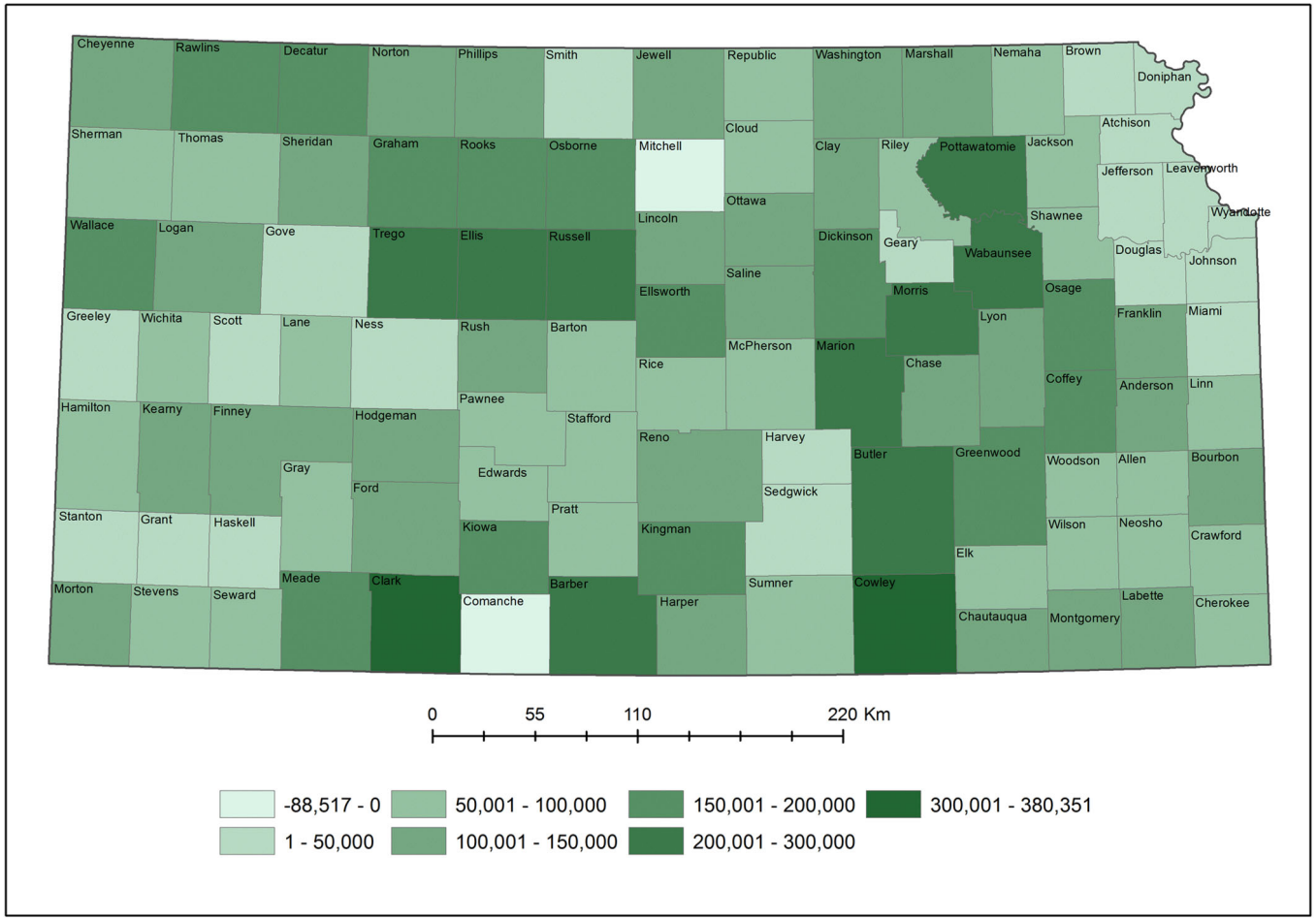


Figure 2.2. A comparison of grassland acres in the FSA data from 2007 to 2015. All except two counties increased in grassland acres reported in the FSA data from 2007 to 2015.

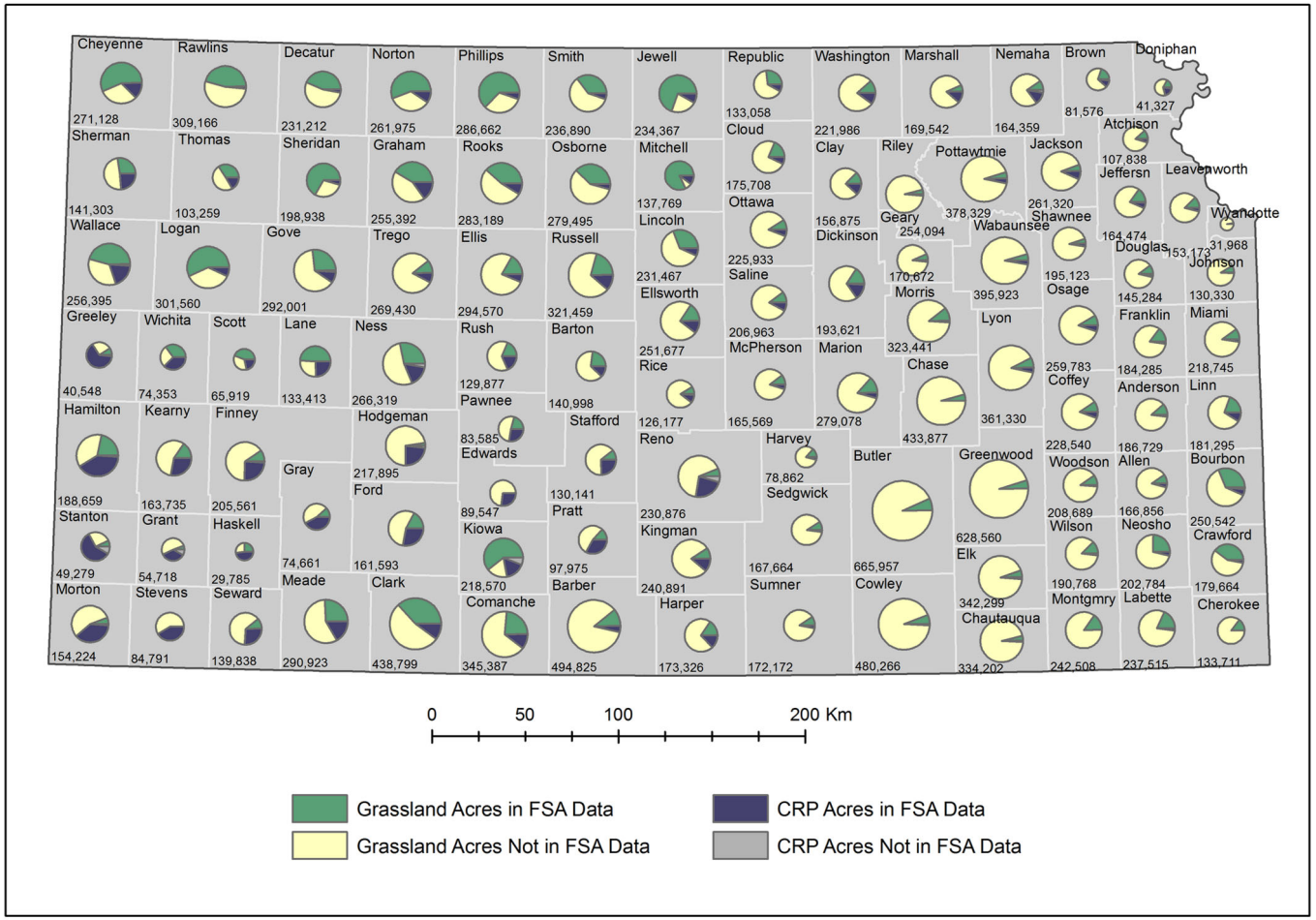


Figure 2.3. Fractions of grassland and CRP acres in the 2004 FSA data (green and blue) and acres not represented in the FSA data derived from supplemental data (land cover data and USDA county-level CRP data) (beige and gray). Pie charts are sized by total grassland acres in each county.

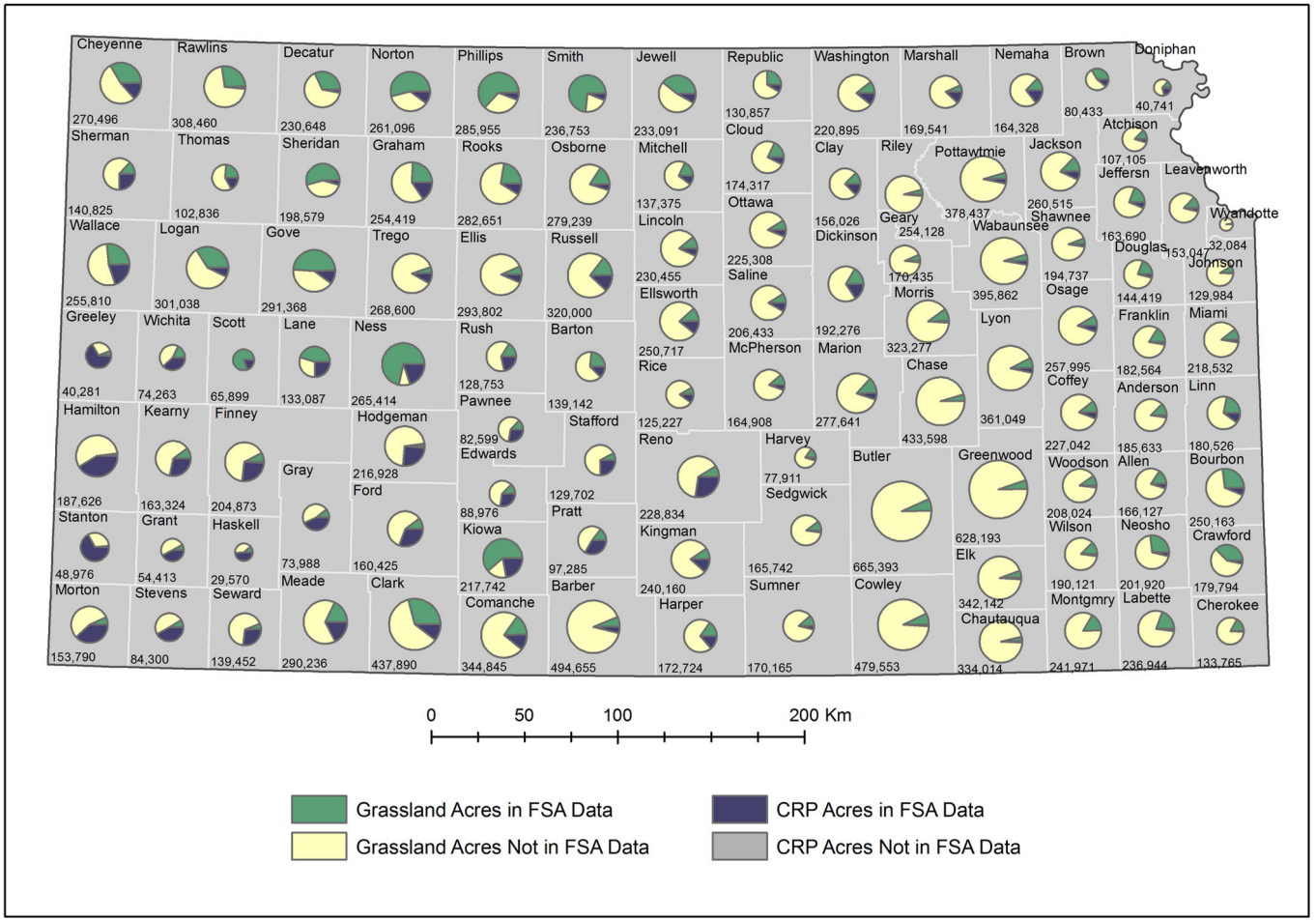


Figure 2.4. Fractions of grassland and CRP acres in the 2005 FSA data (green and blue) and acres not represented in the FSA data derived from supplemental data (land cover data and USDA county-level CRP data) (beige and gray). Pie charts are sized by total grassland acres labeled in each county.

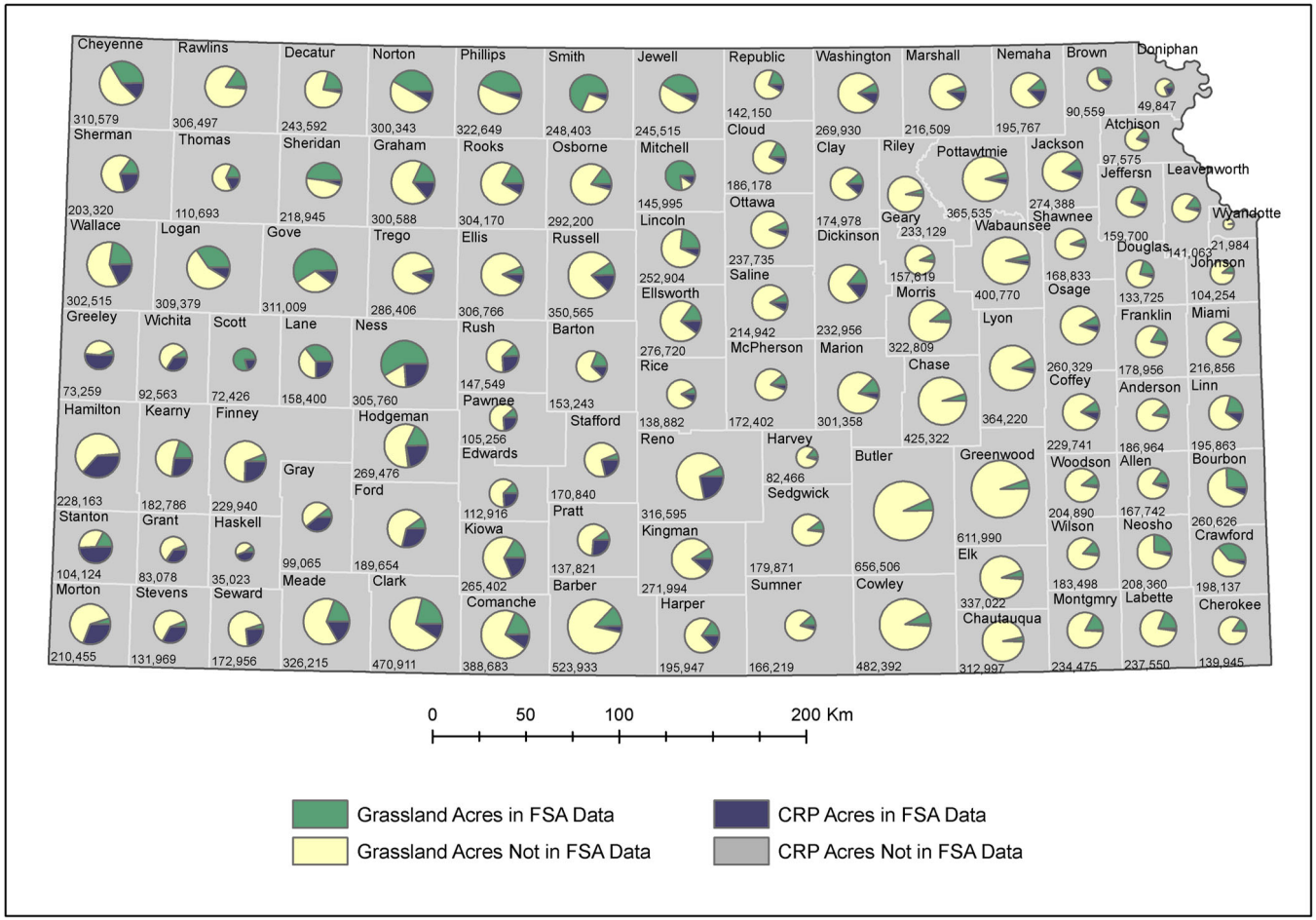


Figure 2.5. Fractions of grassland and CRP acres in the 2006 FSA data (green and blue) and acres not represented in the FSA data derived from supplemental data (land cover data and USDA county-level CRP data) (beige and gray). Pie charts are sized by total grassland acres labeled in each county.

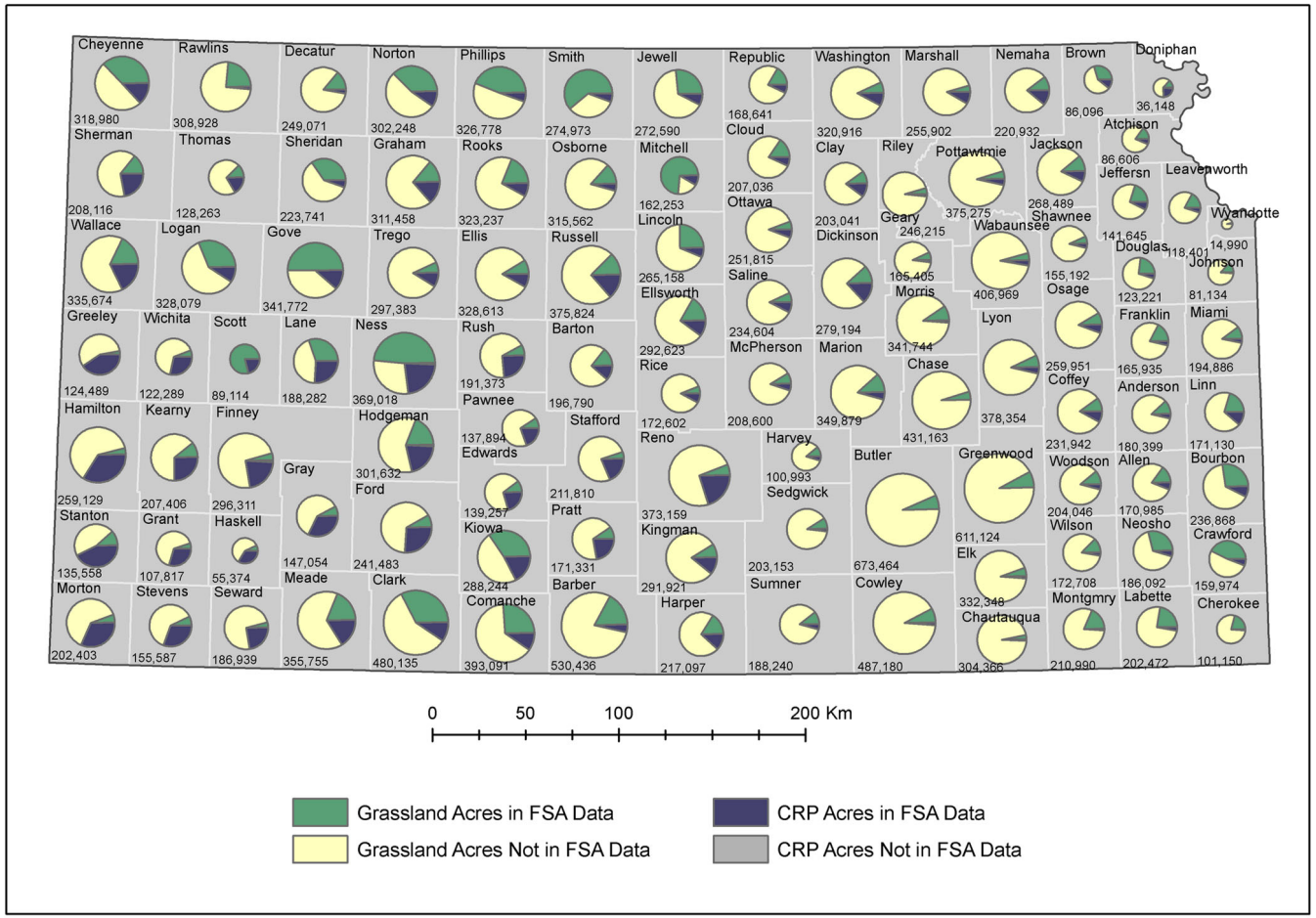


Figure 2.6. Fractions of grassland and CRP acres in the 2007 FSA data (green and blue) and acres not represented in the FSA data derived from supplemental data (land cover data and USDA county-level CRP data) (beige and gray). Pie charts are sized by the total grassland acres labeled in each county.

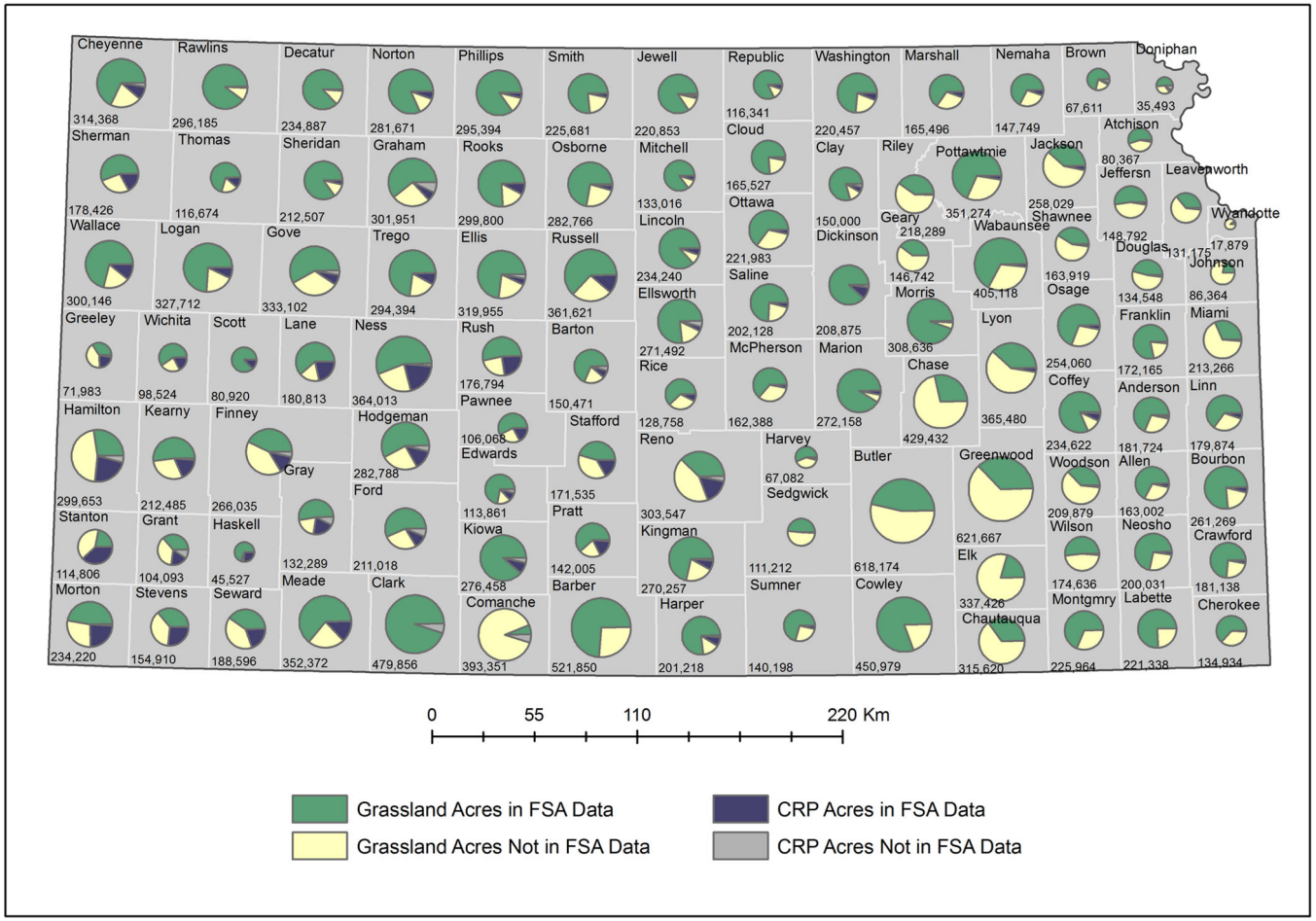


Figure 2.7. Fractions of grassland and CRP acres in the 2015 FSA data (green and blue) versus acres not represented in the FSA data derived from supplemental data (land cover data and USDA county-level CRP data) (beige and gray). Pie charts are sized by the total grassland acres in each county.

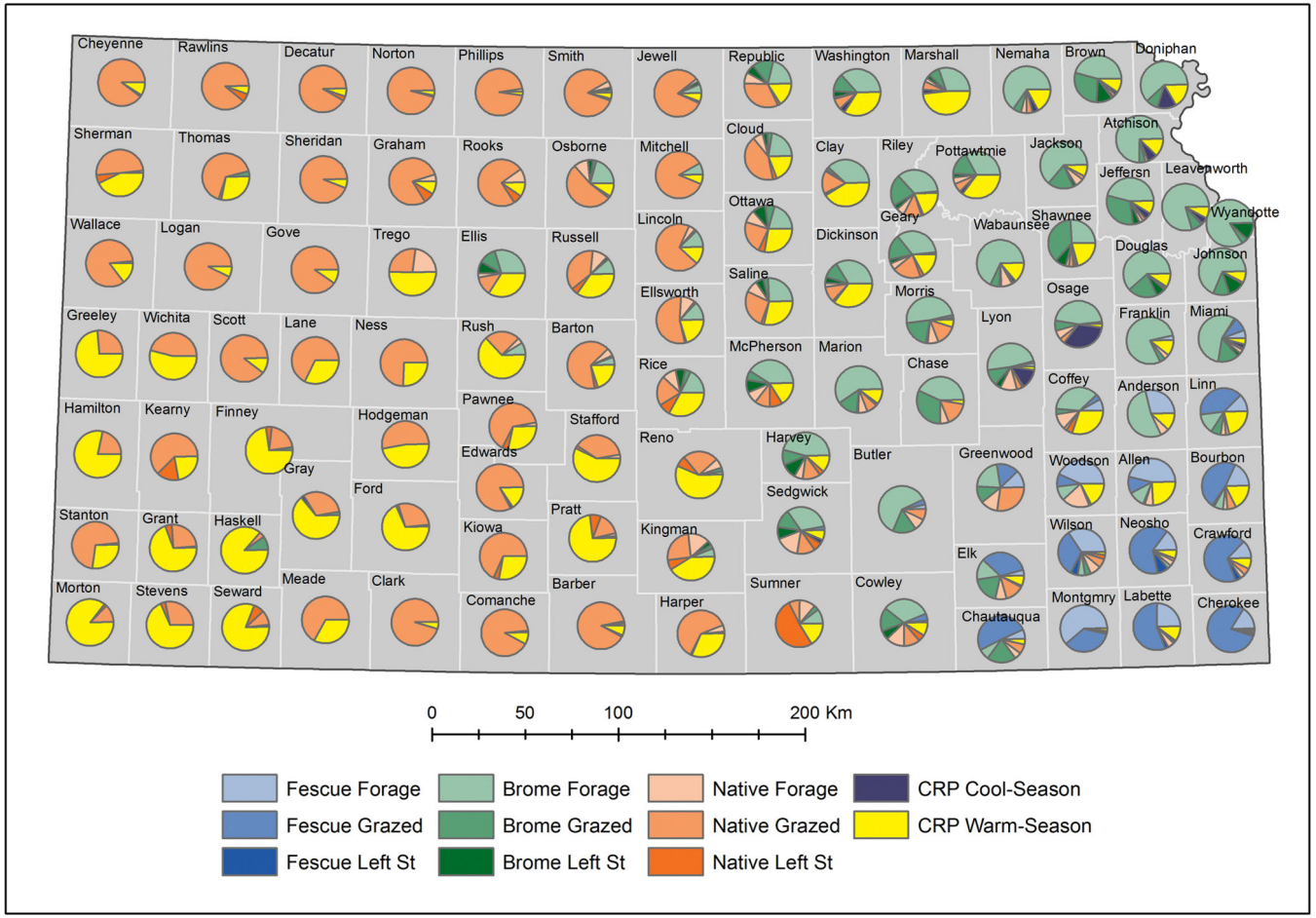


Figure 2.8. County fractions of grassland type and land use in the 2004 FSA data using fixed sized pie charts.

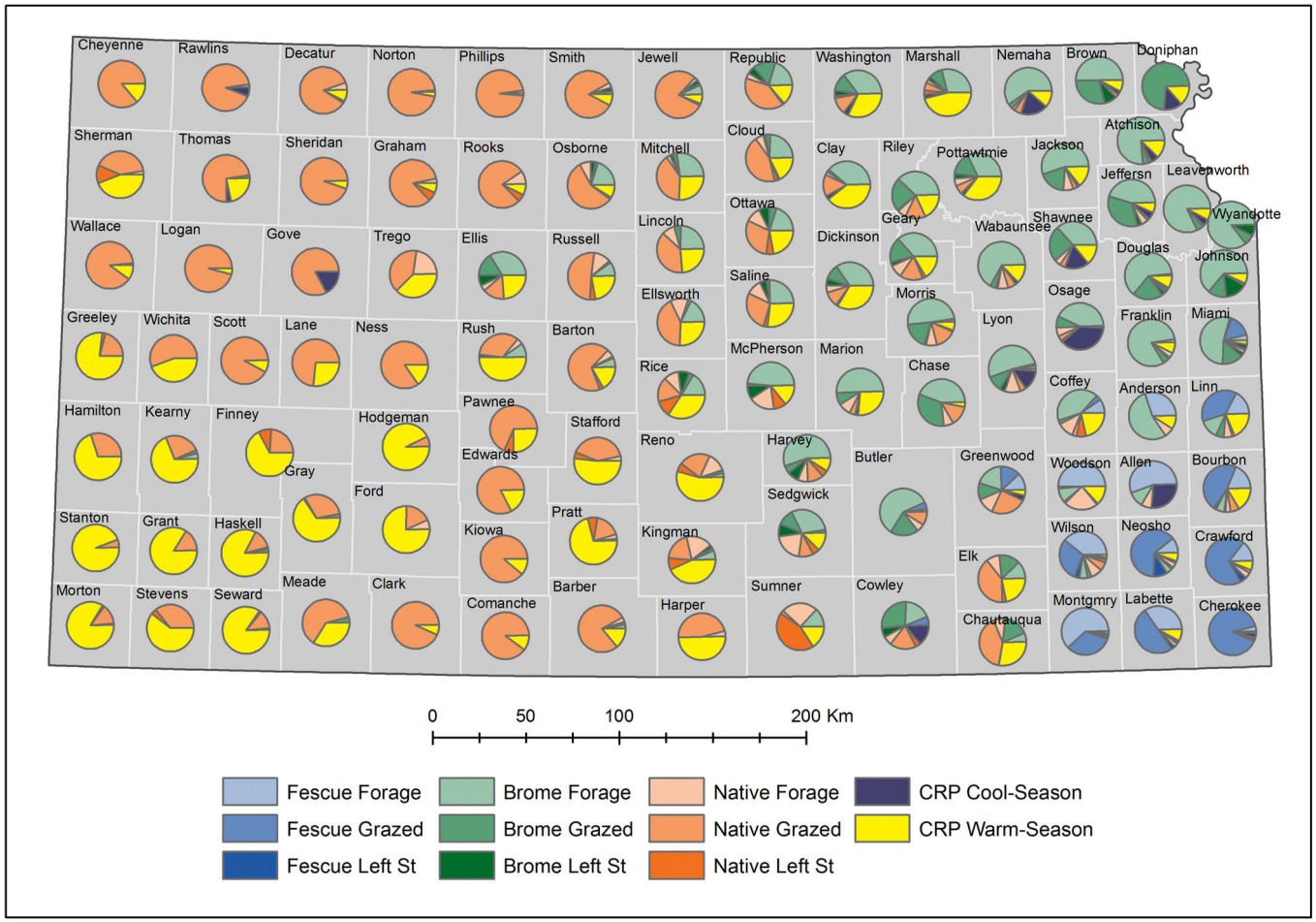


Figure 2.9. County fractions of grassland type and land use in the 2005 FSA data using fixed sized pie charts.

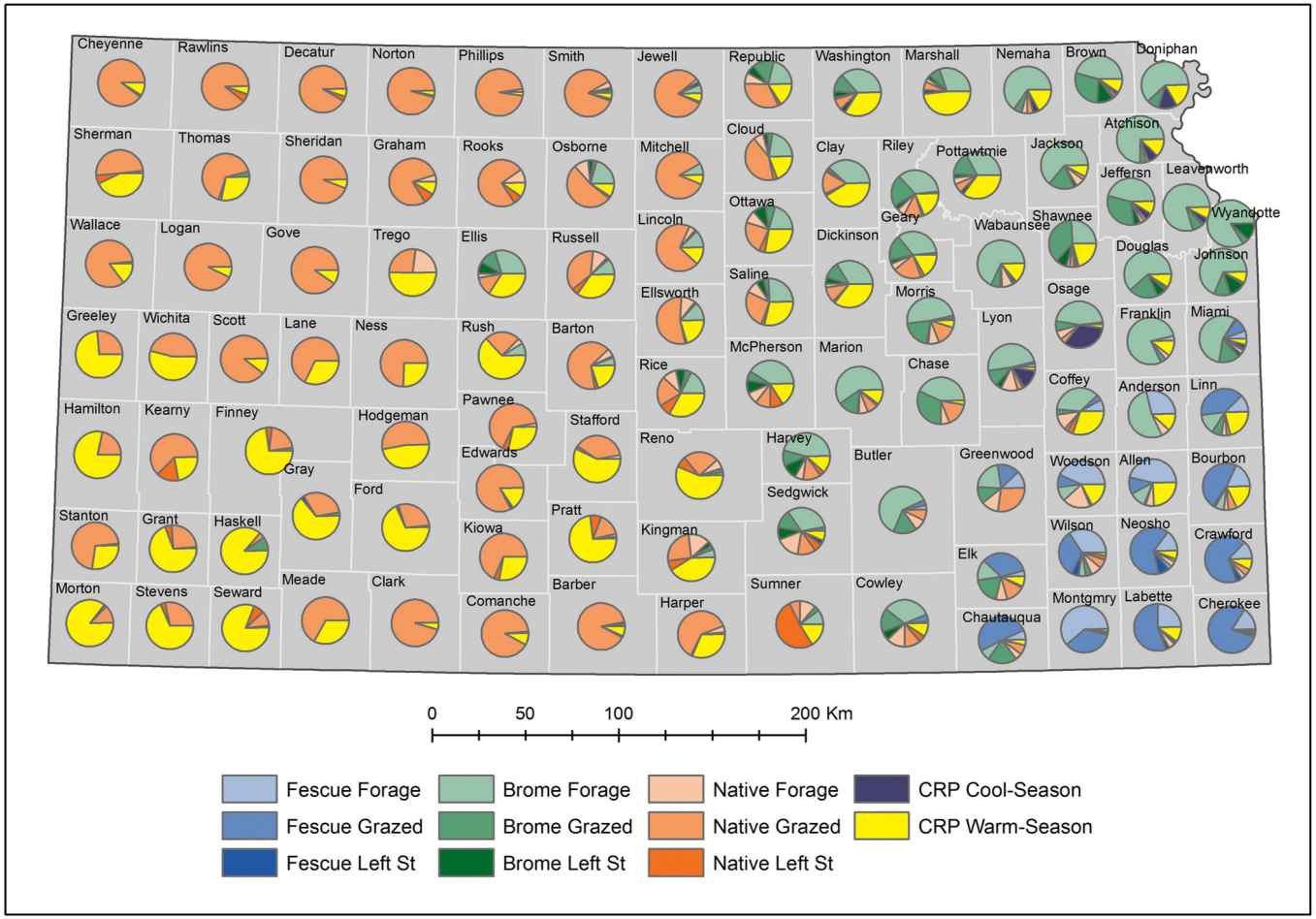


Figure 2.10. County fractions of grassland type and land use in the 2006 FSA data using fixed sized pie charts.

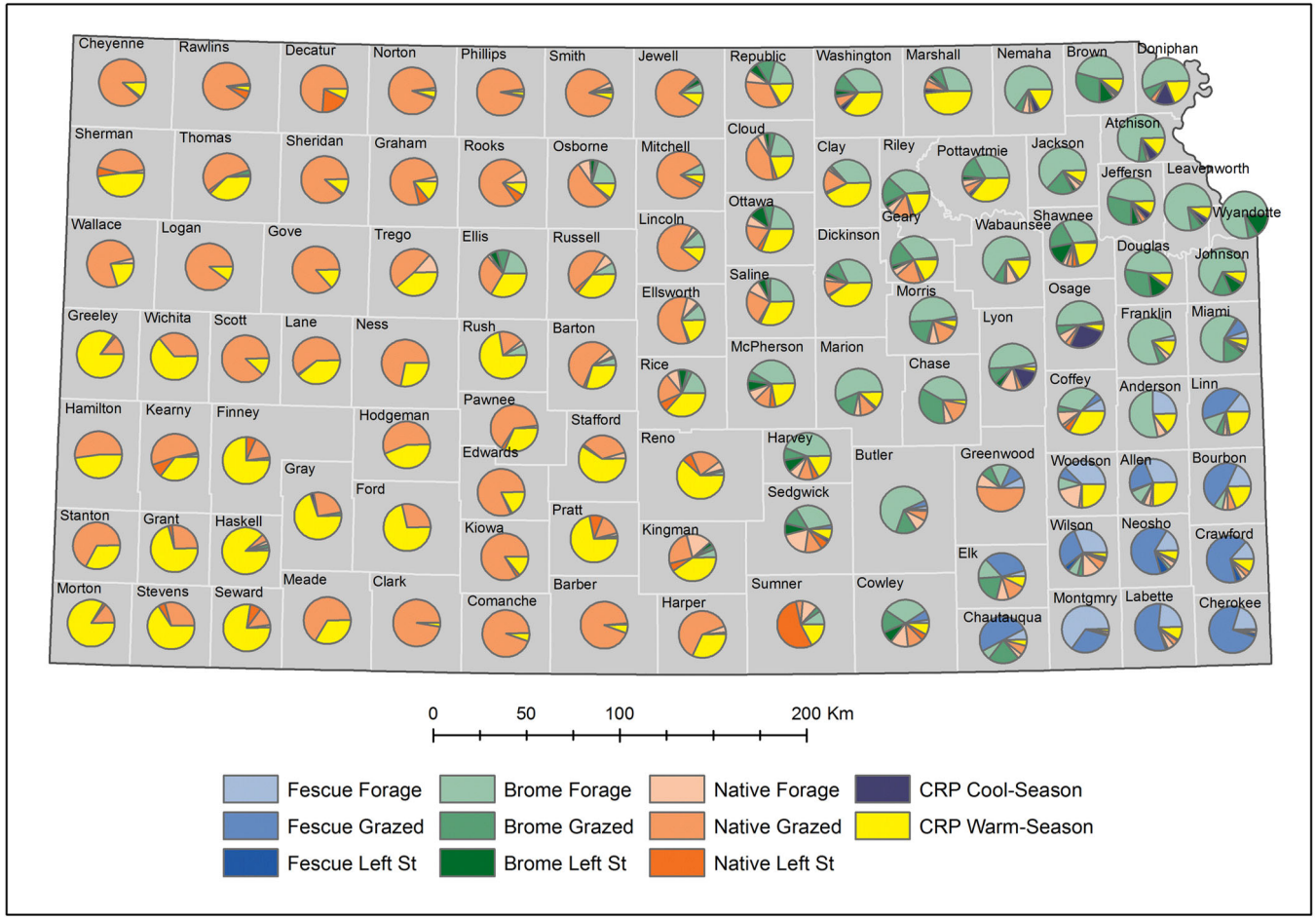


Figure 2.11. County fractions of grassland type and land use in the 2007 FSA data using fixed sized pie charts.

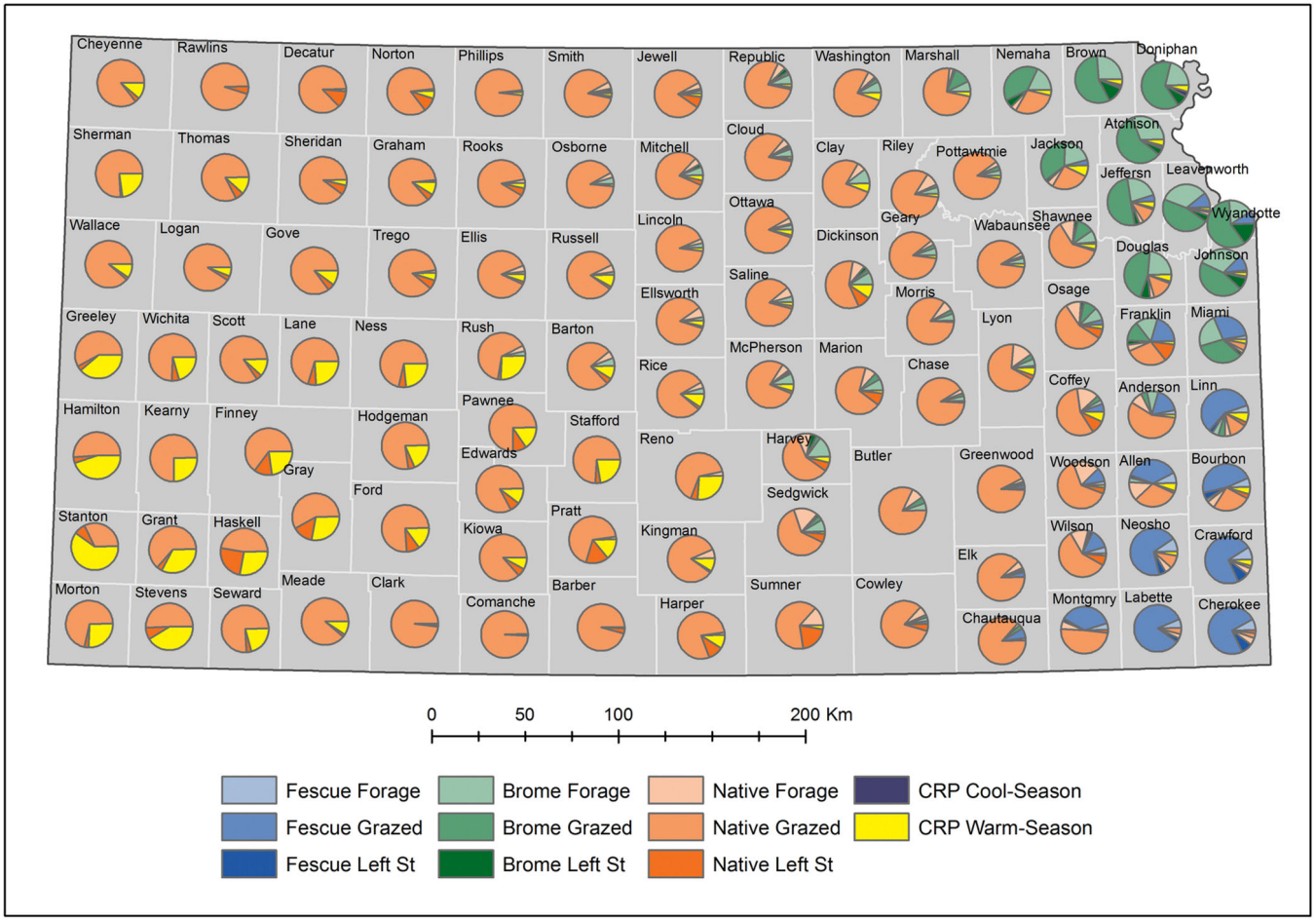


Figure 2.12. County fractions of grassland type and land use in the 2015 FSA data using fixed sized pie charts.

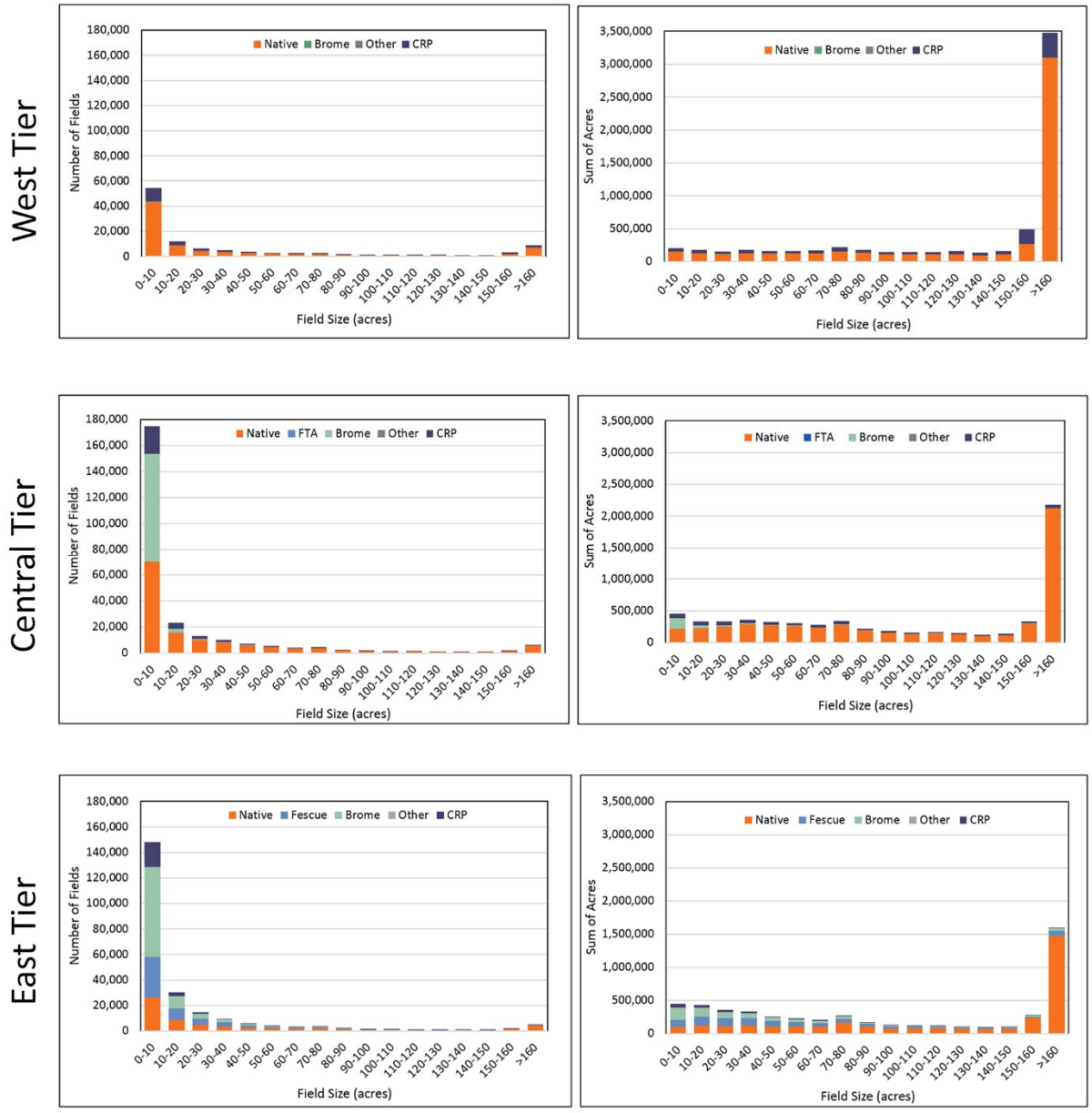


Figure 2.13. Frequency distributions of grassland field size stratified by grassland type in ten acre intervals for each tier in Kansas.

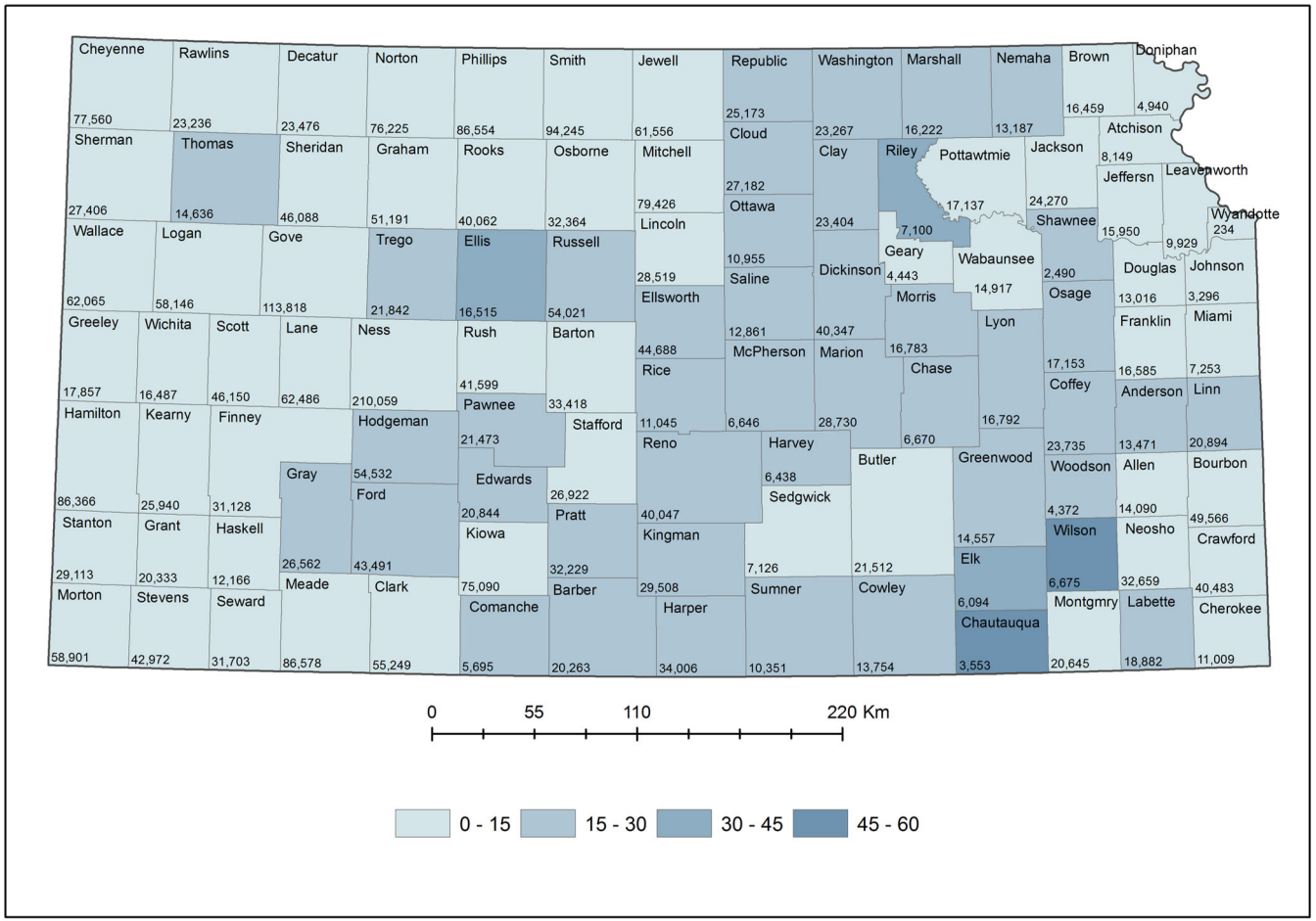


Figure 2.14. Map showing percent change in grassland type using field-level data in the 2004-2007 and 2015 FSA data. The values labeled in each county represent the acreage included in the change analysis.

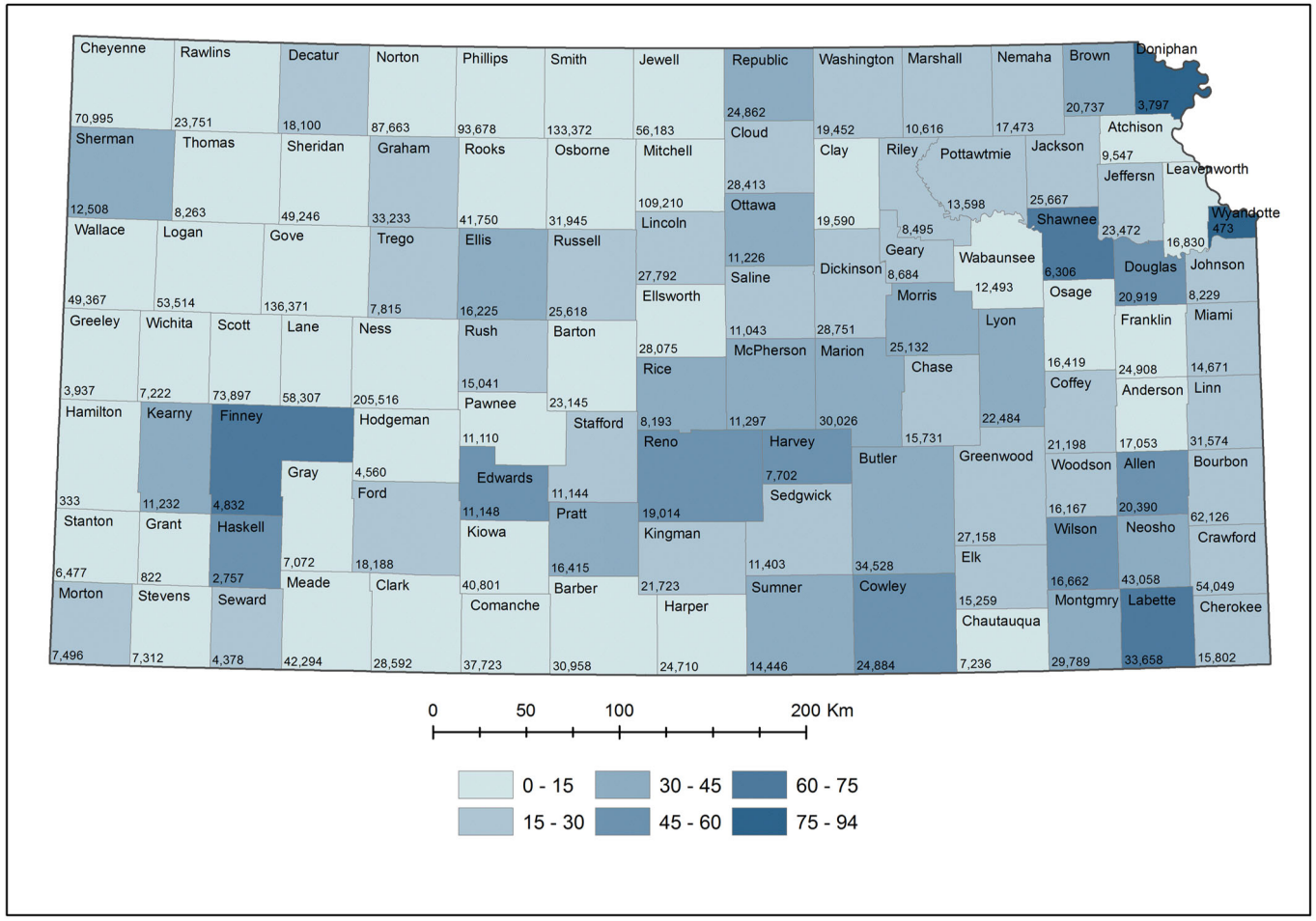


Figure 2.15. Map showing percent change in land use using field-level data in the 2004-2007 and 2015 FSA data. The values labeled in each county represent the acreage included in the change analysis.

Table 2.1. The three types and definitions of intended land use of Kansas grasslands in the FSA data (USDA 2013).

Intended Land Use	Definition
Forage	Intended for harvesting as food for livestock. Does not include crops grown for the intended purpose of grazing by livestock or grown for the intended purpose of grain which may be fed to livestock.
Grazing	Intended solely for pasture for livestock to roam and feed on.
Left Standing	Intended to be left in the field unharvested. Not intended to be mechanically or manually harvested for any purpose, grazed by domesticated livestock, or otherwise harvested in any manner. Typically used for erosion control and nutrient retention.

Table 2.2. The recoding scheme used to assign grassland types in the FSA data to one of seven grassland types used in the analysis.

Grassland Type	FSA Crop Type Name
CRP Warm-Season	Establishment of Permanent Native Grasses
	Rare and Declining Habitat
CRP Cool-Season	Establishment of Introduced grasses and legumes
CRP Other	Bottomland Hardwood Tree Establishment
	Cross Wind Trap Strip
	Diversion
	Duck Nesting Habitat
	Farmable Wetland - Buffer
	Farmable Wetland - Wetland
	Farmable Wetland Program - Aquaculture Wetland
	Farmable Wetland Program - Constructed Wetland
	Farmable Wetland Program - Flooded Prairie Wetland
	Field Windbreak Establishment
	Flood Control Structure
	Grass Contour Strip
	Grass Filter Strips
	Grass Waterway
	Hardwood Tree Planting
	Living Snow Fence
	Longleaf Pine Establishment
	Non-Floodplain Wetland Restoration
	Permanent Wildlife Habitat
	Riparian Buffers
	Salinity Reducing Vegetation Establishment

	Sediment Retention
	Shallow Water Areas for Wildlife
	Shelterbelt Establishment
	State Acres for Wildlife Enhancement
	Tree Planting
	Trees Already Established
	Wetland Buffer (Marginal Pasture)
	Wetland Restoration (Floodplain)
	Wildlife Food Plot
	Wildlife Habitat Buffer
	Wildlife Habitat Corridors
Native Warm-Season	Big Blue
	Big Bluestem
	Buffalo Grass
	Native
	Prairie
Fescue Cool-Season	Arctared Fescue
	Chewing Fescue
	Kentucky Fescue
	Meadow Fescue
	Rough Fescue
	Red Fescue
	Tall Fescue
Brome Cool-Season	Creeping foxtail
	Mountain Brome
	Other Brome
	Polar Brome
	Regar Brome
	Smooth Brome
Grass Other	Aeschynomene
	American Mamegrass
	American Vetch
	Annual Ryegrass
	Bahalia
	Basin Wild Rye
	Blue Grama
	Bluegrass, Alpine
	BlueJoint Reedgrass
	Broadleaf Signal
	Buffel
	California (Para)
	Canadian Wild Ryegrass
	Canary
	Canby

	Coastal Bermuda
	Crabgrass
	Crested Wheat
	Eastern Grama
	Garrison Creeping Fxtl
	Gordo Bluestem
	Grama, Blue Lovington
	Grama, Hairy
	Grama, Side Oats
	Green Panic
	Hybrid Bent
	Hybrid Bermuda
	Illinois Bundle Flower
	Indian
	Intermediate Ryegrass
	Intermediate Wheat
	Johnson
	Jose Tall Wheatgrass
	Kleberg Bluestem
	Leriope
	Little Bluestem
	Magnar
	Mason Sandhill Lovegrass
	Matua
	Maxmillian Sunflower
	Meadow
	Mission
	Mutton
	Napier
	Needle And Thread
	Old World Bluestem
	Oldworld Bluestem
	Orchard
	Other Bent
	Pampas
	Perennial Ryegrass
	Plains Blue Stems
	Prairie Dropseed
	Prarie Sandreed
	Pubescent Wheat
	Red Ratibita
	Redtop
	Reed Canary
	Russian Wild Ryegrass

	Sainfoin
	Sand Bluestem
	Sand Lovegrass
	Secar Bluebunch
	Siberian Wheat
	Side Oats Grama
	Slender Hair
	Small Burnett
	Soft Stem Blurush
	Spike Muhley
	Sprigs Bermuda
	Sudan
	Sun
	Switch
	Thick Spike Wheatgrass
	Timothy
	Tufted Hairgrass
	Turf
	Virginia Wild Rye
	Virginia Wildrye
	Weeping Lovegrass
	Wheat , Slender
	Wheat , Tall
	Wheat Streambank
	White Prairie Clover
	Worm Grass
	Zoysia

Table 2.3. A comparison of statewide grassland acres in the FSA data and annual land cover data.

Program Year	Grassland Acres in FSA Data	Count of Grassland Fields in FSA Data	Grassland Acres from Annual Land Cover Data	Percent Grassland Represented by FSA data
2004	7,546,492	328,386	22,451,617	33.61%
2005	7,054,722	360,225	22,379,843	31.52%
2006	7,323,636	378,287	24,104,610	30.38%
2007	7,655,929	388,135	25,540,048	29.98%
2015	18,019,038	607,500	23,720,380	75.96%

Table 2.4. A comparison of statewide acres of land enrolled in CRP in the FSA data and USDA FSA Program statistics

Program Year	CRP Acres in FSA Data	Count of CRP Fields in FSA Data	USDA Reported Acres Enrolled in CRP	Percent CRP in FSA data
2004	2,646,551	83,531	2,828,911	93.55%
2005	2,833,861	91,358	2,878,784	98.44%
2006	3,028,743	101,998	3,085,227	98.17%
2007	3,208,741	110,411	3,258,989	98.46%
2015	2,077,132	85,311	2,182,877	95.16%

Table 2.5. The number of fields and acres screened for use in evaluating change in grassland type and land use. Criteria were defined as having at least three years of data and reported acres remained nearly constant.

Stratification	Data Assessment	Number of Fields	Acres
Grassland Type	Excluded, criteria not met	36,353	1,031,566
	Included, criteria met	181,667	3,275,514
Land Use	Excluded, criteria not met	24,120	902,693
	Included, criteria met	193,900	3,404,388

Table 2.6. The number of fields, acres and percent of change and no change in grassland type and land use in Kansas for fields meeting the criteria defined in Table 5.

	Trajectory	Number of Fields	Acres	Percent Fields	Percent Acres
Grassland Type	Unlikely Change	5,275	47,663	2.9%	1.5%
	Change	23,653	371,842	13.0%	11.4%
	No Change	152,739	2,856,008	84.1%	87.3%
Intended Land Use	Change	78,037	793,655	23.3%	40.3%
	No Change	115,863	2,610,733	76.6.%	59.8%

References

Banks, E. B. (2012). *Kansas Conservation Reserve Program Technical Guidance Number 82*. Retrieved from NRCS: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/ks/programs/?cid=nrcs142p2_032831

Gao, J., Sheshukov, A. Y., Yen, H., Kastens, J. H., & Peterson, D. L. (2017). Impacts of incorporating dominant crop rotation patterns as primary land use change on hydrologic model performance. *Agriculture, ecosystems & environment, 247*, 33-42. doi:https://doi.org/10.1016/j.agee.2017.06.019

Guo, X., Price, K. P., & Stiles, J. M. (2000). Biophysical and Spectral Characteristics of Cool- and Warm-Season Grasslands under Three Land Management Practices in Eastern Kansas. *Natural Resources Research, 9*(4), 321-331. doi:10.1023/a:1011513527965

KSU. (2018). AgReport. Retrieved from http://www.k-state.edu/agreport/faculty/fall2018/administrativemerger.html

Lamond, R. E., Fritz, J. O., & Ohlenbusch, P. D. (1992). *Smooth Brome Production and Utilization*. Manhattan, Kansas: Kansas State University.

Lin, X., Harrington, J., Ciampitti, I., Gowda, P., Brown, D., & Kisekka, I. (2017). Kansas Trends and Changes in Temperature, Precipitation, Drought, and Frost-Free Days from the 1890s to 2015. *Journal of Contemporary Water Research & Education, 162*(1), 18-30. doi:doi:10.1111/j.1936-704X.2017.03257.x

Peterson, D. L., Egbert, S. L., & Martinko, E. A. (2017). *2015 Kansas Land Cover Patterns Phase I: Final Report.* Lawrence, Kansas: Kansas Biological Survey, University of Kansas.

Shoup, D., Kilgore, G. L., & Brazle, F. K. (2010). *Tall Fescue Production and Utilization*. Manhattan, Kansas: Kansas State University.

Stubbs, M. (2018). Agricultural Disaster Assistance, Congressional Research Service Report. 16.

USDA, F. S. A. (2013a). *FSA Handbook for State and County Offices: Acrage and Compliance Determinations*. (2-CP (Revision 15), Amendment 58). Washington D.C. Retrieved from https://www.fsa.usda.gov/Internet/FSA_File/2cp15-76.pdf.

USDA, F. S. A. (2017). CRP Enrollment and Rental Payments by County, 1986-2017. Retrieved November 13, 2017 https://www.fsa.usda.gov/programs-and-services/conservation-programs/reports-and-statistics/conservation-reserve-program-statistics/index

USDA, N. R. C. S. (2009). Fact Sheet: Grassland Reserve Program.

USDA, N. R. C. S. (2013b). *Summary Report: 2010 National Resources Inventory*. Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa Retrieved from http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1167354.pdf.

Weaver, J. E. (1954). *North American prairie / J.E. Weaver*. Lincoln Neb: Johnsen.

3 Chapter 3 Exploring the Spectral Characteristics and Separability of Four Grassland Type Hierarchies Using Landsat 8 and MODIS NDVI

3.1 *Abstract*

This study used Jeffries-Matusita (JM) distance statistics and spectral profiles to compare the spectral separability of four hierarchies of grassland types in northeastern Kansas. Three remotely sensed datasets from 2015 (three-date multispectral Landsat 8, three-date Landsat 8 NDVI, and 23-period, 16-day composite MODIS NDVI time series) and 2015 reference data from the USDA Farm Service Agency (FSA) were used in the analyses. The results will be used to determine the optimal dataset(s) for regional scale mapping of grassland types, the hierarchy of grassland types to be mapped, and whether land use affects the spectral separability of grassland types. The results show that combining the three datasets maximized the JM distance statistics, and thus the spectral separability of grassland types across all grassland type hierarchies. Individually, the three-date multispectral Landsat 8 dataset had the highest JM distance statistics, followed by MODIS NDVI time series, and three-date Landsat 8 NDVI. Spectral profiles and by-band and by-period JM distance statistics indicate that the spring and fall Landsat near-infrared (NIR) bands and spring and fall NDVI were more important than summer for spectral separability between grassland types. There was variability in JM distance statistics separability comparisons when incorporating land use, indicating that land use does affect spectral separability in some instances. However, the JM distance statistics were not dramatically reduced when land use types were aggregated to coarser grassland types (Level-1 and Level-2), indicating that land use does not negatively affect the spectral separability of functional grassland types. There was moderate to high separability between land enrolled in the

conservation reserve program (CRP) and native grasslands and low separability between fescue and brome. The results suggest that brome and fescue should be combined into one class for grassland type mapping, and that it may be possible to map CRP and native grasslands either as one class or separately.

3.2 *Introduction*

The tallgrass prairies of the Great Plains in North America, considered one of the more biologically diverse grasslands in the world (Risser, 1988), have the greatest reported loss of grassland area with estimates of only 9.4% - 13% of the original tallgrass prairie remaining (Gibson, 2009; Samson *et al.*, 2004). It has been estimated that the tallgrass prairie once occupied 167 million acres, stretching east into western Ohio, west to the eastern third of Kansas and Nebraska, north into southern Manitoba, and south into portions of Texas (Robertson *et al.*, 1997). Kansas has an estimated 18% of its original tallgrass prairie extent, the largest of any state, and the largest contiguous tract of tallgrass prairie in the region known as the Flint Hills. Meanwhile, many other states including Indiana, Illinois, Iowa, and Missouri contain less than a half percent of their original extent (Risser, 1988; Robertson & Schwartz, 1994a). Fragmentation of the tallgrass prairie in the eastern Great Plains began in the early 1800s when European settlers converted "the Great American Desert" into cropland and non-native grasslands for domestic livestock grazing (Samson *et al.*, 2004). Most of the Great Plains and eastern tallgrass prairie remnants now are privately owned and subjected to a variety of land management practices, including grazing and haying for domestic livestock (Owensby, 1993).

Mapping and monitoring the extent, distribution and condition of remaining tallgrass prairie are critical for ensuring preservation and sustainability of these biologically diverse grasslands under global demands for increased food production and biofuels as well as the

stresses of climate change. Such map products have and will continue to be used for conservation and research applications and initiatives for pollinators, upland and migratory birds, and other grassland species. In addition to tallgrass prairie, another grassland type that is of particular interest is land enrolled in the Conservation Reserve Program (CRP). CRP is a United States Department of Agriculture (USDA) Farm Service Agency (FSA) program that began in 1985 and is the largest private-lands conservation program in the U.S. The CRP program offers a 10-15 year contract to landowners where "environmentally sensitive farmland" is taken out of crop production and planted with long-term grass or tree cover in an effort to reduce soil erosion, improve water quality, or improve habitat for wildlife (Ribaudo *et al.*, 1990; Wu & Weber, 2012). Many studies have shown the benefits of CRP to wildlife by providing habitat and landscape connectivity (Hughes *et al.*, 1999; Reynolds *et al.*, 2001; Riffell *et al.*, 2008; Van Pelt *et al.*, 2013). Acreage enrolled in CRP peaked nationally and in Kansas in 2007, when Kansas had over 3 million acres of CRP USDA, 2017). The 2014 Farm Bill set a national cap of 24 million CRP acres resulting in a competitive enrollment process among land owners (Hellerstein, 2017). In 2017, CRP acreage in Kansas declined 37% since the 2007 peak. There are ongoing concerns about losing the environmental services CRP provides with the reduction of allowable enrollments and with commodity prices encouraging land owners to convert these marginal agricultural lands back into cropland production (Gelfand *et al.*, 2011; Johnston, 2014; Wright & Wimberly, 2013).

Accurate and ongoing mapping of the landscape provides tools to monitor the changing landscape, including environmental and socio-economic drivers, and provides the opportunity for conservation planning. Remotely sensed data have been used to map and monitor grasslands, including the tallgrass prairie and land enrolled in CRP. Studies have used remote sensing

technology to monitor and model biophysical characteristics of grasslands including functional distributions (i.e. C3 and C4 grasslands), productivity (biomass and cover) and grassland use that can alter grassland biophysical characteristics and quality. For example, several studies have used remotely sensed data to map or predict distributions and abundance of C3 and C4 grasslands. Tieszen *et al.* (1997) used time series AVHRR Normalized Difference Vegetation Index (NDVI) data to characterize the spatial and temporal distribution of C3 and C4 grasslands in the Great Plains over a five-year period. Davidson and Csillag (2003) also used AVHRR NDVI to compare three approaches to predict the relative abundance of C4 cover in a Canadian mixed-grass prairie. They found a two-date ratio, early season NDVI to late season NDVI, best predicted C4 abundance (Davidson & Csillag, 2003). Meanwhile Foody and Dash (2007) used a 30-week time series of MERIS Terrestrial Chlorophyll Index (MTCI) to map high, medium, and low C3 cover in South Dakota with an overall accuracy of 77%. In addition, Gu and Wylie (2015) leveraged the spatial resolution of Landsat 8 NDVI and the temporal resolution of MODIS NDVI in a rule-based piecewise regression to produce a 30-m grassland productivity map of the Greater Platte River Basin, Nebraska. Zha *et al.* (2003) found that percent vegetation cover in a semi-arid grassland in a western China could be mapped with an accuracy of 89% using calibrated Landsat TM NDVI. With growing interest in using CRP lands or marginal croplands for biofuel feedstock, Porter *et al.* (2014) used Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM) multispectral Landsat 8 and NDVI data to estimate biomass in a CRP pasture to within 8% of the *in situ* measurements. Understanding the distribution, abundance, and productivity of C3 and C4 grasslands is important, as the two grassland types respond differently to environmental change due to grazing intensity, fire frequency, nutrient regimes, and climate change (Tieszen *et al.*, 1997).

Other studies have used remotely sensed data to map thematic grassland classes that are represented by either their dominant functional group or as native and non-native grassland types. Using multi-seasonal ASTER NDVI, Wang *et al.* (2010) mapped cool-season (non-native) and warm-season (native) grasslands in western Missouri with an accuracy of 80%. The authors found that spring and summer NDVI provided the highest separability between these two grasslands types due to their asynchronous phenology, with maximum productivity reached in May and July for cool and warm-season grasslands, respectively. Another study used discriminant analysis and MODIS NDVI time series to spectrally separate native and non-native dry mixed-grass prairie in Alberta, Canada with an overall accuracy of 73% (McInnes *et al.*, 2015). Meanwhile, a pilot study by Peterson *et al.* (2008) found that multi-seasonal Landsat TM data better separated native (warm-season) and non-native (cool-season) grassland types in the Flint Hills ecoregion than coarser resolution MODIS NDVI time series.

Many of these studies and mapping efforts rely on the asynchronous phenology of cool- and warm-season grasslands. However, grasslands are used and managed extensively and intensively. The type, combination, timing, and intensity of land management practices within grassland types alter the biophysical properties of grasslands, including vegetation productivity and composition and soil structure and chemistry, which in turn potentially alters spectral responses that can complicate the ability to accurately map grassland types. Several studies have used remotely sensed data to characterize and monitor land management practices and land use intensity occurring within grasslands. For example, Guo *et al.* (2003) and Guo *et al.* (2000) used multi-seasonal field data and Landsat TM imagery to show that biophysical and spectral characteristics were significantly different among three common land management practices in cool-season (non-native) and warm-season (native) grasslands in Douglas County, Kansas.

Discriminant analysis showed the two grassland types and the three treatments in the two grassland types could be separated with an accuracy of 90.1% and 70.4%, respectively (Price *et al.*, 2002a). Peterson *et al.* (2002b) obtained similar results when using discriminant analysis to separate grazed cool- and warm-season grasslands in the same county. Another study by Lauver and Whistler (1993) found significant differences in the biophysical characteristics (species diversity, plant cover and biomass) of high-quality (hayed) and low quality (overgrazed) tallgrass prairie remnants in Anderson County, Kansas, that were mapped using single-date Landsat TM data and probability thresholding with moderate success (63% overall accuracy). Franke *et al.* (2012) found that multi-temporal RapidEye data and a decision tree classifier could map grassland land use intensity in a 500 km^2 grassland area in Germany with accuracies up to 85.7%. A study by Halabuk *et al.* (2015) used MODIS NDVI and enhanced vegetation index (EVI) to detect haying events in prairie hay meadows in Slovakia with accuracy levels as high as 85%. While these studies provide examples of successful results for grassland mapping and monitoring, they primarily occur on a relatively small scale.

With regard to mapping CRP land, a post-classification trajectory approach has been used to map CRP (Egbert *et al.*, 1998). The trajectory logic identifies CRP when pixels are mapped as cropland in the first temporal period and mapped as grassland subsequently in the second temporal period. Song *et al.* (2005) used a combination of multi-temporal Landsat TM, including multiple indices, image texture, and terrain layers to map CRP land in Texas County, Oklahoma. The authors found that support vector machine (SVM) outperformed the decision tree classifier (DTC) by mapping CRP with an overall accuracy of 96%. A study in southwest Kansas used 203 Landsat TM images from 1984 to 2010 to produce maximum NDVI composites to map land use conversion, specifically cropland to grassland, to identify CRP. The mapping results were

compared to the 2005 Kansas Land Cover Patterns map (KARS, 2008) as reference data and showed 91% agreement (Maxwell & Sylvester, 2012).

While previous studies in eastern Kansas have evaluated the biophysical characteristics of grasslands and have used field and satellite-acquired spectral data to statistically discriminate between grassland types and land management practices, little research has focused on identifying an optimal thematic classification approach for mapping grassland types at a regional scale. Multiple factors must be considered when developing such a land cover classification approach. One key factor is determining what source data or combination thereof maximizes the ability to map the defined grassland types. Another is defining what grassland types will be or can be mapped, meaning the thematic classification scheme. A final factor (that is specific to this study but potentially extensible to others) is understanding the impacts of land management on the spectral separability of grassland types. While the timing and intensities of land management practices vary, there is uncertainty as to if and how land management practices affect the spectral characteristics within and between grassland types. The objective of this study is to use Landsat 8 and MODIS NDVI time series data from 2015 to evaluate the spectral characteristics and separability of spectral profiles for a hierarchy of grassland types and land use in a highly fragmented landscape in northeastern Kansas. Comparison of the spectral and temporal resolutions of multispectral Landsat 8 data, Landsat 8 NDVI, and MODIS NDVI provides a framework for identifying optimal dataset(s) for image classification. In addition, using a hierarchy of grassland types enables identification of an appropriate thematic classification scheme and evaluates the effect of land management on spectral characteristics.

3.3 *Methods*

3.3.1 *Study Area*

Kansas exhibits an east-west precipitation gradient and a north-south temperature gradient, with higher precipitation occurring in the east and lower temperatures occurring in the north. Kansas grassland types largely follow the east-west precipitation gradient with tallgrass prairie in the relatively wet east, mixed prairie in central Kansas, and shortgrass prairie in the semi-arid west. The study area (shown in grey) falls within the Landsat path/row 27/33 (shown in red) in eastern Kansas where the landscape is dominated by a mosaic of agriculture, grasslands and urban areas (Figure 1). Areas falling outside of Kansas were not part of the defined study area.

There is an inherent east-west land use/land cover gradient within the study area. The western edge of the study area is in the Flint Hills, which is the largest remaining tract of native tallgrass prairie in the world. In the Flint Hills, native grasslands dominate and non-native grasslands and croplands are scattered in the river lowlands. Moving eastward from the Flint Hills the landscape becomes highly fragmented and more complex. Cropland becomes prevalent and dominant crops planted include corn, soybeans, grain sorghum, alfalfa, and winter wheat. Grasslands east of the Flint Hills consist both of native, warm-season dominated grasslands and non-native, cool-season grasslands. Cropland, woodlands, and forests occupy the river lowlands and riparian areas.

Warm-season grasslands are either native tallgrass prairie or have been reseeded using a native seed mixture. Warm-season grasslands fix carbon using C4 photosynthesis and are dominated by native bunchgrasses such as big bluestem (*Andropogon gerardii*), little bluestem (*Schizachyrium scoparium*), and indiangrass (*Sorghastrum nutans*) and in high quality prairie native forbs such as leadplant (*Amorpha canescens*), butterfly weed (*Asclepias tuberosa*), and

purple coneflower (*Echinacea angustifolia*). The typical phenology of warm-season grasslands is spring green-up, peak productivity in late spring to early summer when temperatures increase, followed by senescence in fall (Weaver, 1954).

In this study area, cool-season grasslands are defined as non-native grasslands. These grasslands are planted with non-native herbaceous species such as smooth brome (*Bromus inermis*) and tall fescue (*Festuca arundinacea*). Cool-season grasslands fix carbon using C3 photosynthesis. The typical phenology of cool-season grassland is early spring green-up, peak productivity in late spring, a mid-summer semi-dormancy and, with sufficient precipitation, a second, smaller growth period in early fall (Weaver, 1954). Haying and grazing are two common land uses for both grassland types. However, the timing, intensity, and frequency of management practices within each grassland type vary by land owner and by climate conditions in a given year. In addition to grazing, prescribed burning is a common management practice for maintaining the diversity in native warm-season grasslands and preventing woody encroachment.

3.3.2 *Data Sources*

The FSA maintains annual field-level records (referred to as FSA 578 data) of acreage, land cover, and intended land use for all fields participating in USDA program. In Kansas, county-level field offices maintain FSA data, where land owners or producers report land cover and land use information for eligibility for the upcoming USDA program year. Historically these data were maintained by county field offices using photocopies of aerial photos with land cover and land use information annotated on the hardcopy. Today these data are maintained as a geodatabase of field boundaries known as Common Land Units (CLUs), defined by FSA as the smallest land unit with the same ownership, land cover and land use, however some fields can be split into sub-fields. Each CLU is annotated with a number of attributes including crop type, land

use, reported acreage, county FIPS code, farm number, and tract number. In the past, these data have been made available to scientists to use for training and validating several land cover mapping efforts in Kansas (Kennedy, 1999; Mosiman, 2003; Peterson *et al.*, 2005; Wardlow & Egbert, 2008). Through a Memorandum of Use (MOU) with the Kansas FSA office, 2015 CLU and FSA 578 data were acquired for a state-wide land cover mapping project. There are more than one million polygons in the 2015 CLU database with reported crop types and intended land use. Grassland type information is available in the database as a crop type. "Intended Land Use Code" identifies the land use that the land owner intends to use the grassland for during the upcoming year and includes the categories of forage, grazing, and left standing (not grazed or hayed for forage). For record in the FSA 578 data and each feature in the CLU polygon shapefile, a unique identifier was created by concatenating the following attributes: State FIPS code, County FIPS code, Tract Number, and Farm Number (SCTF).

Three datasets of remotely sensed imagery were assembled for the study and include Landsat 8 surface reflectance, Landsat 8 NDVI, and Terra MODIS NDVI time series. Three Landsat 8 surface reflectance images were ordered and acquired using USGS's EarthExplorer (EE) tool https://earthexplorer.usgs.gov/ to represent the spring, summer and fall portions of the growing season for path/row 27/33. The dates of the imagery obtained were 03/30/2015, 06/12/2013, and 11/09/2015 – no cloud-free imagery was available for the summer of 2015, and while not ideal, it is uncommon for grasslands to change in type or use or to be converted to a different land cover from year to year, so this scene from 2013 provided a reasonable proxy. Monthly reports from the High Plains Regional Climate Center (HPRCC) show that both June 2013 and July 2014 were substantially drier than June and July in 2015 (Umphlett, 2013)and thus represents a potential limitation in the summer Landsat data used in this study. Using ERDAS

Imagine, six multispectral bands (bands 2 – 7) were extracted from the three image dates and combined to produce an 18-band multi-seasonal Landsat 8 dataset. Using the same dates listed above, spring, summer, and fall surface reflectance NDVI images were acquired and stacked to create a three-date multi-seasonal Landsat 8 NDVI dataset. Lastly, a biweekly time series of 231-meter Terra MODIS 16-day composite NDVI from the 2015 growing season was downloaded from NASA's EarthData online tool, https://earthdata.nasa.gov/. The data were reprojected from the native Sinusoidal projection to Albers Equal Area projection and clipped to the Landsat WRS path/row (27/33) extent. The MODIS NDVI time series dataset was resampled to 30-meter pixels and snapped to the Landsat 8 pixel grid. The three datasets were stacked to create a 44-band dataset.

Two qualifiers were used to identify MODIS pixels suitable for the spectral separability analysis. Using ESRI ArcGIS, a polygon file of the 231-m MODIS pixel footprints was used to calculate the percentage of grassland in each MODIS pixel using the 2015 Level-I Kansas Land Cover Patterns dataset (KARS, 2008). Next, the MODIS pixel footprints and the 2015 CLU boundaries were intersected to calculate the proportion of the pixel interior to a field. MODIS pixels containing greater than 60% grassland and 60% interior to a CLU were selected for the analysis. The centroids of the MODIS pixels were used to extract reflectance and NDVI values from the 44-band image stack. The centroid was intersected with the USGS's high-resolution National Hydrography Dataset (NHD) waterbody feature layer to exclude point locations that fell within farm ponds that would affect the Landsat reflectance values. In addition, NDVI values profiles and values were inspected and pixels where the cumulative summer NDVI (periods 10 – 14) were less than 2500 were flagged as outliers and were excluded from the analyses. While there were additional Landsat 8 pixels that could be used in the analysis, a one-to-one

correspondence was maintained between MODIS and Landsat data to allow for a direct data comparison. These data were exported to an Excel file and imported into MATLAB software for statistical analysis and plotting spectral profiles.

3.3.3 *Data Analysis*

Four hierarchies of grassland classes were created to determine what level of grassland type exhibited spectral distinction among classes and to evaluate the impact land management has on the spectral separability of grasslands. Table 1 shows the four levels of grassland classes used in the analysis along with the associated sample sizes for the four levels of classes. Level-1 corresponds to functional grassland types where CRP and native grasslands, dominated by warm-season grasses, were aggregated to a single class while fescue and brome, dominated by cool-season grasses, also were aggregated to a single class. Level-2 separates the grassland types into three classes, CRP, native, and cool-season grasslands. Separating CRP from native was based on knowledge of the potential user-base of the land cover product and growing interest regarding CRP land being converted back to cropland. Level-3 separates fescue and brome grassland types. Lastly, level-4 separates grassland types by land use (Forage, Grazed, and Left Standing). The small sample of fescue left standing lacked sufficient degrees of freedom for meaningful comparisons of Jeffries-Matusita (JM) distance for the three datasets with larger numbers of predictors (namely, multispectral Landsat 8, MODIS NDVI, and the combined datasets).

Spectral plots were created for the pairwise comparisons using the median values and a 70% data band, which is bound by the 15^{th} and 85^{th} percentiles. Pairwise JM distance statistics were calculated for grassland classes in the four hierarchies using the three remotely sensed datasets separately (multispectral Landsat 8 – 18 bands, Landsat 8 NDVI – 3 bands, and MODIS

NDVI – 23 bands) and the combined dataset (44 bands). In addition, JM distance statistics were calculated for individual bands to identify any spectral and/or seasonal influences in individual Landsat bands or image dates. JM distance measures the separability between two classes by considering the distance between class centers simultaneously with intra-class spread (variance) and has shown utility in remote sensing applications (Brown *et al.*, 2013; Davis *et al.*, 1978; Kastens *et al.*, 2017; Lin *et al.*, 2017; Masialeti *et al.*, 2010; Richards & Jia, 1999; Swain & King, 1973; Wardlow *et al.*, 2007). Assuming multivariate normal distributions, the JM distance is calculated as:

$$JM_{ij} = 2(1 - e^{-B}) \text{ where}$$

$$B = \frac{1}{8} D^2 + \frac{1}{2} \ln \left(\left| \frac{\Sigma_j + \Sigma_k}{2} \right| \Big/ \sqrt{\left|\Sigma_j\right| \left|\Sigma_k\right|} \right),$$

$$D^2 = \left(\mu_j - \mu_k\right)^T \left(\frac{\Sigma_j + \Sigma_k}{2} \right)^{-1} \left(\mu_j - \mu_k\right), \text{ and where}$$

μ_j and μ_k correspond to class-specific mean spectral profiles, and Σ_j and Σ_k are unbiased estimates for the class-specific covariance matrices. The JM distance ranges between zero and two. A JM distance approaching two suggests highly distinct spectral distributions, or high separability, between two classes, while a JM distance close to zero suggests highly overlapping spectral distributions, or low separability. The JM distances were calculated for each dataset (multispectral Landsat, Landsat 8 NDVI, and MODIS NDVI) separately and then for all three combined. Additionally, using the level-3 hierarchy, the JM distance statistics were calculated for each band in each dataset to identify the spectral bands or temporal periods exhibiting high separability for the pairwise comparisons.

K-means clustering, a commonly used iterative non-parametric unsupervised classification algorithm, was used to evaluate the spectral clustering of the grassland hierarchies using the best dataset determined by JM distance statistics (Brown *et al.*, 2013). K-means clustering was performed using MATLAB software. Thirty replicates were specified and the number of clusters were based on the grassland hierarchies (e.g. two clusters for Level-2 and ten clusters for Level-4). K-means randomly selects observations as initial cluster centroids, calculates the distance from each observation to the initial centroid, observations to the cluster with the lowest distance and recalculates the cluster centroid. This process iterates until the clusters become completely stable. Of the thirty replicates, the replicate with the lowest sum of total distances was selected for the analysis.

3.4 *Results and Discussion*

3.4.1 *Dataset Comparison of Spectral Separability*

Figure 2 shows pairwise JM distance statistics for all grassland class hierarchies and for the three datasets individually and combined. The x-axis is shown in ascending order of JM distance using the three datasets combined (44 bands). As previously stated, the small sample size for Fescue left standing (Fls) lacked sufficient degrees of freedom for meaningful JM distance calculations besides the Landsat NDVI dataset (JM distance is prone to overfitting in instances where the number of observations from a class is somewhat close to the number of bands being considered in the separability calculations, which imparts a favorable bias to the result; when number of bands exceeds the number of observations, JM distance cannot even be calculated). The results show that JM distance statistics were consistently higher for pairwise comparisons when combining the three datasets (multispectral Landsat, Landsat 8 NDVI and MODIS NDVI) versus individual datasets or two datasets combined (Figure 2).

When comparing the three individual datasets, pairwise JM distance statistics were higher using the multi-temporal, multispectral Landsat dataset versus either of the NDVI datasets. This comparison of JM distance statistics suggests the multispectral data bands provide useful information for separating grassland types versus NDVI alone. The multispectral reflectance data include spectral information from the visible and short-wave infrared (SWIR) bands that can provide additional biophysical information about the vegetation including moisture conditions, biophysical stress, and cell structure. Other studies have shown similar results where multispectral Landsat bands outperformed Landsat TM NDVI in cropland mapping in Kansas (Kennedy, 1999; Mosiman, 2003).

Comparing JM distance statistics between Landsat 8 NDVI and MODIS NDVI suggests that the higher temporal resolution of MODIS NDVI provides more spectral separability of grassland types than the three-date, higher spatial resolution of Landsat 8 NDVI. This result could be influenced by the two qualifiers used for extracting MODIS pixels by seeking to avoid pixels "contaminated" by other land cover types. Including all MODIS pixels would have allowed more mixed pixels to be included in the analysis and likely could have produced a different result than shown here.

3.4.2 *Seasonal Spectral Separability of Grassland Types*

Figure 3 shows the JM distance statistics calculated for each of the eighteen multispectral Landsat bands, the three dates of Landsat 8 NDVI and 23-periods in the MODIS NDVI using the Level-3 hierarchy. As expected, JM distance statistics by band or period were relatively low for all datasets; however, there were specific spectral bands and temporal periods that were more spectrally distinct than others when comparing between grassland types. The 18-band Landsat dataset shows the JM statistic was higher in the fall near-infrared (NIR) band

(Fa5), followed by the spring NIR band (Sp5) and the fall red band (Fa3; Figure 3). Grassland types were more separable in the spring followed by fall for both Landsat 8 NDVI and MODIS NDVI. More specifically for MODIS NDVI, biweekly periods 6-8 (Mar 22–Apr 6, Apr 7–Apr 22, Apr 23–May 8) were higher, followed by periods 20-23 (Nov 1–Nov 16, Nov 17–Dec 2, Dec 3–Dec 18, Dec 19–Dec 31). JM statistics were near zero for the summer periods or image dates. The MODIS temporal statistics of JM distance could be used to target additional image acquisition dates of Landsat imagery. Also, given that summer data provided the least separation among grassland types, increasing the number of Landsat dates in the spring and fall may provide a better mapping approach versus the spring/summer/fall Landsat 8 image triplicate used in this study.

The spectral profiles in Figure 4 illustrate seasonal differences between functional grassland types (warm- and cool-season grasslands) in the study area. In the spring, warm-season grasslands have lower reflectance in the Landsat NIR band (Sp5) and NDVI for Landsat 8 data. The spectral plot of MODIS NDVI shows that the onset of the growing season occurred approximately one period earlier for cool-season than warm-season grasslands. Summer NDVI was slightly lower for warm-season grasslands, but there was significant overlap in the summer NDVI distributions. In the fall, cool-season NDVI remained higher than warm-season, with the typical late-season flush for cool-season grasses. These results support the seasonality of cool- and warm-season grasslands shown by previous research (Foody & Dash, 2007; Guo *et al.*, 2003; Peterson *et al.*, 2002a; Wang *et al.*, 2010).

3.4.3 *Spectral Separability and Grassland Hierarchies*

There were multiple trends in the JM distance statistics across the grassland hierarchies. Generally speaking, pairwise comparisons incorporating land use (level-4 hierarchy) had higher

JM statistics between grassland types than within grassland types (Figure 2). Within grassland type and across the three datasets, there was consistently low separability between forage and grazing land use (e.g. pairwise comparison of Ffg (fescue forage) and Fgz (fescue grazed); Figure 3). Grasslands that were used for forage and grazing in 2015 had low separability, potentially due to the variability in land owner decisions on the timing, frequency, and intensity of use. Market prices and current or prior year climate conditions are among the dynamic variables that factor into land owner decisions. Interestingly, Landsat 8 and MODIS NDVI JM distance was higher between left standing and grazed and left standing and forage for both brome and native grasslands.

Spectral profile statistics (median and 70% data band) of land use within grassland type for native, fescue and brome for the three datasets are shown in Figures 6-8, respectively. The profile distributions show substantial overlap between land use within the grassland types. There were slight seasonal variations in left standing with lower NIR and NDVI in the spring that may result from increased dead vegetation matter remaining from the previous year, and generally higher late summer NDVI that likely result from biomass accumulation over the growing season and lower fall NDVI values during late senescence (periods 19-23) as shown by MODIS profiles.

Since CRP and native grasslands are both dominated by warm-season grasslands, it was anticipated that JM distance would be lower for these comparisons. The statistics, shown in Figure 9, were fairly consistent across the grassland hierarchies and datasets. JM distance was consistently low (less than one) using either the Landsat 8 or MODIS NDVI dataset and was surprisingly consistently high (greater than 1.6) using multispectral Landsat 8 and the combined datasets. The spectral profiles in the top row of Figure 10 show that the distributions of CRP and

native grasses largely overlap across the datasets. For the multispectral Landsat 8 dataset the largest differences occurred in the fall image. Native grasslands had slightly higher spectral reflectance across the six fall bands. By-band statistics show the Landsat fall green (Fa2) and NIR (Fa4) bands were higher than other fall multispectral bands. CRP land is released for grazing or haying under severe drought conditions. The last known time CRP land were released for grazing and haying in the study area was July 2012 (USDA, 2012). Therefore, two-three years of accumulated senesced vegetation were on CRP lands from the 2014 and 2015 growing season that could result in lower NIR spectral reflectance. The higher separability between CRP and native grasslands using Landsat 8 indicates there is utility in the variety of information found in the multispectral bands of Landsat 8 and in using a fall image date. Additionally, given the consistency in the JM distance across CRP and native land use, the results indicate that land use within native grasslands do not largely impact the spectral separability of these two grassland types.

JM distance for CRP and hierarchies of cool-season grassland types are shown in Figure 11. The statistics for multispectral Landsat 8 and the combined datasets were consistently high between CRP and across the hierarchies of cool-season grassland types. CRP land is typically planted using a few dominant warm-season grass species and are largely left unmanaged. Only in severe drought conditions are land owners permitted to graze or hay CRP land. Given the often multi-year accumulation of senesced vegetation residue from grasses in CRP, vegetation structure in CRP is visually distinct and may be a factor in the spectral separability from cool-season grasslands. This result indicates that grazing and forage land use in cool-season grasslands does not suppress the phenological differences between the two functional grassland types and/or that the unique vegetation structure of CRP land influences the separability. As

figure 11 shows, CRP and left standing had relatively lower JM-distance statistics when using Landsat 8 NDVI. The spectral profiles in Figure 10 show that CRP has higher spring and fall red spectral reflectance (Sp4, Fa4) and lower spring and fall NIR spectral reflectance (Sp5, Fa5) and lower Landsat and MODIS NDVI values than brome and fescue, highlighting the asynchronous phenology of warm-season and cool-season grasslands.

Comparisons between the different hierarchies of native and fescue and native and brome are shown in Figures 12 and 13, respectively. There was variability in the statistics with the lowest JM distance between native grazed versus fescue grazed and native grazed versus brome grazed when using the multispectral Landsat 8 or combined dataset. There was more variability in the statistics derived using the NDVI datasets. For Landsat 8 NDVI, values were lowest for the left standing land use. As mentioned previously, left standing may suppress the NDVI values and potentially lessen the spectral separability between these classes. When aggregated to level-3, Native-Fescue (NF) or Native-Brome (NB), level-2 hierarchy Native-Cool, or a level-1 hierarchy, Warm-Cool (NC), the JM values using the combined dataset remained relatively high (>1.75), suggesting that these classes were spectrally separable even under different land use. This again supports the notion that phenological differences remain distinct between the functional grassland types regardless of land use.

Lastly, the results indicate relatively low separability between brome and fescue across the datasets (Figure 14, bottom row). As the spectral profile plots illustrate, the spectral reflectance and temporal response were very similar between brome and fescue which, given that both grassland types are cool-season, was not surprising. These results suggest that when using these datasets, fescue and brome should be grouped into one class for grassland type mapping and that JM distance statistics indicate the grouping does not affect their spectral separability

from CRP or native grassland. It is possible that other remotely sensed data, such as hyperspectral could better separate brome and fescue grassland types.

K-Means clustering was used to examine the potential of the four hierarchies to be reflected in unsupervised spectral clustering. Tables 2 and 3 show the results from K-means along with the percentage of the samples represented in each cluster for two grassland hierarchies. At a Level-1 hierarchy, there was more separation of the class types. Class 1 primarily represented cool-season grasslands and Class 2, warm-season. The classes for potential Level-2 to Level-4 hierarchies were not as well separated. For example, Table 3 shows the confusion in class two for the Level-2 hierarchy. Results from K-means runs with more than three classes (potentially resembling Level-3 and Level-4 class schemes) are not interesting, and thus are not shown.

3.5 *Conclusions*

This study used a data-driven approach to identify the optimal datasets for separating the spectral characteristics of grassland types, determined whether land management practices impact spectral separability of grassland types, and identified what grassland hierarchy should be used for the thematic classification scheme of mapping grassland types. While these findings are limited to the study area and the datasets available, it is anticipated the results will be applicable to similar grassland landscapes in the Great Plains.

The results show that combining the multispectral Landsat 8, Landsat 8 NDVI, and MODIS NDVI datasets resulted in the highest JM distance statistics across all grassland class hierarchies. While the formulations underlying JM distance guarantee improved separability upon addition of more bands, the gains observed with the combined dataset were generally substantial and thus believed to be meaningful. Individually, the three-date multispectral data had

higher JM distance statistics than either NDVI dataset. When including land use, JM distance statistics were lower within grassland types than among grassland types, with the exception of left standing. While JM distance was high for many level-4 pairwise comparisons, it remained relatively high in comparison for functional grassland types (i.e. lower classification levels), indicating land use does not have a highly negative impact on the spectral separability of level-1 and level-2 grassland classes.

The results indicate that brome and fescue were not spectrally distinct and, at least when using inputs like those examined here, should be aggregated as a single class for thematic classification. Meanwhile, CRP and native grasslands had moderately high separability statistics, even though the spectral profiles appeared to largely overlap. These results suggest CRP may be able to be mapped separately or could be aggregated with native grassland into a single warm-season grassland type. However, the three-class K-means clustering did not separate CRP as a separate class. The temporal JM distance statistics indicate the spring and fall were more important for separating cool- and warm-season grasslands than summer when the distributions overlapped more. Future research could leverage the temporal JM distance statistics gained from the 23-period MODIS NDVI time series to increase the density of Landsat imagery during times where spectral separability of functional grassland types is highest.

3.6 *Figures and Tables*

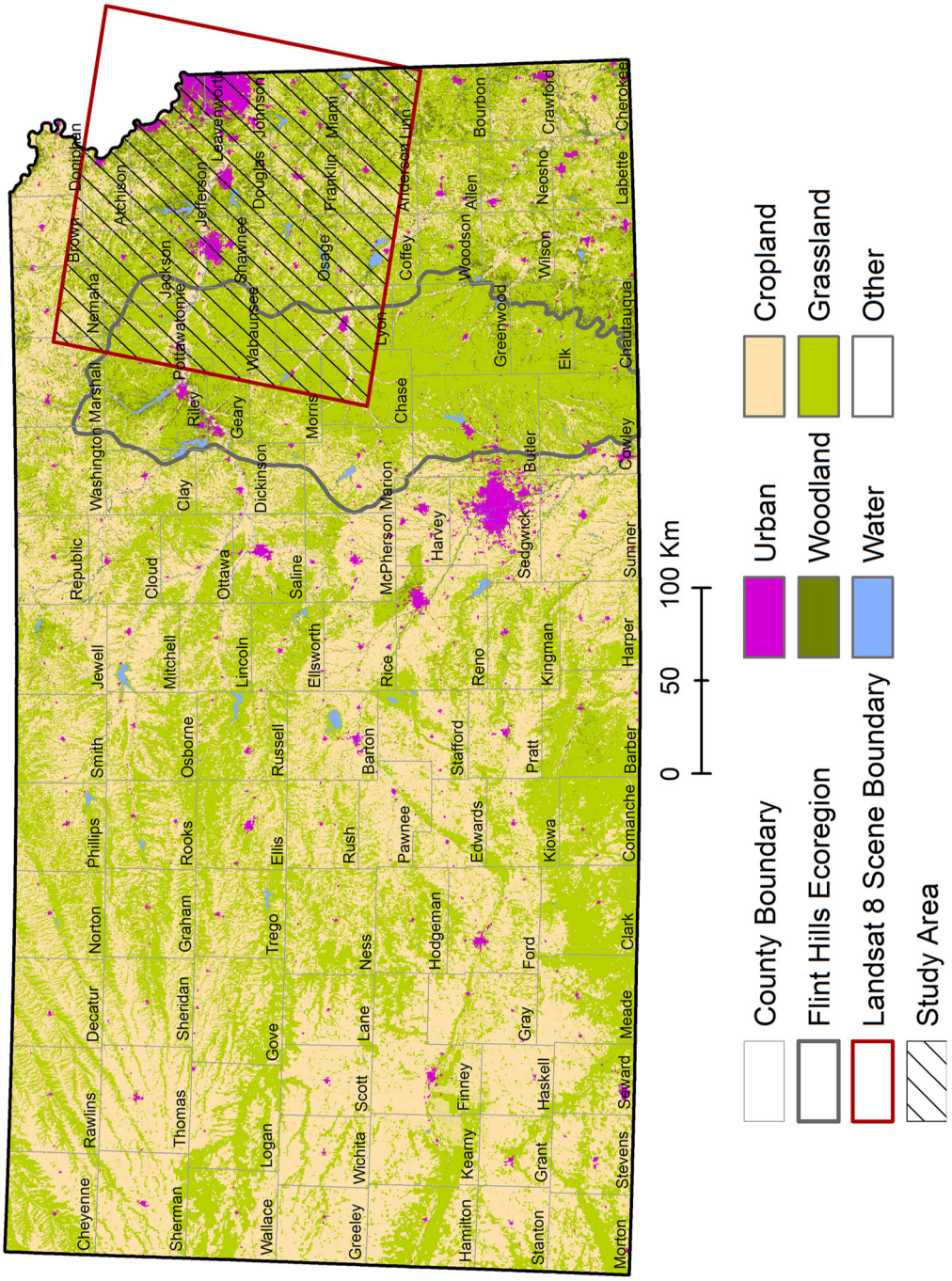


Figure 3.1. The study area with Landsat 8 WRS path/row 27/33 and the Flint Hills ecoregion overlaid. The study area provides a landscape gradient with large tracts of tallgrass prairie in the west and a highly fragmented landscape in the east. The portion of path/row 27/33 were clipped to the Kansas state boundary for the analysis.

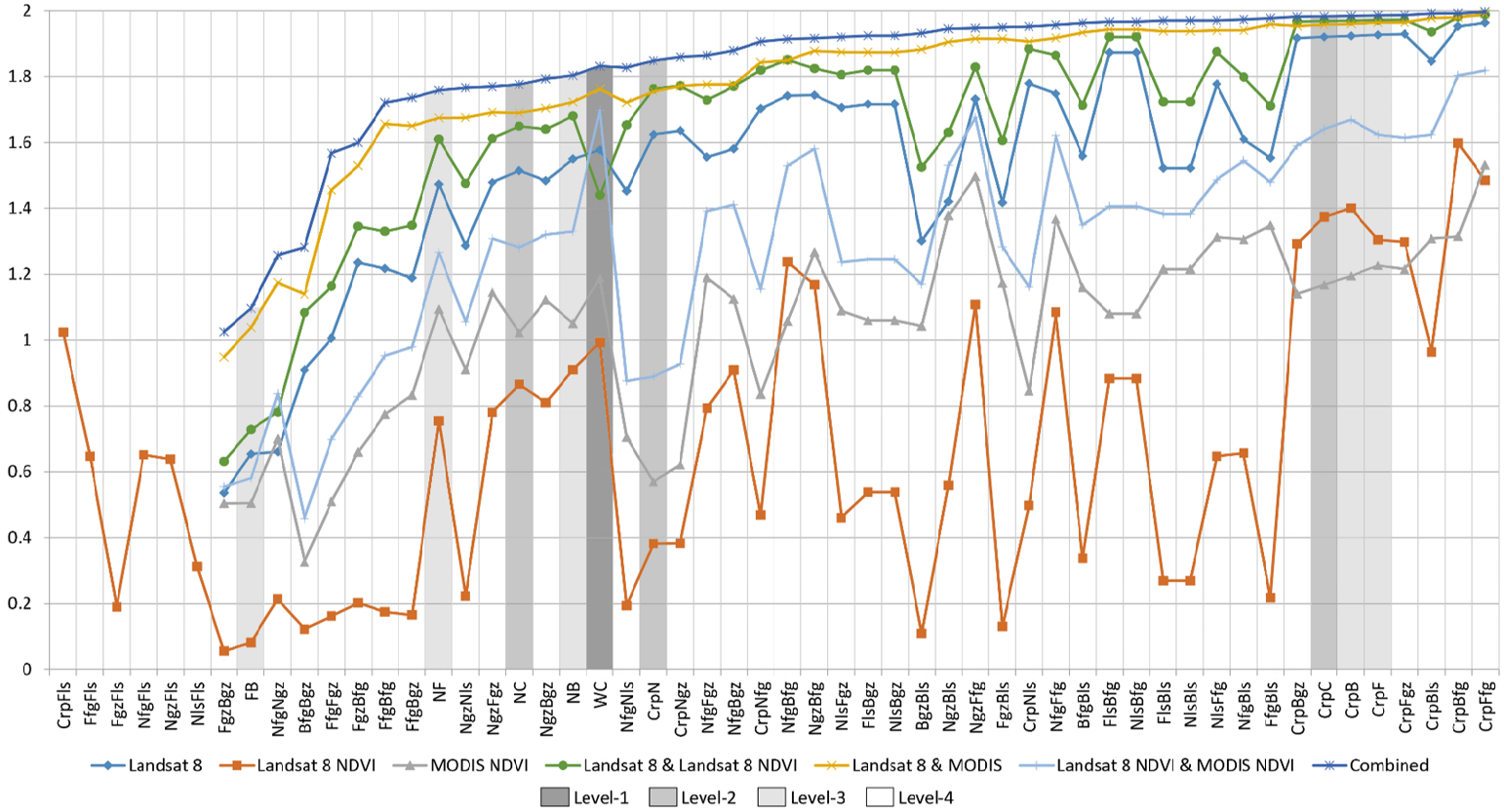


Figure 3.2. Pairwise JM distance statistics for individual datasets, two datasets combined, and all three datasets combined. JM statistics were sorted in ascending order using all three datasets combined, which had the highest values for all pairwise combinations. Individually, the three-date Landsat 8 multispectral dataset had higher JM distance statistics. While JM distances were high for some level-four grassland type comparisons, level 1-2 hierarchies were not greatly impacted by land use.

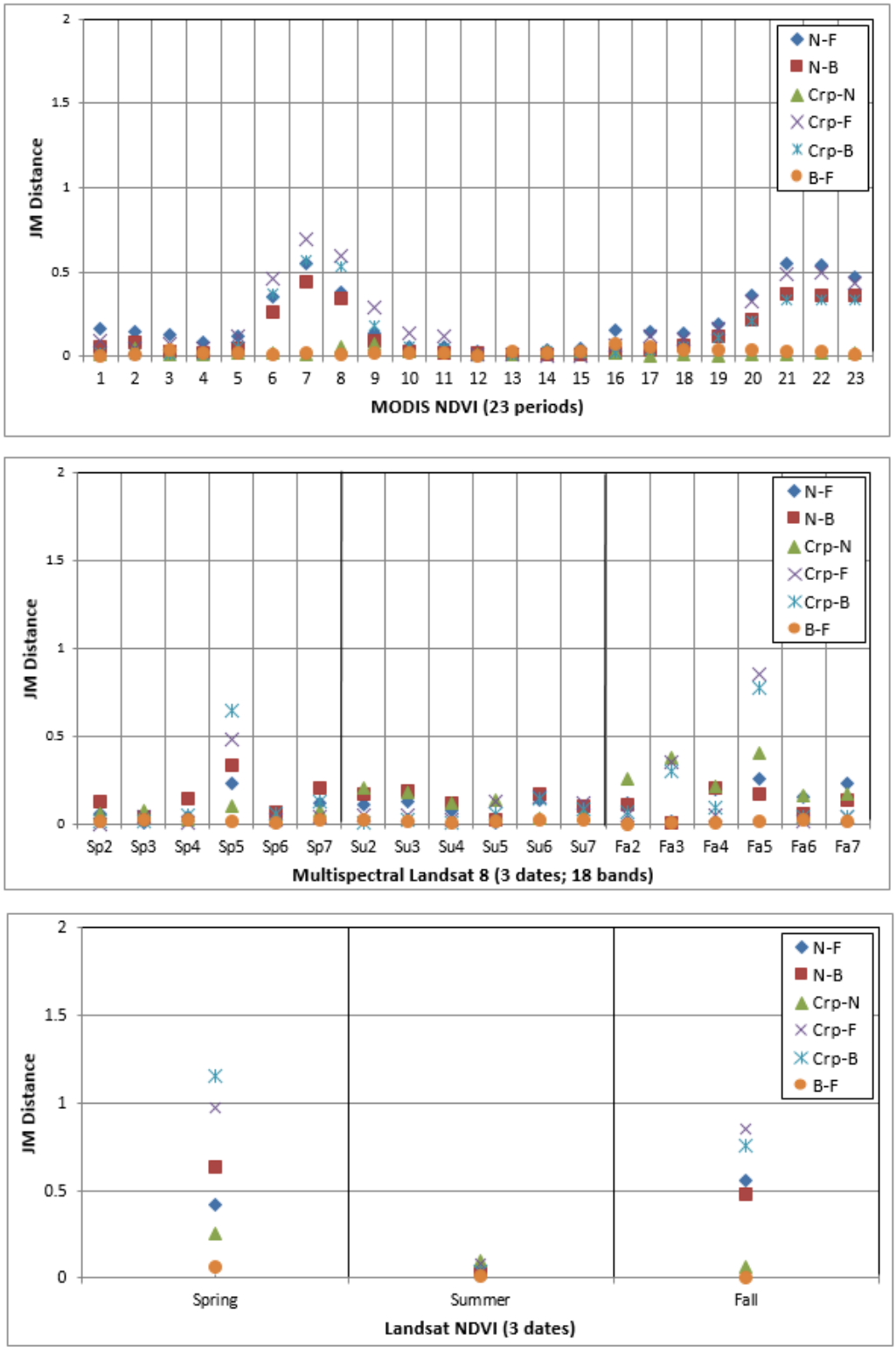


Figure 3.3. JM distance statistics for level three grassland classes using individual bands from the three-date multispectral Landsat 8 bands (top); individual dates of Landsat 8 NDVI (middle), and individual periods from MODIS NDVI 16-day composites (bottom). Spring and fall provided more separability than summer.

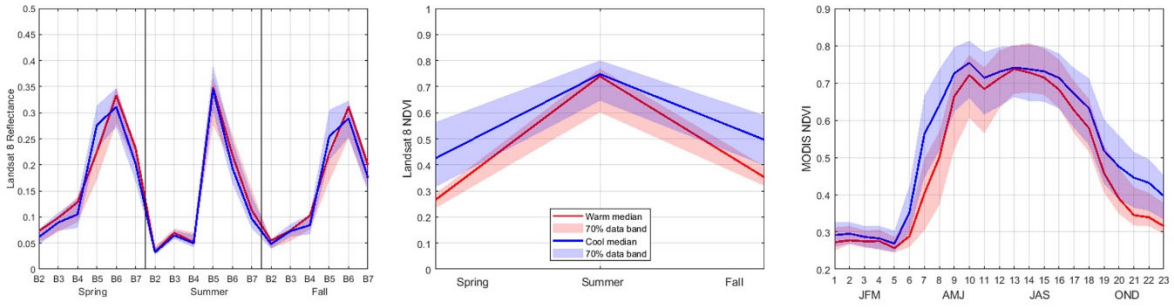


Figure 3.4. Spectral profiles for warm- and cool-season grassland types in the study area using three-date multispectral Landsat 8 bands (left), three-date Landsat 8 NDVI (center), and 23-period MODIS NDVI time series from 2015. Spectral differences were highest in the spring Landsat 8 NIR band (B5) and spring and both fall NDVI's.

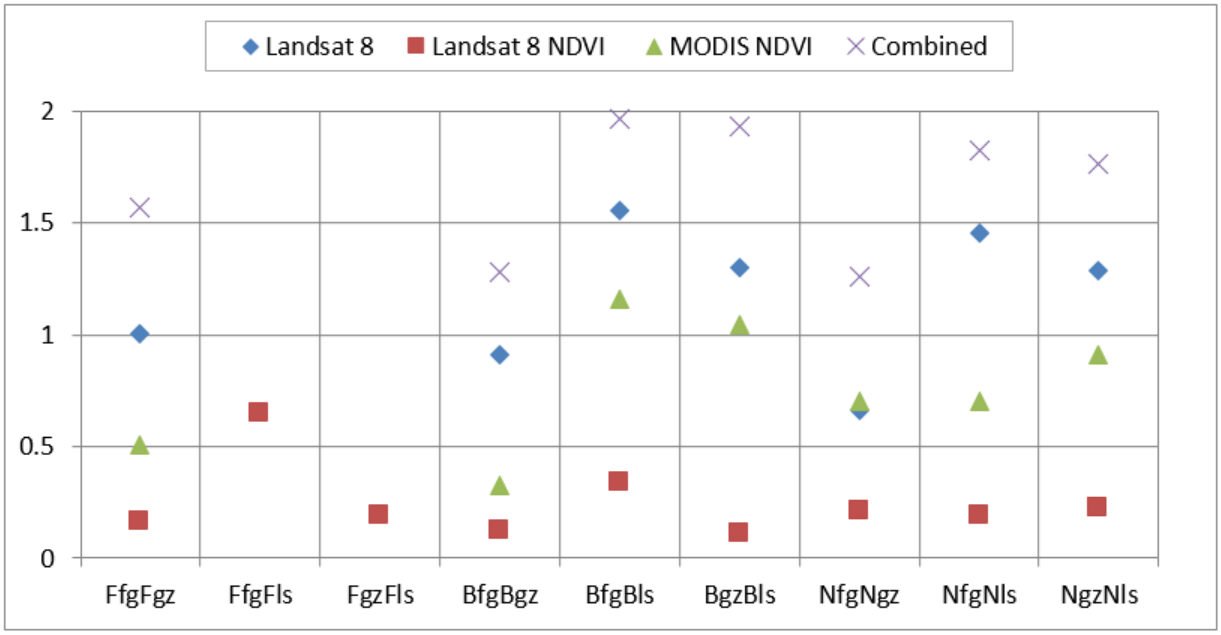


Figure 3.5. The JM distance statistics within the three grassland types: fescue (F), brome (B) and native (N) between three land use: forage (fg), grazed (gz) and left standing (ls). Pairwise comparisons containing left standing consistently had higher JM distance statistics while comparisons of forage and grazing had consistently lower JM distance statistics. Class Fls (fescue left standing) was excluded from three dataset comparisons due to the small sample size and resulting in inadequate degrees of freedom.

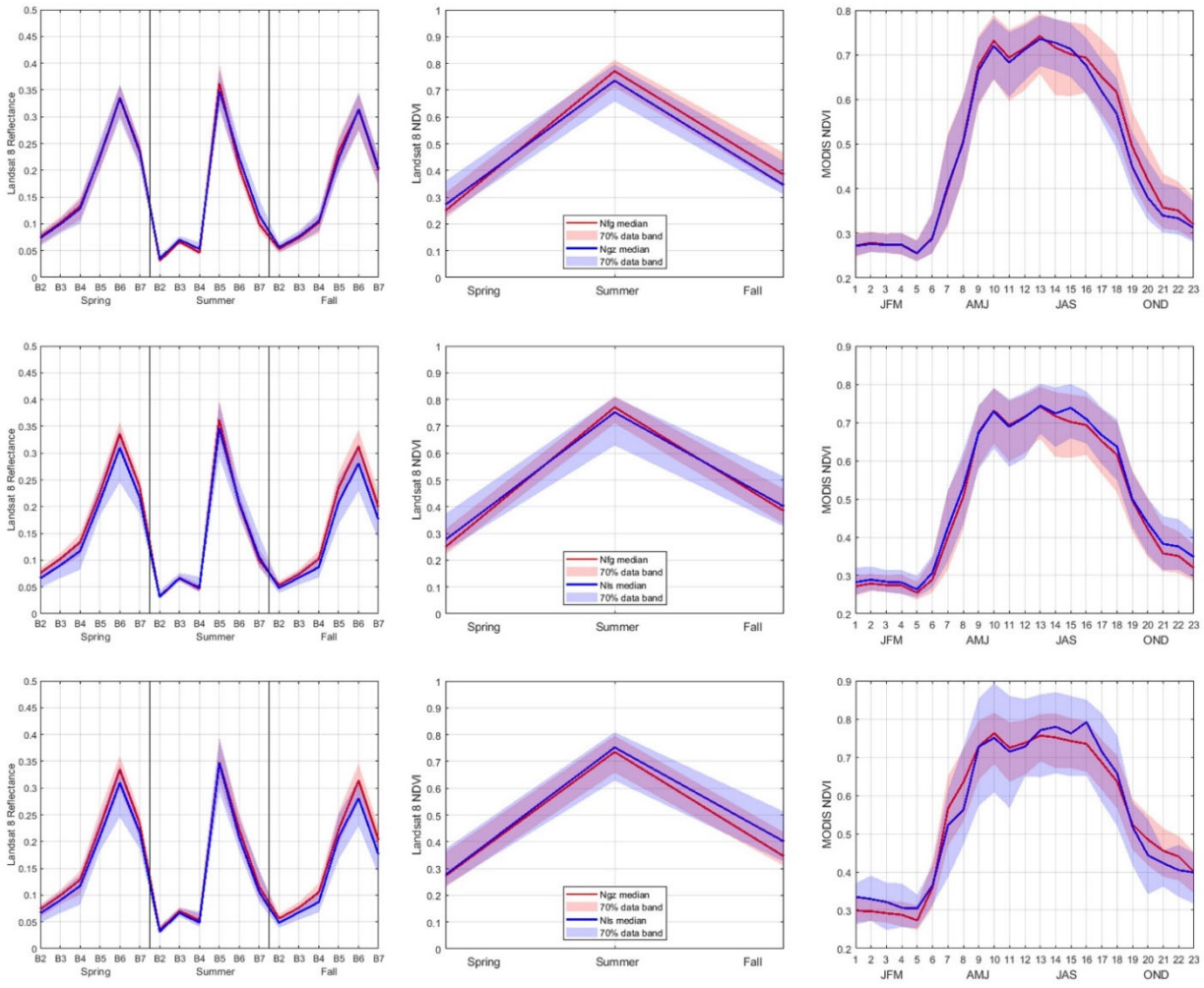


Figure 3.6. Spectral profiles comparing, pairwise, three land use categories within native grasslands using three-date multispectral Landsat 8 bands (left), three-date Landsat 8 NDVI (center), and 23-period MODIS NDVI time series from 2015. Nfg = Native forage; Ng z = Native grazed; Nls = Native left standing.

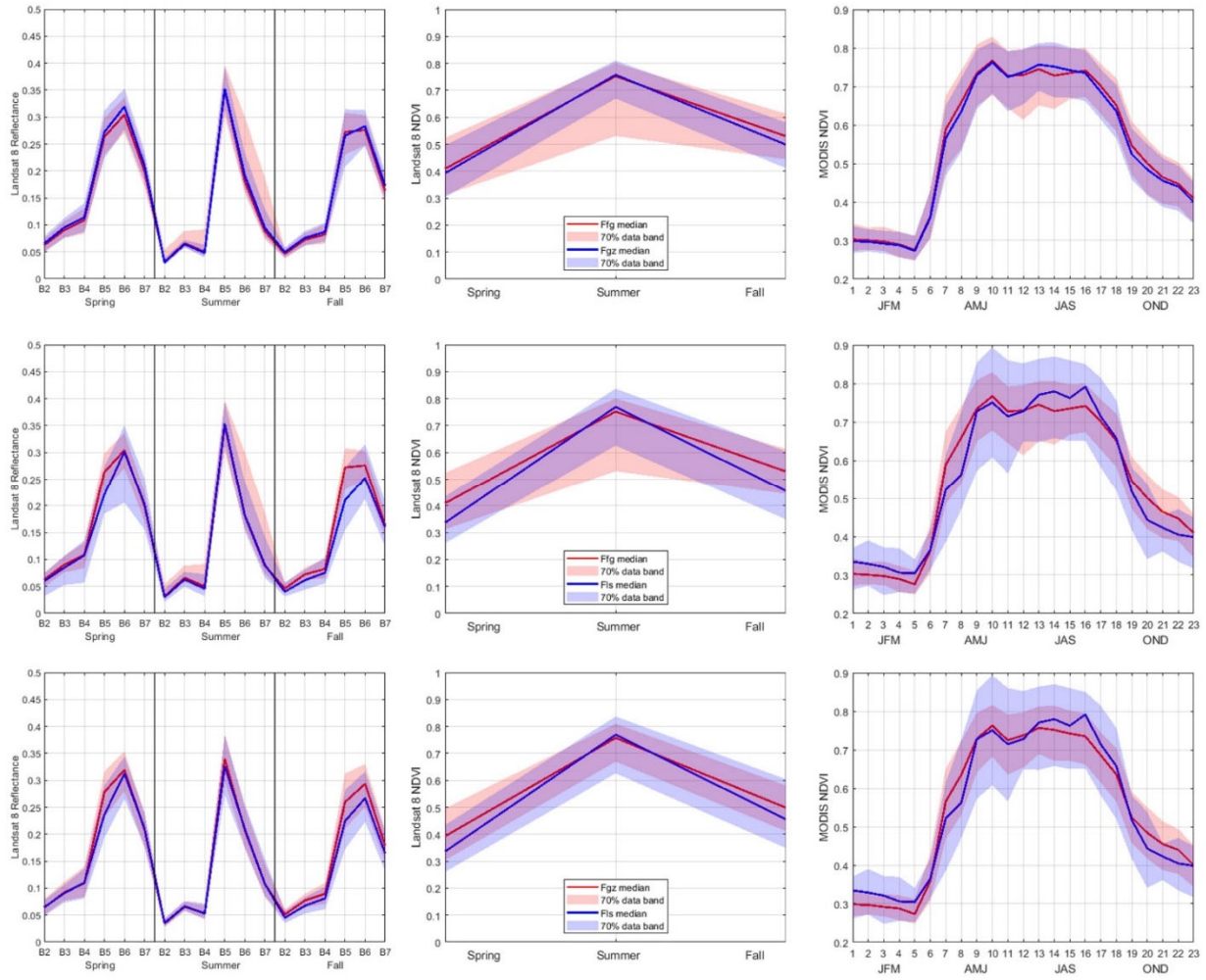


Figure 3.7. Spectral profiles comparing, pairwise, three land use categories within fescue grasslands using three-date multispectral Landsat 8 bands (left), three-date Landsat 8 NDVI (center), and 23-period MODIS NDVI time series from 2015. Ffg = Fescue forage; Fg z = Fescue grazed; Fls = Fescue left standing.

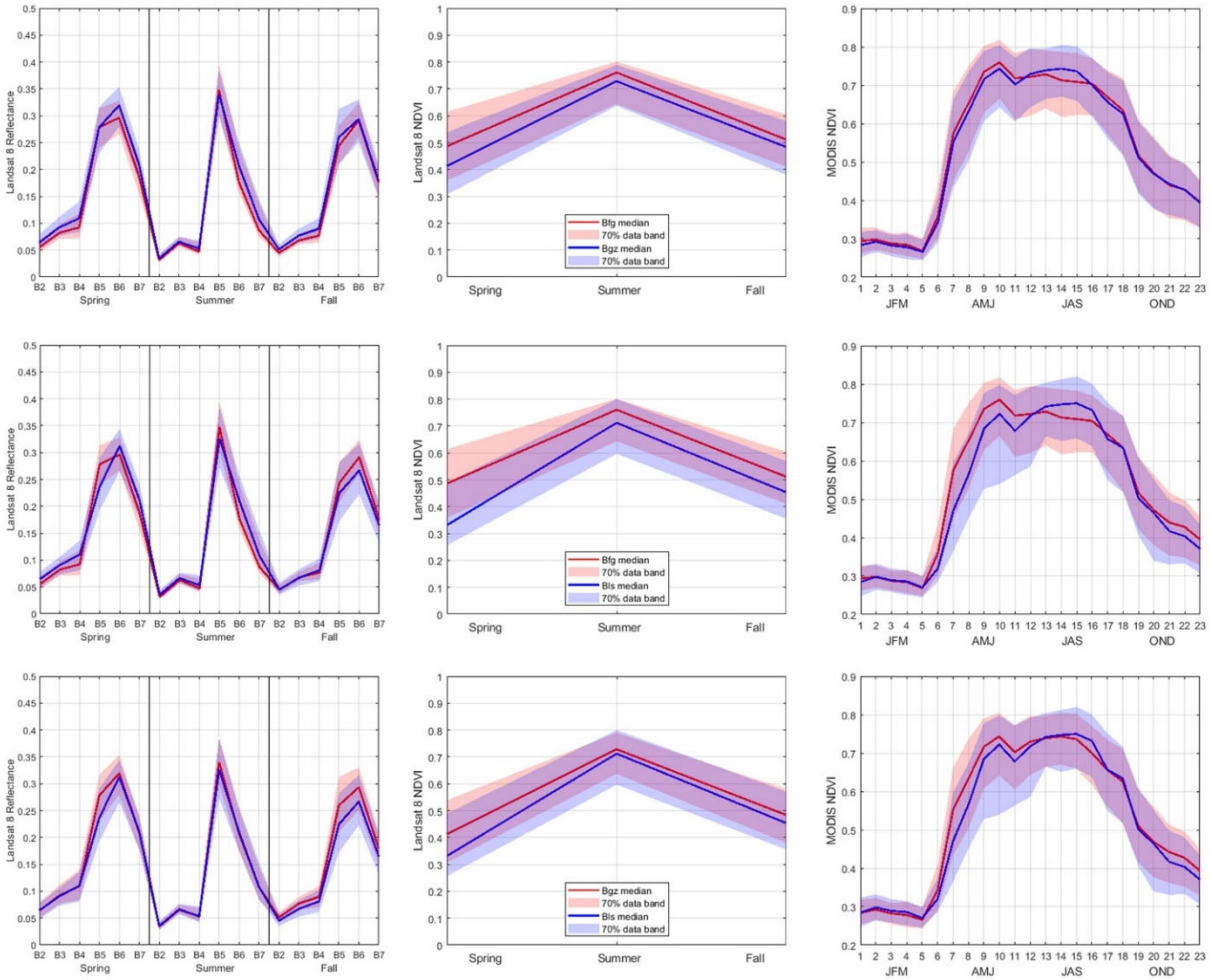


Figure 3.8. Spectral profiles comparing three land use within brome grasslands using three-date multispectral Landsat 8 bands (left), three-date Landsat 8 NDVI (center), and 23-period MODIS NDVI time series from 2015. Bfg = Brome forage; Bgz = Brome grazed; Bls = Brome left standing.

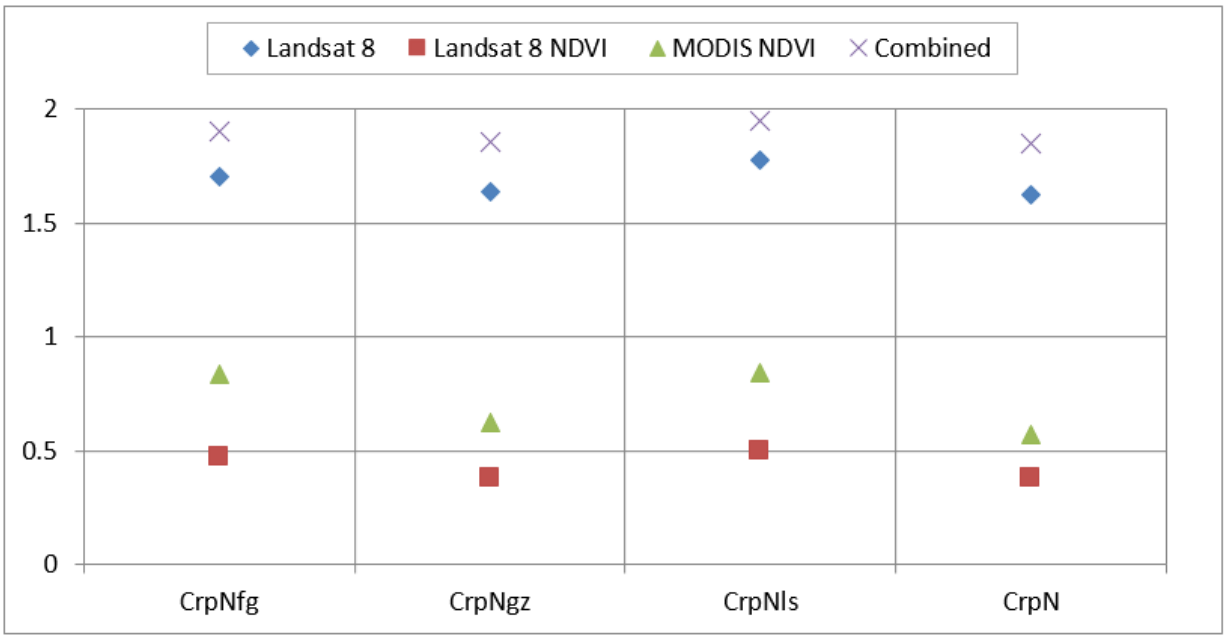


Figure 3.9. The pairwise JM distance statistics between CRP (Crp) and native (N) and between CRP and the three land use in native grasslands separately (N). JM distance statistics were fairly consistent across the three land use and when land use was aggregated as shown far right.

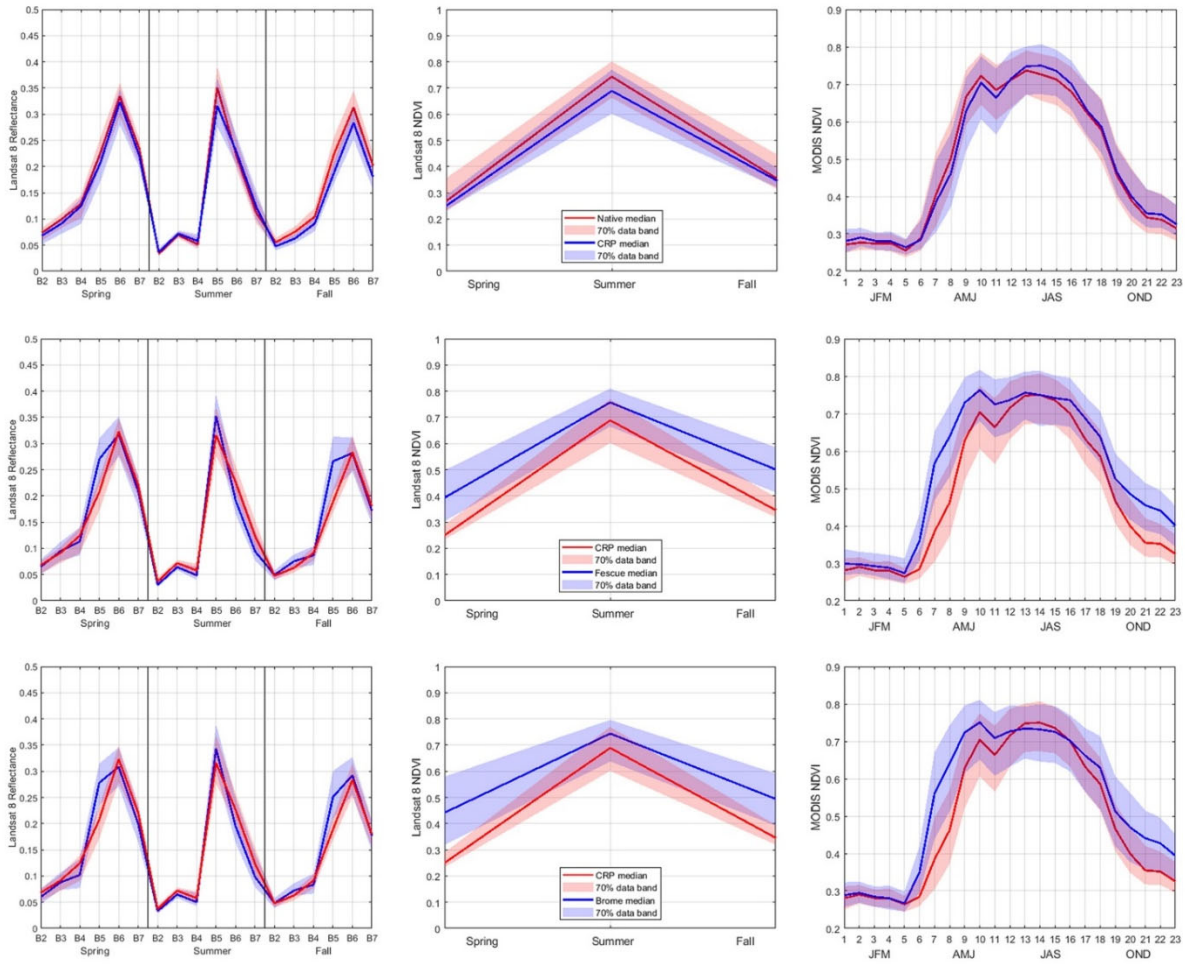


Figure 3.10. Median spectral profiles and the 70% data band comparing level three grassland classes, CRP and native (top), CRP and fescue (middle), and CRP and brome (bottom). Three-date multispectral Landsat 8 is in the first column, three-date Landsat 8 NDVI in the second, and 23 16-day composites of MODIS NDVI in the third column. Visually the distributions between CRP and brome and CRP and fescue were more separable than CRP and native. Spring and fall NIR bands (Band 5) and spring and fall NDVI showed the least overlap in the spectral and temporal distributions with cool-season, fescue, and brome having higher reflectance in the NIR bands and higher NDVI than warm-season native and CRP.

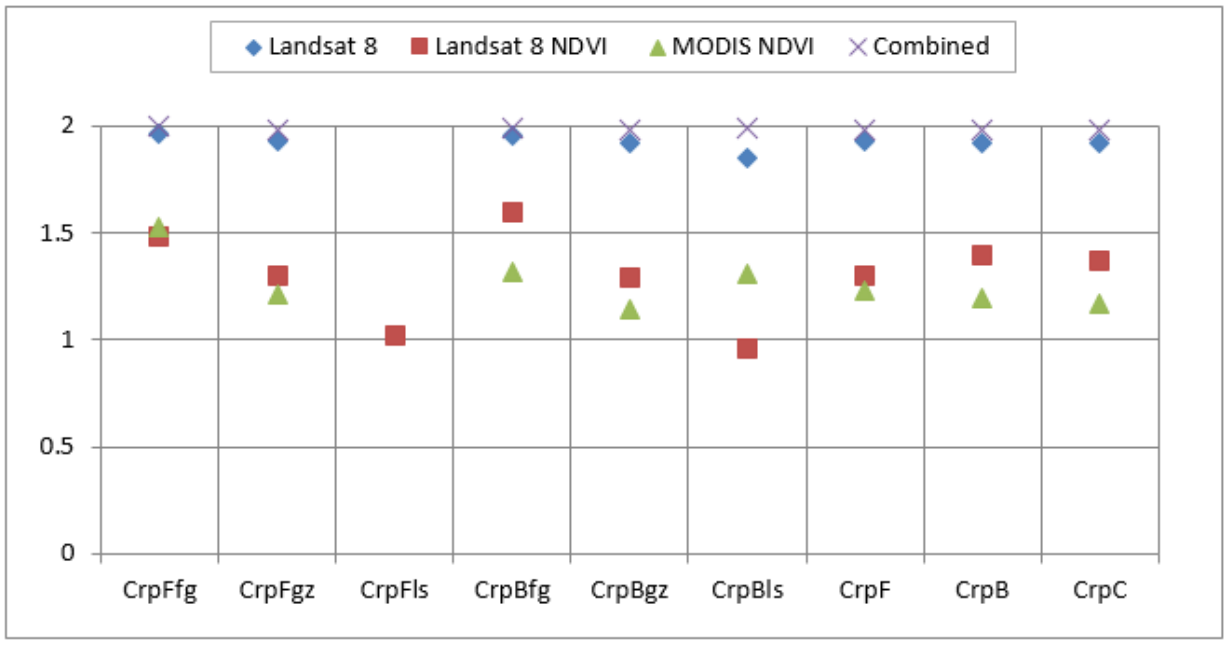


Figure 3.11. Pairwise JM distance statistics between CRP (Crp) and different land management levels of fescue (F) brome (B). Ffg = three land use separately. JM distance statistics fluctuated using NDVI datasets, and JM distances were maintained when fescue and brome were aggregated to level-two and level-one hierarchies. Ffg = Fescue forage; Fgz = Fescue grazed; Fls = Fescue left standing; Bfg = Brome forage; Bgz = Brome grazed; Bls = Brome left standing; C = cool-season (fescue and brome combined).

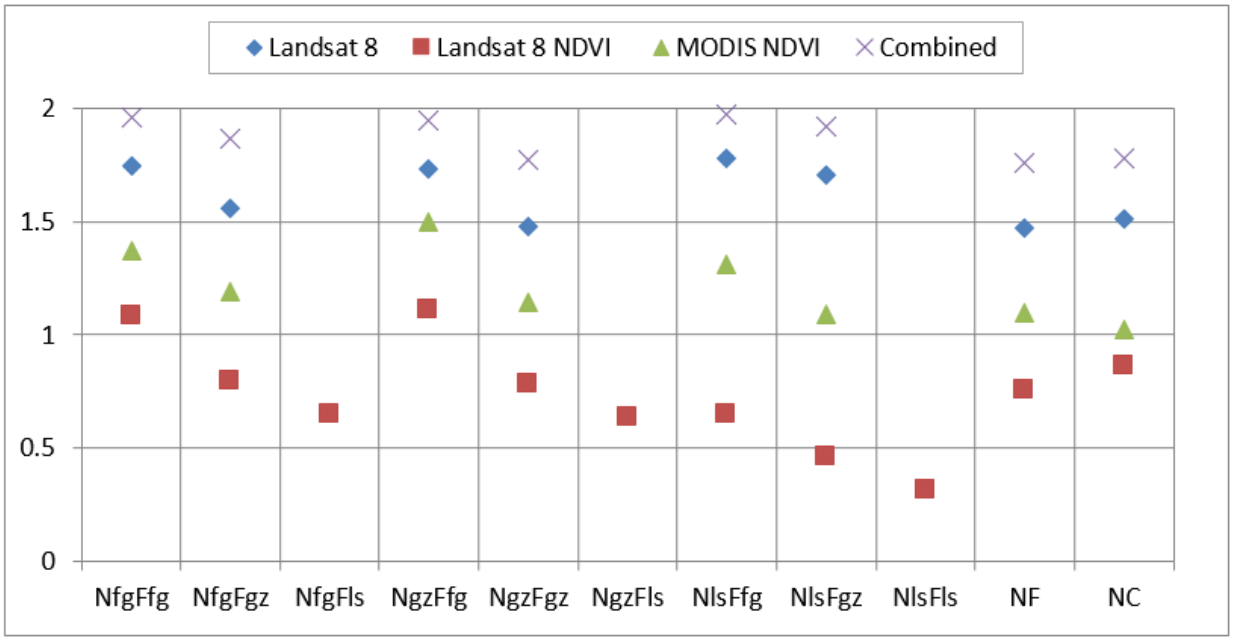


Figure 3.12. Pairwise JM distance statistics between native (N) and fescue (F) using three grassland type hierarchies. JM distance statistics fluctuated more when using NDVI data. For Landsat 8 and the combined dataset, JM distance remained relatively high when native and fescue were aggregated to level-two and level-one hierarchies. Ffg = Fescue forage; Fgz = Fescue grazed; Fls = Fescue left standing; C = cool-season (fescue and brome combined).

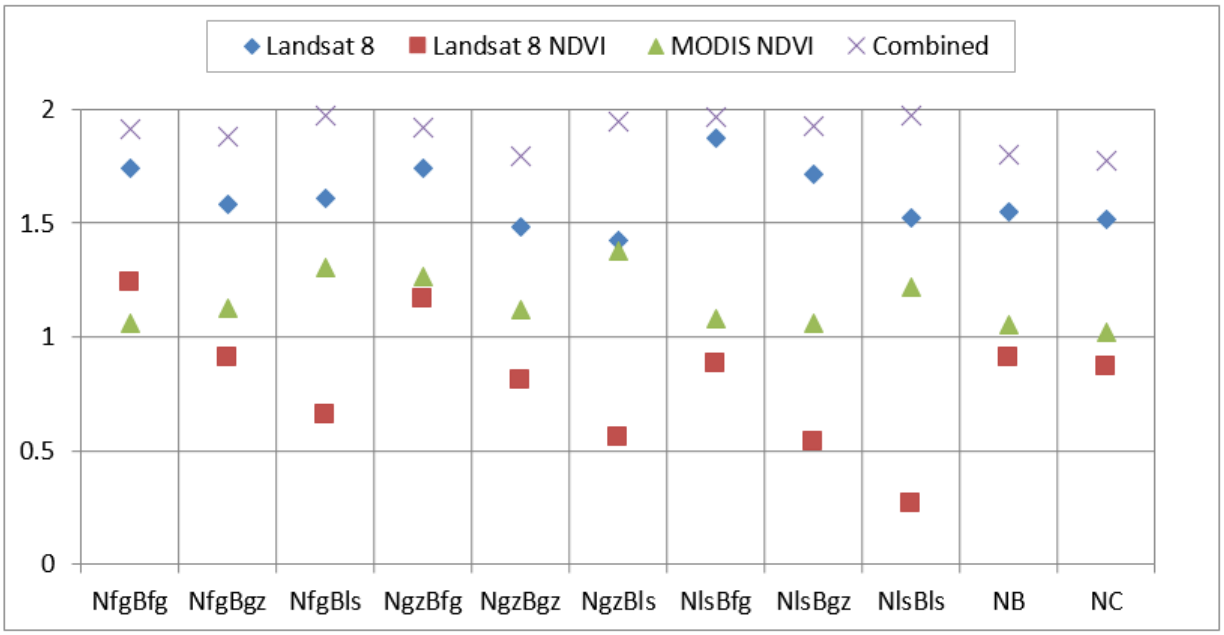


Figure 3.13. Pairwise JM distance statistics between native (N) and brome (B) using three grassland type hierarchies. JM distance statistics fluctuated more when using NDVI data. For Landsat 8 and the combined dataset, JM distance remained relatively high when native and brome were aggregated to level-two and level-one hierarchies. Bfg = Brome forage; Bgz = Brome grazed; Bls = Brome left standing; C = cool-season (fescue and brome combined).

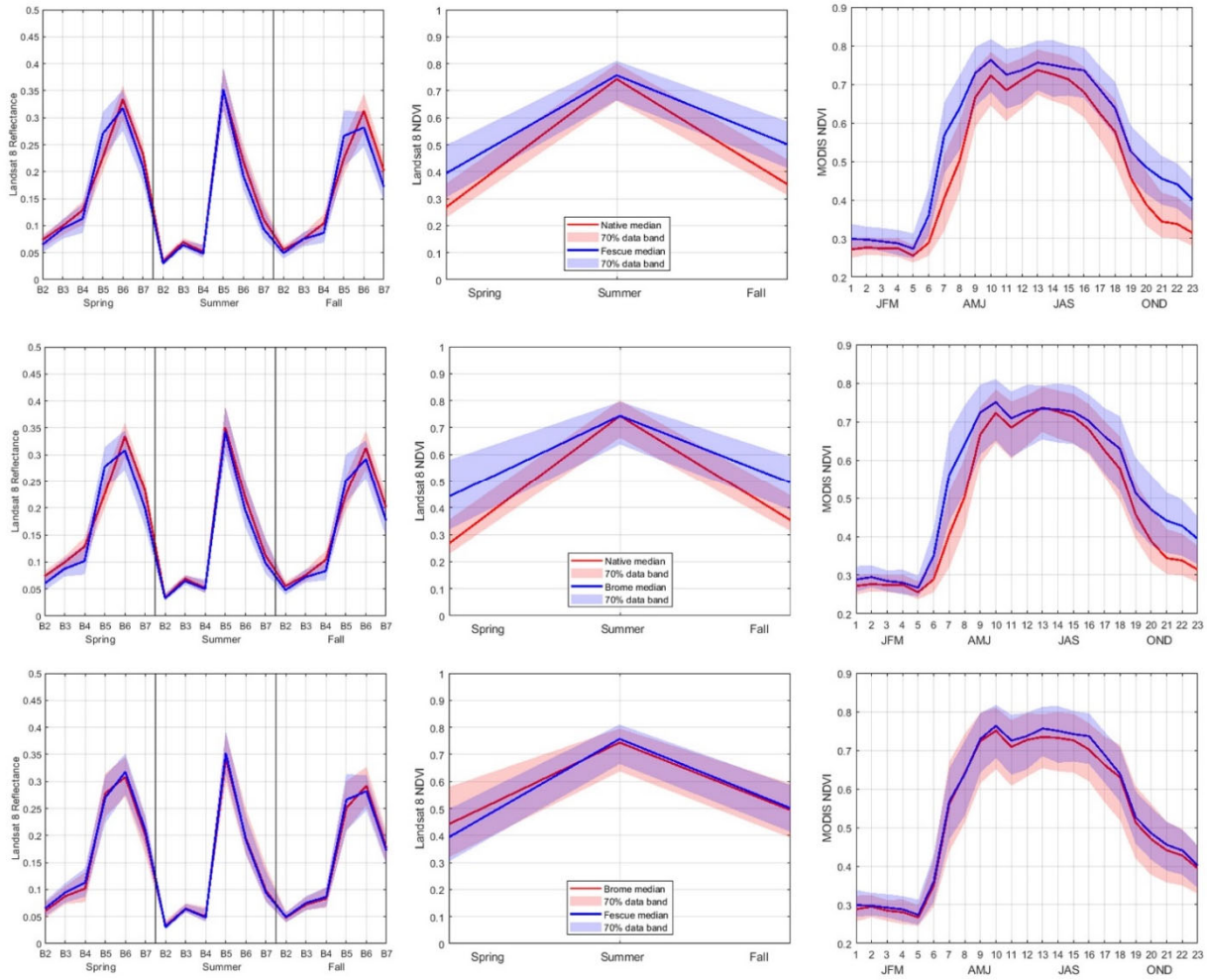


Figure 3.14. Median spectral profiles and the 70% data band comparing level three grassland classes, native and fescue (top), native and brome (middle), and fescue and brome (bottom). Three-date multispectral Landsat 8 is in the first column, three-date Landsat 8 NDVI in the second, and 23-date, 16-day composites of MODIS NDVI in the third column. Visually the distributions between native and brome and native and fescue were more separable than brome and fescue, which illustrates differences in phenology characteristics between functional grasslands types. Spring and fall NIR and SWIR Landsat bands (Bands 5 and 6) and spring and fall NDVI showed the least overlap in the spectral and temporal distributions between native and the two cool-season grasses, fescue and brome. There was little distinction between the distributions for brome and fescue spectral plots.

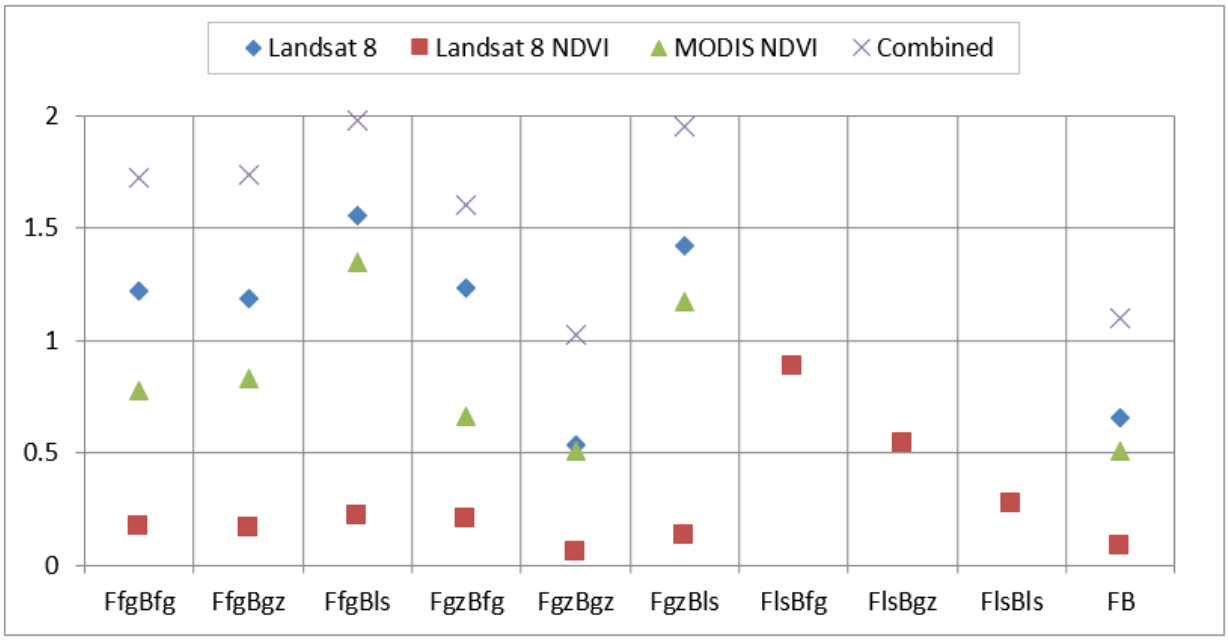


Figure 3.15. Pairwise JM distance statistics between fescue (F) and brome (B) using two grassland type hierarchies. JM distance statistics were relatively low and fluctuated compared to other grassland type comparisons. There were several moderate to high JM values at the level-four hierarchy. When aggregated, the JM distance dropped to 1.1, indicating that land use has a larger impact on separating these two cool-season grassland types.

Table 3.1. The four levels of grassland type hierarchies used to test spectral separability using JM distance for three remotely sensed datasets. Level one separates functional grassland types while level four separates land use within a grassland type.

Level-1	**Level-2**	**Level-3**	**Level-4**	**Sample Size**
Warm-Season (W)	CRP (Crp)	CRP (Crp)	CRP (Crp)	746
	Native (N)	Native (N)	Native Forage (Nfg)	1,414
			Native Grazed (Ngz)	6,739
			Native Left Standing (Nls)	301
Cool-Season (C)	Cool-Season (C)	Fescue (F)	Fescue Forage (Ffg)	265
			Fescue Grazed (Fgz)	2,349
			Fescue Left Standing (Fls)**	24
		Brome (B)	Brome Forage (Bfg)	2,948
			Brome Grazed (Bgz)	3,805
			Brome Left Standing (Bls)	103
Total Sample Size				*18,694*

Table 3.2. K-means clustering using two classes for the Level-1 grassland type hierarchy. Class one was dominated by cool-season grasslands while class two was dominated by warm-season grasslands.

Level-1 Hierarchy	**Class 1**	**Class 2**
Warm-season Grassland	730 (16%)	4,134 (79%)
Cool-season Grassland	3,818 (84%)	1,119 (21%)

Table 3.3. K-means clustering using three classes for the Level-2 grassland type hierarchy. Class one was dominated by cool-season grasslands, class two was a mix of cool- and warm-season and class three was dominated by warm-season grasslands.

Level-2 Hierarchy	**Class 1**	**Class 2**	**Class 3**
CRP	3 (<1%)	67 (2%)	145 (4%)
Warm-season Grassland	241 (8%)	1,487 (46%)	2,921 (82%)
Cool-season Grassland	2,766 (92%)	1,684 (52%)	487 (14%)

4 Chapter 4: Random Forest and Landsat 8 and MODIS NDVI Data for Mapping Grassland Types at a Regional Scale

4.1 *Abstract*

This study evaluates random forest (RF) models using four classification hierarchies of grasslands and four datasets in two adjacent Landsat scenes (path/rows). The data-driven results will be used to inform classification methodology (data and thematic classification scheme) for regional-scale mapping of grasslands in Kansas. Random forest models were built using multi-spectral Landsat 8, Landsat 8 NDVI, and MODIS NDVI time series, both separately and combined. Training and test data samples were obtained from the Farm Service Agency (FSA) 578 data. For the Level-1 grassland hierarchy that maps the two functional grassland types, cool- and warm-season grassland, out-of-bag (OOB) error (12.5%) indicated that using the combined dataset was the optimal classification scheme. The Level-2 hierarchy that separates land enrolled in the Conservation Reserve Program (CRP) as a thematic class had overall OOB error estimates ranging from 14-18%; CRP had low producer's accuracy levels and was largely mapped as warm-season grasslands. Path/rows 27/33 and 28/33 had OOB overall accuracy levels of 87% and 92%, respectively. User's and producer's accuracy levels indicate that cool-season grasslands were mapped more accurately in path/row 27/33 where that class is more dominant than in 28/33. Using test data (withheld verification data) unexpectedly increased overall accuracy levels by 4% and 6% over OOB accuracies, which may have resulted from varying data proportions between OOB and test data and warrants a more detailed evaluation of the RF structure.

4.2 *Introduction*

Grasslands cover 40.5% of the earth's surface, more than either forest or cropland (Gibson, 2009). While expansive, grasslands are considered the most threatened biome due to land conversion and intensive land use (Samson *et al.*, 2004). Globally, the conversion of grassland to cropland represents the leading cause of landscape fragmentation and lost grassland extent (Gibson, 2009). In addition, the quality of remaining native grassland has been modified or degraded by invading non-native species, fire suppression, and overgrazing by domestic livestock (Weaver, 1954; Gibson, 2009; Risser, 1988).

The tallgrass prairies of the Great Plains in North America, considered one of the more biologically diverse grasslands in the world (Risser, 1988), have the greatest reported loss of grassland area with estimates of only 9.4% - 13% of the original estimated 167 million acres remaining (Gibson, 2009; Samson *et al.*, 2004). The majority of the original tallgrass prairie was converted to cropland and non-native forage for livestock during European settlement. While many states have less than a half percent of the original extent of tallgrass prairie, Kansas has the largest percent of any state and the largest contiguous tract of remnant tallgrass prairie, known as the Flint Hills (Risser, 1988; Robertson & Schwartz, 1994b). The Flint Hills tallgrass prairie has persisted due to the shallow rocky substrate that prevented conversion to cropland. Today, the majority of tallgrass prairie remnants are privately owned and are managed using a variety of practices to maximize vegetation productivity for grazing and forage for livestock.

In addition to tallgrass prairie, land enrolled in the Conservation Reserve Program (CRP) is a grassland type that is of particular interest. CRP is a United States Department of Agriculture (USDA) Farm Service Agency (FSA) program that began in 1985 and is the largest private-lands conservation program in the U.S. Typically 10-15 year contracts are offered to landowners where

marginal agricultural land is taken out of crop production and planted with grass or tree cover in an effort to provide enhanced wildlife habitat, improved water quality and reduced soil erosion (Ribaudo *et al.*, 1990; Wu & Weber, 2012). In 2007 Kansas CRP acreage peaked at 3 million acres, while in 2017, Kansas only had a reported 1.1 million acres. Due to recent national enrollment limits of 27 million acres set by the 2018 Farm Bill, crop commodity prices and ongoing interests in biofuel production, research has shown that CRP land frequently has been converted back into cropland production (Hendricks & Er, 2018; Johnston, 2014; Wright & Wimberly, 2013). There are ongoing concerns about further CRP land conversion to cropland and the loss of environmental services CRP land provides.

Accurate and ongoing land use/land cover mapping provides tools to monitor the changing landscape, including environmental and socio-economic drivers, and provides the opportunity for conservation planning. Remotely sensed data have been used for decades to map and monitor grasslands, including tallgrass prairie and land enrolled in CRP. Studies have used remote sensing technology to monitor and model biophysical characteristics of grasslands including the distributions and abundance of functional grasslands (i.e. C3 and C4) (Davidson & Csillag, 2003; Foody & Dash, 2007; Peterson *et al.*, 2002b; Tieszen *et al.*, 1997), grassland productivity (biomass and cover) (Gu & Wylie, 2015; Porter *et al.*, 2014; Zha *et al.*, 2003) and grassland use and condition that can alter grassland biophysical characteristics and quality. Many of these studies and mapping efforts rely on the asynchronous phenology of cool- and warm-season grasslands. However, grasslands are used and managed extensively and intensively, largely to support livestock production. The type, combination, timing, and intensity of land management practices within grassland types alter the biophysical properties of grasslands, including vegetation productivity and composition and soil structure and chemistry,

which in turn potentially alter spectral responses that can complicate the ability to accurately map grassland types. Several studies have used remotely sensed data to characterize and monitor land management practices and land use intensity occurring within grasslands (Franke *et al.*, 2012; Guo *et al.*, 2003; Guo *et al.*, 2000; Halabuk *et al.*, 2015; Lauver & Whistler, 1993; Peterson *et al.*, 2002c; Price *et al.*, 2002b). And with regard to CRP land, several approaches have been used identify CRP land including a post-classification trajectory approach (Egbert *et al.*, 1998; Maxwell & Sylvester, 2012; Song *et al.*, 2005).

While previous studies have evaluated the biophysical characteristics of grasslands and have used field and satellite-acquired spectral data to statistically discriminate between grassland types and land management practices, little research has focused on identifying an optimal thematic classification approach for mapping grassland types at a regional scale. Multiple factors must be considered when developing such a land cover classification approach; one key factor is determining what source data or combination thereof maximizes the ability to map the defined grassland types. Another is defining what grassland types can be mapped, meaning the thematic classification scheme. Many times the thematic classification scheme is developed *a priori*, without knowing if the classes are spectrally distinct for accurate mapping results. The objective of this study is to implement a data-driven approach using Landsat 8, Landsat 8 NDVI and MODIS NDVI time series data from 2015 to determine both the optimal source imagery and the optimal thematic classification scheme for mapping grasslands in northeastern Kansas. A comparison of the spectral and temporal resolutions of Landsat 8 multispectral data, Landsat 8 NDVI, and MODIS NDVI provides a framework for identifying relative strengths and weaknesses of these datasets for grassland classification. In addition, using a hierarchy of

grassland types enables identification of an appropriate thematic classification scheme and evaluation of the effect of land management on spectral characteristics.

4.3 *Study Area and Data Sources*

4.3.1 *Study Area*

There is an inherent east-west land use/land cover gradient within the study area (Figure 1). The study area encompasses the Flint Hills, which, as previously mentioned, is the largest remaining tract of native tallgrass prairie in the world. Native grasslands dominate the Flint Hills with some non-native grasslands and croplands scattered in the river lowlands. Moving eastward from the Flint Hills the landscape is more fragmented where cropland becomes prevalent and grasslands consist of both native, warm-season grasslands, and non-native, cool-season grasslands.

Warm-season grasslands are either native tallgrass prairie or have been reseeded using a native seed mixture. Warm-season grasslands fix carbon using C4 photosynthesis and are dominated by native bunchgrasses such as big bluestem (*Andropogon gerardii*), little bluestem (*Schizachyrium scoparium*), and indiangrass (*Sorghastrum nutans*) and native forbs such as leadplant (*Amorpha canescens*), butterfly weed (*Asclepias tuberosa*), and purple coneflower (*Echinacea angustifolia*). The typical phenology of warm-season grasslands is spring green-up, peak productivity in late spring to early summer when temperatures increase, followed by senescence in fall (Weaver, 1954).

Cool-season grasslands in the study area are defined as non-native grasslands that are predominately planted with either smooth brome (*Bromus inermis*) or tall fescue (*Festuca arundinacea*). Cool-season grasslands fix carbon using C3 photosynthesis and have a typical phenology of early spring green-up, peak productivity in late spring, a mid-summer semi-

dormancy and, with sufficient precipitation, a second, smaller growth period in early fall (Weaver, 1954). Haying for forage and grazing are two of the common land uses for both grassland types. However, given that most of the land is privately owned, the timing, intensity, and frequency of management practices within each grassland type vary by land owner and by economic and climate conditions in a given year. In addition to grazing, prescribed burning is a commonly used management practice to maintain species diversity in native grasslands as well as prevent woody encroachment.

4.3.2 *Data Sources*

The USDA-FSA maintains annual field-level records of acreage, land cover, and intended land use for all fields participating in USDA programs, referred to as FSA 578 data. In Kansas, county-level field offices maintain FSA 578 data, where land owners or producers report land cover and land use information for eligibility for the upcoming USDA program year. Historically these data were maintained by county field offices using photocopies of aerial photos with land cover and land use information annotated on the hardcopy. Today these data are maintained as a geodatabase of field boundaries known as Common Land Units (CLUs), which typically represent the smallest land unit with the same ownership, land cover, and land use. (Some fields have been observed to be subdivided into smaller sub-CLU units, however.) Each CLU is attributed with information including crop type, land use, reported acreage, county FIPS code, farm number, and tract number. In the past, these data have been made available to scientists to use for training and validating several land cover mapping efforts in Kansas (Kennedy, 1999; Mosiman, 2003; Peterson *et al.*, 2005; Wardlow & Egbert, 2008). Through a Memorandum of Use (MOU) with the Kansas FSA office, 2015 CLU and FSA 578 data were acquired for a state-wide land cover mapping project. There are over a million polygons in the

2015 Kansas CLU database with reported crop types and intended land use. Grassland type information is included in the database as a crop type. “Intended Land Use Code” identifies the land use that the land owner intends to use the grassland for during the upcoming year and includes the categories of forage, grazing, and left standing (not grazed or hayed). For each feature in the CLU database, a unique identifier was created by concatenating the following attributes: State FIPS code, County FIPS code, Tract Number, and Farm Number (SCTF).

Three datasets of remotely sensed imagery were assembled for the study and include Landsat 8 surface reflectance, Landsat 8 NDVI, and Terra MODIS NDVI time series. Two Landsat path/row areas from the Landsat Worldwide Reference System (WRS), 27/33 and 28/33, were used in the analysis to determine the generality of the mapping approach. Three Landsat 8 surface reflectance images for each path/row were ordered and acquired using USGS’s EarthExplorer (EE) tool https://earthexplorer.usgs.gov/ to represent the spring, summer and fall portions of the growing season for both path/rows. The dates of the imagery obtained for 27/33 were 03/30/2015, 06/12/2013, and 11/09/2015. The dates of imagery obtained for 28/33 were 03/21/2015, 07/24/2014, and 10/15/2015. The fall image for path/row 28/33 contained 12% cloud cover in the northwest portion of the image. Only the cloud-free portions of the study area were included in the analyses. Near cloud-free imagery were unavailable for the summer of 2015 for either path/row; however, it is uncommon for grasslands to change structure or composition from year to year, and the out-of-year summer dates represented the best available data. Monthly reports from the High Plains Regional Climate Center (HPRCC) show that both June 2013 and July 2014 were substantially drier than June and July in 2015 (Umphlett, 2013, 2014, 2015a, 2015b), suggesting a potential limitation of the anachronistic summer Landsat data used in this study. Using ERDAS Imagine, six multispectral bands (bands 2–7) were extracted from the three

image dates and combined to produce an 18-band multi-seasonal Landsat 8 dataset for each path/row. Using the same dates listed above, spring, summer, and fall surface reflectance Normalized Difference Vegetation Index (NDVI) images were acquired and stacked to create a three-date multi-seasonal Landsat 8 NDVI dataset. Lastly, a biweekly time series of 231-meter Terra MODIS 16-day composite NDVI from the 2015 growing season was downloaded from NASA's EarthData online tool, https://earthdata.nasa.gov/. The MODIS time series dataset was reprojected from the native Sinusoidal projection to the Albers Equal Area projection and clipped to the two Landsat WRS path/row (27/33 and 28/33) extents. The MODIS NDVI time series dataset was then resampled to 30-meter pixels using bilinear interpolation and snapped to the Landsat 8 footprint. The three datasets were stacked to create a 44-band imagery dataset.

Since the large size of MODIS pixels increases the chances of mixed pixels (which are comprised by a mixture of two or more land cover types), two qualifiers were used as a measure of pixel purity to identify MODIS pixels suitable for image classification training. First using ESRI ArcGIS, a polygon file of the original 231-m MODIS pixel footprint was used to calculate the percent of grassland in each MODIS pixel using the 30-m 2015 Level I Kansas Land Cover Patterns dataset (KARS, 2017). Second, the MODIS pixel footprint and the 2015 CLU boundary were intersected to calculate the percent of each pixel that fell within a field boundary. MODIS pixels containing 60% or more grassland that were 60% or more inside a CLU boundary were extracted and used in the analysis. The centroids of the MODIS pixels were used to extract reflectance and NDVI values from the 44-band image stack. The centroid was intersected with the USGS's high-resolution National Hydrography Dataset (NHD) waterbody feature layer to exclude point locations that fell within farm ponds that would affect the Landsat reflectance values. In addition, NDVI data values for all points were screened for negative values and those

point locations were excluded from the analyses. While there were additional Landsat 8 pixels that met the purity criteria and could have been used in the analysis, a one-to-one correspondence was maintained between MODIS and Landsat data to allow for a direct data comparison. These data were exported to an Excel file and imported into MATLAB software for statistical analysis and plotting spectral profiles.

Four hierarchies of grassland classes were created using the FSA 578 to determine what level of grassland type could be mapped. Table 1 shows the four levels of grassland classes used in the analysis along with the abbreviations used for the classes. Level-1 corresponds to functional grassland types where CRP and native grasslands, dominated by warm-season grasses, were aggregated to a single class while fescue and brome, dominated by cool-season grasses, also were aggregated to a single class. Level-2 separates the grassland types into three classes, CRP, native, and cool-season grasslands. Separating CRP from native was based on knowledge of the needs of the potential user-base of the land cover product in light of growing interest regarding CRP land being converted back to cropland. Level-3 separates fescue and brome grassland types. And lastly, Level-4 separates grassland types by land use (Forage, Grazed, and Left Standing).

The number of training sites for cool and warm season grasslands were selected in an attempt to represent approximate proportions of acres and counts of features (i.e. fields) in the landscape within each path/row. Tables 2 and 3 show the proportion of field counts (column 3) and the proportion of area (column 4) stratified by Level-4 grassland types for path/row 27/33 and 28/33, respectively. These two proportions were averaged (column five) to determine the number of training sites used in the random forest (RF) classification for each path/row. For example, the average proportion of native grazed (Ngz) was 36.8% in path/row 27/33, so 36.8%

of the total training sample sites were Ngz. This strategy was used in an attempt to incorporate the high frequency of small fields in fragmented landscapes. These proportions were based on fields represented in the 2015 FSA and CLU datasets, which represent 76% of the grassland acres in Kansas as shown in Chapter 1. As the table shows, the proportions vary between path/rows. Path/row 27/33 had a roughly 50-50% split between cool- and warm-season grassland types while path/row 28/33 had a 16%-84% split.

4.3.3 *Data Analysis*

Once the training data were extracted, supervised classifications using the RF classifier (Breiman *et al.*, 1984) were run using the training dataset for each grassland type hierarchy and for each of the four predictor datasets for each path/row. The "treebagger" function in MATLAB was used to develop ten forests (unique RF models) for each run. Each RF model contained a classification ensemble consisting of one thousand constructed decision trees. Each tree was built using a bootstrap sample containing 63.2% of the training data. The remaining 36.8% of the training data, referred to as "out-of-bag" (OOB) samples, were used to calculate unbiased estimates for predictive error and predictor importance of that tree. The default for the number of predictors used at each split were used (the square root of the total number of predictors). The OOB errors for the ten forests were calculated and plotted to assess model performance and stability as a function of the number of trees grown (maximum of 1000 trees). Predictor importance was estimated using the OOB permutated predictor delta error where for each predictor variable, data values were permutated while other predictors remained unchanged. The forest ensemble was retrained and the change in model OOB accuracy (delta error) was calculated and averaged across the trees (sub-models) and normalized by dividing by the standard deviation, which is referred to as Mean Decrease in Accuracy (MDA). A small change

in MDA indicates low importance ranking while a large increase in MDA indicates high predictor importance ranking.

The mean OOB errors from RF were also compared across the four grassland hierarchies and four datasets for each path/row to determine to evaluate the performance of each dataset for each grassland type hierarchy. Once the optimal dataset and grassland hierarchy were selected, the models for each path/row were applied to the image data to produce maps. OOB samples and independent validation data samples were used to assess model performance and accuracy. Using the OOB error and independent validation data, traditional map accuracy values were calculated and compared, including overall, user's and producer's accuracy levels, and the Kappa statistic. In addition, probabilistic mapping disagreements (i.e. quantity, allocation, and total disagreement) were assessed by rescaling the OOB and validation contingency tables to reflect map proportions (Kastens *et al.*, 2017; Pontius Jr & Millones, 2011).

The two Landsat path/rows in the study area were adjacent and provided an approximately 54 by 158 kilometer overlap (see overlap area in Figure 1). Validation samples from the overlap area were used to compare mapped proportions and model performance between the independently processed path/rows.

4.4 *Results*

4.4.1 *OOB Error and Grassland Hierarchies*

The OOB error across ten RF models for the four datasets consistently showed that using a forest of 1,000 trees was more than adequate for model stability. Figure 2 shows an example of OOB error as a function of the number of trees grown for Level-2 mapping in path/row 27/33. The results suggest that for this study the number of trees could be substantially reduced to improve data processing efficiency.

As expected, OOB error estimates consistently increased from Level-1 to Level-4 grassland hierarchies across all datasets and for both path/rows, meaning that the Level-1 grassland hierarchy had the lowest OOB error and Level-4 had the highest OOB error. The combined dataset consistently had the lowest OOB error versus the other independent datasets. The results from Chapter 2 support this result where the highest spectral separabilities using Jeffries-Matusita (JM) distance to compare different class pairs from the various grassland hierarchies was obtained using the combined dataset. Comparing the three independent datasets (multispectral Landsat 8, Landsat 8 NDVI and MODIS NDVI), the multispectral Landsat dataset had the lowest OOB error for both path/rows. This result may indicate that the spectral resolution of the multispectral Landsat data provides significant information for separating grassland types beyond NDVI. Other portions of the electromagnetic spectrum (EMS) represented in Landsat 8 bands correspond to several biophysical properties of vegetation. Jensen (1983) illustrated the correspondence of the vegetation spectral reflectance curve and the associated biophysical characteristics across the EMS, and the USGS (2018) lists the mapping utility for each Landsat band as it corresponds to various vegetation properties. Table 5 combines Jenson's and USGS information to show that Landsat bands 2-4 in the visible portion of the EMS correspond to leaf pigments with two chlorophyll absorption regions; Landsat band 5 in the NIR region corresponds to cell structure, with higher reflectance corresponding to more vegetation biomass; and Landsat 8 bands 6-7 in the shortwave-infrared region correspond to vegetation water content, where shortwave radiation is absorbed with increasing water content.

Table 6 shows OOB error estimates for Level-2 grassland mapping. Overall accuracy levels for both path/rows were above 85%; however, the low producer's accuracy indicates that CRP has high omission error. The error matrix for path/row 27/33 shows 78% of CRP samples

were mapped as warm-season grasslands (Table 7), and similar results were obtained for path/row 28/33. Given that CRP was a relatively small class and that JM distance statistics were low between CRP and native warm-season grassland (Chapter 2, Figure 3), the misclassification of CRP as warm-season grasslands was not surprising. So, while mapping CRP is of special interest to end-users, the high omission error indicates CRP mapping would provide inadequate representation for general use. However, the user's accuracy levels (85% and 78%) indicate confidence in what was mapped as CRP, which could have utility for specific uses, such as finding CRP samples for targeted field campaigns or other research.

4.4.2 *Level-1 Mapping Results*

Figures 3 and 4 show the RF results using the combined dataset and Level-1 grassland hierarchy along with corresponding mapped pixel counts, acres, and areal proportions and CLU reported proportions and acres. The regional distribution of grassland types was as expected, with large tracts of warm-season grassland in the Flint Hills region, a larger dominance of cool-season grassland in the eastern half of 27/33, and small interlaced fields of cool-season grassland in the river lowlands in 28/33. For path/row 27/33, 56% percent of the area was mapped as warm-season and 44% as cool-season. This differs from the 62% and 38% indicated by CLU data. Meanwhile, for path/row 28/33 (Figure 4) the mapped proportions correspond well with the CLU area proportions. However, it should be noted that the CLU represents only an estimated 58% of the total grassland area in 28/33 and 47% in 27/33. Interestingly, the 27/33 mapped proportions correspond with the average proportion that are accounted for by a high frequency of small fields, whereas the 28/33 mapped proportions correspond with CLU area proportions.

4.4.3 *Predictor Importance*

The estimated predictor importance shown in Figures 5 and 6 show that some predictors in the combined dataset ranked higher than others and that predictor importance varies by path/row. For 27/33 these estimates of predictor importance correspond to the by-predictor (band or date) JM statistics shown in Figure 3 of Chapter 2. They correspond since both are using a single predictor versus the JM distance where all bands are combined and distributions are evaluated in multidimensional space. For MODIS NDVI, periods 6-8 and 20-23 had the highest predictor importance and JM distance. For multispectral Landsat, the spring and fall NIR bands had the highest predictor importance and JM distance.

For path/row 28/33, spring MODIS NDVI periods (6-8) ranked high for estimated predictor importance, but not periods 20-23 (November 1st-December 18th). Also, the spring Landsat NDVI ranked high along with the Landsat Fall NIR band. Spring Landsat 8 dates were of similar importance between the two path/rows; however 27/33 had a fall date that was roughly a month later than that for path/row 28/33. Even so, predictor importance for fall/winter periods for MODIS NDVI did not rank high for 28/33. The different rankings between path/rows could correspond to the proportion of the two functional grassland types being mapped within each path/row. The higher proportion of cool-season grasslands in 27/33 could result in higher predictor importance rankings for fall dates. And while some of the important predictors in each path/row may provide redundant information (e.g. Landsat near-infrared (NIR) (B4) and Landsat NDVI), at each binary split in the RF, a single predictor variable from the random subset of predictors is used versus the multidimensional data vector, and therefore RF is not severely impacted by correlated predictors or the inclusion of weak predictors (Cánovas-García & Alonso-Sarría, 2015).

4.4.4 *Accuracy Assessment*

The OOB estimates of producer's accuracy levels were 86% and 98% for warm-season grasslands for the two path/rows. Producer's accuracy levels differed more substantially for cool-season grasslands between 27/33 (89%) and 28/33 (69%). For 28/33, 25% of the cool-season OOB error samples were classified as warm-season grasslands (results not shown). The increased omission error in mapping cool-season grasslands in 28/33 could be due to its more lopsided class proportion compared to 27/33, with a smaller class typically being more challenging to model (and map) effectively. Comparatively, cool-season grassland was a small class in 28/33, but a dominant class in 27/33. The lower user's accuracy for cool-season grasslands in 28/33 could also result from the sampling design used for allocating training samples. Using the average of the proportion of acres and count of fields roughly doubled the proportion of cool-season training sites in the training sample (15.8%) versus using a solely area-based sample allocation (7.1%) (Table 3). The proportions in path/row 27/33 went from a 60-40 split using an area proportion to a 50-50 split for warm- and cool-season grasslands using the averaged proportion. This shift in proportions did not appear to affect the user's and producer's accuracy levels in 27/33. In a review of RF in remote sensing applications, Belgiu and Drăguţ (2016) highlight research showing that a proportionally allocated training sample scheme provides optimal classification results and that RF is sensitive to the proportions of training samples used (Colditz, 2015; Millard & Richardson, 2015). However, Jin *et al.* (2014) showed that proportionally allocated training samples increased user's accuracy of an under-represented class while equally allocated training samples increased producer's accuracy of an under-represented class. Other research has shown that imbalanced training data maximizes class accuracy levels for the majority class, but at the expense of the minority class and that balanced

training can be used to reduce errors in minority classes (Chen *et al.*, 2004; Mellor *et al.*, 2015). For this study the intention was to balance between high frequency small fields and lower frequency large fields.

Comparing overall accuracy levels between OOB estimates and test data, the test data showed unexpected 6% and 4% increases in accuracy for 27/33 and 28/33, respectively. The increased overall accuracy with using test data could relate to differences between training and test data proportions. For example, for 28/33 there was an increase in the proportion of Native grazed (Ngz) in the testing data (82%) versus the proportion used in the training data (69%) and a decrease in all Level-4 cool-season proportions, with 4.4% cool-season grasslands represented in the test data compared to 14.8% used in the training data. The proportions used for training and test data for 27/33 were more closely aligned, with three small classes not represented in the test data.

According to Breiman (1996), the OOB error estimate is an unbiased estimate of the classification error and provides as good a measure of error as if using independent test data of the same sample size as the RF training data. Other research has shown OOB error to overestimate error in RF based on several interacting factors, including sample size, sample proportions (balanced vs. imbalanced), subsample proportions. number of predictors, and correlations between predictors (Janitza & Hornung, 2018). For example, Janitza and Hornung showed that for small, imbalanced sample sizes, OOB error bias increased due to the extreme imbalance of bootstrap subsamples used to build trees that were preferential to the dominant class. And that the number of predictors can affect the bias of the OOB error depending upon sample size and whether the samples were balanced or imbalanced. Others argue that OOB error can be overestimated due to differences in the distribution of the bootstrap sample (63.2% of the

training data) used to build the tree versus the OOB samples (36.8% of the training data) uses for testing (Efron & Tibshirani, 1994). It could be argued that the test error is more accurate since the entire forest is used versus the OOB error estimate that uses a subsample of trees for OOB prediction of each training sample. There are several potential explanations as to why the test error was lower than the OOB error, there appear to be several potential explanations including complex interactions among forest parameters and nuances between training and test data that warrant require further exploration.

The OOB quantity disagreement was higher for path/row 28/33 (6.2%) than 27/33 (0.8%) while the allocation disagreement was higher for 27/33 (11.7%) than for 28/33 (1.6%) (Table 8). The differences between OOB and test data disagreement result from the different class proportions in the training and test data sets used as reference data for calculating allocation and quantity disagreement against proportions mapped. As previously mentioned, the proportions for warm-season grassland increased from 69% in the training to 82% in the test data. Total disagreement for OOB and test data were lower for 28/33 which was likely influenced by the dominance of warm-season grasslands in 28/33.

4.4.5 *Path/row Model Comparison*

Table 9 shows the acreage and proportion of agreement and disagreement in mapping the two Level-1 classes in the overlap area between the two adjacent path/rows. The results show there was 94% agreement in the mapping overall, while there was 5% disagreement (73,249 acres) where grasslands that were mapped as warm-season in path/row 28/33 were mapped as cool-season in path/row 27/33. Overall accuracy levels and Kappa were higher for path/row 27/33 than 28/33 (Table 10). Additionally, user's and producer's accuracy levels for cool-season grasslands were higher for 27/33 (89% and 88%, respectively) than in path/row 28/33 (73% and

63%) (Table 10). And the total disagreement (quantity + allocation disagreement) was lower in 27/33 (4.6%) than in 28/33 (7.0%). These values indicate that the model from path/row 27/33 mapped cool-season grassland more accurately than 28/33, where cool-season grasslands constituted a small class.

4.5 *Conclusions*

The RF results indicate that a Level-1 classification using the combined dataset of multispectral Landsat 8, Landsat 8 NDVI and MODIS NDVI produced the lowest OOB error of all the datasets. The OOB error for Level-2 classification was only 1-2% higher; however, the CRP class had high omission error and was largely mapped as native warm-season grassland, further supporting using the two-class Level-1 classification scheme. Predictor importance rankings for the Level-1 classification varied by path/row, and were likely influenced by the proportions of cool-season grassland class in each path/row. The spatial distributions of the mapped classes appeared reasonable, e.g. cool-season occupying river lowlands and large tracts of warm-season in the Flint Hills. Overall accuracy levels were greater than 87% using OOB and test data. However, path/row 28/33 had lower user's and producer's accuracy levels for cool-season grasslands and a lower Kappa statistic. In evaluating the overlap area, there was 94% map agreement between the two path/rows, but again the cool-season class in 28/33 had lower user's and producer's accuracy levels, which was expected due to the small proportion of cool-season grassland represented in the path/row. These results will be used to formulate a methodology for mapping functional grassland types at a regional scale. Future work could evaluate whether smaller classes could be mapped more accurately by increasing the density of spring and fall Landsat dates, adding additional training data from Landsat (that were restricted by MODIS in

this study), and by combining path/rows that may provide additional training data samples for under-represented classes, like cool-season grasslands in path/row 28/33.

4.6 *Figures and Tables*

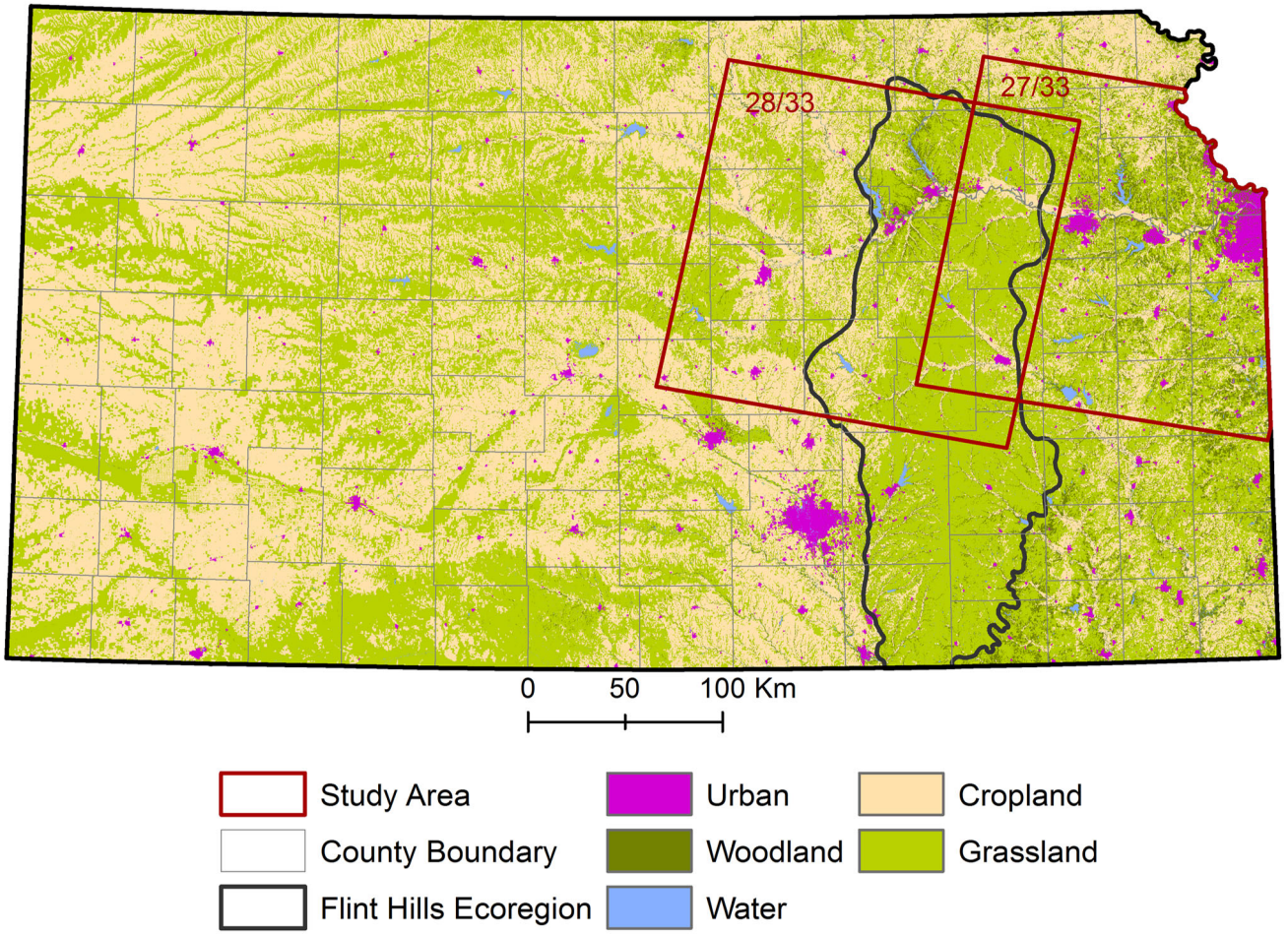


Figure 4.1. The study area comprised of two adjacent Landsat 8 path/rows, 28/33 and 27/33 (shown in red). The Flint Hills ecoregion is overlaid a 2015 land cover map (KARS, 2017).

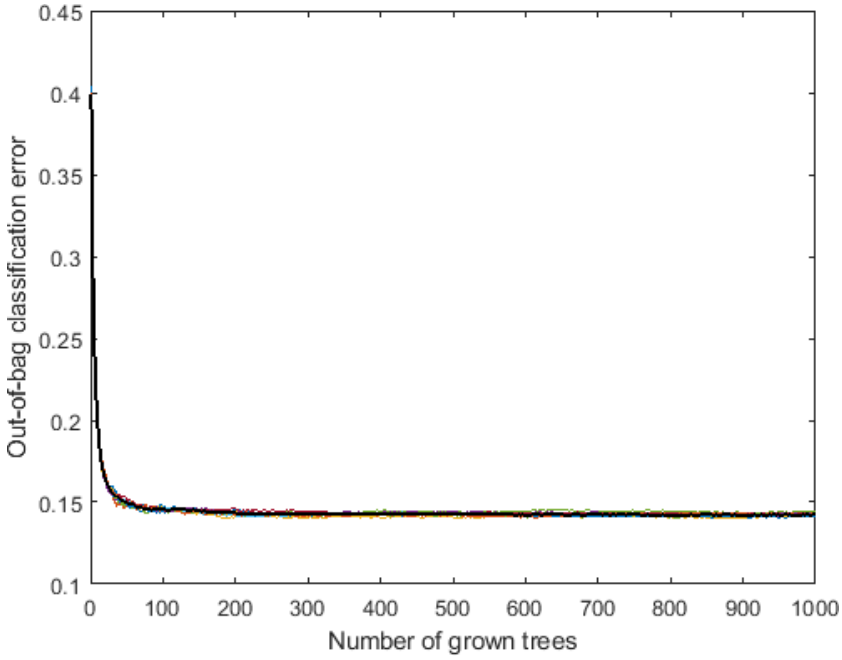
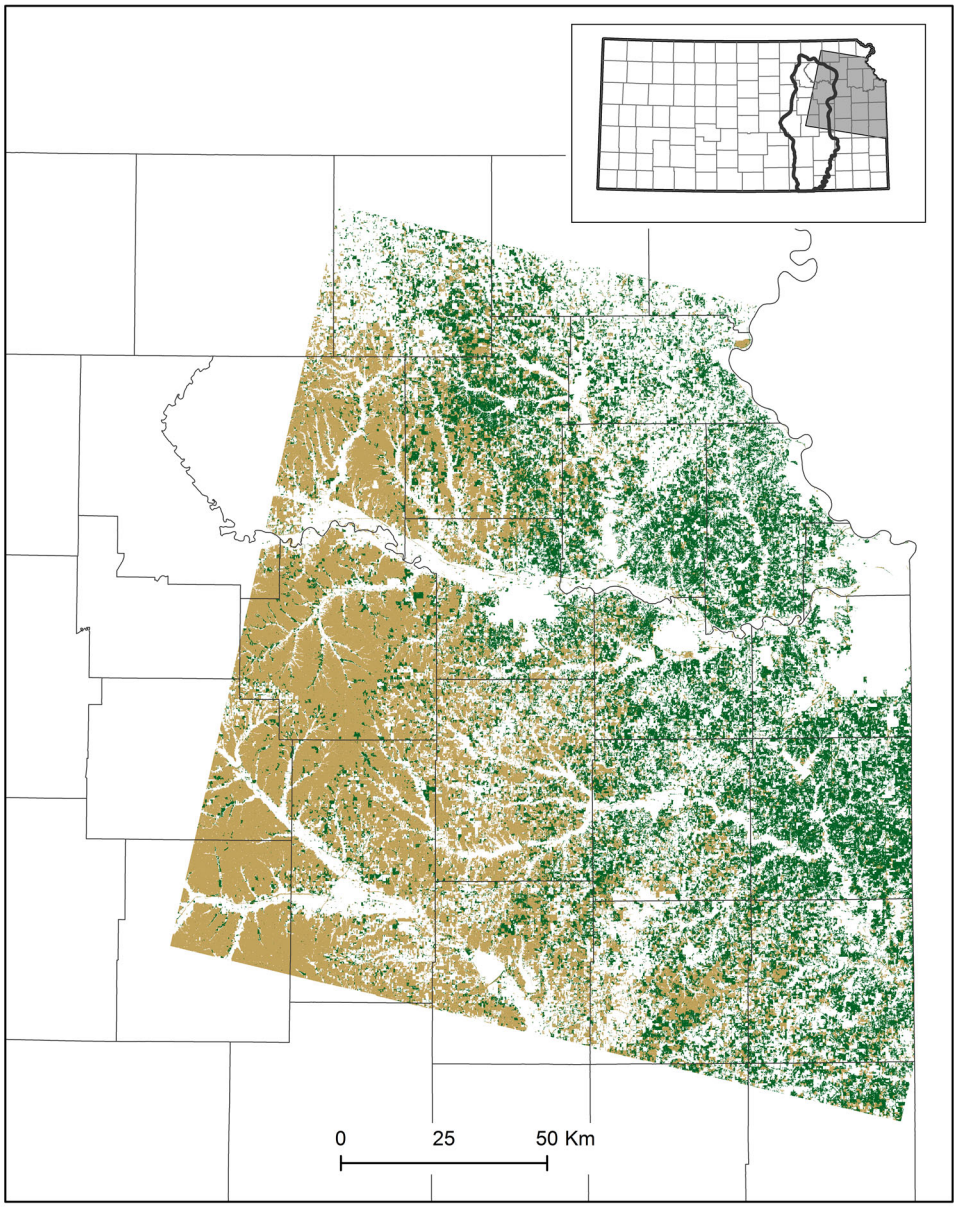
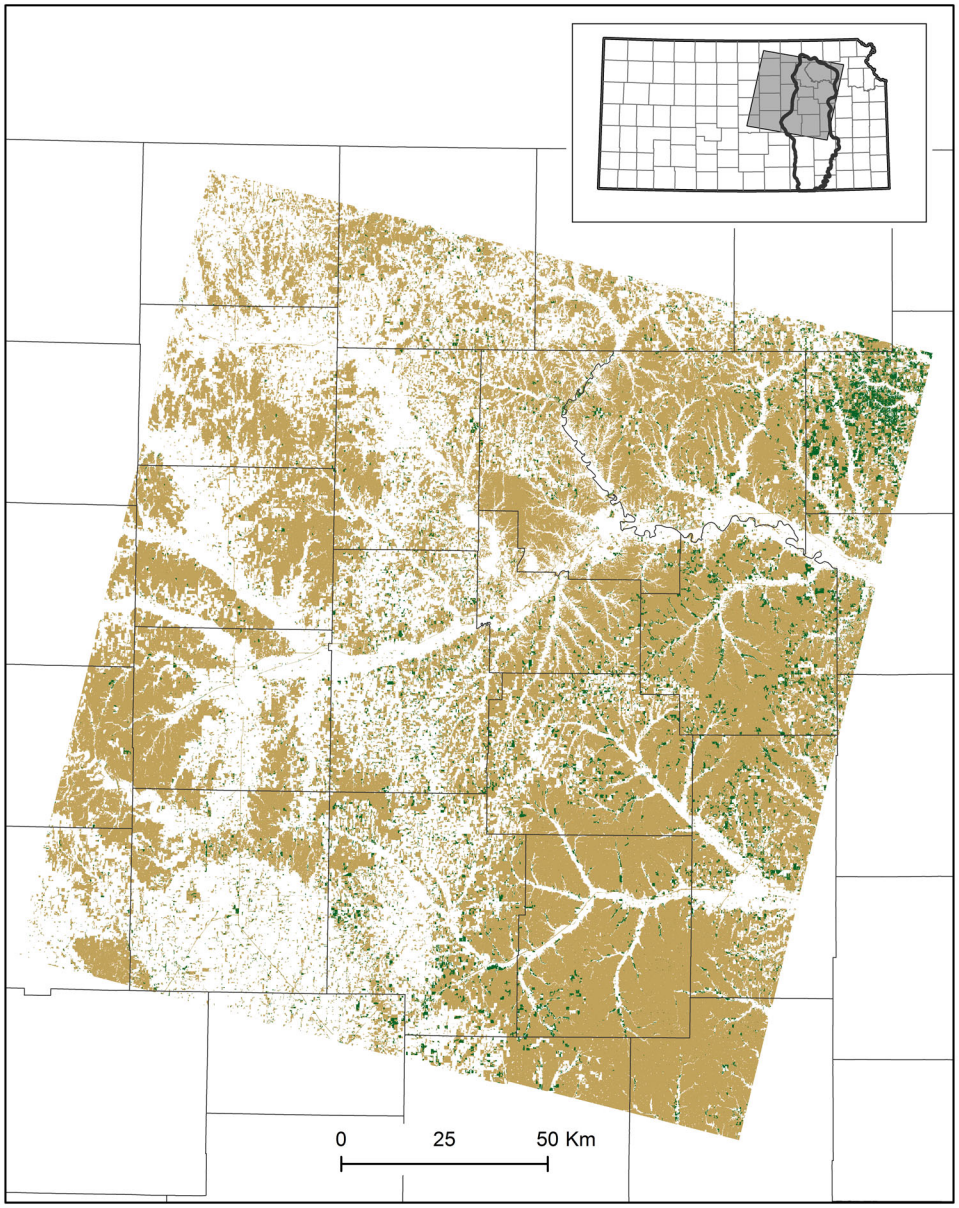


Figure 4.4.1. OOB error as function of the number of trees grown for Level-2 mapping in path/row 27/33. Models were consistently stable with respect to OOB error when forest size reached about 200-500 trees.



	Pixel Count (30-meter)	Area (Acres)	Mapped Proportions	CLU Area Proportions	CLU Area (Acres)
Warm-season	9,180,099	2,041,604	55.8%	61.9%	1,081,137
Cool-season	7,276,736	1,618,306	44.2%	38.1%	665,648
Total	*16,456,835*	*3,659,910*			*1,746,785*

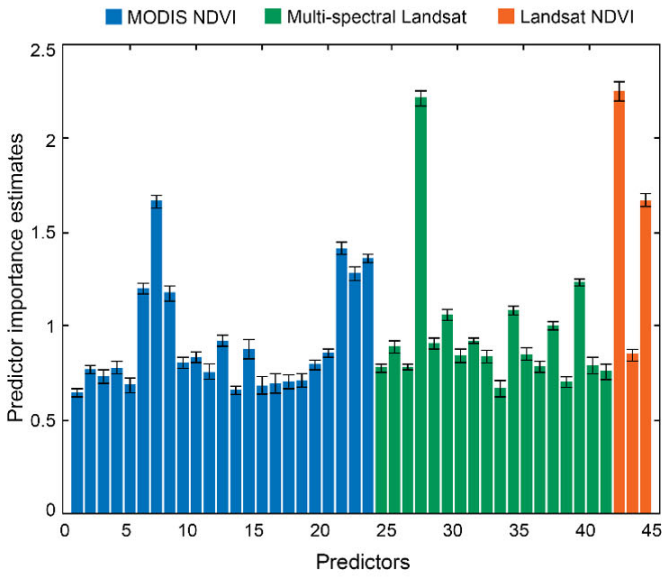
Figure 4.4.2. Level-1 map for path/row 27/33 using the RF classifier and the combined dataset.



County Boundary Warm-season Grassland Cool-season Grassland

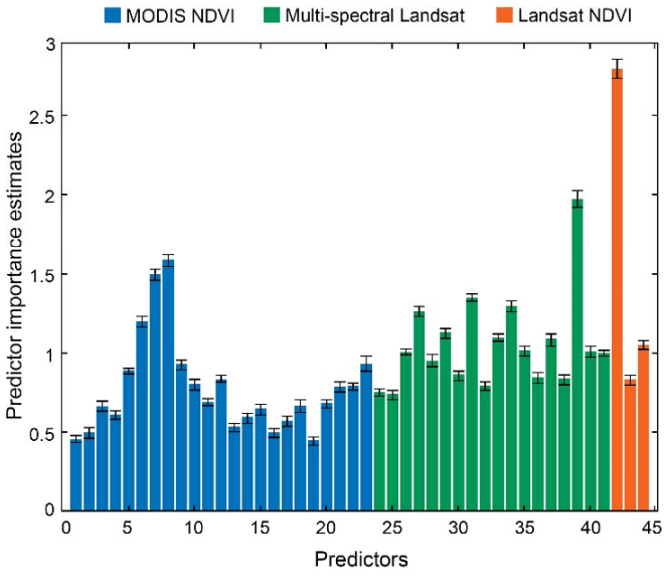
	Pixel Count (30-meter)	Area (Acres)	Mapped Proportions	CLU Area Proportions	CLU Area (Acres)
Warm-season	19,371,392	4,308,091	93.2%	92.9%	2,483,241
Cool-season	1,417,297	315,199	6.8%	7.1%	189,965
Total	*20,788,689*	*4,623,290*			*2,673,206*

Figure 4.4.3. Level-1 map for path/row 28/33 using the RF classifier and the combined dataset.



Predictors	MODIS NDVI	Predictors	Landsat 8
1	Period 1	24	Spring, Band 2
2	Period 2	25	Spring, Band 3
3	Period 3	26	Spring, Band 4
4	Period 4	27	Spring, Band 5
5	Period 5	28	Spring, Band 6
6	Period 6	29	Spring, Band 7
7	Period 7	30	Summer, Band 2
8	Period 8	31	Summer, Band 3
9	Period 9	32	Summer, Band 4
10	Period 10	33	Summer, Band 5
11	Period 11	34	Summer, Band 6
12	Period 12	35	Summer, Band 7
13	Period 13	36	Fall, Band 2
14	Period 14	37	Fall, Band 3
15	Period 15	38	Fall, Band 4
16	Period 16	39	Fall, Band 5
17	Period 17	40	Fall, Band 6
18	Period 18	41	Fall, Band 7
19	Period 19	42	Spring NDVI
20	Period 20	43	Summer NDVI
21	Period 21	44	Fall NDVI
22	Period 22		
23	Period 23		

Figure 4.4.4. Predictor importance estimates for path/row 27/33. Spring and fall MODIS and Landsat NDVI and spring Landsat NIR band ranked highest.



Predictors	MODIS NDVI	Predictors	Landsat 8
1	Period 1	24	Spring, Band 2
2	Period 2	25	Spring, Band 3
3	Period 3	26	Spring, Band 4
4	Period 4	27	Spring, Band 5
5	Period 5	28	Spring, Band 6
6	Period 6	29	Spring, Band 7
7	Period 7	30	Summer, Band 2
8	Period 8	31	Summer, Band 3
9	Period 9	32	Summer, Band 4
10	Period 10	33	Summer, Band 5
11	Period 11	34	Summer, Band 6
12	Period 12	35	Summer, Band 7
13	Period 13	36	Fall, Band 2
14	Period 14	37	Fall, Band 3
15	Period 15	38	Fall, Band 4
16	Period 16	39	Fall, Band 5
17	Period 17	40	Fall, Band 6
18	Period 18	41	Fall, Band 7
19	Period 19	42	Spring NDVI
20	Period 20	43	Summer NDVI
21	Period 21	44	Fall NDVI
22	Period 22		
23	Period 23		

Figure 4.4.5. Predictor importance estimates for path/row 28/33. Spring MODIS and Landsat NDVI and fall Landsat NIR band ranked highest.

Table 4.4.1 The four levels of grassland type hierarchies and abbreviations used to evaluate OOB error using the four remotely sensed datasets. Level-1 separates functional grassland types, Level-2 separates the warm-season grasslands into two classes, Level-3 separates the cool-season grasslands into two classes and Level-4 separates land use within Level-3 grassland types.

Level-1	Level-2	Level-3	Level-4
Warm-Season (W)	CRP (Crp)	CRP (Crp)	CRP (Crp)
	Native (N)	Native (N)	Native Forage (Nfg)
			Native Grazed (Ngz)
			Native Left Standing (Nls)
Cool-Season (C)	Cool-Season (C)	Fescue (F)	Fescue Forage (Ffg)
			Fescue Grazed (Fgz)
			Fescue Left Standing (Fls)
		Brome (B)	Brome Forage (Bfg)
			Brome Grazed (Bgz)
			Brome Left Standing (Bls)

Table 4.4.2. For path/row 27/33, the proportions of the number of fields, acres, and the average proportion. The last two columns show the training sample allocation using the average proportion and the testing data sample count.

Level-1 Grassland Type	Level-4 Grassland Type	Proportion of Fields (Count)	Proportion Area (Acres)	Proportion Average	Training Data Points	Test Data Points
Warm	CRP	2.4%	1.9%	2.2%	215	522
	Nfg	8.0%	5.6%	6.8%	665	743
	Ngz	21.8%	51.8%	36.8%	3,682	3,047
	Nls	4.7%	2.6%	3.7%	302	0
Cool	Ffg	2.1%	0.9%	1.5%	147	119
	Fgz	11.9%	10.0%	11.0%	1,098	1,265
	Fls	0.8%	0.4%	0.6%	32	0
	Bfg	25.3%	10.1%	17.7%	1,770	1,174
	Bgz	19.9%	15.8%	17.9%	1,787	1,999
	Bls	3.1%	0.9%	2.0%	103	0
				Total	*9,801*	*8,869*

Table 4.4.3. For path/row 28/33, the proportions of the number of fields, acres and the average proportion. The last two columns show the training sample allocation using the average proportion and the testing data sample count.

Level-1 Grassland Type	**Level-4 Grassland Type**	**Proportion of Fields (Count)**	**Proportion Area (Acres)**	**Proportion Average**	**Training Data Points**	**Test Data Points**
Warm	CRP	2.3%	1.5%	1.9%	189	138
	Nfg	11.9%	5.3%	8.6%	813	1,151
	Ngz	55.2%	84.4%	69.8%	6,603	9,113
	Nls	5.9%	1.7%	3.8%	360	191
Cool	Ffg	0.1%	0.0%	0.0%	0	1
	Fgz	0.4%	0.2%	0.3%	30	15
	Fls	0.0%	0.0%	0.0%	1	0
	Bfg	15.5%	3.8%	9.7%	918	260
	Bgz	6.9%	2.8%	4.8%	454	219
	Bls	1.7%	0.3%	1.0%	95	0
				Total	*9,460*	*11,088*

Table 4.4.4. The OOB error for each path/row for each grassland hierarchy and each predictor dataset. OOB error increased with grassland hierarchy, and the combined dataset had the lowest OOB error.

Path/Row	Grassland Hierarchy	Combined	Landsat 8	Landsat NDVI	MODIS
27/33	Level-1	12.5%	13.6%	15.8%	15.8%
	Level-2	14.3%	15.5%	17.8%	17.6%
	Level-3	22.4%	24.0%	29.7%	26.7%
	Level-4	35.6%	37.1%	47.4%	43.7%
28/33	Level-1	8.3%	8.8%	11.4%	14.6%
	Level-2	13.1%	13.6%	17.5%	20.1%
	Level-3	13.3%	13.8%	17.7%	20.3%
	Level-4	23.9%	24.2%	30.8%	32.3%

Table 4.4.5. Landsat 8 bands used in this study and the associated wavelengths, with suggested mapping utility and vegetation biophysical characteristics (Source: Jensen, 1983; USGS, 2018).

Landsat 8 Bands	Wavelength	Mapping Utility USGS (2018)	Vegetation Characteristics Jensen (1983)
Band 2 – Blue	0.452 - 0.512	Bathymetric mapping, distinguishing soil from vegetation, and deciduous from coniferous vegetation.	Leaf Pigments; chlorophyll absorption
Band 3 - Green	0.533 - 0.590	Emphasizes peak vegetation, which is useful for assessing plant vigor.	Leaf Pigments
Band 4 - Red	0.636 - 0.673	Discriminates vegetation slopes	Leaf Pigments; chlorophyll absorption
Band 5 - Near Infrared (NIR)	0.851 - 0.879	Emphasizes biomass content and shorelines.	Cell structure
Band 6 - Short-wave Infrared (SWIR) 1	1.566 - 1.651	Discriminates moisture content of soil and vegetation; penetrates thin clouds.	Water content; Water absorption
Band 7 - Short-wave Infrared (SWIR) 2	2.107 - 2.294	Improved moisture content of soil and vegetation and thin cloud penetration.	Water content; Water absorption

Table 4.4.6. Accuracy levels for the three grassland classes in the Level-2 hierarchy. The CRP class had low producer's accuracy levels and was mostly mapped as warm-season grassland.

Path/Row	Accuracy Levels	OOB Estimates		
		CRP	***Warm-season***	***Cool-season***
27/33	User's Accuracy	85%	85%	86%
	Producer's Accuracy	13%	86%	89%
	Overall Accuracy	85.8%		
	Kappa	0.72		
28/33	User's Accuracy	78%	87%	87%
	Producer's Accuracy	20%	97%	70%
	Overall Accuracy	86.8%		
	Kappa	0.64		

Table 4.4.7. The error matrix for the Level-2 hierarchy for 27/33 illustrates that CRP was largely mapped as warm-season grassland.

		Reference Class			
		CRP	***Warm-season***	***Cool-season***	**Total**
Predicted Class	***CRP***	28	2	3	33
	Warm-season	**168**	3,976	527	4,671
	Cool-season	19	671	4,407	5,097
	Total	215	4,649	4,937	9801

Table 4.4.8. Accuracy levels for both path/rows calculated using both OOB estimates and test data.

Path/Row	Accuracy Levels	OOB Estimates		Test Samples	
		Cool	*Warm*	*Cool*	*Warm*
27/33	User's	87%	88%	91%	94%
	Producer's	89%	86%	95%	90%
	Overall	87%		93%	
	Kappa	0.75		0.85	
	Proportion Correct	87.5%		92.8%	
	Quantity Disagreement	0.8%		0.6%	
	Allocation Disagreement	11.7%		6.6%	
	Total Disagreement	12.5%		7.2%	
28/33	User's	88%	92%	56%	99%
	Producer's	69%	98%	70%	97%
	Overall	92%		96%	
	Kappa	0.73		0.60	
	Proportion Correct	92.1%		95.6%	
	Quantity Disagreement	6.2%		1.7%	
	Allocation Disagreement	1.6%		2.7%	
	Total Disagreement	7.9%		4.4%	

Table 4.4.9.Co-occurrence matrix showing acres and percentages of mapped classes in the overlap area.

	28/33	
27/33	*Warm-season* *Acres (%)*	*Cool-season* *Acres (%)*
Warm-season	1,212,152 (84%)	15,559 (1%)
Cool-season	73,249 (5%)	137,931 (10%)

References

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

Breiman, L. (1996). Out-of-bag estimation, ftp. stat. berkeley. edu/pub/users/breiman. *OOBestimation. ps, 199*(6).

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and decision trees. *Wadsworth, Belmont, 378*.

Brown, J. C., Kastens, J. H., Coutinho, A. C., de Castro Victoria, D., & Bishop, C. R. (2013). Classifying multiyear agricultural land use data from Mato Grosso using time series MODIS vegetation index data. *Remote Sensing of Environment, 130*, 39-50.

Cánovas-García, F., & Alonso-Sarría, F. (2015). Optimal Combination of Classification Algorithms and Feature Ranking Methods for Object-Based Classification of Submeter Resolution Z/I-Imaging DMC Imagery. *Remote Sensing, 7*(4), 4651.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley, 110*, 1-12.

Colditz, R. R. (2015). An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing, 7*(8), 9655-9681.

Davidson, A., & Csillag, F. (2003). A comparison of three approaches for predicting C4 species cover of northern mixed grass prairie. *Remote Sensing of Environment, 86*(1), 70-82. doi:http://dx.doi.org/10.1016/S0034-4257(03)00069-5

Davis, S. M., Landgrebe, D. A., Phillips, T. L., Swain, P. H., Hoffer, R. M., Lindenlaub, J. C., & Silva, L. F. (1978). Remote sensing: the quantitative approach. *New York, McGraw-Hill International Book Co., 1978. 405 p., 1*.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*: CRC press.

Egbert, S. L., Lee, R. Y., Price, K. P., Boyce, R., & Nellis, M. D. (1998). Mapping conservation reserve program (CRP) grasslands using multi-seasonal thematic mapper imagery. *Geocarto International, 13*(4), 17-24.

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*(2), 89-93. doi:http://dx.doi.org/10.1016/j.ecoinf.2007.03.009

Franke, J., Keuck, V., & Siegert, F. (2012). Assessment of grassland use intensity by remote sensing to support conservation schemes. *Journal for Nature Conservation, 20*(3), 125-134. doi:http://dx.doi.org/10.1016/j.jnc.2012.02.001

Gelfand, I., Zenone, T., Jasrotia, P., Chen, J., Hamilton, S. K., & Robertson, G. P. (2011). Carbon debt of Conservation Reserve Program (CRP) grasslands converted to bioenergy production. *Proceedings of the National Academy of Sciences, 108*(33), 13864-13869.

Gibson, D. J. (2009). *Grasses and grassland ecology*: Oxford University Press.

Gu, Y., & Wylie, B. K. (2015). Developing a 30-m grassland productivity estimation map for central Nebraska using 250-m MODIS and 30-m Landsat-8 observations. *Remote Sensing of Environment, 171*, 291-298.

Guo, X., Price, K. P., & Stiles, J. (2003). Grasslands Discriminant Analysis Using Landsat TM Single and Multitemporal Data. *Photogrammetric Engineering & Remote Sensing, 69*(11), 1255-1262. doi:10.14358/PERS.69.11.1255

Guo, X., Price, K. P., & Stiles, J. M. (2000). Biophysical and Spectral Characteristics of Cool- and Warm-Season Grasslands under Three Land Management Practices in Eastern Kansas. *Natural Resources Research, 9*(4), 321-331. doi:10.1023/a:1011513527965

Halabuk, A., Mojses, M., Halabuk, M., & David, S. (2015). Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series. *Remote Sensing, 7*(5), 6107.

Hellerstein, D. M. (2017). The US conservation reserve program: the evolution of an enrollment mechanism. *Land Use Policy, 63*, 601-610.

Hendricks, N. P., & Er, E. (2018). Changes in cropland area in the United States and the role of CRP. *Food Policy, 75*(C), 15-23.

Hughes, J. P., Robel, R. J., Kemp, K. E., & Zimmerman, J. L. (1999). Effects of Habitat on Dickcissel Abundance and Nest Success in Conservation Reserve Program Fields in Kansas. *The Journal of Wildlife Management, 63*(2), 523-529. doi:10.2307/3802638

Janitza, S., & Hornung, R. (2018). On the overestimation of random forest's out-of-bag error. *PloS one, 13*(8), e0201904.

Jensen, J. R. (1983). Biophysical remote sensing. *Annals of the Association of American Geographers, 73*(1), 111-132.

Jin, H., Stehman, S. V., & Mountrakis, G. (2014). Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado. *International Journal of Remote Sensing, 35*(6), 2067-2081.

Johnston, C. A. (2014). Agricultural expansion: land use shell game in the US Northern Plains. *Landscape ecology, 29*(1), 81-95.

KARS (Cartographer). (2008). 2005 Kansas Land Cover Patterns. Retrieved from http://kars.ku.edu/geonetwork/srv/en/main.home

KARS (Cartographer). (2017). 2015 Kansas Land Cover Patterns. Retrieved from http://kars.ku.edu/geonetwork/srv/en/main.home

Kastens, J. H., Brown, J. C., Coutinho, A. C., Bishop, C. R., & Esquerdo, J. C. D. (2017). Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PloS one, 12*(4), e0176168.

Kennedy, S. S. (1999). *Spectral discrimination of irrigated/non-irrigated crop types using landsat thematic mapper imagery in Haskell county, Kansas.* University of Kansas, Geography.

Lauver, C. L., & Whistler, J. L. (1993). A hierarchical classification of Landsat TM imagery to identify natural grassland areas and rare species habitat. *Photogrammetric Engineering and Remote Sensing, 59*(5), 627-634.

Lin, X., Harrington, J., Ciampitti, I., Gowda, P., Brown, D., & Kisekka, I. (2017). Kansas Trends and Changes in Temperature, Precipitation, Drought, and Frost-Free Days from the 1890s to 2015. *Journal of Contemporary Water Research & Education, 162*(1), 18-30. doi:doi:10.1111/j.1936-704X.2017.03257.x

Masialeti, I., Egbert, S., & Wardlow, B. D. (2010). A comparative analysis of phenological curves for major crops in Kansas. *GIScience & Remote Sensing, 47*(2), 241-259.

Maxwell, S. K., & Sylvester, K. M. (2012). Identification of "ever-cropped" land (1984–2010) using Landsat annual maximum NDVI image composites: Southwestern Kansas case study. *Remote Sensing of Environment, 121*, 186-195.

McInnes, W. S., Smith, B., & McDermid, G. J. (2015). Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(4), 1395-1403. doi:10.1109/JSTARS.2015.2416713

Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 155-168.

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing, 7*(7), 8489-8515.

Mosiman, B. N. (2003). *Exploring multi-temporal and transformation methods for improved cropland classification using Landsat Thematic Mapper imagery.* University of Kansas, Geography.

Owensby, C. E. (1993). *Introduction to range management*: Custom Academic Publishing Company.

Peterson, D., Price, K., & Martinko, E. (2002a). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing, 23*(23), 5015-5030.

Peterson, D. L., Price, K. P., & Martinko, E. A. (2002b). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing, 23*(23), 5015-5030.

Peterson, D. L., Price, K. P., & Martinko, E. A. (2002c). Investigating grazing intensity and range condition of grasslands in northeastern Kansas using Landsat thematic mapper data. *Geocarto International, 17*(4), 15-26.

Peterson, D. L., Whistler, J. L., Egbert, S. L., & Martinko, E. A. (2008). *2005 Kansas Land Cover Patterns: Phase II-Final Report*. Retrieved from Kansas Biological Survey:

Peterson, D. L., Whistler, J. L., Lomas, J. M., Dobbs, K. E., Jakubauskas, M. E., Egbert, S. L., & Martinko, E. A. (2005). *2005 Kansas Land Cover Patterns Phase I: Final Report* (150). Retrieved from Kansas Biological Survey, University of Kansas:

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32*(15), 4407-4429.

Porter, T. F., Chen, C., Long, J. A., Lawrence, R. L., & Sowell, B. F. (2014). Estimating biomass on CRP pastureland: A comparison of remote sensing techniques. *Biomass and Bioenergy, 66*, 268-274. doi:http://dx.doi.org/10.1016/j.biombioe.2014.01.036

Price, K. P., Guo, X., & Stiles, J. M. (2002a). Comparison of Landsat TM and ERS-2 SAR data for discriminating among grassland types and treatments in eastern Kansas. *Computers and Electronics in Agriculture, 37*(1), 157-171.

Price, K. P., Guo, X., & Stiles, J. M. (2002b). Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas.

International Journal of Remote Sensing, 23(23), 5031-5042. doi:10.1080/01431160210121764

Reynolds, R. E., Shaffer, T. L., Renner, R. W., Newton, W. E., & Bruce, D. J. B. (2001). Impact of the Conservation Reserve Program on Duck Recruitment in the U.S. Prairie Pothole Region. *The Journal of Wildlife Management, 65*(4), 765-780. doi:10.2307/3803027

Ribaudo, M. O., Colacicco, D., Langner, L. L., Piper, S., & Schaible, G. D. (1990). Natural Resources and Users Benefit from the Conservation Reserve Program. Agricultural Economic Report 627. *US Department of Agriculture, Economic Research Service, Washington, DC.*(Report 627).

Richards, J., & Jia, X. (1999). Remote sensing digital image analysis: An introduction Springer-Verlag. *Berlin Heidelberg*.

Riffell, S., Scognamillo, D., & Burger, L. W. (2008). Effects of the Conservation Reserve Program on northern bobwhite and grassland birds. *Environmental Monitoring and Assessment, 146*(1), 309-323. doi:10.1007/s10661-007-0082-8

Risser, P. G. (1988). *Diversity in and among grasslands*: National Academy of Sciences Washington, DC.

Robertson, K. R., Anderson, R. C., & Schwartz, M. W. (1997). The tallgrass prairie mosaic *Conservation in highly fragmented landscapes* (pp. 55-87): Springer.

Robertson, K. R., & Schwartz, M. (1994a). Prairies. *The changing Illinois environment: critical trends, 3*, 1-32.

Robertson, K. R., & Schwartz, M. W. (1994b). Prairies. *The changing Illinois environment: critical trends, 3*, 1-32.

Samson, F. B., Knopf, F. L., & Ostlie, W. R. (2004). Great Plains ecosystems: past, present, and future. *Wildlife Society Bulletin, 32*(1), 6-15.

Song, X., Fan, G., & Rao, M. (2005). Automatic CRP mapping using nonparametric machine learning approaches. *IEEE Transactions on Geoscience and Remote Sensing, 43*(4), 888-897. doi:10.1109/TGRS.2005.844031

Swain, P., & King, R. (1973). Two effective feature selection criteria for multispectral remote sensing.

Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K., & DeJong, D. D. (1997). NDVI, C3 and C4 production, and distributions in Great Plains grassland land cover classes. *Ecological applications, 7*(1), 59-78.

Umphlett, N. (2013). *June 2013 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

Umphlett, N. (2014). *July 2014 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

Umphlett, N. (2015a). *July 2015 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

Umphlett, N. (2015b). *June 2015 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

USDA, F. (2012). FSA Bulletin: CRP Emergency Haying and Grazing Released in 91 Kansas Counties [Press release]. Retrieved from https://content.govdelivery.com/accounts/USFSA/bulletins/496494

USDA, F. S. A. (2017). CRP Enrollment and Rental Payments by County, 1986-2017. Retrieved November 13, 2017 https://www.fsa.usda.gov/programs-and-services/conservation-programs/reports-and-statistics/conservation-reserve-program-statistics/index

USGS. (2018). Landsat Missions. Retrieved from https://landsat.usgs.gov/what-are-best-spectral-bands-use-my-study

Van Pelt, W. E., Kyle, S., Pitman, J., Klute, D., Beauprez, G., Schoeling, D., . . . Haufler, J. (2013). The lesser prairie-chicken range-wide conservation plan. *Western Association of Fish and Wildlife Agencies, Cheyenne, Wyoming, USA*.

Wang, C., Jamison, B. E., & Spicci, A. A. (2010). Trajectory-based warm season grassland mapping in Missouri prairies with multi-temporal ASTER imagery. *Remote Sensing of Environment, 114*(3), 531-539.

Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. *Remote Sensing of Environment, 112*(3), 1096-1116.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment, 108*(3), 290-310.

Wright, C. K., & Wimberly, M. C. (2013). Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences, 110*(10), 4134-4139. doi:10.1073/pnas.1215404110

Wu, J., & Weber, B. (2012). Implications of a reduced conservation reserve program. The conservation crossroads of agriculture. *Council on Food, Agriculture and Resource Economics. Washington, D.C., USA.*

Zha, Y., Gao, J., Ni, S., Liu, Y., Jiang, J., & Wei, Y. (2003). A spectral reflectance-based approach to quantification of grassland cover from Landsat TM imagery. *Remote Sensing of Environment, 87*(2-3), 371-375.

5 Conclusions

The goal of this research was to use a data-driven approach to develop a classification approach, i.e., which combination of remotely sensed imagery and thematic classification scheme, to most accurately map dominant grassland types at a regional scale. To achieve this research goal there were three main objectives.

1. Identify the dominant land use within the two grassland types (warm- and cool-season grasslands) using United States Department of Agriculture (USDA) Farm Service Agency (FSA) 578 data and characterize the static or dynamic nature of land use in grassland types in Kansas.
2. Determine the spectral separability of four hierarchies of grassland types and land use using multi-seasonal Landsat 8 spectral bands, Landsat 8 Normalized Difference Vegetation Index (NDVI), and Moderate Resolution Imaging Spectrometer (MODIS) NDVI time series.
3. Determine the optimal combination of data, and the appropriate thematic resolution, for mapping grassland type by comparing modeling performance using a Random Forest (RF) modeling approach.

5.1 *Major Conclusions and Findings*

5.1.1 *Objective 1.*

The research in Chapter 2 used multiple years of FSA 578 data to characterize grassland types and land use across Kansas and to evaluate both the dynamic and static nature of grassland type and land use over time to inform a methodology for land cover mapping of grassland types. The assessment and analysis of multiple years of FSA 578 data showed variability in degree of

data completeness, meaning that the 2004-2007 FSA 578 data were not all-inclusive of total grassland acres in Kansas. Even so, the data were sufficient to identify several regional trends in grassland type, land use, and field size. Eastern Kansas was found to have more grassland types, with the inclusion of non-native brome and fescue, a larger number of small fields, and more variability in land use, which together creates a more fragmented and complex landscape that could impact mapping grasslands in that region. Western Kansas had larger fields that primarily consisted of grazed native grassland and land enrolled in the Conservation Reserve Program (CRP), creating a comparatively simpler landscape for mapping grasslands. The inclusion of 2015 data provided a more complete representation of grassland type and land use in Kansas compared to 2004-2007 data, which possibly was the result of three new FSA programs that were implemented in the interim. These data and results will be used to inform a grassland mapping approach for Kansas, including training data allocation for image classification.

5.1.2 *Objective 2.*

The research in Chapter 3 identified the evaluated remote sensing datasets (among those tested) for separating the spectral characteristics of grassland types, determined whether land management practices impact spectral separability of grassland types, and identified what grassland hierarchy should be used for the thematic classification scheme of mapping grassland types. While these findings are limited to the study area, it is anticipated the results will be applicable to similar grassland landscapes in the Great Plains. The results show that combining the Landsat 8 multispectral, Landsat 8 NDVI, and MODIS NDVI datasets resulted in the highest JM distance statistics across all grassland class hierarchies. While the formulations underlying JM distance guarantee improved separability upon addition of more bands, the gains observed with the combined dataset were generally substantial and thus are believed to be meaningful.

Individually, the three-date Landsat multispectral data had higher JM distance statistics than either NDVI dataset. When including land use, JM distance statistics were lower within grassland types than among grassland types, with the exception of the 'left standing' category. While JM distance was high for many Level-4 pairwise comparisons, it remained relatively high in comparison for functional grassland types (i.e. lower classification levels), indicating land use does not have a highly negative impact on the spectral separability of Level-1 and Level-2 grassland classes. The results indicate that brome and fescue, both of which are non-native cool-season grasses, were not spectrally distinct and, at least when using inputs like those examined here, should be aggregated as a single class for thematic classification. Meanwhile, CRP and native grasslands demonstrated moderately high separability even though the spectral profiles appeared to largely overlap. These results suggest it may be possible to map CRP separately, or CRP could be aggregated with native grassland into a single warm-season grassland type. The temporal JM distance statistics indicate that spring and fall were more important for separating cool- and warm-season grasslands than summer where the distributions overlapped more.

5.1.3 *Objective 3.*

The research in Chapter 4 evaluated random forest (RF) models using four classification hierarchies of grasslands and four datasets in two adjacent Landsat scenes (path/rows 27/33 and 28/33) in eastern Kansas. The RF results indicate that a Level-1 classification using the combined dataset of multispectral Landsat 8, Landsat 8 NDVI and MODIS NDVI produced the lowest Out-of-bag (OOB) error of all the datasets. The OOB error for Level-2 classification was only 1-2% higher; however, the CRP class had high omission error and was largely mapped as native warm-season grassland, further supporting using the two-class Level-1 classification scheme. Predictor importance rankings for the Level-1 classification varied by path/row, and

were likely influenced by the proportion of the cool-season grassland class in each path/row. The spatial distributions of the mapped classes were qualitatively as expected, and overall accuracy levels were greater than 87% using OOB and independent test data. However, path/row 28/33 had lower user's and producer's accuracy levels for cool-season grasslands and a lower Kappa statistic. In evaluating the overlap area, there was 94% agreement between the two path/rows, but again the cool-season class in path/row 28/33 had lower user's (73%) and producer's accuracy levels (63%), which may be expected due to the small proportion of cool-season grasslands represented in that path/row. These results will be used to formulate a methodology for mapping functional grassland types at a regional scale.

5.2 *Future Research Directions*

As with all research of this nature, in-depth analysis of available datasets and processing methods exposes paths for future research; foremost among these for this research is the need for further refinement of grassland mapping protocols. The results obtained in this research demonstrate the utility of Landsat 8 and MODIS data to map grassland types using remotely sensed data; however, further exploratory analyses could provide refinements to the mapping protocol in an effort to increase the accuracy levels and/or the ability to map higher grassland hierarchies (e.g. Level-2). Future research thrusts could include the following.

1. This dissertation research indicated that the best classification approach was obtained using the combined dataset of multispectral Landsat 8, Landsat 8 NDVI, and MODIS NDVI to map a Level-1 grassland hierarchy; however, more research is needed that would maximize the utility of multispectral Landsat 8 and Landsat 8 NDVI data. Future research should test whether increasing the number of spring and fall Landsat 8 dates would improve the spectral separability and mapping accuracy levels for

Level-1 and Level-2 grassland hierarchies. Chapter 3 showed JM distance separability statistics were higher for the multispectral Landsat 8 data than MODIS time series NDVI. Furthermore, the temporal JM distance statistics indicated higher separability for several individual Landsat 8 bands and NDVI compared to individual periods of MODIS NDVI, and predictor importance estimates from Chapter 4 confirmed that similar Landsat 8 bands and Landsat NDVI dates ranked high in the RF models. The temporal (by-period) JM-distance statistics for MODIS also could be used to identify optimal acquisition dates for Landsat 8 data.

2. As described in Chapter 4, single corresponding MODIS and Landsat pixels were used for training and test data in this research so as to provide a one-to-one comparison of MODIS and Landsat 8 data. The relatively coarse spatial resolution of MODIS NDVI restricted the number of sufficiently pure pixels that could be used for training and testing the RF modeling and thus restricted the number of Landsat 8 pixels as well. Future research would evaluate the utility of a data fusion approach as used in this research versus using all available pixels from multispectral Landsat 8 and Landsat NDVI. It is possible that the inclusion of additional thousands of Landsat pixels could improve the RF OOB errors for Level-1 and Level-2 hierarchies by increasing the quantity of information available for RF training and testing, including increasing samples for minority classes and from smaller fields.
3. The results showed that in path/row 28/33 the minority class, cool-season grasslands, had relatively low user's and producer's accuracy levels. The goal of RF modeling is to minimize the overall error rate. However, when there is a large majority class, the prediction focuses more on the accuracy of the majority class at the expense of

minority class (Chen *et al.*, 2004; Mellor *et al.*, 2015). Other studies have used methods to improve the accuracy of mapping a minority class, including balanced training data, where the majority class is down-sampled or a balanced random forest is employed where a stratified bootstrap training set forces the inclusion of samples from minority classes (Chen *et al.*, 2004; Jin *et al.*, 2014). Exploring options to handle minority classes could improve the mapping accuracies for those classes without severely diminishing accuracies for larger classes.

4. The RF models for mapping grassland types shown in Chapter 4 models were developed by path/row. Future research could determine if combining path/rows for grassland mapping provided gains in accuracy levels and/or mapping efficiencies. This approach would require different image dates of Landsat 8 imagery by path/row that may or may not affect the mapping of grassland types. The Cropland Data Layer (CDL) and the National Land Cover Database (NLCD) map, both national mapping programs, utilize Landsat data and regional mapping zones that encompass multiple Landsat path/rows. The research would compare the mapping results and accuracy levels of the combined path/row RF models with individual path/row results shown in Chapter 4.
5. Inventorying, downloading, and preprocessing of remotely sensed imagery, especially when using a data fusion approach (i.e. Landsat and MODIS), is a time-consuming process. Furthermore, depending upon hardware capabilities, significant processing time is required to create and apply RF models by path/row in MATLAB. Google Earth Engine (GEE) is a cloud-based platform that utilizes an application programming interface (API) and a web-based interactive development environment

that allows access to a large catalog of imagery data and rapid cloud processing (Gorelick *et al.*, 2017). Users can utilize their own data and/or access the GEE's large catalogue of preprocessed geospatial data including remotely sensed imagery from a variety of sensors as well as environmental and climate data. Parallel processing allows large volumes of data to be analyzed rapidly. These efficiencies provide opportunity for numerous exploratory or data mining exercises, including the research topics listed above, and also provide opportunity to map larger spatial extents more rapidly. Several unsupervised and supervised image classification algorithms are available through GEE including K-means, Support Vector Machine, Decision Tree Classifier, and Random Forest (Gorelick *et al.*, 2017), and numerous studies have used GEE for land cover mapping and land cover change applications (Huang *et al.*, 2017; Patel *et al.*, 2015; Shelestov *et al.*, 2017; Simonetti *et al.*, 2015). Future work could evaluate the efficiencies gained using GEE versus the approach used in this research.

References

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley, 110*, 1-12.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27.

Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., . . . Zheng, Y. (2017). Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sensing of Environment, 202*, 166-176.

Jin, H., Stehman, S. V., & Mountrakis, G. (2014). Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado. *International Journal of Remote Sensing, 35*(6), 2067-2081.

Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 155-168.

Patel, N. N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F. R., . . . Trianni, G. (2015). Multitemporal settlement and population mapping from Landsat using Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation, 35*, 199-208.

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring Google earth engine platform for Big Data Processing: Classification of multi-temporal satellite imagery for crop mapping. *Frontiers in Earth Science, 5*, 17.

Simonetti, D., Simonetti, E., Szantoi, Z., Lupi, A., & Eva, H. (2015). First results from the phenology-based synthesis classifier using Landsat 8 imagery. *IEEE Geoscience and remote sensing letters, 12*(7), 1496-1500.

6 Acronyms and Definitions

ASD: Agricultural Statistics District

BACC–FLUD: Biofuels and Climate Change - Farmers' Land Use Decisions

C-CAP: Coastal Change Analysis Program

CDL: Cropland Data Layer

CLU: A Common Land Unit is the smallest unit of land that has a permanent, contiguous boundary, a common land cover and land management, a common owner and a common producer in agricultural land associated with USDA farm programs. CLU boundaries typically are delineated from relatively permanent features such as fence lines, roads, and/or waterways.

CRP: Conservation Reserve Program

DTC: decision tree classifier

EE: EarthExplorer

ELAP: Farm-Raised Fish Program

ESRI: Environmental Systems Research Institute

ETM: Enhanced Thematic Mapper

EVI: Enhanced Vegetation Index

GEE: Google Earth Engine

FIPS: Federal Information Processing Standard – uniquely identifies states and counties in the United States with a two and three digit code, respectively.

FSA: Farm Service Agency

GAP: Gap Analysis Program

GRP: Grassland Reserve Program

JM: Jeffries-Matusita

KARS: Kansas Applied Remote Sensing Program

KSU: Kansas State University

LFP: Livestock Forage Program

LIP: Livestock Indemnity Program

MDA: Mean decrease in accuracy

MODIS: Moderate-Resolution Imaging Spectroradiometer

MOU: Memorandum of Use

MRLC: Multi-Resolution Land Characteristics

NASS: National Agricultural Statistics Service

NDVI: Normalized Difference Vegetation Index

NIR: Near-infrared

NLCD: National Land Cover Dataset

OLI: Operational Land Imager

OOB: Out-of-bag

RF: Random Forest

SCTF: A concatenation of State, County, Tract and Field attributes from the FSA and CLU data to create a unique identifier for linking field-level data.

SVM: Support vector machine

SWIR: Short-wave infrared

TM: Thematic Mapper

Tract Number: A tract of land is generally a single field or multiple fields connected in the same section of a township with common ownership. One tract of land could have any number of fields. Identifies a tract that belongs to a farm number.

USDA: United States Department of Agriculture

USGS: United States Geological Survey

WRS: Worldwide Reference System

References

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: from observation to management–a review. *Journal of Plant Ecology*. doi:10.1093/jpe/rtw005

Banks, E. B. (2012). *Kansas Conservation Reserve Program Technical Guidance Number 82*. Retrieved from NRCS: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/ks/programs/?cid=nrcs142p2_032831

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

Breiman, L. (1996). Out-of-bag estimation, ftp. stat. berkeley. edu/pub/users/breiman. *OOBestimation. ps, 199*(6).

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and decision trees. *Wadsworth, Belmont, 378*.

Brown, J. C., Kastens, J. H., Coutinho, A. C., de Castro Victoria, D., & Bishop, C. R. (2013). Classifying multiyear agricultural land use data from Mato Grosso using time series MODIS vegetation index data. *Remote Sensing of Environment, 130*, 39-50.

Cánovas-García, F., & Alonso-Sarría, F. (2015). Optimal Combination of Classification Algorithms and Feature Ranking Methods for Object-Based Classification of Submeter Resolution Z/I-Imaging DMC Imagery. *Remote Sensing, 7*(4), 4651.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley, 110*, 1-12.

Colditz, R. R. (2015). An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing, 7*(8), 9655-9681.

Cooper, G. (2015). University of Wisconsion Study Based on Shaky Foundation of Faulty Data and Conclusion. Retrieved from http://ethanolrfa.org/2015/04/university-of-wisconsin-study-based-on-shaky-foundation-of-faulty-data-and-conclusions/

Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V. E., . . . Michael, F. (2015). Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation, 37*, 7-16. doi:http://dx.doi.org/10.1016/j.jag.2014.11.005

Davidson, A., & Csillag, F. (2003). A comparison of three approaches for predicting C4 species cover of northern mixed grass prairie. *Remote Sensing of Environment, 86*(1), 70-82. doi:http://dx.doi.org/10.1016/S0034-4257(03)00069-5

Davis, S. M., Landgrebe, D. A., Phillips, T. L., Swain, P. H., Hoffer, R. M., Lindenlaub, J. C., & Silva, L. F. (1978). Remote sensing: the quantitative approach. *New York, McGraw-Hill International Book Co., 1978. 405 p., 1*.

Drummond, M. A. (2007). Regional dynamics of grassland change in the western Great Plains. *Great Plains Research*, 133-144.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*: CRC press.

Egbert, S. L., Lee, R. Y., Price, K. P., Boyce, R., & Nellis, M. D. (1998). Mapping conservation reserve program (CRP) grasslands using multi-seasonal thematic mapper imagery. *Geocarto International, 13*(4), 17-24.

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*(2), 89-93. doi:http://dx.doi.org/10.1016/j.ecoinf.2007.03.009

Franke, J., Keuck, V., & Siegert, F. (2012). Assessment of grassland use intensity by remote sensing to support conservation schemes. *Journal for Nature Conservation, 20*(3), 125-134. doi:http://dx.doi.org/10.1016/j.jnc.2012.02.001

Gao, J., Sheshukov, A. Y., Yen, H., Kastens, J. H., & Peterson, D. L. (2017). Impacts of incorporating dominant crop rotation patterns as primary land use change on hydrologic model performance. *Agriculture, ecosystems & environment, 247*, 33-42. doi:https://doi.org/10.1016/j.agee.2017.06.019

Gelfand, I., Zenone, T., Jasrotia, P., Chen, J., Hamilton, S. K., & Robertson, G. P. (2011). Carbon debt of Conservation Reserve Program (CRP) grasslands converted to bioenergy production. *Proceedings of the National Academy of Sciences, 108*(33), 13864-13869.

Gibson, D. J. (2009). *Grasses and grassland ecology*: Oxford University Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27.

Gu, Y., & Wylie, B. K. (2015). Developing a 30-m grassland productivity estimation map for central Nebraska using 250-m MODIS and 30-m Landsat-8 observations. *Remote Sensing of Environment, 171*, 291-298.

Guo, X., Price, K. P., & Stiles, J. (2003). Grasslands Discriminant Analysis Using Landsat TM Single and Multitemporal Data. *Photogrammetric Engineering & Remote Sensing, 69*(11), 1255-1262. doi:10.14358/PERS.69.11.1255

Guo, X., Price, K. P., & Stiles, J. M. (2000). Biophysical and Spectral Characteristics of Cool- and Warm-Season Grasslands under Three Land Management Practices in Eastern Kansas. *Natural Resources Research, 9*(4), 321-331. doi:10.1023/a:1011513527965

Halabuk, A., Mojses, M., Halabuk, M., & David, S. (2015). Towards Detection of Cutting in Hay Meadows by Using of NDVI and EVI Time Series. *Remote Sensing, 7*(5), 6107.

Hellerstein, D. M. (2017). The US conservation reserve program: the evolution of an enrollment mechanism. *Land Use Policy, 63*, 601-610.

Hendricks, N. P., & Er, E. (2018). Changes in cropland area in the United States and the role of CRP. *Food Policy, 75*(C), 15-23.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., . . . Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens, 81*(5), 345-354.

Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., . . . Zheng, Y. (2017). Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sensing of Environment, 202*, 166-176.

Hughes, J. P., Robel, R. J., Kemp, K. E., & Zimmerman, J. L. (1999). Effects of Habitat on Dickcissel Abundance and Nest Success in Conservation Reserve Program Fields in Kansas. *The Journal of Wildlife Management, 63*(2), 523-529. doi:10.2307/3802638

Janitza, S., & Hornung, R. (2018). On the overestimation of random forest's out-of-bag error. *PloS one, 13*(8), e0201904.

Jensen, J. R. (1983). Biophysical remote sensing. *Annals of the Association of American Geographers, 73*(1), 111-132.

Jin, H., Stehman, S. V., & Mountrakis, G. (2014). Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado. *International Journal of Remote Sensing, 35*(6), 2067-2081.

Johnston, C. A. (2014). Agricultural expansion: land use shell game in the US Northern Plains. *Landscape ecology, 29*(1), 81-95.

KARS (Cartographer). (2008). 2005 Kansas Land Cover Patterns. Retrieved from http://kars.ku.edu/geonetwork/srv/en/main.home

KARS (Cartographer). (2017). 2015 Kansas Land Cover Patterns. Retrieved from http://kars.ku.edu/geonetwork/srv/en/main.home

Kastens, J. H., Brown, J. C., Coutinho, A. C., Bishop, C. R., & Esquerdo, J. C. D. (2017). Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PloS one, 12*(4), e0176168.

Kennedy, S. S. (1999). *Spectral discrimination of irrigated/non-irrigated crop types using landsat thematic mapper imagery in Haskell county, Kansas.* University of Kansas, Geography.

KSU. (2018). AgReport. Retrieved from http://www.k-state.edu/agreport/faculty/fall2018/administrativemerger.html

Lamond, R. E., Fritz, J. O., & Ohlenbusch, P. D. (1992). *Smooth Brome Production and Utilization*. Manhattan, Kansas: Kansas State University.

Lauver, C. L., & Whistler, J. L. (1993). A hierarchical classification of Landsat TM imagery to identify natural grassland areas and rare species habitat. *Photogrammetric Engineering and Remote Sensing, 59*(5), 627-634.

Lin, X., Harrington, J., Ciampitti, I., Gowda, P., Brown, D., & Kisekka, I. (2017). Kansas Trends and Changes in Temperature, Precipitation, Drought, and Frost-Free Days from the 1890s to 2015. *Journal of Contemporary Water Research & Education, 162*(1), 18-30. doi:doi:10.1111/j.1936-704X.2017.03257.x

Masialeti, I., Egbert, S., & Wardlow, B. D. (2010). A comparative analysis of phenological curves for major crops in Kansas. *GIScience & Remote Sensing, 47*(2), 241-259.

Maxwell, S. K., & Sylvester, K. M. (2012). Identification of "ever-cropped" land (1984–2010) using Landsat annual maximum NDVI image composites: Southwestern Kansas case study. *Remote Sensing of Environment, 121*, 186-195.

McInnes, W. S., Smith, B., & McDermid, G. J. (2015). Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(4), 1395-1403. doi:10.1109/JSTARS.2015.2416713

Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 155-168.

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing, 7*(7), 8489-8515.

Mosiman, B. N. (2003). *Exploring multi-temporal and transformation methods for improved cropland classification using Landsat Thematic Mapper imagery.* University of Kansas, Geography.

Owensby, C. E. (1993). *Introduction to range management*: Custom Academic Publishing Company.

Patel, N. N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F. R., . . . Trianni, G. (2015). Multitemporal settlement and population mapping from Landsat using Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation, 35*, 199-208.

Peterson, D., Price, K., & Martinko, E. (2002a). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing, 23*(23), 5015-5030.

Peterson, D. L., Egbert, S. L., & Martinko, E. A. (2017). *2015 Kansas Land Cover Patterns Phase I: Final Report.* Lawrence, Kansas: Kansas Biological Survey, University of Kansas.

Peterson, D. L., Price, K. P., & Martinko, E. A. (2002b). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing, 23*(23), 5015-5030.

Peterson, D. L., Price, K. P., & Martinko, E. A. (2002c). Investigating grazing intensity and range condition of grasslands in northeastern Kansas using Landsat thematic mapper data. *Geocarto International, 17*(4), 15-26.

Peterson, D. L., Whistler, J. L., Egbert, S. L., & Martinko, E. A. (2008). *2005 Kansas Land Cover Patterns: Phase II-Final Report*. Retrieved from Kansas Biological Survey:

Peterson, D. L., Whistler, J. L., Lomas, J. M., Dobbs, K. E., Jakubauskas, M. E., Egbert, S. L., & Martinko, E. A. (2005). *2005 Kansas Land Cover Patterns Phase I: Final Report* (150). Retrieved from Kansas Biological Survey, University of Kansas:

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32*(15), 4407-4429.

Porter, T. F., Chen, C., Long, J. A., Lawrence, R. L., & Sowell, B. F. (2014). Estimating biomass on CRP pastureland: A comparison of remote sensing techniques. *Biomass and Bioenergy, 66*, 268-274. doi:http://dx.doi.org/10.1016/j.biombioe.2014.01.036

Price, K. P., Guo, X., & Stiles, J. M. (2002a). Comparison of Landsat TM and ERS-2 SAR data for discriminating among grassland types and treatments in eastern Kansas. *Computers and Electronics in Agriculture, 37*(1), 157-171.

Price, K. P., Guo, X., & Stiles, J. M. (2002b). Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing, 23*(23), 5031-5042. doi:10.1080/01431160210121764

Reynolds, R. E., Shaffer, T. L., Renner, R. W., Newton, W. E., & Bruce, D. J. B. (2001). Impact of the Conservation Reserve Program on Duck Recruitment in the U.S. Prairie Pothole Region. *The Journal of Wildlife Management, 65*(4), 765-780. doi:10.2307/3803027

Ribaudo, M. O., Colacicco, D., Langner, L. L., Piper, S., & Schaible, G. D. (1990). Natural Resources and Users Benefit from the Conservation Reserve Program. Agricultural Economic Report 627. *US Department of Agriculture, Economic Research Service, Washington, DC.*(Report 627).

Richards, J., & Jia, X. (1999). Remote sensing digital image analysis: An introduction Springer-Verlag. *Berlin Heidelberg*.

Riffell, S., Scognamillo, D., & Burger, L. W. (2008). Effects of the Conservation Reserve Program on northern bobwhite and grassland birds. *Environmental Monitoring and Assessment, 146*(1), 309-323. doi:10.1007/s10661-007-0082-8

Risser, P. G. (1988). *Diversity in and among grasslands*: National Academy of Sciences Washington, DC.

Robertson, K. R., Anderson, R. C., & Schwartz, M. W. (1997). The tallgrass prairie mosaic *Conservation in highly fragmented landscapes* (pp. 55-87): Springer.

Robertson, K. R., & Schwartz, M. (1994a). Prairies. *The changing Illinois environment: critical trends, 3*, 1-32.

Robertson, K. R., & Schwartz, M. W. (1994b). Prairies. *The changing Illinois environment: critical trends, 3*, 1-32.

Samson, F. B., Knopf, F. L., & Ostlie, W. R. (2004). Great Plains ecosystems: past, present, and future. *Wildlife Society Bulletin, 32*(1), 6-15.

Schuster, C., Schmidt, T., Conrad, C., Kleinschmit, B., & Förster, M. (2015). Grassland habitat mapping by intra-annual time series analysis – Comparison of RapidEye and TerraSAR-X satellite data. *International Journal of Applied Earth Observation and Geoinformation, 34*, 25-34. doi:http://dx.doi.org/10.1016/j.jag.2014.06.004

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring Google earth engine platform for Big Data Processing: Classification of multi-temporal satellite imagery for crop mapping. *Frontiers in Earth Science, 5*, 17.

Shoup, D., Kilgore, G. L., & Brazle, F. K. (2010). *Tall Fescue Production and Utilization*. Manhattan, Kansas: Kansas State University.

Simonetti, D., Simonetti, E., Szantoi, Z., Lupi, A., & Eva, H. (2015). First results from the phenology-based synthesis classifier using Landsat 8 imagery. *IEEE Geoscience and remote sensing letters, 12*(7), 1496-1500.

Song, X., Fan, G., & Rao, M. (2005). Automatic CRP mapping using nonparametric machine learning approaches. *IEEE Transactions on Geoscience and Remote Sensing, 43*(4), 888-897. doi:10.1109/TGRS.2005.844031

Stubbs, M. (2018). Agricultural Disaster Assistance, Congressional Research Service Report. 16.

Swain, P., & King, R. (1973). Two effective feature selection criteria for multispectral remote sensing.

Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K., & DeJong, D. D. (1997). NDVI, C3 and C4 production, and distributions in Great Plains grassland land cover classes. *Ecological applications, 7*(1), 59-78.

Umphlett, N. (2013). *June 2013 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

Umphlett, N. (2014). *July 2014 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

Umphlett, N. (2015a). *July 2015 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

Umphlett, N. (2015b). *June 2015 Climate Summary*. Retrieved from Lincoln, Nebraska: https://hprcc.unl.edu/climatesummaries.php

USDA, F. (2012). FSA Bulletin: CRP Emergency Haying and Grazing Released in 91 Kansas Counties [Press release]. Retrieved from https://content.govdelivery.com/accounts/USFSA/bulletins/496494

USDA, F. S. A. (2013a). *FSA Handbook for State and County Offices: Acrage and Compliance Determinations*. (2-CP (Revision 15), Amendment 58). Washington D.C. Retrieved from https://www.fsa.usda.gov/Internet/FSA_File/2cp15-76.pdf.

USDA, F. S. A. (2017). CRP Enrollment and Rental Payments by County, 1986-2017. Retrieved November 13, 2017 https://www.fsa.usda.gov/programs-and-services/conservation-programs/reports-and-statistics/conservation-reserve-program-statistics/index

USDA, N. R. C. S. (2009). Fact Sheet: Grassland Reserve Program.

USDA, N. R. C. S. (2013b). *Summary Report: 2010 National Resources Inventory*. Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa Retrieved from http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1167354.pdf.

USGS. (2018). Landsat Missions. Retrieved from https://landsat.usgs.gov/what-are-best-spectral-bands-use-my-study

Van Pelt, W. E., Kyle, S., Pitman, J., Klute, D., Beauprez, G., Schoeling, D., . . . Haufler, J. (2013). The lesser prairie-chicken range-wide conservation plan. *Western Association of Fish and Wildlife Agencies, Cheyenne, Wyoming, USA*.

Wang, C., Jamison, B. E., & Spicci, A. A. (2010). Trajectory-based warm season grassland mapping in Missouri prairies with multi-temporal ASTER imagery. *Remote Sensing of Environment, 114*(3), 531-539.

Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. *Remote Sensing of Environment, 112*(3), 1096-1116.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment, 108*(3), 290-310.

Weaver, J. E. (1954). *North American prairie / J.E. Weaver*. Lincoln Neb: Johnsen.

Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., & Coulston, J. (2014). The multi-resolution land characteristics (MRLC) consortium—20 years of development and integration of USA national land cover data. *Remote Sensing, 6*(8), 7424-7441.

Wickham, J., Stehman, S. V., Gass, L., Dewitz, J. A., Sorenson, D. G., Granneman, B. J., . . . Baer, L. A. (2017). Thematic accuracy assessment of the 2011 national land cover database (NLCD). *Remote Sensing of Environment, 191*, 328-341.

Wright, C. K., & Wimberly, M. C. (2013). Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences, 110*(10), 4134-4139. doi:10.1073/pnas.1215404110

Wu, J., & Weber, B. (2012). Implications of a reduced conservation reserve program. The conservation crossroads of agriculture. *Council on Food, Agriculture and Resource Economics. Washington, D.C., USA.*

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment, 122*, 2-10.

Zha, Y., Gao, J., Ni, S., Liu, Y., Jiang, J., & Wei, Y. (2003). A spectral reflectance-based approach to quantification of grassland cover from Landsat TM imagery. *Remote Sensing of Environment, 87*(2-3), 371-375.