2019

Augmenting Land Cover/Land Use Classification by Incorporating Information from Land Surface Phenology: An Application to Quantify Recent Cropland Expansion in South Dakota

Lan Hoang Nguyen

AUGMENTING LAND COVER/LAND USE CLASSIFICATION BY

INCORPORATING INFORMATION FROM LAND SURFACE PHENOLOGY: AN

APPLICATION TO QUANTIFY RECENT CROPLAND EXPANSION IN SOUTH

DAKOTA

BY

LAN HOANG NGUYEN

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Geospatial Science and Engineering

Specialization in Remote Sensing Geography

South Dakota State University

2019

AUGMENTING LAND COVER/LAND USE CLASSIFICATION BY INCORPORATING

INFORMATION FROM LAND SURFACE PHENOLOGY: AN APPLICATION TO

QUANTIFY RECENT CROPLAND EXPANSION IN SOUTH DAKOTA

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Geospatial Science and Engineering and it is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Geoffrey M. Henebry, Ph.D.   Date: 11JUN2019

Dissertation Advisor

Bob Watrel, Ph.D.   Date:

Head, Department of Geography and Geospatial Sciences

Kinchel Doerner, Ph.D.

Dean, Graduate School   Date:

I dedicate this dissertation to my lovely wife, Quynh Anh Truong, and my son, Andy D. Truong-Nguyen.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

ABBREVIATIONS

AA	Accuracy Assessment
AEA	Albert Equal Area
AGDD	Accumulate Growing Degree-Days
ARD	Analysis Ready Data
AVHRR	Advanced Very High Resolution Radiometer
BRDF	Bidirectional Reflectance Distribution Function
CDL	Cropland Data Layer
CI	Confidence Interval
CONUS	Contiguous United States
CxQ	Convex Quadratic
DOY	Day of Year
ETM+	Enhanced Thematic Mapper plus
EVI	Enhanced Vegetation Index
EVI2	2-band Enhanced Vegetation Index
GDD	Growing Degree-Days
GI	Gini Importance
HLS	Harmonized Landsat Sentinel-2
HPLM	Hybrid Piecewise Logistic Model
LCC	Land Capability Class
LCLU	Land Cover/Land Use
LCMAP	Land Change Monitoring, Assessment, and Projection
LSP	Land Surface Phenology

LST	Land Surface Temperature
MODIS	MODerate resolution Imaging Spectroradiometer
MSI	Multi-Spectral Instrument
NAD83	North American Datum 1983
NAIP	National Agriculture Imagery Program
FSA	Farm Service Agency
NASS	National Agricultural Statistics Service
NBAR	Nadir BRDF-Adjusted Reflectance
NDVI	Normalized Difference Vegetation Index
NGP	Northern Great Plains
NIR	Near Infrared
NLCD	National Land Cover Database
NRCS	Natural Resources Conservation Service
NRI	National Resources Inventory
OA	Overall Accuracy
OLI	Operational Land Imager
PA	Producer's Accuracy
RFC	Random Forest Classifier
SLC	Scanline Corrector
SR	Surface Reflectance
SSURGO	Soil Survey Geographic Database
SWIR	Shortwave Infrared
TM	Thematic Mapper

UA	User's Accuracy
USDA	United States Department of Agriculture
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
V-NIR-SWIR	Visible-Near Infrared-Shortwave Infrared
WELD	Web-Enabled Landsat Data
WGS84	World Geodetic System 1984

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

AUGMENTING LAND COVER/LAND USE CLASSIFICATION BY
INCORPORATING INFORMATION FROM LAND SURFACE PHENOLOGY: AN
APPLICATION TO QUANTIFY RECENT CROPLAND EXPANSION IN SOUTH
DAKOTA

LAN HOANG NGUYEN

2019

Understanding rapid land change in the U.S. NGP region is not only critical for management and conservation of prairie habitats and ecosystem services, but also for projecting production of crops and biofuels and the impacts of land conversion on water quality and rural transportation infrastructure. Hence, it raises the need for an LCLU dataset with good spatiotemporal coverage as well as consistent accuracy through time to enable change analysis. This dissertation aims (1) to develop a novel classification method, which utilizes time series images from comparable sensors, from the perspective of land surface phenology, and (2) to apply the land cover/land use dataset generated from the phenometrically-based classification approach to quantify crop expansion in South Dakota.

A novel classification approach from the perspective of land surface phenology (LSP) uses rich time series datasets. First, surface reflectance products at 30 m spatial resolution from Landsat Collection-1, its newer structure—Landsat Analysis Ready Data, and the Harmonized Landsat Sentinel-2 (HLS) data are used to construct vegetation index time series, including the Enhanced Vegetation Index (EVI), and the 2-band EVI (EVI2), and various spectral variables (spectral band and normalized ratio composites). MODIS

Level-3 Land Surface Temperature & Emissivity 8-day composite products at 1 km spatial resolution from both the Aqua and Terra satellites are used to compute accumulated growing degree-days (AGDD) time series. The EVI/EVI2 and AGDD time series are then fitted by two different land surface phenology models: the Convex Quadratic model and the Hybrid Piecewise Logistic Model. Suites of phenometrics are derived from the two LSP models and spectral variables and input to Random Forest Classifiers (RFC) to map land cover of sample areas in South Dakota. The results indicate that classifications using only phenometrics can accurately map major crops in the study area but show limited accuracy for non-vegetated land covers. RFC models using the combined spectral-phenological variables can achieve higher accuracies than those using either spectral variables or phenometrics alone, especially for the barren/developed class. Among all sampling designs, the "same distribution" models—proportional distribution of the sample is like proportional distribution of the population—tends to yield best land cover prediction. A "same distribution" random sample dataset covering approximately 0.25% or more of the study area appears to achieve an accurate land cover map.

To characterize crop expansion in South Dakota, a trajectory-based analysis, which considers the entire land cover dataset generated from the LSP-based classifications, is proposed to improve change detection. An estimated cropland expansion of 5,447 km2 (equivalent to 14% of the existing cropland area) occurred between 2007 and 2015, which matches more closely the reports from the National Agriculture Statistics Service—NASS (5,921 km2) and the National Resources Inventory—NRI (5,034 km2) than an estimation from a bi-temporal change approach (8,018 km2). Cropland gains were mostly concentrated in 10 counties in northern and central South Dakota. An evaluation of land

suitability for crops using the Soil Survey Geographic Database—SSURGO indicates a

scarcity in high-quality arable land available for cropland expansion.

CHAPTER 1

INTRODUCTION

1.1   Background

Despite low population, the US Northern Great Plains (NGP) region has been undergoing substantial land cover/land use change (LCLUC) over the past two decades. Federally mandated policies spurring the demand for biofuels, particularly corn-based ethanol and biodiesel from soybean (*Schnepf & Yacobucci, 2013*), coupled with federally subsidized crop insurance led to a tripling in market prices for corn and soybean between 2002 and 2012 (*Johnston, 2013*). Increases in commodity prices resulted in one of the most significant land change episodes in recent US history: the conversion of grasslands and wetlands to croplands, primarily in the eastern Dakotas, western Minnesota, and southern Iowa (*Faber et al., 2012; Johnston, 2013, 2014; Wright & Wimberly, 2013*). *Johnston (2013)* estimated wetlands losses of 5,000-6,000 ha per year since 1979 due to cropland expansion. *Wright & Wimberly (2013)* estimated a net loss of 530,000 ha in grass-dominated lands in the Western Corn Belt from 2006 to 2011 based on analysis of the USDA's Cropland Data Layer (CDL). Remaining intact grassland habitats are also at risk of conversion to cropland under projected future demands for food and fuel since ~84% of the area is privately owned (*WWF, 2016*).

Several efforts have been made to characterize land change dynamics in the Great Plains. Much of the extant literature has used either the National Resources Inventory (NRI), a periodic statistical survey of land use and natural resource conditions on US non-federal lands (*Claassen et al., 2011; Rashford et al., 2011*) that lacks spatial detail, or post-classification change detection methods using publicly available LCLU databases, especially the USDA Cropland Data Layer (*Faber et al., 2012; Wright & Wimberly,*

2013a,b; Johnston, 2013, 2014; Lark et al., 2015; WWF, 2016) due to its finer spatiotemporal resolution and high number of thematic classes.

The use of CDL can be justified by its high overall accuracy of 85% to 95% for major crops, and often 97% producer's and user's accuracies for corn and soybean (*Boryan et al., 2011*). However, it is important to realize that the CDL is meant to monitor agricultural land cover annually; and it has undergone substantial methodological changes over time (*Mueller & Seffrin, 2006; Boryan et al., 2011*). In addition, the conventional classification methods, such as ones first applied to the CDL and the more general National Land Cover Database (NLCD), were developed in an era of data scarcity and limited computational power and thus focused on using just a few cloud/snow-free scenes.

Understanding rapid land change is not only critical for management and conservation of prairie habitats and ecosystem services, but also for projecting production of crops and biofuels and the impacts of land conversion on rural infrastructure, such as roads and water quality. Hence, there is a need for a land cover/land use dataset with good spatiotemporal coverage as well as consistent accuracy through time. In the current era of abundant earth observations, a better approach would be able to take advantage of all available useful data to accurate map different land cover types.

Late fall 2016, the U.S. Geological Survey (USGS) announced a science initiative called Land Change Monitoring, Assessment, and Projection (LCMAP) to fulfill the demand of "even higher quality data, additional land cover and land change variables, more detailed legends, and most importantly, more frequent land change information" (*Young, 2017*). LCMAP utilizes the global Landsat archive that dates back to 1972 to characterize historical, near real-time land change at any location across the entire Landsat record using

the Continuous Change Detection and Classification algorithm (*Zhu & Woodcock, 2014*). Although the future of LCMAP is bright, it will be quite some time before those data become generally available. Furthermore, LCMAP is restricted to the few classes of Anderson Level 1 plus a "transition" class; thus, it will not be suitable for studies needing higher levels of detail, for example, monitoring crop types and rotation.

Shifts in land cover or environmental conditions, management practice, disturbance may lead to a spatiotemporal variation of land surface phenology (LSP), *i.e.*, seasonal patterns of reflectance from the vegetated land surface as observed using remote sensing (*Henebry & de Beurs, 2013*). Several methods have been used to simulate the temporal variation of vegetation index time series (*Jönsson & Eklundh, 2003*; *Zhang et al., 2003; de Beurs & Henebry, 2004; Beck et al., 2006; Qader et al., 2016; Roy & Yan, 2018*). Among those, the double logistic curve (and its modifications) is one of the more commonly used fitting methods. Application of the logistic curve seems better suited to landscapes dominated by woody vegetation (*Zhang et al., 2003; Ahl et al., 2006; Beck et al., 2006; Baumann et al, 2017*). However, characterizing phenology for herbaceous vegetation is more difficult than for woody vegetation due to strong interannual variation in grasslands and croplands (*Schwartz & Reed, 1999*). Another LSP fitting approach originates from traditional phenological models that relate to the progression of thermal time during the growing season to events in plant development. Studies have shown that the temporal development of remotely sensed vegetation indices in temperate and boreal ecosystems can be well approximated as a quadratic function of accumulated growing degree-days (AGDD) because AGDD captures well the seasonal course of insolation at middle to higher

latitudes (*de Beurs & Henebry, 2004; Henebry & de Beurs, 2013; Krehbiel & Henebry 2016; Krehbiel et al., 2016, 2017*).

Modeled vegetation index time series can provide a larger number of sequentially related predictor variables to be exploited by classification techniques. *Zhong et al. (2011)* mapped multiple crops in San Joaquin Valley, California using phenometrics derived by fitting the double asymmetric sigmoid functions to the smoothed 8-d NDVI at 250-m resolution calculated from MOD09Q1 product. The classification accuracies were between 70% to 80% for all study years. *Xue et al. (2014)* computed "phenological markers" (timings of phenological events) from the seasonal and trend components of the 16-d EVI/NDVI time series at 250-m from MOD13Q1 product. Using only those time markers, the authors were able to map land covers in Nanjing City, China using different classifiers with overall accuracies of between 88% and 98%. Also using MODIS EVI/NDVI time series, *Yan et al. (2015)* calculated various statistical composites as well as amplitude and phase information of harmonic components derived from Fourier transform to be used in LCLU classification. The entire Northeast China was classified with an overall accuracy of 84% and kappa statistics of 0.79. Using similar data and method to *Zhong et al. (2011)*, *Qader et al. (2016)* accurately mapped broad dominant cover classes in Iraq and characterized changes from 2002 to 2012. The overall accuracy for 2003, 2006, and 2013 classifications were 94%, 91% (using Google Earth images), and 88.5% (ground truths), respectively. Both *Jia et al. (2014)* and *Kong et al. (2016)* attempted to produce land cover maps at finer resolution by fusing Landsat NDVI at 30-m and GF-1 NDVI at 16-m resolution, respectively, with MODIS NDVI at 250-m resolution. Phenometrics were then computed from the fused datasets to be used in land cover classifications with spectral data.

Both studies found that phenometrics significantly improved classification accuracy compared to those using only spectral data (overall accuracy increased approximate 10%). So far, all extant studies were based on MODIS surface reflectance or vegetation index time series which have fine temporal but coarse spatial resolution (250 m), and none of those study used thermal data. A convex quadratic (CxQ) LSP model linking a vegetation index, such as the NDVI or the enhanced vegetation index (EVI), to thermal time as measured by accumulated growing degree-days (AGDD) was first applied to compare spring green-up dynamics before and after the collapse of the Soviet Union (*de Beurs & Henebry, 2004*) to detect significant change in a noisy AVHRR time series. *Krehbiel et al. (2016, 2017)* using a similar approach with Landsat (for NDVI) and MODIS (for AGDD) time series detected the conversion of croplands to residential areas near Omaha, NE and Minneapolis-St. Paul, MN. Although *de Beurs & Henebry (2004)* and *Krehbiel et al. (2016, 2017)* successfully detected land cover changes in their study area through significant shifts in phenological metrics, these studies did not quantify areal changes or generate spatially explicit LCLU change maps.

1.2   Research questions

Understanding rapid land change is not only critical for management and conservation of prairie habitats and ecosystem services, but also for projecting production of crops and biofuels and the impacts of land conversion on water quality and rural transportation infrastructure. Hence, it raises the need for an LCLU dataset with good spatiotemporal coverage as well as consistent accuracy through time to enable change analysis. There are four major questions I want to address:

First, *how well does land cover mapping perform if phenological metrics alone are used for input to the classification algorithm*? Multi-temporal classification has proved superior to classification relying on just a few scenes (*Franklin et al., 2015*). Therefore, a current rapid increase of accessible Earth Observation data (particularly from the Landsat archive and its augmentation by newer sensors, such as Sentinel 2A and 2B) coupled with improved computing power is leading to the emergence of methods for generation annual land cover products from time series data. Generally, vegetation index (VI) time series at different pixels are often not observed at the same set of days due to variation in data quality. This limitation prevents a direct comparison between annual VI patterns to distinguish between land cover types. Thus, a common approach to utilize multi-temporal images is to generate composited images (*Hansen et al., 2011; Zhang & Roy, 2017; Teluguntla et al., 2018*). However, since the composites images may still contain gaps (if the composited period is too short) and mixed spectral signals (from multiple observations), the use of composited images in LCLU classification is generally limited to one of two options: (1) use multi-year data of short-period composites (*e.g.,* monthly) to generate a single land cover map with fairly high details (*e.g.,* NLCD Level-2 land cover/land use classes; *Zhang & Roy, 2017*), or (2) use long-period composites (*e.g.,* quarterly) or statistics from multiple short-period composites to generate annual land cover map with only few broad categories (*Hansen et al., 2011; Teluguntla, 2018*).

To exploit multi-temporal data more fully, I propose a novel approach to map land cover map accurately in a timely manner using land surface phenology modeling. First, I filtered the entire annual vegetation index and AGDD time series using simple functional forms (*e.g.*, Convex Quadratic Model; *de Beurs & Henebry 2004* or Hybrid Piecewise

Logistic Model; *Zhang, 2015)* so that land surface phenology at every pixel was described by a set of phenological metrics (phenometrics). Those phenometrics then were used as the only input for land cover/land use classification. I hypothesized that classification using only phenometrics could produce consistent and accurate land cover maps. Output of this study was cross-compared with the CDL. The newly generated land cover maps were also validated using point reference data and compared with reports from the U.S. Department of Agriculture.

Second, *what is the rate and spatial pattern of crop expansion in South Dakota over the past decade?* Agriculture is the leading industry in South Dakota, contributing approximate $21 billion to the State's economy each year (about 20% of our state's economic activity, *SDDA, 2018*). Over the past decade, the growing demand for biofuel production increased agricultural activities in South Dakota, leading to the conversion of grassland to cropland. Although shifts in land cover may impact a wide range of stakeholders and interest groups, and society in general (*Reitsma et al., 2014*), there is not yet a comprehensive land change analysis available for South Dakota. Most extant studies take a "bi-temporal snapshot" approach (*Decision Innovation Solutions, 2013; Wright & Wimberly, 2013; Reitsma et al., 2014*) that only compares data between two isolated points in time and disregards intermediate-year data. The bi-temporal approach does not capture the regular rotation of lands into and out of cultivation and the approach can be affected by misclassification error at either or both time points. Thus, bi-temporal change detection can potentially inflate (or deflate) reported rates of conversion. On the other hand, *Lark et al. (2015)* and *Arora & Wolter (2018*) provided a continuous picture of land change in South Dakota. However, while *Lark et al. (2015)* only examined changes from 2008 to 2012, the

long-term analysis from *Arora & Wolter (2018)* covered only a portion of the state (Landsat WRS-2 Path 30, Rows 28-29). Comprehensive land change analysis in South Dakota over the past decade will be critical for management and conservation of prairie habitats and ecosystem services, as well as for projecting production of crops and of biofuels and assessing the impacts of conversion on rural infrastructure, such as roads and water use and water quality.

LSP-based classification is a simple and consistent way to map land cover/land use. However, from initial results, three major challenges of the LSP-based classification were identified. First, the proposed method only performs well with vegetated land covers, especially crops. Estimated areas for non-crop covers can be unreliable (*e.g.*, overestimation of grassland and underestimation of urban/built up area). A better land cover map would enable detection of changes not in only cropland but in other cover types. Second, LSP model fitting may fail due to few valid observations being available as a result of obscuring cloud cover and/or sensor artifacts. Those failed models create gaps in the outputs preventing direct comparison of land cover maps for change detection. Although the gaps can be filled using temporal and/or spatial context, the filling process itself can be complicated. The gap filling often works only for small gaps, and/or when classification task only retrieves a few broad land cover classes (*e.g.*, I only mapped three land covers in South Dakota: cropland, grassland, and others). For example, if before- and after-year covers are crop, current-year cover is very likely crop. However, if before- and after-year covers are corn, it still not enough information to confirm that current-year cover is corn since rotation with soybean is likely but not guaranteed. Finally, the classification accuracy may vary due to the chosen model as one may be more suitable for some certain vegetation

types than others. Above challenges of LSP-based classification prohibit the creation of complete, high detailed land cover maps over a large area with diversity land cover. Thus, it will significantly reduce the use of those maps in change analysis as it only able to focus on the most accurate class and in small area where coverage is complete (or gaps are small enough to be filled). An effort to overcome those limitations leads me to ask the third and fourth research questions. Third, *how will the classification accuracy be impacted by selecting different LSP models to fit annual time series?* Fourth, *can LSP-based classification be improved for non-vegetated surfaces by incorporating information from all other spectral bands?*

1.3   Data and Method

1.3.1   Study Area

Among the States composing the NGP region, we selected South Dakota as our study area for four reasons: (1) it has continental climate with high seasonal and interannual variation, pronounced gradients in precipitation (east-west) and temperature (north-south); (2) the State has strong livestock, dairy, and row crop production industries that make it economically viable to convert grasslands to croplands (*Reitsma et al., 2015*); (3) unique high resolution reference datasets were available for training and validation; and (4) South Dakota is the hot spot of grassland conversion to cropland between 2006 and 2011 as reported by *Wright and Wimberly (2013)*.

1.3.2   Input Data

Several satellite products were used in this research. Surface reflectance products at 30 m resolution from Landsat Collection-1 (from Landsat 5, 7 & 8) (*USGS, 2016*), its newer structure—Landsat Analysis Ready Data (ARD) (*USGS, 2018*), and Harmonized

Landsat Sentinel-2 (HLS) (*Claverie et al., 2018*) were used to construct vegetation index time series, including Enhanced Vegetation Index (EVI) and 2-band EVI (EVI2), and spectral variables (spectral band and normalized ratio composites). Collections 5 and 6 of the MODIS level-3 global Land Surface Temperature (LST) & Emissivity 8-day composite products at 1 km resolution (from both Aqua and Terra satellites) (*NASA LP-DAAC, 2013; Wan et al., 2015a&b*) were used to compute accumulated growing degree-days (AGDD) time series. The Cropland Data Layer (CDL; *Boryan et al., 2011*) was used to generate sample datasets and cross-compare with outputs from this study. Beside the CDL, a rich reference point dataset derived from high spatial resolution imagery (only cover 3 years: 2006, 2012 & 2014; *Reitsma et al, 2015; 2016*) was also used to evaluate accuracy of the land cover maps newly generated in this study.

1.3.3 Technical Approach

Research question #1

First, an annual time series of accumulated growing degree-days (AGDD) was built from MODIS 8-day composites of land surface temperatures. Using the EVI time series derived from Landsat Collection 1's surface reflectance, a downward convex quadratic model to each year's progression of AGDD (derived from Collection 5 LST) was then fit at each pixel (*i.e.*, $EVI = \alpha + \beta \times AGDD - \gamma \times AGDD^2$). Phenological metrics derived from fitted model and the goodness of fit then are submitted to a random forest classifier (RFC) to characterize LCLU for four sample counties (Roberts, Bon Homme, Codington, Walworth: located within the Landsat's sidelaps to achieve more observations) in South Dakota in three years (2006, 2012, 2014) when reference point datasets are available for training and validation. To answer research question #1, accuracy of RFC models and

predicted land cover maps were evaluated by testing data generated from the CDL. To examine the sensitivity of the RFC to sample size and design, land cover classifications were performed under different sample selection scenarios.

Research question #2

To characterize land changes in the study area, a fine spatiotemporal resolution land cover dataset with just three broad categories ("cropland", "grassland", and "others") was generated using the phenometrically-based classification developed to answer the research question #1. To overcome the limitations of the bi-temporal change detection, a trajectory-based approach—centers on the logistic regression—was proposed that considers the entire land cover/land use time series to determine if there was actual land change at a particular location. Crop expansion in South Dakota between 2007 and 2015 were then summarized for each county or each NASS reporting district to answer research question #2. The results were compared against various official data sources released by the United States Department of Agriculture.

Research questions #3 and #4

First, several annual time series of remotely sensed data were built, including: accumulated growing degree-days from the Collection 6 MODIS 8-day land surface temperature product, 2-band Enhanced Vegetation Index (EVI2), and spectral statistics from the Harmonized Landsat Sentinel-2 as well as from the U.S. Landsat Analysis Ready Data surface reflectance products. Then at each pixel, EVI2 time series were simulated using two land surface phenology models: Convex Quadratic model (CxQ) and Hybrid Piecewise Logistic Model (HPLM). Phenometrics and spectral variables were submitted separately and together to Random Forest Classifiers to depict land cover/land use in

Roberts County, South Dakota. Four classification scenarios using different sets of input variables were performed: (1) only CxQ phenometrics, (2) only HPLM phenometrics, (3) only spectral variables, and (4) the combined spectral-phenological variables. Comparisons between classification scenarios, which answer research questions #3 and #4, were conducted based on conventional accuracy (Congalton & Green, 2008) metrics and two alternatives of kappa statistic (Pontius & Millones, 2011).

1.4   Significance of the research

My dissertation research will advance the researcher's toolkit for land cover mapping and change analysis as well as shine fresh light on what has been a controversial issue since 2013: the conversion of mixed-grass prairie to commodity crops, particularly to corn and soybean, due to the increasing demand for biofuels, animal feed, and exports

1.5   Thesis structure

This research dissertation has five chapters, including this introductory first chapter. Chapter Two explores the accuracy of land cover classification using phenometrics generated from the Convex Quadratic model solely (research question #1). The chapter also examined performance of Random Forest Classifiers (RFC) under multiple sampling designs (no-controlled versus controlled samples) and increasing sample sizes. Chapter Three provides a comprehensive analysis of cropland expansion in South Dakota between 2007 and 2015 (research question #2). To overcome the limitation of the conventional bi-temporal method, the trajectory-based change detection approach, which utilized the entire land cover time series to separate between true changes and mis-classifications as well as rotations of land use, was proposed. Chapter Four further explores land surface phenology-based classification demonstrated in Chapter Two. In the chapter,

I evaluated performance of land cover classification using (1) only phenological metrics derived from two different land surface phenology models, (2) only spectral composited bands and ratios, and (3) combined phenological-spectral variables. Finally, Chapter Five presents the main research summaries and recommendations.

1.6 References

Ahl, D.E., Gower, S.T., Burrows, S.N., Shabanov, N.V., Myneni, R.B., Knyazikhin, Y., 2006. Monitoring spring canopy phenology of a deciduous broadleaf forest using MODIS. Remote Sens. Environ. 104 (1), 88–95.

Arora, G., & Wolter, P. T. (2018). Tracking land cover change along the western edge of the US Corn Belt from 1984 through 2016 using satellite sensor data: observed trends and contributing factors. J. Land Use Sci. 1-22.

Baumann, M., Ozdogan, M., Richardson, A.D., Radeloff, V.C., 2017. Phenology from Landsat when data is scarce: using MODIS and dynamic time-warping to combine multi-year Landsat imagery to derive annual phenology curves. Int. J. Appl. Earth Obs. Geoinf. 54, 72–83.

Beck, P.S., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens. Environ. 100 (3), 321–334.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department Of Agriculture, National Agricultural Statistics Service, Cropland Data Layer program. Geocarto Int. 26 (5), 341–358.

Claassen, R., Carriazo, F., Cooper, J.C., Hellerstein, D., and Ueda, K. (2011). Grassland to cropland conversion in the Northern Plains. Economic Research Report, 120.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., ... & Justice, C. 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145-161.

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

de Beurs, K.M., Henebry, G.M., 2004. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. Remote Sens. Environ. 89 (4), 497–509.

Decision Innovation Solutions. (2013). 2013 Multi-State Land Use Study: Estimated Land Use Changes 2007-2012. Urbandale, IA 50322: Decision Innovation Solutions.

Faber, S., Rundquist, S., Male, T., 2012. Plowed under: How Crop Subsidies Contribute to Massive Habitat Losses. Environmental Working Group, Washington, DC.

Franklin, S.E., Ahmed, O.S., Wulder, M.A., White, J.C., Hermosilla, T., Coops, N.C., 2015. Large area mapping of annual land cover dynamics using multi-temporal change detection and classification of Landsat time-series data. Can. J. Remote. Sens. 41, 293–314.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T.R., 2011. Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. Remote Sens. Lett. 2, 279–288.

Henebry, G.M., & de Beurs, K.M. (2013). Remote Sensing of Land Surface Phenology: A Prospectus. In: (M.D. Schwartz, ed.) Phenology: An Integrative Environmental Science, 2e. Springer. pp. 385-411.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., & Wang, X. 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. 6(11), 11518-11532.

Johnston, C.A. (2013). Wetland losses due to row crop expansion in the Dakota Prairie Pothole Region. Wetlands, 33(1), 175-182.

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

Jönsson, P., & Eklundh, L. (2003). Seasonality extraction from time-series of satellite sensor data. In Frontiers of Remote Sensing Information Processing (pp. 487-500).

Kong, F., Li, X., Wang, H., Xie, D., Li, X., & Bai, Y. 2016. Land cover classification based on fused data from GF-1 and MODIS NDVI time series. Remote Sens. 8(9), 741.

Krehbiel, C., Henebry, G.M., 2016. A comparison of multiple datasets for monitoring thermal time in urban areas over the US Upper Midwest. Remote Sens. 8 (4), 297.

Krehbiel, C., Zhang, X., Henebry, G.M., 2017. Impacts of thermal time on land surface phenology in urban areas. Remote Sens. 9 (5), 499.

Krehbiel, C.P., Jackson, T., Henebry, G.M., 2016. Web-enabled Landsat data time series for monitoring urban heat island impacts on land surface phenology. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 9 (5), 2043–2050.

Lark, T.J., Salmon, J.M., & Gibbs, H.K. (2015). Cropland expansion outpaces agricultural and biofuel policies in the United States. Environ. Res. Lett. 10(4), 044003.

Mueller, R., and R. Seffrin. (2006). New methods and satellites: A program update on the NASS cropland data layer acreage program. The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Science, 36, 97-102.

NASA LP DAAC, 2013. MODIS terra/aqua land surface temperature/emissivity 8-day L3 global 1 km SIN grid. Version 5. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. https://lpdaac.usgs.gov, Accessed date: 15 January 2017.

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

Qader, S. H., Dash, J., Atkinson, P. M., and Rodriguez-Galiano, V. (2016). Classification of vegetation type in Iraq using satellite-based phenological parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9(1), 414-424.

Rashford, B.S., Walker, J.A., and Bastian, C.T. (2011). Economics of grassland conversion to cropland in the Prairie Pothole Region. Conserv. Biol. *25*(2), 276-284.

Reitsma K.D., Dunn B.H., Mishra U., Clay S.A., DeSutter T., Clay D.E. (2015) Land-use change impact on soil sustainability in a climate and vegetation transition zone. Agron. J. 107(6), 2363-2372.

Reitsma, K. D., Clay, D. E., Carlson, C. G., Dunn, B. H., Smart, A. J., Wright, D. L., & Clay, S. A. (2014). Estimated South Dakota land use change from 2006 to 2012. iGrow Agronomy.

Reitsma, K.D., Clay, D.E., Clay, S.A., Dunn, B.H., and Reese, C. (2016). Does the US cropland data layer provide an accurate benchmark for land-use change estimates? Agron. J. 108(1), 266-272.

Roy, D. P., & Yan, L. (2018). Robust Landsat-based crop time series modelling. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.06.038

Schnepf, R., and Yacobucci, B.D. (2013). Renewable Fuel Standard (RFS): Overview and Issues. Congressional Research Service, Washington, DC.

Schwartz, M.D., Reed, B.C., 1999. Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. Int. J. Remote Sens. 20 (17), 3451–3457.

SDDA–South Dakota Department of Agriculture. (2018). South Dakota Agriculture - Facts & Impact. Available from: https://sdda.sd.gov/education-outreach/agriculture-industry/. Assessed: 10/20/2018.

Teluguntla, P., Thenkabail, P., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., ... & Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J. Photogramm. Remote Sens. 144, 325-340.

USGS — U.S. Geological Survey. 2016. Landsat update special issue 4: Landsat Collection 1 level-1 data processing starting soon. Retrieved from: https://landsat.usgs.gov.

USGS — U.S. Geological Survey. 2018a. Landsat collections: U.S. Geological Survey Fact Sheet 2018–3049. https://doi.org/10.3133/fs20183049.

Wan, Z., Hook, S., Hulley, G. 2015a. MOD11A2 MODIS/Terra Land Surface Temperature/ Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD11A2.006

Wan, Z., Hook, S., Hulley, G. 2015b. MYD11A2 MODIS/Aqua Land Surface Temperature/ Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MYD11A2.006

Wright, C.K., and Wimberly, M.C. (2013). Recent land use change in the Western Corn Belt threatens grasslands and wetlands. Proc. Natl. Acad. Sci. 110(10), 4134-4139.

WWF — World Wildlife Fund. (2016). Plow Print Report. Retrieved from: http://www.worldwildlife.org/projects/plowprint-report

Xue, Z., Du, P., & Feng, L. 2014. Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7(4), 1142-1156.

Yan, E., Wang, G., Lin, H., Xia, C., & Sun, H. 2015. Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. Int. J. Remote Sens. 36(2), 489-512.

Young, S.M. (2017). Land Change Monitoring, Assessment, and Projection (LCMAP) revolutionizes land cover and land change research. U.S. Geological Survey General Information Product 172.

Zhang, H.K., & Roy, D.P. 2017. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. Remote Sens. Environ. 197, 15-34.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., ... Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84 (3), 471–475.

Zhong, L., Hawkins, T., Biging, G., Gong, P., 2011. A phenology-based approach to map crop types in the San Joaquin Valley, California. Int. J. Remote Sens. 32 (22), 7777–7804.

Zhu, Z., and Woodcock, C.E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152-171.

CHAPTER 2

CHARACTERIZING LAND COVER/LAND USE FROM MULTIPLE YEARS OF
LANDSAT AND MODIS TIME SERIES: A NOVEL APPROACH USING LAND
SURFACE PHENOLOGY MODELING AND RANDOM FOREST CLASSIFIERS

2.0   Abstract

Over the last 20 years, substantial amounts of grassland have been converted to other land uses in the Northern Great Plains. Most of land cover/land use (LCLU) assessments in this region have been based on the U.S. Department of Agriculture - Cropland Data Layer (USDA - CDL), which may be inconsistent. Here, we demonstrate an approach to map land cover utilizing multi-temporal Earth Observation data from Landsat and MODIS. We first built an annual time series of accumulated growing degree-days (AGDD) from MODIS 8-day composites of land surface temperatures. Using the Enhanced Vegetation Index (EVI) derived from Landsat Collection 1's surface reflectance, we then fit at each pixel a downward convex quadratic model to each year's progression of AGDD (i.e., $EVI = \alpha + \beta \times AGDD - \gamma \times AGDD2$). Phenological metrics derived from fitted model and the goodness of fit then are submitted to a random forest classifier (RFC) to characterize LCLU for four sample counties in South Dakota in three years (2006, 2012,

2014) when reference point datasets are available for training and validation. To examine the sensitivity of the RFC to sample size and design, we performed classifications under different sample selection scenarios. The results indicate that our proposed method accurately mapped major crops in the study area but showed limited accuracy for non-vegetated land covers. Although all RFC models exhibit high accuracy, estimated land cover areas from alternative models could vary widely, suggesting the need for a careful examination of model stability in any future land cover supervised classification study. Among all sampling designs, the "same distribution" models (proportional distribution of the sample is like proportional distribution of the population) tend to yield best land cover prediction. RFC used only the most eight important variables (e.g., three fitted parameter coefficients [α, β, and γ]; maximum modeled EVI; AGDD at maximum modeled EVI; the number of observations used to fit CxQ model; and the number of valid observations) have slightly higher accuracy compared to those using all variables. By summarizing annual image time series through land surface phenology modeling, LCLU classification can embrace both seasonality and interannual variability, thereby increasing the accuracy of LCLU change detection.

2.1   Introduction

Since 2000, the US Northern Great Plains region (NGP, embraces five U.S. States: North Dakota, South Dakota, Nebraska, Minnesota, and Iowa) has experienced substantial land cover/land use (LCLU) change, especially the grassland conversion to cropland (*Faber et al., 2012; Johnston, 2014; Wright & Wimberly, 2013a*), due to contribution of many factors including government policy (*Lark et al., 2015*), an aging workforce of farmers, commodity prices (*Claassen et al., 2011*), energy development (*Preston & Kim,*

2016; Singh et al., 2017), and rental agreements. *Johnston (2013)* estimated wetlands losses of 5,000-6,000 ha per year over the past decade due to cropland expansion. *Wright & Wimberly (2013b)* found a net loss of 530,000 ha in grass-dominated lands from 2006 to 2011 in the Western Corn Belt (area of the Midwest U.S. dominated by agriculture, mostly corn and soybean cultivation) at the rate of 1.0% to 5.4% annually. Those changes have caused habitat fragmentation and threaten the loss of regional biodiversity (*Stephens et al., 2008; Meehan et al., 2010; Mutter et al., 2015*; *Otto et al., 2016*; *Wimberly et al., 2018*), increased soil erosion and water pollution (*Vache et al., 2002; Montgomery, 2007*), and increased net carbon debt (*Fargione et al., 2008; Searchinger et al., 2008*).

Reliable data sources to quantify land cover changes come from the periodic statistical surveys by the USDA, such as the National Resources Inventory (NRI) by the Natural Resources Conservation Service (NRCS), and Quick Stats Database by the National Agricultural Statistics Service (NASS). However, changes detected from those datasets lack sufficient spatial and temporal details preventing identification of where and when transitions occur (*Claassen et al., 2011*; *Rashford et al., 2011*). These limitations prohibit the use of NRI and NASS data for comprehensive analysis of LCLU change. Thus, to better understand land cover changes, researchers have turned to geospatial datasets such as the National Wetland Inventory, the National Land Cover Datasets (NLCD), and, especially, the USDA Cropland Data Layer (CDL) due to finer spatiotemporal resolution and high number of classes (*Faber et al., 2012; Wright & Wimberly, 2013a,b*; *Johnston, 2013, 2014*; *Lark et al., 2015*; *WWF, 2016*). The relatively high spatiotemporal resolution of the CDL enables tracking of not only land cover transitions but also crop rotations on a field basis. Use of the CDL can be justified by overall accuracy of 85% to 95% for major

crops, and often 97% producer and user accuracies for corn and soybean (*Boryan et al.,
2011*). However, recent studies have raised concerns about CDL-based analysis, especially
when using a simple bi-temporal comparison approach with at least one study period before
2010 (*Reitsma et al., 2016; Lark et al, 2017*). It is important to aware that the CDL is meant
to monitor agricultural land cover annually rather than tracking changes over time; and it
has undergone substantial methodological changes over time (*Mueller & Seffrin, 2006*;
Boryan et al., 2011). Independent validation of the 2006 and 2012 South Dakota CDL
(*Reitsma et al., 2016*) revealed inconsistencies in the accuracy of generalized cropland and
grassland classes. Reitsma *et al.* (2016) reported that the producer and user accuracies were
dependent on location and land use. For example, in southeastern South Dakota, the
producer accuracies for cropland and grassland were 88.4% and 38.9%, respectively. The
importance of land cover knowledge raises the need for a more consistent land cover/land
use product.

It is worth to note that the conventional classification methods, such as ones first
applied to the CDL and the more general NLCD, were developed in an era of data scarcity
and limited computational power. These methods focused on comparing just a few scenes
and were challenged by classification error between spectrally-similar covers.
Additionally, in areas with frequent morning cloud cover, collecting even a few cloud-free
scenes over a year could be difficult, if not impossible. In the current era, a rapid increase
of accessible Earth Observation data (particularly from the Landsat archive and its
augmentation by newer sensors, such as Sentinel 2A and 2B) coupled with improved
computing and storage capability is leading to the emergence of methods for generation

annual land cover products from time series data as multi-temporal classification has been proved as superior to classification relying on just a few scenes (*Franklin et al., 2015*).

Shifts in land cover or environmental conditions, management practice, disturbance may lead to a spatiotemporal variation of land surface phenology (LSP), *i.e.*, seasonal patterns of reflectance from the vegetated land surface as observed using remote sensing (*Henebry & de Beurs, 2013*). Several methods have been developed to detect LSP that estimate timings of phenophase transitions during the growing season based either on pre-defined thresholds of vegetation indices (*Lloyd, 1990; Reed et al., 1994; White et al., 1997*) or on features of fitted curves (*Zhang et al., 2003; de Beurs & Henebry, 2004*). Among many proposed LSP models, the logistic curve (and its modifications) is one of the more commonly used fitting methods. Application of the logistic curve seems more suitable to landscapes dominated by woody vegetation (*Zhang et al., 2003; Beck et al., 2006; Ahl et al., 2006; Baumann et al, 2017*). Characterizing phenology can be more difficult for herbaceous than for woody vegetation due to strong interannual variation in grasslands and croplands (*Schwartz & Reed, 1999*). Another LSP fitting approach originates from traditional phenological models that relate to the progression of thermal time during the growing season to events in plant development. Studies have shown that the temporal development of remotely sensed vegetation indices in temperate ecosystems can be well approximated as a quadratic function of accumulated growing degree-days (AGDD) because AGDD captures well the seasonal course of insolation at mid-latitudes (*de Beurs & Henebry, 2004; Henebry & de Beurs, 2013; Krehbiel & Henebry 2016; Krehbiel et al., 2016, 2017*).

LSP-based classifications that use time series data may be more accurate than traditional classification methods. Phenological time series can provide a larger number of sequentially related predictor variables to be exploited by machine learning classifiers. Over the past decade, LCLU change studies have included LSP information to map vegetated land cover and changes at regional to global scales (*de Beurs & Henebry, 2004; Friedl et al., 2010; Gu et al., 2010; Zhong et al., 2011; Clerici et al., 2012; Zhu & Woodcock, 2014; Krehbiel et al., 2016; Qader et al., 2016*).

A downward convex quadratic (CxQ) LSP model linking a vegetation index (such as the NDVI) to thermal time as measured by AGDD was first applied to compare spring green-up dynamics before and after the collapse of the Soviet Union (*de Beurs & Henebry 2004*) to detect land cover change in a noisy AVHRR time series. Using a similar approach with Landsat time series for NDVI and MODIS for AGDD, *Krehbiel et al. (2016, 2017)* characterized the conversion of croplands to residential areas near Omaha, NE and Minneapolis-St. Paul, MN. Although *de Beurs & Henebry (2004)* and *Krehbiel et al. (2016, 2017)* successfully detected land cover changes in their study area through significant shifts in the value of phenometrics, neither studies quantified areal changes and generate spatially explicit LCLU change maps. On the other hand, *Zhong et al. (2011)* differentiated crop types in San Joaquin Valley, CA with an overall accuracy of above 70% for all three study years using only phenological metrics derived from MODIS NDVI time series. Using a very similar approach, *Qader et al. (2016)* accurately (generally >90%) mapped broad dominant vegetation cover classes (cropland, grassland, and shrubland) in Iraq and characterized changes from 2002 to 2012. Compared to *Zhong et al. (2011)*, *Qader et al. (2016)* sacrificed the classification detail by fewer classes for higher accuracy. However,

both land cover change maps were produced at 250 meters using MODIS NDVI data as input and neither used thermal data.

Understanding rapid land change is not only critical for management and conservation of prairie habitats and ecosystem services, but also for projecting production of crops and biofuels and the impacts of land conversion on rural infrastructure, such as roads and water quality. Hence, there is a need for a LCLU dataset with good spatiotemporal coverage as well as consistent accuracy through time. Taking advantage of abundant satellite observations, we developed a novel approach to mapping LCLU using as input only information derived from LSP modeling. First, we fitted a downward CxQ LSP model at each pixel using Landsat EVI and MODIS AGDD co-registered time series. Through the model fitting, we described phenological characteristics of each land cover type (timing of the growing season, peak EVI, timing of the peak EVI, *etc.*) as a simple functional form to allow cross-comparison. Fitted parameter coefficients and phenometrics at each pixel are then submitted to a random forest classifier (RFC) to generate a map depicting LCLU. We applied this approach to representative cropland areas in South Dakota for three years (2006, 2012, 2014) that reference point datasets are available for training and validation. Our objective is to examine how well we can classify land cover using only information from the land surface phenology modeling process. By utilizing the entire annual time series, we have multiple views of the surface that should help to better characterize LSP and thereby improve the classification.

We started the classification with three broad categories: "cropland", "grassland", and "others" as used by the reference point datasets. Previous studies have suggested that classification by machine learning algorithms are sensitive to a sample size and design

(e.g., amount of each land cover type in the sample dataset) (*Jin et al, 2014*; *Colditz, 2015*; *Millard & Richardson, 2015*; *Lyons et al., 2018*). Accordingly, we generated multiple RFC models for each county-year using different sample datasets constructed from the CDL and the reference points to understand in which scenarios do RFC models perform best. We first examined how RFC models respond to increasing sample size with no control in sampling design. Next, RFC performances were evaluated using one of the following controls: (1) the proportional distribution among classes in the sample dataset was equivalent to the proportional distribution found in the re-classified CDL ("same distribution"), or (2) each class in the sample dataset contained the same number of pixels ("same size"). We also examined the effect of variable importance on RFC performance. Lastly, we demonstrate the capability of LSP-based classification to map major crops in two representative counties in South Dakota: Roberts and Codington.

2.2   Study Area and Data

2.2.1   Study Area
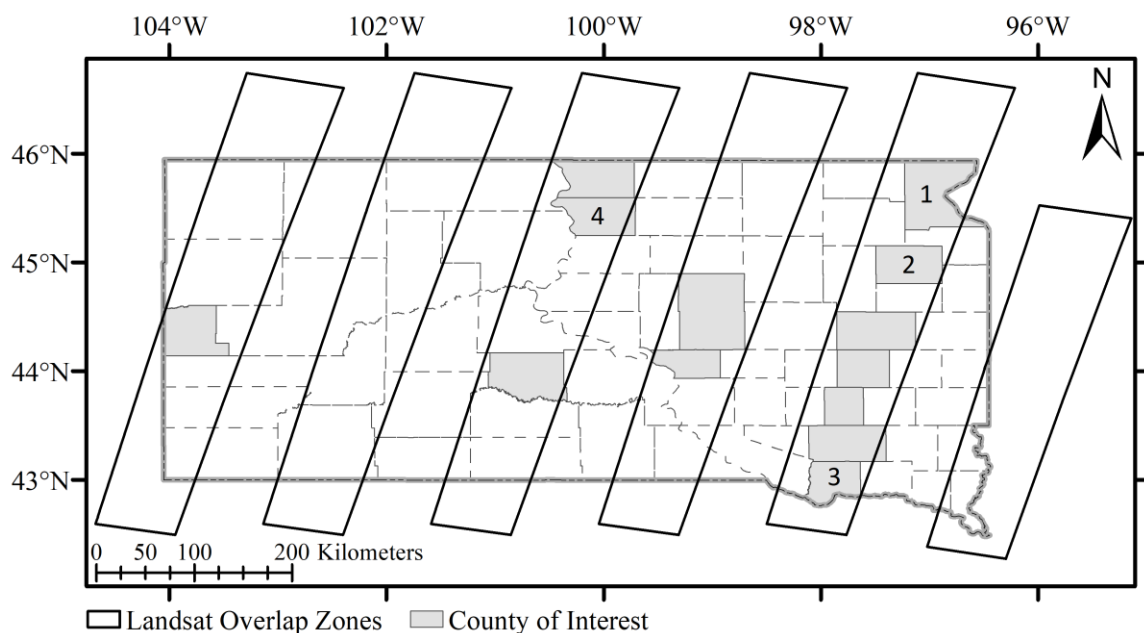
Figure 2.1. South Dakota counties that fall mostly (>90%) within Landsat overlap zones are indicated in grey. Four counties (1-Roberts, 2-Codington, 3-Bon Homme, and 4-Walworth) are the focus of this investigation.

Among the States composing the NGP region, we selected South Dakota as our study area for three reasons: (1) it has continental climate with high seasonal and interannual variation, pronounced gradients in precipitation (east-west) and temperature (north-south); (2) the State has strong livestock, dairy, and row crop production industries that make it economically viable to convert grasslands to croplands (*Reitsma et al., 2015*); and (3) unique high resolution reference datasets were available for training and validation.

Initially, we intended to focus on all 13 counties in South Dakota that have at least 90% of county area within overlap zones of Landsat paths (grey counties in Figure 2.1) to maximize the potential number of cloud-free Landsat observations. However, due to limitations in the training dataset, we restricted our focus to four counties: Bon Homme, Codington, Roberts, and Walworth. In each of these four focal counties, we characterized LCLU for three years (2006, 2012, 2014) separately. According to 2014 CDL data, cropland is a dominant land cover in Roberts, Codington and Bon Homme, accounting for approximately 56%, 56% and 67% of county area, respectively (Figure S2.1-S2.4). Major crops in these counties include corn, soybean, spring wheat (in Roberts and Codington), winter wheat (in Bon Homme), and alfalfa. More than 80% of croplands were used for corn and soybean production. On the other hand, grassland is the predominant land cover in Walworth, covering 49% of the county area. Croplands, which consist mostly of corn/soybean, sunflower and winter wheat, cover 41% of the county.

2.2.2 Data

Reference Point Dataset for Training

In each NASS region of South Dakota, 1,600 random points were generated (*Reitsma et al., 2015* and unpublished data) and laid over high-resolution imagery obtained from the National Agricultural Imagery Program (NAIP; USDA-FSA, 2013). The NAIP data were collected during the growing season at 2-meter resolution for 2006 and 1-meter resolution for 2012 and 2014. Information from three visible bands was used to construct a natural color image. Based on various features presented in these high-resolution images, an analyst classified the dominant land use in 2006, 2012, and 2014 at each location into one of five broad categories—cropland, grassland, habitat, not-agriculture, or water. A total of 43,200 points (1,600 points in each NASS region × 9 regions × 3 years) were manually classified in South Dakota across the three years. For field validation, we randomly selected 100 points across the entire state for ground-level observations. The image-based classification and field observations matched in every case, giving us high confidence in the accuracy of the image-based classification.

There are 270 points in Roberts, 172 points in Codington, 122 points in Walworth, and 157 points in Bon Homme (see Figures S2.1-S2.4 for maps of point locations). Although the original reference dataset consisted of five broad types, we are only interested in the croplands and grasslands; thus, we regrouped the reference dataset into just three categories: "cropland", "grassland", and "others".

Cropland Data Layer

The USDA Cropland Data Layer is a crop-specific land cover raster created annually for the continental United States by the NASS using moderate resolution satellite imagery and extensive agricultural ground observations. The CDL, which was first produced in 1997 for North Dakota, has covered the entire Great Plains yearly since 2006,

with approximately 130 classes and a spatial resolution of 30 m at best (*Boryan et al., 2011*). For this study, we regrouped the CDL land covers into three classes to match the reference dataset. "Cropland" category contains the CDL's field crops (including alfalfa), vegetables, and tree fruits. All grass-like classes, including native grassland, pasture/hay, other hay/non-alfalfa, were combined to create a broad "grassland" category. The remaining CDL classes were put into the "others" category. Crosswalk details appear in Table S2.1.

Landsat Collection-1 Surface Reflectance Product

In 2016, the U.S. Geological Survey (USGS) implemented the new organizational method for the Landsat archive called Collections to ensure consistent quality through time and across instruments for Landsat Level-1 products (USGS, 2016). This data structure provides a consistent archive of known data quality to support time-series analyses and data "stacking". Collection 1 initially consists of Level-1 products generated from Landsat 4-5 TM, Landsat 7 ETM+, and Landsat 8 OLI/ TIRS instruments that are assigned to one of three categories: Tier-1, Tier-2, or Real-Time. Here, we used only the surface reflectance product from Landsat scenes in Tier-1 (WRS-2 Paths:28-33 & Rows:29-34), because those images are produced with the highest quality data available and, thus, are better suited for time series analysis and modeling.

MODIS Land Surface Temperature (LST) Product

We used the Collection 5 MODIS level-3 global Land Surface Temperature and Emissivity 8-day composite products at 1000 m resolution from both Aqua (MYD11A2) and Terra (MOD11A2) satellites (NASA LP-DAAC, 2013). Both the "LST_Day_1km" and "LST_Night_1km" scientific datasets from the M{O|Y}D11A2 products were used.

Overpass time for daytime (nighttime) observations are about 1030 (2230) local solar time for Terra and about 1330 (0130) local solar time for Aqua. The MODIS LST are provided in a sinusoidal grid format and display the mean clear-sky LST in Kelvin observed during an 8-day time frame. To use these LST products along with Landsat data, we re-projected the MODIS products to Albers Equal Area Conic projection, and then resampled the layer to 30 m using bilinear interpolation. The LST time series was converted from Kelvin to degrees Celsius for calculation of the thermal time used in the LSP modeling.

2.3   Methods

2.3.1   Construction of Landsat EVI and MODIS AGDD time series

The EVI was calculated from Landsat Collection-1 surface reflectance product (courtesy of the U.S. Geological Survey) as described in the Landsat Spectral Indices Product Guide (*USGS, 2017*). The outputs were re-projected into Albers Equal Area Conic projection to match with other datasets (MODIS LST, CDL). We then applied cloud/snow masks delivered with the product to remove "bad" observations (snow, high confidence cloud, or cloud shadow pixels). EVI values outside the valid range (from 0 to 1) were also excluded. The remaining "good" EVI values at each pixel were then stacked in chronological order from the first day of the year (DOY=1) to the final day of the year (DOY=365, or 366 in leap years). A number of valid observations was significantly lower in 2012 compared to 2006 and 2014 due to the end of Landsat 5 TM sensor operations.

The AGDD were calculated from MODIS LST as follow:

$$GDD_t = \max\left\{\frac{T_{max,t}+T_{min,t}}{2}, 0\right\} \qquad \text{(Equation 2.1)}$$

$$AGDD_t = AGDD_{t-1} + 8 \times GDD_t \qquad \text{(Equation 2.2)}$$

where GDD$_t$ are the growing degree-days for compositing period (t), T$_{max,t}$ and T$_{min,t}$ are the highest and lowest LST values from available MODIS observations during the compositing period. Accumulation of GDD starts at the beginning of the calendar year and continues to the end of the year, but the temporal resolution of the LST composites is 8 d. Thus, we multiply the composite-calculated GDD by 8 to rescale it.

While good EVI observations from Landsat may occur at any DOY, there are only 46 MODIS composites each year. Although each pixel of the MODIS LST composite contains the best possible estimate during an 8 d period, each composite is linked to a single DOY, starting from DOY 1 to DOY 361 at 8 d intervals. Here we linked each EVI value from a specific DOY to the AGDD value associated with the "nearest" compositing period. For example, EVI of DOYs from 11 to 14 were coupled with AGDD composite 3 which is tied to DOY 17. If EVI was obtained right in the middle of the two MODIS compositing periods, paired AGDD is calculated as the average of the two AGDD compositing periods. For example, EVI of DOY 5 is paired with mean AGDD of composite 1 and 2 (DOY 1 and DOY 9). A complete list of possible EVI×AGDD pairs in a year appears in Table S2.2. As multiple EVI values may be linked to a specific AGDD compositing period, only the highest EVI of each available DOY will be used for model fitting.

2.3.2 Convex Quadratic (CxQ) Model for Land Surface Phenology

The CxQ model for land surface phenology first linked the NDVI to AGDD (*de Beurs & Henebry 2004; Henebry & de Beurs 2013; Krehbiel et al. 2016, 2017*). Here, we used the EVI—Enhanced Vegetation Index (*Huete et al., 2002*)—for LSP modeling:

$$EVI = \alpha + \beta \times AGDD - \gamma \times AGDD^2 \qquad \text{(Equation 2.3)}$$

where α, β, γ (alpha, beta, and gamma, respectively) are the parameter coefficients to be fitted. The negative sign on gamma in Equation 2.3 is to show that we seek a fitted curve that is downward arching, since the EVI values will go up and then down over the growing season. Only a negative gamma coefficient will produce this shape and, thus, we retained only fits with a negative gamma coefficient. Occasionally, we may retrieve model fits with positive gamma coefficient for non-vegetated pixels (developed or barren areas). However, those fits were both very few and were deemed failed model fit.

Using the EVI and AGDD time series, we fitted the CxQ model at each pixel over the growing season in each year as described in Equation 2.3 (Figure 2.2). To identify the breadth of the fitting window of the growing season within year, we first calculated the change in EVI (ΔEVI) and AGDD (ΔAGDD) at each step and the rate-of-change from these two variables (ΔEVI/ΔAGDD, which measures how quickly EVI changes as a function of thermal time). We then looked for two transition points (Figure 2.2a). Our approach to detecting the two transition points originates from *Zhang et al (2003)*, where MODIS EVI time series were fitted as a double logistic function of time. On the DOY axis, the two transition points are those with the highest rate-of-change in EVI (ΔEVI/Δt, which measures how quickly EVI changes as a function of calendar time). If data points are distributed equally on the time (t) axis (Δt remains unchanged), we expect to have extreme rates-of-change (the two transition points) around the maturity phase (highest EVI increasing rate) and the senescence phase (highest EVI decreasing rate) of vegetation as absolute changes in EVI (ΔEVI) are highest around these times. In our case, Landsat observations were not equally distributed along the thermal time axis and vary each year. In addition, we calculated the rate-of-change as ΔEVI/ΔAGDD. As the thermal time axis

(AGDD) was used instead of the DOY axis, our calculated rates-of-change can be very high at the beginning or the end of the year due to small changes in temperature. Therefore, we cannot use only the ΔEVI or the ΔEVI/ΔAGDD values to detect transition points. Here, we coupled the rate-of-change with EVI change to detect the transition points. We first identified two highest and two lowest values of ΔEVI/ΔAGDD (Figure 2.2b). Then, among the two highest rates-of-change, the one with maximum ΔEVI (largest increase in EVI) was labeled as the first transition point. Similarly, between the two lowest rates-of-change, the one with minimum ΔEVI (largest decrease in EVI) was labeled as the second transition point (Figure 2.2c). As shown in Figure 2.2b, the lowest point is not the one selected as that point does not have high ΔEVI (high ΔEVI/ΔAGDD values are due to small ΔAGDD). The two transition points consist of one point with highest ΔEVI/ΔAGDD value (Figure 2.2b) and one point with second lowest ΔEVI/ΔAGDD but lowest in ΔEVI (Figure 2.2b and 2.2c). From the two transition points, we searched outward to find two Landsat observations nearest in thermal time that fall below a certain EVI threshold (Figure 2.2a). If the highest observed EVI (y_{max}) is above 0.65, the EVI threshold is 0.3. The EVI threshold is 0.2 for y_{max} ranging from 0.4 to 0.65; and there is no cut-off threshold if y_{max} is smaller than 0.40. In theory, only three points are required to retrieve the unique solution for a quadratic fit. However, to achieve a better fitted model, we restricted the fitting to pixels with at least five valid data points per year. Hence, we failed to produce proper quadratic fits for many pixels, mostly non-vegetated surfaces, and in the Scan Line Corrector (SLC) data gaps of Landsat 7 (cf. maps for 2012 in Figures 2.3 and S2.12-S2.14. Those pixels were then flagged as no-data.

From each fitted model, we derived a suite of 16 metrics including fitted parameter coefficients, derived phenometrics, and model fit statistics (Table 2.1) that describe the shape and location of the fitted curve in the EVI×AGDD space, the goodness of model fit, and other relevant characteristics. These metrics, which were distilled from the pixel time series for each year, were then used as input for the LCLU classification task.

Table 2.1. Fitted parameter coefficients, derived phenometrics, and other metrics from the CxQ model.

Parameters	Meaning
α, β, γ	Fitted parameter coefficients of CxQ model (Equation 2,1)
TTP	Thermal time to peak (AGDD at the max fitted EVI) (TTP = -β/2×γ)
PH	Peak height EVI (max fitted EVI) (PH = α − β²/4×γ)
HTV	Value of EVI at half-TTP (HTV = α + β×TTP/2 + γ×TTP²/4)
ymax	Highest observed EVIs
r²	Coefficient of determination of the fitted model
lpos, rpos	Observation index of start and end of the fitting window
o_fit	Number of observations used to fit the CxQ model
o_per	Ratio of "o_fit" to the total number of observations
minx, maxx	AGDD at left and right ends of the fitted curve in the first quadrant
peaks	Number of high EVI values (≥0.8*ymax) outside the fitting window
jumps	Number of times that ΔEVI ≥ 0.2

Figure 2.2. Fitting the CxQ model to EVI and AGDD time series for a 2006 sample corn pixel in Roberts County (N45.8090, W96.9028): (a) EVI and AGDD time series, (b) rate-of-change (ΔEVI/ΔAGDD) and (c) EVI changes (ΔEVI) along the thermal time axis, and (d) the fitted CxQ curve. In (b) and (c), ΔEVI and ΔEVI/ΔAGDD between two days were paired with AGDD value of the later day. The two transitions are displayed as filled squares (a, b, c) and the fitting window is in between the two unfilled square (a). The model was fitted only for valid observations (filled circles).

2.3.3   Land cover/land use classification using a Random Forest Classifier (RFC)

2.3.3.1 Blended sample dataset

To increase the size of sample data, we blended reference points with the CDL. Note that overall accuracy of the CDL can be low at some times and places. For example, overall accuracy of crop classes (the most accurate classes) in South Dakota are only 61.2% in 2006, 74.7% in 2012, and 86.8% in 2014. To increase accuracy of sample data selected from the CDL, we first extracted core patches from the CDL, *i.e.*, center pixels surrounded by eight pixels (Queen's neighborhood) of the same type, to reduce misclassification that happens more frequently at the edge. We then compared LCLU of those CDL pixels over three study years (2006, 2012, 2014) with nearby reference points. Only the CDL core pixels that display the same land cover with the reference point and remain unchanged over the three study years were selected for training and validation as it is highly likely that LCLU types of those pixels from the CDL are correct. Despite known issues with the CDL as discussed above, we used it to enhance the reference dataset for three reasons. First, the major crops are mapped with very high accuracy. Second, by collapsing CDL classes to create broader categories and searching for consistent land cover core patches over years, the chance to select "good" pixels (those that show true land cover information) for the sample datasets is high. Third, the CDL is the only publicly available LCLU cropland dataset with fine spatiotemporal resolution.

2.3.3.2 Classification and accuracy assessment

The RFC (*Breiman, 2001*) is an ensemble classification algorithm that constructs a set of decision trees to make a prediction. Each tree is created using a randomly selected subset of training samples and variables (Nvar). By growing the forest up to a user-defined

number of trees (Ntree), the RFC creates a set of trees with high variance but low bias. The final classification result is generated by averaging the class assignment probabilities calculated across all produced trees. New un-labelled data inputs are evaluated against all decision trees created in the ensemble and each tree votes for a class membership. The membership class receiving maximum votes is selected. Conventional accuracy metrics (overall accuracy-OA, user accuracy-UA, producer accuracy – PA, and kappa statistics) were used to evaluate model performance. Each RFC run was performed using default values of Nvar (the square root of the number of predictor variables $= \sqrt{16} = 4$) and Ntree (500) using 'randomForest' package in R program (*Liaw & Wiener, 2002*).

2.3.3.3 Scenarios for constructing the sample dataset

To examine the sensitivity of the RFC to sample size and design, we generated sample datasets in different scenarios to use in the classification process. For each scenario, we generated 20 RFC runs in which 2/3 of the sample data were randomly selected for the training phase and the remainder reserved for the validation step.

(a) Increasing sample size with no control on sampling design

We first want to examine how RFC performed without any control on sampling design. This scenario served as a baseline to evaluate whether or not putting restrictions on the sampling design would help improve the RFC model performance. In this scenario, we selected CDL core pixels at various distances from the reference points, from 300 m to 1500 m at 100 m increments. All selected pixels within each buffer were used in the classification. Table 2.2 shows the size of sample datasets at 1500 m buffer in the initial list of study areas. To enable consistent comparisons of model performance, we excluded

four counties (italics in Table 2.2) that have no data for training/validation in at least one land cover category.

(b) Increasing sample size with control on sampling design

RFC has been found to perform better with larger sample datasets (*Deng & Wu, 2013; Du et al., 2015*). The tradeoffs for better performance are higher costs in data collection and longer times in computation. A previous study suggested that the sample dataset should represent about 0.25% of the total study area for accurate classification using RFC (*Colditz, 2015*). RFC is also sensitive to the design of the sample dataset. In this scenario, we built a sample dataset from the 1500 m data pool so that it followed one of two options: either (1) the proportional distribution among classes in the sample dataset was equivalent to the proportional distribution found in the re-classified CDL layer ("Same Distribution") or (2) each class in the sample dataset contained the same number of pixels ("Same Size"). Both "Same Distribution" and "Same Size" sample datasets were generated to represent from 0.15% to 0.35% of the total county area (in 0.05% increments) of the total county area. Table 2.3 presents the maximum size of sample datasets constructed from 1500 m data pool for the two options as a percentage of the corresponding county area. Five counties highlighted in italics had sample datasets covering less than 0.35% of their county areas in at least one of three years. To compare these sample designs, we examined only the four counties that had sufficient sample pixels support both designs: Bon Homme, Codington, Roberts, and Walworth.

Table 2.2. Number of pixels in each land cover group at 1500 m buffer. Counties with zero pixels in at least one land cover group are highlighted in italics.

County. year	Crop	Grass	Others	County. year	Crop	Gras	Others
Roberts.06	42,447	11,552	19,526	Campbell.06	4,954	59,724	5,397
Roberts.12	38,218	11,461	17,547	Campbell.12	3,519	78,272	19
Roberts.14	42,707	11,556	18,601	Campbell.14	4,975	78,342	1,138
Codington.06	42,471	6,448	9,789	Walworth.06	13,580	107,057	5,572
Codington.12	37,237	6,425	7,903	Walworth.12	10,292	115,960	2,284
Codington.14	42,508	6,445	9,355	Walworth.14	13,613	115,711	2,959
Kingsbury.06	99,221	11,905	8,409				
Kingsbury.12	89,253	11,892	2,448	*Hand.06*	82,887	150,825	0
Kingsbury.14	100,400	11,901	1,622	*Hand.12*	77,461	150,601	0
Miner.06	37,458	9,224	1,289	*Hand.14*	83,445	150,228	0
Miner.12	32,681	9,165	1,218	*Hanson.06*	51,212	4,695	0
Miner.14	37,698	9,232	1,304	*Hanson.12*	45,742	4,695	0
Hutchinson.06	154,241	16,586	1,562	*Hanson.14*	51,413	4,689	0
Hutchinson.12	150,543	16,289	1,478	*Jones.06*	1,417	469,531	0
Hutchinson.14	155,510	16,233	1,541	*Jones.12*	1,405	475,311	0
Bon_Homme.06	46,719	2,538	13,220	*Jones.14*	1,416	477,774	0
Bon_Homme.12	44,574	2,529	9,715	*Lawrence.06*	0	0	356,281
Bon_Homme.14	47,176	2,521	12,875	*Lawrence.12*	0	0	356,979
Buffalo.06	8,121	195,519	10	*Lawrence.14*	0	0	356,851
Buffalo.12	7,522	196,922	12				
Buffalo.14	7,943	199,560	16				

Table 2.3. Maximum size of sample datasets constructed from 1500 m buffer data pool (Table 2.2) as a percentage to county area. Counties with the sample size covering less than 0.34% of their areas for at least one year are highlighted in italics.

	Same Distribution				Same Size		
County	**2006**	**2012**	**2014**	**County**	**2006**	**2012**	**2014**
Roberts	0.85	1.39	1.36	Roberts	1.10	1.17	1.11
Codington	1.06	1.04	1.12	Codington	1.00	1.10	1.02
Kingsbury	1.76	1.20	0.81	*Kingsbury*	1.09	0.34	0.21
Miner	0.51	0.95	1.05	*Miner*	0.24	0.24	0.24
Hutchinson	0.52	0.87	0.92	*Hutchinson*	0.20	0.20	0.20
Bon Homme	0.64	0.57	0.58	Bon Homme	0.48	0.49	0.47
Buffalo	0.01	0.03	0.03	*Buffalo*	0.00	0.00	0.00
Campbell	0.82	0.02	0.15	*Campbell*	0.75	0.00	0.16
Walworth	1.65	0.34	0.36	Walworth	0.83	0.37	0.44

2.3.3.4 Major crops mapping

To understand classification errors of alternative RFC models (section 2.3.3.3) and to examine the capability of LSP-based classification in characterizing major crops, we conducted land cover mapping for Codington and Roberts in 2012 and 2014 because they have sufficient sample data with good quality derived from the CDL. Three broad land use classes from the previous processing were divided into finer classes. "Cropland" had four classes: corn, soybean, wheat, and other crops. "Others" was separated into water, developed, and other natural (mostly contains forest and wetland). "Grassland" remained as grassland.

2.3.4   Cross-comparison between RFC model outputs and the Cropland Data Layer

We compared estimated areas from RFC models with those from re-classified CDL, in addition to using conventional accuracy metrics to evaluate the performance of classifications. It is important to note that the CxQ modeling may fail due to a lack of valid Landsat observations yielding "no-data" (section 2.3.2, Figure 2.3). These "no-data" pixels were not classified in RFC models and generated gaps in the output. We masked the reclassified CDL by "valid pixels" from RFC to allow cross-comparison between the two datasets. Due to variation in quantity and quality of Landsat data over time, the RFC outputs contain gaps in different places across the years. Therefore, direct comparison of land cover area between annual RFC outputs is not appropriate and do not indicate actual changes in the study area. However, since estimation from RFC and CDL each year are based on the same 'valid pixels' raster, we can compare changes in land cover areas as estimated by RFC and by the re-classified CDL. This indirect measurement gave us an idea about how well RFC models performed.
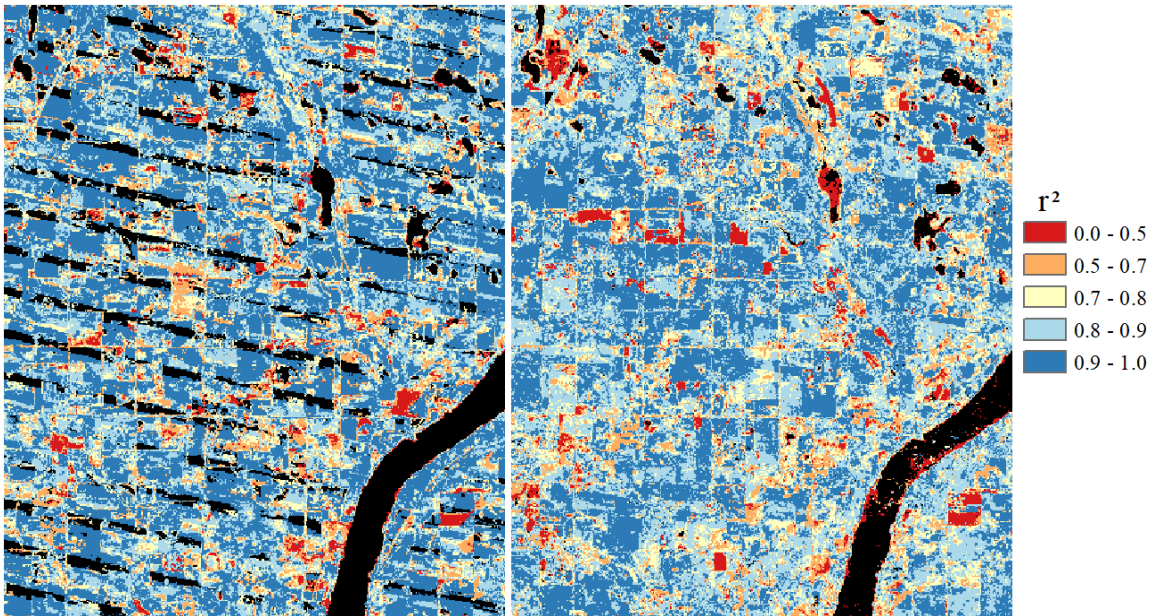
Figure 2.3. Coefficients of determination (r^2) of the CxQ models for a sample area in Roberts County in 2012 (left) and 2014 (right). Black pixels indicate that no CxQ model was fitted due to a lack of valid observations. There were almost no observations from Landsat 5 in 2012 (end of operation), and the Scan Line Corrector (SLC) issue of Landsat 7 causes missing data generating the diagonal stripes.

2.4   Results

2.4.1   Increasing Sample Size with No Control on Designs

We examined how the RFC responses to different sample sizes created by collecting pixels within 300 m to 1500 m from reference points. Table 2.4 contains the accuracy metrics means, estimated from running the RFC 260 times (20 trials x 13 sample scenarios) for each count-year. Accuracy metrics indicate good performance of the RFC for the four counties in 2006, 2012, and 2014. Performance in 2012 is slightly weaker than the other two years, possibly due to drier, warmer weather and early crop planting in eastern SD in 2012. Standard deviation values are always less than 5% of mean values for all cases indicating that different sample sizes did not significantly affect RFC accuracy (Table 2.4).

Figure 2.4 presents predicted areas (mean and standard deviation from 20 runs) for each land cover type by the RFC based on sample datasets collected within 300 m to 1500 m (in 100-m increments) of reference points. Although accuracy metrics of those models are all quite good with only small variations, their predicted values can be very different from each other and from the CDL value, especially for the "grassland" and "others" categories. Among the three classes, predicted cropland areas are the most consistent among different sample scenarios and they are closest to CDL values. There is no surprise that cropland is the best predicted class due to more pronounced phenological characteristics and the higher accuracy of crops in the CDL. The "others" class is the least accurate group; again, not surprising as there are many non-vegetated surfaces included in this heterogeneous residual category. Figure 2.4 also indicates that larger sample size, in some cases, does not lead to better prediction: predicted values departed from the baseline at longer buffer distances. Across all trials, the RFC generally overestimated land cover class with the largest area.

Table 2.4. Accuracy assessment of the RFC models with no control on design. OA is overall accuracy; PA is producer's accuracy; UA(σ) is user's accuracy and its standard deviation in percent to mean value; and kappa is Cohen's kappa.

Year	kappa	OA(σ)	Cropland		Grassland		Others	
			PA	UA	PA	UA	PA	UA
(a) **Bon Homme**								
2006	0.96	0.98 (0.08%)	0.98	0.99	0.91	0.95	0.99	0.96
2012	0.87	0.94 (0.64%)	0.93	0.98	0.93	0.91	0.96	0.84
2014	0.98	0.99 (0.07%)	0.99	0.99	0.93	0.95	0.99	0.98
(b) **Codington**								
2006	0.98	0.99 (0.06%)	0.99	0.99	0.96	0.94	0.98	0.98
2012	0.81	0.90 (1.08%)	0.86	0.99	0.89	0.87	0.97	0.77
2014	0.97	0.98 (0.37%)	0.99	0.99	0.92	0.88	0.97	0.98
(c) **Roberts**								
2006	0.96	0.97 (0.18%)	0.98	0.98	0.97	0.96	0.97	0.96
2012	0.87	0.92 (0.59%)	0.88	0.99	0.94	0.94	0.97	0.83
2014	0.95	0.97 (0.26%)	0.98	0.99	0.95	0.92	0.95	0.96
(d) **Walworth**								
2006	0.93	0.96 (0.39%)	0.95	0.98	0.98	0.96	0.90	0.96
2012	0.89	0.94 (1.13%)	0.76	0.99	0.99	0.97	0.92	0.82
2014	0.96	0.98 (0.18%)	0.99	0.99	0.99	0.97	0.91	0.98
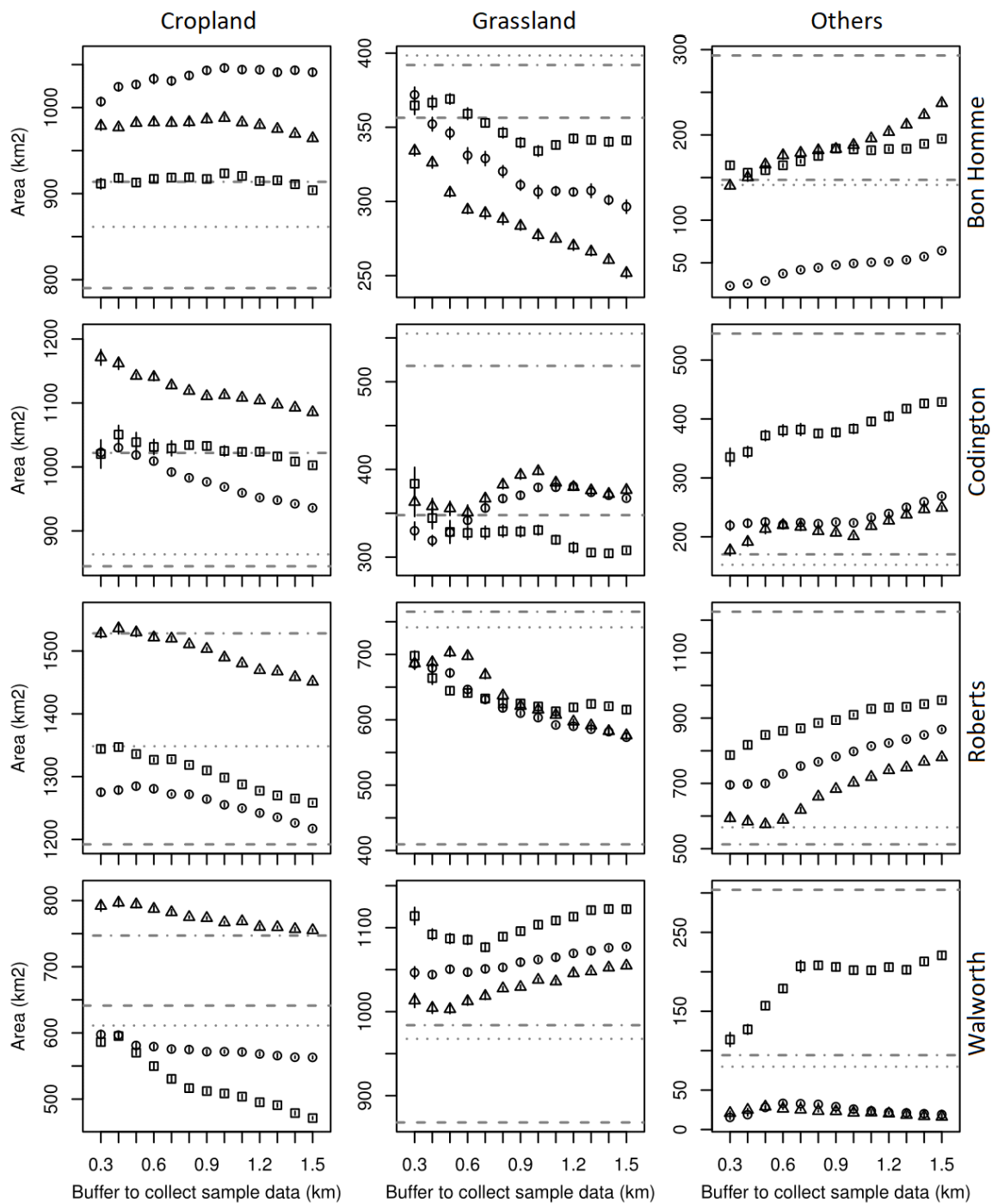
Figure 2.4. Predicted areas for each land cover type (mean ± 1σ) by the RFC in 2006 (square), 2012 (circle), and 2014 (triangle). Dashed, dotted, and dot-dash lines present estimated areas using re-classified 2006, 2012, and 2014 CDL, respectively, as baselines for evaluation.

2.4.2 "Same Distribution" vs. "Same Size" Models

Table 2.5. Accuracy assessment of RFC for "Same Distribution" and "Same Size" models. OA is overall accuracy; PA is producer's accuracy; UA is user's accuracy; and kappa is Cohen's kappa. Accuracy metrics of less than 0.7 are highlighted in **bold**.

| | SAME DISTRIBUTION | | | | | | | | SAME SIZE | | | | | | | |
| Year | | **Crop** | | **Grass** | | **Others** | | | | **Crop** | | **Grass** | | **Others** | |
	kappa	OA	PA	UA	PA	UA	PA	UA	kappa	OA	PA	UA	PA	UA	PA	UA
	(a) Bon Homme															
2006	0.95	0.97	0.98	0.99	0.95	0.95	0.93	0.93	0.94	0.96	0.97	0.98	0.95	0.95	0.95	0.94
2012	0.90	0.94	0.96	0.95	0.97	0.93	0.78	0.93	0.90	0.93	0.90	0.92	0.98	0.94	0.91	0.94
2014	0.96	0.97	0.99	0.99	0.97	0.96	0.90	0.92	0.94	0.96	0.98	0.99	0.96	0.95	0.95	0.95
	(b) Codington															
2006	0.95	0.97	0.98	0.99	0.94	0.94	0.95	0.95	0.95	0.96	0.98	0.99	0.96	0.95	0.95	0.95
2012	0.91	0.95	0.98	0.99	0.96	0.92	0.77	0.85	0.88	0.92	0.96	0.98	0.91	0.88	0.89	0.90
2014	0.93	0.96	0.99	0.99	0.96	0.92	0.77	0.89	0.88	0.92	0.98	0.99	0.92	0.86	0.86	0.92
	(c) Roberts															
2006	0.91	0.95	0.95	0.96	0.92	0.92	0.94	0.94	0.92	0.95	0.95	0.96	0.97	0.95	0.92	0.93
2012	0.91	0.94	0.98	0.97	0.94	0.92	0.85	0.89	0.89	0.93	0.96	0.96	0.93	0.92	0.89	0.90
2014	0.91	0.94	0.97	0.98	0.96	0.92	0.84	0.87	0.89	0.93	0.95	0.98	0.94	0.90	0.89	0.90
	(d) Walworth															
2006	0.87	0.92	0.95	0.95	0.94	0.93	0.78	0.82	0.86	0.90	0.94	0.95	0.90	0.89	0.88	0.87
2012	0.90	0.95	0.99	0.99	0.99	0.93	**0.23**	0.78	0.82	0.88	0.98	0.99	0.82	0.83	0.84	0.83
2014	0.92	0.96	0.99	0.99	0.98	0.94	**0.40**	0.84	0.86	0.90	0.99	0.99	0.85	0.86	0.86	0.86

Table 2.5 contains the RFC models accuracy metrics using "Same Distribution" and "Same Size" sample datasets (mean values calculated from 5x20=100 models). The high accuracy values indicate that the RFC performed well in almost every case, except

producer's accuracies of "others" class for 2012 and 2014 in Walworth County (bold in Table 2.5). Both producer's and user's accuracies of cropland are consistently higher than the other two classes in all years and every county. The "Same Distribution" RFC tends to perform better than 'Same Size' models, as shown in higher overall accuracy as well as kappa scores.

Figure 2.5 provides the RFC performance of "Same Distribution" and "Same Size" sample scenarios for Codington. In both cases, these RFC models seemed to have better cropland and grassland area estimates than the models from Figure 2.4 (*i.e.*, increasing sample size due to larger buffer and no control on sample designs) as predicted values of control scenarios are located closer to the baseline value from CDL. Overall, "Same Size" models tend to underestimate cropland area compared to "Same Distribution" models. However, there is no clear improvement between "Same Distribution" and "Same Size" in cropland area estimation, as one scenario can be better in some cases but worse in the others (Figure 2.5, S2.1-S2.3). However, "Same Distribution" models clearly perform better for the other two groups: the "Same Distribution" values (squares) tend to locate closer to the CDL baseline than the "Same Size" values (circles).
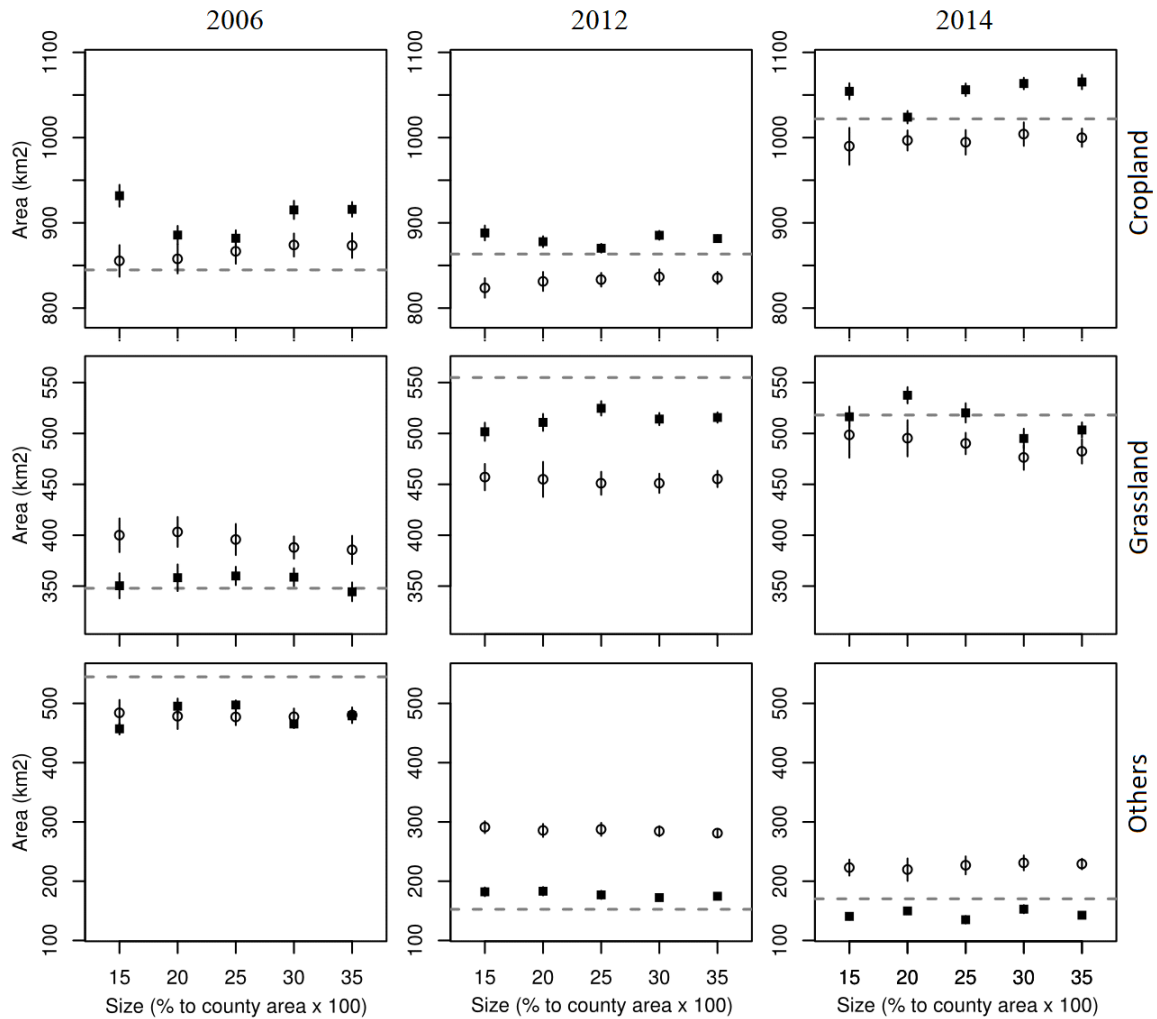
Figure 2.5. Area of each land cover type (mean ± 1σ) estimated by 'Same Distribution' (square) and 'Same Size' (circle) RFC models for Codington. Dashed lines present areas of corresponding land cover types estimated by re-classified CDL.

2.4.3   Influence of Variable Importance on Model Performance

Published studies have indicated that although the RFC can work with a large amount of input variable, high dimensional input may weaken performance of the RFC (*Millard & Richardson, 2013, 2015*). To better understand the output of previous analyses and to improve the RFC performance, we evaluated variable importance from 720 RFC models (4 counties x 3 years x 3 sample sizes x 20 runs) in the 'Same Distribution' models

(since these had performed better in earlier evaluation scenarios) in seven groups including four by-county groups (Roberts, Codington, Walworth, and Bon Homme) and three by-year groups (2006, 2012, 2014). In each group, we calculated the sum of "Mean Decrease Gini" to evaluate variable importance. To treat variable importance in each group equally, we graded each variable by its "Mean Decrease in Gini" from 16 to 1 in descending order of importance. A higher summation of grade over the seven groups (rescaled to 0-1) indicated a more overall important variable. From the seven groups, we also counted a number of times each variable placed in the top eight ranked important variables. The top eight important variables appear in Table 2.6. It is no surprise that the three parameter coefficients of the quadratic curve are three most important variables as shape of the fitted curve is defined by these coefficients. Other phenometrics were calculated from the fitted values of α, β, γ, including TTP, PH, HTV, xmin, xmax. Histograms of α, β, γ fitted values for each county-year are presented in Table S2.8-S2.11.

We compared performance of the RFC with all 16 variables ('full') and the top eight important variables ('reduced') in estimation of cropland area for all study counties in three years due to much higher and consistent accuracy of cropland prediction over the other two land cover type. Figure 2.6 indicates that the 'reduced variable' models might perform slightly better than the 'full variable' models, but differences between the two are not striking. In 2006 and 2012, cropland areas seem to be underestimated by the reduced models compared to the full one: unfilled circles tend to fall closer to the dashed line (Figure 2.6). In 2014, estimations from both sets of models were very similar. Sample land cover maps produced by "reduced variables" models are provided in Figures S2.12-S2.15.

Table 2.6. Top eight important variables for the Random Forest Classification.

Variables	Meaning	Count (max of 7)	Grade (range: 0-1)
α	Constant component of the CxQ model	7	0.92
β	Linear component of the CxQ model	7	0.91
γ	Quadratic component of the CxQ curve	7	0.82
o_per	Ratio of 'o_fit' to the total number of observations	5	0.49
minx	AGDD at the left end of the fitted curve	5	0.47
ttp	AGDD at the max fitted EVI	5	0.46
ph	Max fitted EVI	4	0.37
o_fit	Number of observations used to fit the CxQ model	4	0.34

Considering that the reduced variable models were more likely to return the better estimation of cropland area, we used the output of those models to compare changes of cropland areas as predicted by our study with the re-classified CDL layers. Table 2.7 & 2.8 present estimated cropland for the full spatiotemporal coverage by RFC and CDL as well as differences in cropland areas between different periods. Cropland area estimates were similar for the 'reduced variable' and the 'full variable' models. The changes in cropland areas were very different between RFC and CDL for the 2006-2012 period, possibly due to (1) lower accuracy of the 2006, 2012 CDLs than the 2014 CDL (*USDA-NASS 2018*), and (2) the crop failures due to widespread drought in 2012. The changes in cropland areas were more similar for the 2012-2014 period (Table 2.8). In the 2012-2014, the CDL tends to underestimate the increase of cropland compared to RFC estimation possibly due to area of crop failure areas in 2012 detected by the RFC as "grassland" or "others".

Table 2.7. Comparison between cropland areas (in km^2) estimated by the "full variable" and the "reduced variable" RFC models and those from re-classified CDL for the four study counties.

	2006		2012		2014	
	Full	**Reduced**	**Full**	**Reduced**	**Full**	**Reduced**
BonHomme_RFC	843	796	898	881	940	936
BonHomme_CDL	790	790	861	861	914	914
Diff w/CDL (%)	6.7%	0.8%	4.3%	2.3%	2.9%	2.4%
Codington_RFC	906	860	881	879	1053	1062
Codington_CDL	845	845	863	863	1022	1022
Diff w/CDL (%)	7.3%	1.8%	2.0%	1.8%	3.0%	3.9%
Roberts_RFC	1175	1159	1228	1211	1453	1447
Roberts_CDL	1192	1192	1348	1348	1528	1528
Diff w/CDL (%)	-1.5%	-2.8%	-9.0%	-10.2%	-4.9%	-5.3%
Walworth_RFC	549	540	644	633	794	799
Walworth_CDL	641	641	611	611	474	747
Diff w/CDL (%)	-14.5%	-15.8%	5.31	3.7	6.2%	6.9%

Table 2.8. Changes in cropland (in km2) estimated by the "full variable" and "reduced variable" RFC models and percent difference from changes in cropland estimated by the re-classified CDL.

	Δ2006-2012		Δ2012-2014	
	Full	**Reduced**	**Full**	**Reduced**
BonHomme_RFC	55.33	84.63	41.86	54.47
BonHomme_CDL	71.23	71.23	52.30	52.30
Diff w/CDL(%)	-22.3%	18.8%	-20.0%	4.2%
Codington_RFC	-25.45	18.93	172.00	182.86
Codington_CDL	18.60	18.60	158.60	158.60
Diff w/CDL (%)	-236.8%	1.8%	8.5%	15.3%
Roberts_RFC	52.80	51.79	224.79	236.11
Roberts_CDL	156.19	156.19	179.66	179.66
Diff w/CDL (%)	-66.2%	-66.8%	25.1%	31.4%
Walworth_RFC	94.95	93.74	150.04	165.35
Walworth_CDL	-30.16	-30.16	136.01	136.01
Diff w/CDL (%)	-414.8%	-410.8%	10.3%	21.6%
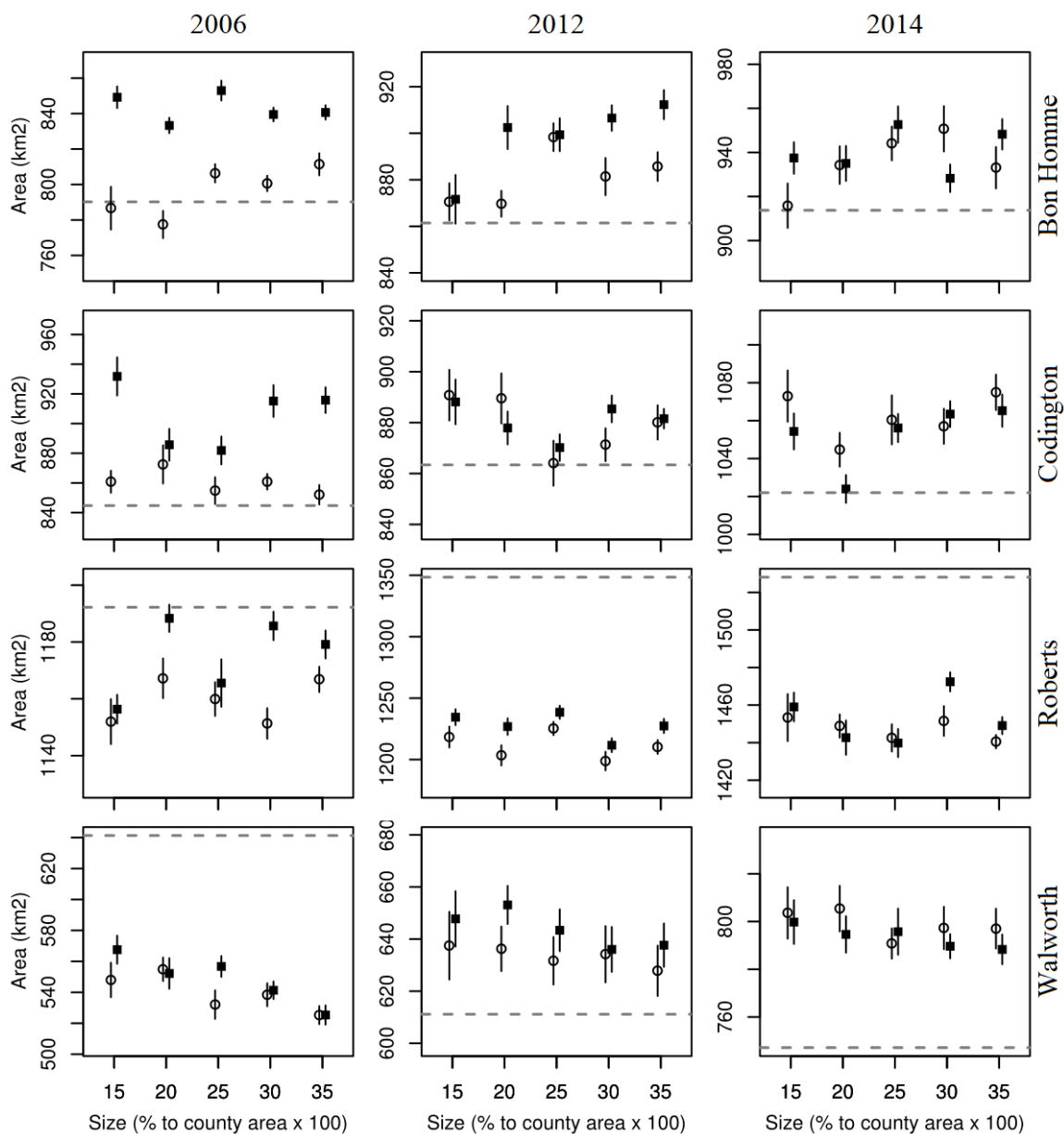
Figure 2.6. Comparison between cropland areas (mean ± 1σ) estimated by "full variable" (square) and "reduced variable" (circle) RFC models. Dashed lines present cropland areas estimated by re-classified CDL.

2.4.4   Mapping of Major Crops

We performed land cover classification with more detailed cover types (corn, soybean, wheat, other crops, grassland, other natural, developed, and water) for Roberts and Codington counties. Tables 2.9 and 2.10 show accuracy metrics and estimated land cover areas of RFC models. Overall accuracies for all four scenarios are greater than 0.75. The 2014 models performed much better than the 2012 models, not a surprise considering that the 2014 CDL has higher accuracy than the 2012 CDL. Among the land cover types, "water" has the highest user's/producer's accuracy despite the poor CxQ fit, since the EVI time series of water pixels always nearly flat. On the other hand, "other natural" and "developed" were classified with lower accuracy due to their complex phenological characteristics. While "other natural" consists of minor vegetation classes with quite different LSP pattern, e.g., forest and wetland, "developed" encompasses by heterogeneous surfaces and materials, including both vegetated and non-vegetated surfaces. Estimated land cover areas are most accurate for corn and soybean due to their distinctive phenological characteristics and high CxQ fits. Many vegetated pixels within the urban boundary are identified as either grassland or other vegetation leading to underestimation of the urban area. On the other hand, RFC models overestimated grassland area because other land cover types could be misclassified as "grassland", e.g., green spaces within the urban area, failed crop pixels or dry wetland pixels.

Table 2.9. Accuracy assessment of RFC models (mean of 60 models with sample dataset covering 0.15%, 0.25%, 0.35% of the total area) for Codington and Roberts counties. PA is producer's accuracy; UA is user's accuracy; kappa is Cohen's kappa.

	Codington 2012		Codington 2014		Roberts 2012		Roberts 2014	
Overall Accuracy	0.763		0.871		0.796		0.844	
kappa	0.696		0.837		0.743		0.804	
	PA	UA	PA	UA	PA	UA	PA	UA
Corn	0.80	0.77	0.90	0.89	0.86	0.80	0.90	0.89
Soybean	0.66	0.71	0.89	0.91	0.80	0.89	0.89	0.91
Wheat	0.69	0.76	0.93	0.90	0.74	0.81	0.92	0.88
Other Crops	0.62	0.78	0.77	0.82	0.57	0.82	0.67	0.87
Water	0.89	0.88	0.95	0.96	0.93	0.92	0.96	0.96
Developed	0.29	0.51	0.43	0.68	0.27	0.47	0.30	0.51
Other Natural	0.30	0.63	0.28	0.61	0.66	0.63	0.66	0.66
Grassland	0.92	0.79	0.93	0.84	0.89	0.81	0.91	0.83

Table 2.10. Land cover areas (in km2) estimated by RFC and by reclassified CDL and percent difference in area estimated for Codington (C) and Roberts (R) counties.

	Corn	Soybean	Wheat	Other Crops	Water	Developed	Other Natural	Grassland	Total Area
C.RFC.2012	417	296	59	44	50	51	17	637	1571
C.CDL.2012	419	307	76	63	17	98	36	555	1571
%Diff	-0.3	-3.6	-21.8	-29.6	200.7	-48.2	-54.3	14.8	--
C.RFC.2014	454	366	101	66	39	54	12	620	1721
C.CDL.2014	419	430	91	82	39	103	27	518	1709
%Diff	8.2	-15.0	11.1	-19.9	-1.0	-48.0	-56.1	19.8	--
R.RFC.2012	613	504	68	17	43	65	510	836	2656
R.CDL.2012	637	597	79	36	40	126	398	741	2654
%Diff	-3.7	-15.6	-14.3	-54.3	7.4	-48.3	28.0	12.8	--
R.RFC.2014	624	688	96	29	42	83	424	820	2806
R.CDL.2014	604	788	90	50	51	129	330	765	2807
%Diff	3.3	-12.7	6.5	-41.0	-16.5	-35.6	28.7	7.1	--

2.5 Discussion

Sensitivity of Random Forest Classification to Sample Size, Design and Quality

Our findings confirm previous findings that RFC is sensitive to sample size (*Deng & Wu, 2013; Du et al., 2015; Millard & Richardson, 2015*). However, while previous studies indicated that larger sample size increases RFC accuracy, our results show that is not always the case. Sometimes, larger sample dataset may reduce RFC accuracy depending on the quality of the sample dataset. Our study showed that a sample dataset

covering 0.15% - 0.35% of the study area is adequate to achieve an accurate land cover classification, similar to the findings of Colditz (2015).

As published studies suggest that the proportional distribution of classes in the sample dataset should represent the proportional distribution of those classes in the target population (*Dalponte et al., 2015; Millard & Richardson, 2013; 2015*), we tested RFC performance with different sample designs. Our results also showed that sample dataset created by controlling the proportional distribution of data in each land cover group ("Same Distribution") yielded better RFC performance than those with no control ("Same Size"). However, we did not find substantial differences between 'Same Distribution' and 'Same Size' models, possibly because the performance of the "Same Size" models were already quite good.

We leveraged the CDL dataset to augment the number of training samples; thus, our results strongly depend on the accuracy of the CDL layers. Although we sought pixels with correct land cover information by using reference data, avoiding edge pixels and matching information for multiple years, there is a good chance that the training sample dataset still contains a considerable but unquantifiable amount of classification error. The overall accuracy of crop classes (the most accurate classes) in South Dakota are only 61.2% in 2006, 74.7% in 2012 and 86.8% in 2014 (*USDA-NASS, 2018*). The 2006 and 2012 CDLs have lower accuracies than the 2014 CDL, possibly due to the start of a new CDL classification system in 2006 (*Boryan et al., 2011*), and the unusually early spring, subsequent drought (*Ault et al. 2013*) as well as the lack of observations due to the end of Landsat 5 operation in 2012

Accuracy Metrics and Prediction Consistency of RFC

Generally, land cover/land use studies must work with limited samples due to the high cost of conducting field observations. Thus, most attention focuses on retrieving an accurate classification model using a single sample dataset. The RFC performance depends on the actual sample dataset in hand and how it is used to train the RFC. So, an important question is how the model accuracy and predicted land cover map might change were we to have the opportunity to collect the second sample dataset independent of the first. By performing multiple RFC models on different sample sizes and scenarios, we were able to test something that has received relatively little attention in the literature (but see *Jin et al, 2014; Colditz, 2015; Millard & Richardson, 2015; Lyons et al., 2018*). Our results show that even if the RFC models have quite good accuracy, their predicted values still can be quite different from each other and from the real population. This pattern of performance suggests that an ensemble RFC output generated over multiple scenarios is important, even if there is just one training sample dataset available.

Here, we have assumed that RFC models with outputs closer to the re-classified CDL are better. However, this assumption may not be true due to low accuracies in the CDL, even in some crop classes. Despite the differences between our estimates with those from CDL, the performance of RFC in our study is quite consistent through multiple scenarios and runs. It means that predicted land cover maps derived from different sample selections would be very similar to each other, and the only option to get a better model would be to improve the quality of sample data.

Weaknesses and Prospects for the Proposed Classification Approach

Due to variation in data quality, EVI time series at different locations are often not observed at the same set of days, preventing a direct comparison between annual EVI

patterns. Here, we distilled the entire annual AGDD and EVI time series into a simple functional form so that land surface phenology at every pixel was described by the same set of parameters (though different fitted parameter coefficients). The attempt to use a single function for every land cover was only partially successful. The CxQ models fitted well most vegetated surfaces, especially for cropland pixels. Median r2 in Roberts County were approximately 0.77 in 2006, and 0.90 in 2012 and 2014 (Figure S2.16). In contrast, non-vegetated pixels had low or sometimes negative EVI leading to very poor model fits or even model failure that generates gaps in the spatial coverage. The quality of the model fits was not consistent between land cover types: higher quality for crops, medium quality for non-crop vegetation including grassland, and lower quality for non-vegetated surfaces (Figure 2.3). The quality of model fits also varied within a class due to seasonal and interannual climatic variation as well as gaps resulting from the Scan Line Corrector failure on Landsat 7 ETM+. These inconsistencies in CxQ LSP model fitting reduces the strength of using just the LSP coefficients and phenometrics to map land cover and track changes. Another limitation of LSP-based approach was the gap creation during the fitting process arising from model failure due to low number of good Landsat observations in some years and over some areas (cf. section 2.3.2). Because the gaps were produced in different places across the years, a direct comparison between yearly RFC outputs was not appropriate in most cases, without spatio-temporal interpolation, something that we chose not to do. Lastly, except for croplands with very distinctive and consistent phenological shapes for the CxQ model to capture, much non-crop vegetation can exhibit similar phenological curves and phenometrics, making it difficult to map that those land cover types accurately. For example, CxQ curves fitted for grasslands can appear like ones fitted to wetlands

during drier weather. Many lawns or recreational fields in urban areas can be misclassified as grassland, which is not completely wrong, but requires contextual information to be classified more accurately as "developed".

Despite all limitations as discussed above, the future of LSP-based classification is promising for several reasons. First, although LSP-based classification may not work well for non-cropped surfaces, this method may still be useful for detailed crop mapping. Our sample analyses demonstrated that we could discriminate quite well between three major commodity crop types: corn, soybean, and wheat. Second, here we used only LSP-based variables as input to the RFC; however, our approach could be expanded to take advantage of additional time series from other spectral bands and specific indices tailored to reveal different aspects of the dynamic land surface. Third, although the LSP-fitting process generated gaps in the output, bringing together complementary sensor datastreams, such as from Sentinel 2A and 2B, could increase the temporal density of observations and thereby increase the quality of the time series to be modeled. In areas with persistent cloud cover, it may be possible to leverage "CubeSats" with datastreams at even higher spatial and temporal but lower spectral resolution to infill gaps intelligently using physical-based modeling (*Houborg & McCabe, 2018*).

2.6   Conclusions

We have introduced a new approach to characterize land cover/land use using Landsat image time series, *first*, by modeling the seasonality of the vegetated land surface at the pixel scale using Landsat and MODIS time series linked with a simple parametric model of land surface phenology and, *second*, by submitting the per-pixel fitted model coefficients and associated phenometrics to a random forest classifier (RFC). The LSP-

based classification was demonstrated through a comparative analysis on four counties in South Dakota over three separate years. Our results showed that the classification based only on LSP could accurately differentiate major commodity crops, but it had limited accuracy for non-vegetated classes. To evaluate the differential performance of RFC models arising from the training sample dataset, we created several distinct training sample datasets from reference points and CDL layers using different sample sizes and designs. Among all sampling designs, the "Same Distribution" models tended to yield marginally better predictions of land cover. In addition, the RFC models used only the most eight important variables (*i.e*., the three fitted parameter coefficients (α, β, γ); the maximum model EVI; the AGDD at the maximum modeled EVI; the number of observations used to fit CxQ model; and the number of valid observations) had slightly higher accuracy compared to those using all variables. Although all the RFC models achieved quite good accuracy metrics, there were substantial differences in predicted land cover areas between alternative RFC models. As each sample dataset will yield different prediction, we suggest that future supervised classification studies should carefully examine output stability. Ideally, the final prediction should be an ensemble of all outputs retrieved by performing classification with multiple independent samples. However, even if only one sample dataset is available, the final output will still benefit from iterating training/validation and prediction multiple times.

Although the current results show limitations of the LSP-based classification for non-vegetated surfaces, the prospect for using land surface phenology variables in land cover/land use classification is promising. First, the LSP-based classification offered an accurate, consistent way to identify cropland areas and changes between major crops (*e.g*.,

corn-soybean rotation) across years. Second, it should be possible to combine LSP variables with other indices and spectral information to improve classification accuracy and consistency, and advance the monitoring of land change using Landsat and comparable sensor datastreams and archives.

2.7 Acknowledgements

2.8 References

Ahl, D.E., Gower, S.T., Burrows, S.N., Shabanov, N.V., Myneni, R.B., Knyazikhin, Y., 2006. Monitoring spring canopy phenology of a deciduous broadleaf forest using MODIS. Remote Sens. Environ. 104 (1), 88–95.

Ault, T.R., Henebry, G.M., de Beurs, K.M., Schwartz, M.D., Betancourt, J.L., Moore, D., 2013. The false spring of 2012, earliest in North American record. EOS Trans. Am. Geophys. Union 94 (20), 181–182.

Baumann, M., Ozdogan, M., Richardson, A.D., Radeloff, V.C., 2017. Phenology from Landsat when data is scarce: using MODIS and dynamic time-warping to combine multi-year Landsat imagery to derive annual phenology curves. Int. J. Appl. Earth Obs. Geoinf. 54, 72–83.

Beck, P.S., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens. Environ. 100 (3), 321–334.

de Beurs, K.M., Henebry, G.M., 2004. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. Remote Sens. Environ. 89 (4), 497–509.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department Of Agriculture, National Agricultural Statistics Service, Cropland Data Layer program. Geocarto Int. 26 (5), 341–358.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Claassen, R., Carriazo, F., Cooper, J.C., Hellerstein, D., Ueda, K., 2011. Grassland to cropland conversion in the Northern Plains. In: Economic Research Report. 120.

Clerici, N., Weissteiner, C.J., Gerard, F., 2012. Exploring the use of MODIS NDVI-based phenology indicators for classifying forest general habitat categories. Remote Sens. 4 (6), 1781–1803.

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 Sens. 7 (8), 9655–9681.

Dalponte, M., Orka, H.O., Gobakken, T., Gianelle, D., Næsset, E., 2013. Tree species classification in boreal forests with hyperspectral data. IEEE Trans. Geosci. Remote Sens. 51 (5), 2632–2645.

Deng, C., Wu, C., 2013. The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. ISPRS J. Photogramm. Remote Sens. 86, 100–110.

Du, P., Samat, A., Waske, B., Liu, S., Li, Z., 2015. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS J. Photogramm. Remote Sens. 105, 38–53.

Faber, S., Rundquist, S., Male, T., 2012. Plowed under: How Crop Subsidies Contribute to Massive Habitat Losses. Environmental Working Group, Washington, DC.

Fargione, J.E., Cooper, T.R., Flaspohler, D.J., Hill, J., Lehman, C., McCoy, T., ... Tilman, D., 2009. Bioenergy and wildlife: threats and opportunities for grassland conservation. Bioscience 59 (9), 767–777.

Franklin, S.E., Ahmed, O.S., Wulder, M.A., White, J.C., Hermosilla, T., Coops, N.C., 2015. Large area mapping of annual land cover dynamics using multi-temporal change detection and classification of Landsat time-series data. Can. J. Remote. Sens. 41, 293–314.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114 (1), 168–182.

Gu, Y., Brown, J.F., Miura, T., Van Leeuwen, W.J., Reed, B.C., 2010. Phenological classification of the United States: a geographic framework for extending multi-sensor timeseries data. Remote Sens. 2 (2), 526–544.

Henebry, G.M., de Beurs, K.M., 2013. Remote sensing of land surface phenology: a prospectus. In: Schwartz, M.D. (Ed.), Phenology: An Integrative Environmental Science, 2e. Springer, pp. 385–411.

Houborg, R., McCabe, M.F., 2018. A cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sens. Environ. 209, 211–226.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1–2), 195–213.

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. Int. J. Remote Sens. 35 (6), 2067–2081.

Johnston, C.A., 2013. Wetland losses due to row crop expansion in the Dakota Prairie Pothole Region. Wetlands 33 (1), 175–182.

Johnston, C.A., 2014. Agricultural expansion: land use shell game in the US Northern Plains. Landsc. Ecol. 29 (1), 81–95.

Krehbiel, C., Henebry, G.M., 2016. A comparison of multiple datasets for monitoring thermal time in urban areas over the US Upper Midwest. Remote Sens. 8 (4), 297.

Krehbiel, C.P., Jackson, T., Henebry, G.M., 2016. Web-enabled Landsat data time series for monitoring urban heat island impacts on land surface phenology. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 9 (5), 2043–2050.

Krehbiel, C., Zhang, X., Henebry, G.M., 2017. Impacts of thermal time on land surface phenology in urban areas. Remote Sens. 9 (5), 499.

Lark, T.J., Salmon, J.M., Gibbs, H.K., 2015. Cropland expansion outpaces agricultural and biofuel policies in the United States. Environ. Res. Lett. 10 (4), 044003.

Lark, T.J., Mueller, R.M., Johnson, D.M., Gibbs, H.K., 2017. Measuring land-use and land cover change using the US Department of Agriculture's cropland data layer: cautions and recommendations. Int. J. Appl. Earth Obs. Geoinf. 62, 224–235.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.

Lloyd, D., 1990. A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. Int. J. Remote Sens. 11 (12), 2269–2279.

Lyons, M.B., Keith, D.A., Phinn, S.R., Mason, T.J., Elith, J., 2018. A comparison of resampling methods for remote sensing classification and accuracy assessment. Remote Sens. Environ. 208, 145–153.

Meehan, T.D., Hurlbert, A.H., Gratton, C., 2010. Bird communities in future bioenergy landscapes of the Upper Midwest. Proc. Natl. Acad. Sci. 107 (43), 18533–18538.

Millard, K., Richardson, M., 2013. Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. Can. J. Remote. Sens. 39 (4), 290–307.

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 Sens. 7 (7), 8489–8515.

Montgomery, D.R., 2007. Soil erosion and agricultural sustainability. Proc. Natl. Acad. Sci. 104 (33), 13268–13272. Mueller, R., Seffrin, R., 2006. New methods and satellites: a program update on the NASS cropland data layer acreage program. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 36, 97–102.

Mutter, M., Pavlacky, D.C., Lanen, N.J., Grenyer, R., 2015. Evaluating the impact of gas extraction infrastructure on the occupancy of sagebrush-obligate songbirds. Ecol. Appl. 25 (5), 1175–1186.

NASA LP DAAC, 2013. MODIS terra/aqua land surface temperature/emissivity 8-day L3 global 1 km SIN grid. Version 5. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. https://lpdaac.usgs.gov, Accessed date: 15 January 2017.

Otto, C.R., Roth, C.L., Carlson, B.L., Smart, M.D., 2016. Land-use change reduces habitat suitability for supporting managed honey bee colonies in the Northern Great Plains. Proc. Natl. Acad. Sci. 113 (37), 10430–10435.

Preston, T.M., Kim, K., 2016. Land cover changes associated with recent energy development in the Williston Basin; Northern Great Plains, USA. Sci. Total Environ. 566, 1511–1518.

Qader, S.H., Dash, J., Atkinson, P.M., Rodriguez-Galiano, V., 2016. Classification of vegetation type in Iraq using satellite-based phenological parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 9 (1), 414–424.

Rashford, B.S., Walker, J.A., Bastian, C.T., 2011. Economics of grassland conversion to cropland in the Prairie Pothole Region. Conserv. Biol. 25 (2), 276–284.

Reed, B.C., Brown, J.F., Vanderzee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., 1994. Measuring phenological variability from satellite imagery. J. Veg. Sci. 5 (5), 703–714.

Reitsma, K.D., Dunn, B.H., Mishra, U., Clay, S.A., Desutter, T., Clay, D.E., 2015. Land-use change impact on soil sustainability in a climate and vegetation transition zone. Agron. J. 107 (6), 2363–2372.

Reitsma, K.D., Clay, D.E., Clay, S.A., Dunn, B.H., Reese, C., 2016. Does the US cropland data layer provide an accurate benchmark for land-use change estimates? Agron. J. 108 (1), 266–272.

Schwartz, M.D., Reed, B.C., 1999. Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. Int. J. Remote Sens. 20 (17), 3451–3457.

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., ... Yu, T.H., 2008. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319 (5867), 1238–1240.

Singh, N., Kline, K.L., Efroymson, R.A., Bhaduri, B., O'Banion, B., 2017. Uncertainty in estimates of bioenergy-induced land use change: the impact of inconsistent land cover data sets and land class definitions. In: Bioenergey and Land Use Change. 231. pp. 143.

Stephens, S.E., Walker, J.A., Blunck, D.R., Jayaraman, A., Naugle, D.E., Ringelman, J.K., Smith, A.J., 2008. Predicting risk of habitat conversion in native temperate grasslands. Conserv. Biol 22 (5), 1320–1330.

USDA-FSA (United States Department of Agriculture, Farm Service Agency), 2013. National Agriculture Imagery Program (NAIP). Retrieved from: https://www.fsa.usda.gov/programs-and-services/aerial-photography/index.

USDA-NASS (U.S. Department of Agriculture, National Agriculture Statistics Service), 2018. Cropland data layer metadata. Retrieved from: https://www.nass.usda.gov

USGS (U.S. Geological Survey), 2016. Landsat update special issue 4: Landsat Collection 1 level-1 data processing starting soon. Retrieved from: https://landsat.usgs.gov

USGS (U.S. Geological Survey), 2017. Product guide: Landsat surface reflectance-derived spectral indices. Retrieved from: https://landsat.usgs.gov

Vaché, K.B., Eilers, J.M., Santelmann, M.V., 2002. Water quality modeling of alternative agricultural scenarios in the US Corn Belt. J. Am. Water Resour. Assoc. 38 (3), 773–787.

White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. Glob. Biogeochem. Cycles 11 (2), 217–234.

Wimberly, M.C., Narem, D.M., Bauman, P.J., Carlson, B.T., Ahlering, M.A., 2018. Grassland connectivity in fragmented agricultural landscapes of the north-central United States. Biol. Conserv. 217, 121–130.

World Wildlife Fund (WWF), 2016. Plow print report. Retrieved from:

Wright, C.K., Wimberly, M.C., 2013a. Recent land use change in the Western Corn Belt threatens grasslands and wetlands. Proc. Natl. Acad. Sci. 110 (10), 4134–4139.

Wright, C.K., Wimberly, M.C., 2013b. Reply to Kline et al.: Cropland data layer provides a valid assessment of recent grassland conversion in the Western Corn Belt. Proc. Natl. Acad. Sci. 110 (31), E2864.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., ... Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84 (3), 471–475.

Zhong, L., Hawkins, T., Biging, G., Gong, P., 2011. A phenology-based approach to map crop types in the San Joaquin Valley, California. Int. J. Remote Sens. 32 (22), 7777–7804.

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.

2.9   Supplementary

Table S1.1. Reclassification of CDL land use/land cover classes.

CDL Code	CDL Class	New Code	New Class	CDL Code	CDL Class	New Code	New Class
0	Background	NA	NoData	81	Clouds/No Data	3	Others
1	Corn	1	Cropland	82	Developed	3	Others
2	Cotton	1	Cropland	83	Water	3	Others
3	Rice	1	Cropland	87	Wetlands	3	Others
4	Sorghum	1	Cropland	88	Nonag/Undefined	3	Others
5	Soybeans	1	Cropland	92	Aquaculture	3	Others
6	Sunflower	1	Cropland	111	Open Water	3	Others
10	Peanuts	1	Cropland	112	Perennial Ice/Snow	3	Others
11	Tobacco	1	Cropland	121	Developed/Open Space	3	Others
12	Sweet Corn	1	Cropland	122	Developed/Low Intensity	3	Others
13	Pop or Orn Corn	1	Cropland	123	Developed/Med Intensity	3	Others
14	Mint	1	Cropland	124	Developed/High Intensity	3	Others

CDL Code	CDL Class	New Code	New Class	CDL Code	CDL Class	New Code	New Class
21	Barley	1	Cropland	131	Barren	3	Others
22	Durum Wheat	1	Cropland	141	Deciduous Forest	3	Others
23	Spring Wheat	1	Cropland	142	Evergreen Forest	3	Others
24	Winter Wheat	1	Cropland	143	Mixed Forest	3	Others
25	Other Small Grains	1	Cropland	152	Shrubland	3	Others
26	Dbl Crop WinWht/Soybeans	1	Cropland	176	Grass/Pasture	2	Grassland
27	Rye	1	Cropland	190	Woody Wetlands	3	Others
28	Oats	1	Cropland	195	Herbaceous Wetlands	3	Others
29	Millet	1	Cropland	204	Pistachios	1	Cropland
30	Speltz	1	Cropland	205	Triticale	1	Cropland
31	Canola	1	Cropland	206	Carrots	1	Cropland
32	Flaxseed	1	Cropland	207	Asparagus	1	Cropland
33	Safflower	1	Cropland	208	Garlic	1	Cropland
34	Rape Seed	1	Cropland	209	Cantaloupes	1	Cropland
35	Mustard	1	Cropland	210	Prunes	1	Cropland

CDL Code	CDL Class	New Code	New Class	CDL Code	CDL Class	New Code	New Class
36	Alfalfa	1	Cropland	211	Olives	1	Cropland
37	Other Hay/Non Alfalfa	2	Grassland	212	Oranges	1	Cropland
38	Camelina	1	Cropland	213	Honeydew Melons	1	Cropland
39	Buckwheat	1	Cropland	214	Broccoli	1	Cropland
41	Sugarbeets	1	Cropland	216	Peppers	1	Cropland
42	Dry Beans	1	Cropland	217	Pomegranates	1	Cropland
43	Potatoes	1	Cropland	218	Nectarines	1	Cropland
44	Other Crops	1	Cropland	219	Greens	1	Cropland
45	Sugarcane	1	Cropland	220	Plums	1	Cropland
46	Sweet Potatoes	1	Cropland	221	Strawberries	1	Cropland
47	Misc Vegs & Fruits	1	Cropland	222	Squash	1	Cropland
48	Watermelons	1	Cropland	223	Apricots	1	Cropland
49	Onions	1	Cropland	224	Vetch	1	Cropland
50	Cucumbers	1	Cropland	225	Dbl Crop WinWht/Corn	1	Cropland
51	Chick Peas	1	Cropland	226	Dbl Crop Oats/Corn	1	Cropland

CDL Code	CDL Class	New Code	New Class	CDL Code	CDL Class	New Code	New Class
52	Lentils	1	Cropland	227	Lettuce	1	Cropland
53	Peas	1	Cropland	229	Pumpkins	1	Cropland
54	Tomatoes	1	Cropland	230	Dbl Crop Lettuce/Durum Wht	1	Cropland
55	Caneberries	1	Cropland	231	Dbl Crop Lettuce/Cantaloupe	1	Cropland
56	Hops	1	Cropland	232	Dbl Crop Lettuce/Cotton	1	Cropland
57	Herbs	1	Cropland	233	Dbl Crop Lettuce/Barley	1	Cropland
58	Clover/Wildflowers	1	Cropland	234	Dbl Crop Durum Wht/Sorghum	1	Cropland
59	Sod/Grass Seed	1	Cropland	235	Dbl Crop Barley/Sorghum	1	Cropland
60	Switchgrass	1	Cropland	236	Dbl Crop WinWht/Sorghum	1	Cropland
61	Fallow/Idle Cropland	3	Others	237	Dbl Crop Barley/Corn	1	Cropland
63	Forest	3	Others	238	Dbl Crop WinWht/Cotton	1	Cropland
64	Shrubland	2	Grassland	239	Dbl Crop Soybeans/Cotton	1	Cropland
65	Barren	2	Grassland	240	Dbl Crop Soybeans/Oats	1	Cropland
66	Cherries	1	Cropland	241	Dbl Crop Corn/Soybeans	1	Cropland
67	Peaches	1	Cropland	242	Blueberries	1	Cropland

CDL Code	CDL Class	New Code	New Class	CDL Code	CDL Class	New Code	New Class
68	Apples	1	Cropland	243	Cabbage	1	Cropland
69	Grapes	1	Cropland	244	Cauliflower	1	Cropland
70	Christmas Trees	1	Cropland	245	Celery	1	Cropland
71	Other Tree Crops	1	Cropland	246	Radishes	1	Cropland
72	Citrus	1	Cropland	247	Turnips	1	Cropland
74	Pecans	1	Cropland	248	Eggplants	1	Cropland
75	Almonds	1	Cropland	249	Gourds	1	Cropland
76	Walnuts	1	Cropland	250	Cranberries	1	Cropland
77	Pears	1	Cropland	254	Dbl Crop Barley/Soybeans	1	Cropland

Table S2.2. Pairing day-of-year (DOY) of EVI and AGDD time series. EVI value is paired with AGDD value for the nearest composing period. For example, EVI of DOYs from 11 to 14 were coupled with AGDD composite 3 which is tied to DOY 17. If EVI was obtained right in the middle of the two MODIS compositing periods, paired AGDD is calculated as the average of the two AGDD compositing periods. For example, EVI of DOY 5 is paired with mean AGDD of composite 1 and 2 (DOY 1 and DOY 9).

AGDD 8-d Composite	AGDD DOY	EVI DOY start	EVI DOY end	AGDD 8-d Composite	AGDD DOY	EVI DOY start	EVI DOY end
1	1	1	4	24	185	182	188
2	9	6	12	25	193	190	196
3	17	14	20	26	201	198	204
4	25	22	28	27	209	206	212
5	33	30	36	28	217	214	220
6	41	38	44	29	225	222	228
7	49	46	52	30	233	230	236
8	57	54	60	31	241	238	244
9	65	62	68	32	249	246	252
10	73	70	76	33	257	254	260
11	81	78	84	34	265	262	268
12	89	86	92	35	273	270	276
13	97	94	100	36	281	278	284
14	105	102	108	37	289	286	292
15	113	110	116	38	297	294	300
16	121	118	124	39	305	302	308
17	129	126	132	40	313	310	316
18	137	134	140	41	321	318	324
19	145	142	148	42	329	326	332
20	153	150	156	43	337	334	340
21	161	158	164	44	345	342	348
22	169	166	172	45	353	350	356
23	177	174	180	46	361	358	364
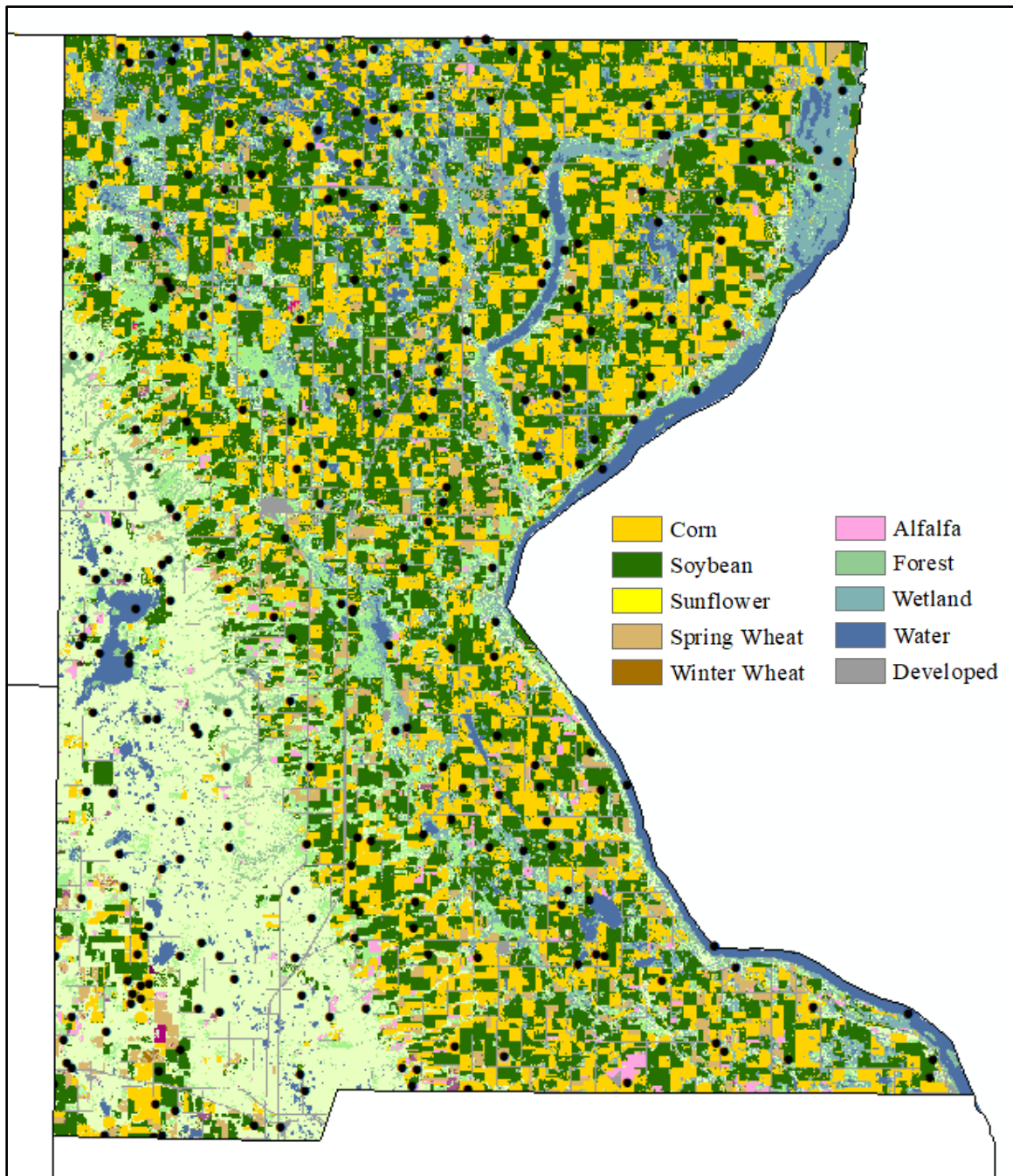
Figure S2.1. Roberts County 2014 CDL and reference points (black circles).
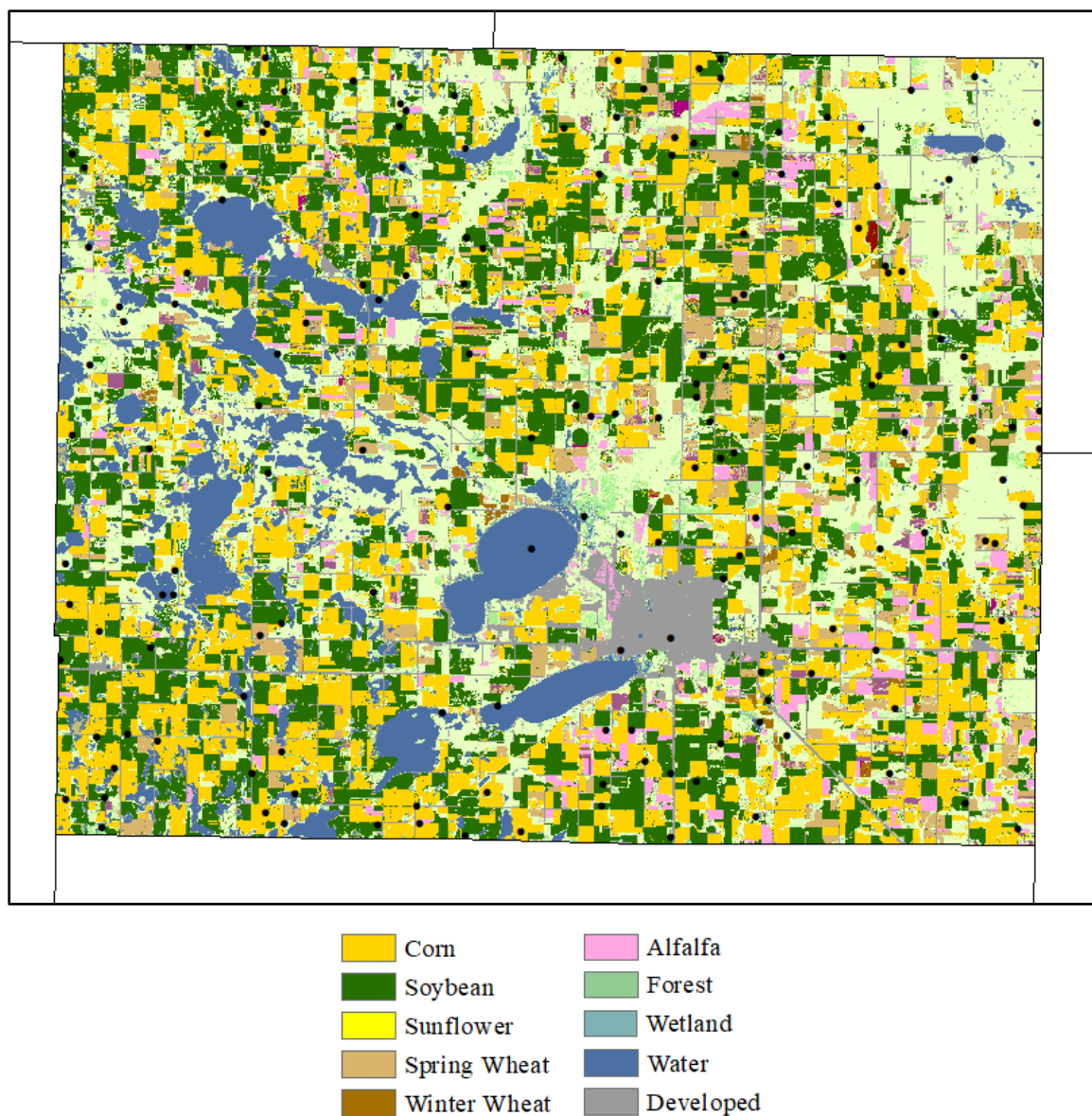
Figure S2.2. Codington County 2014 CDL and reference points (black circles).

	Corn		Alfalfa
	Soybean		Forest
	Sunflower		Wetland
	Spring Wheat		Water
	Winter Wheat		Developed
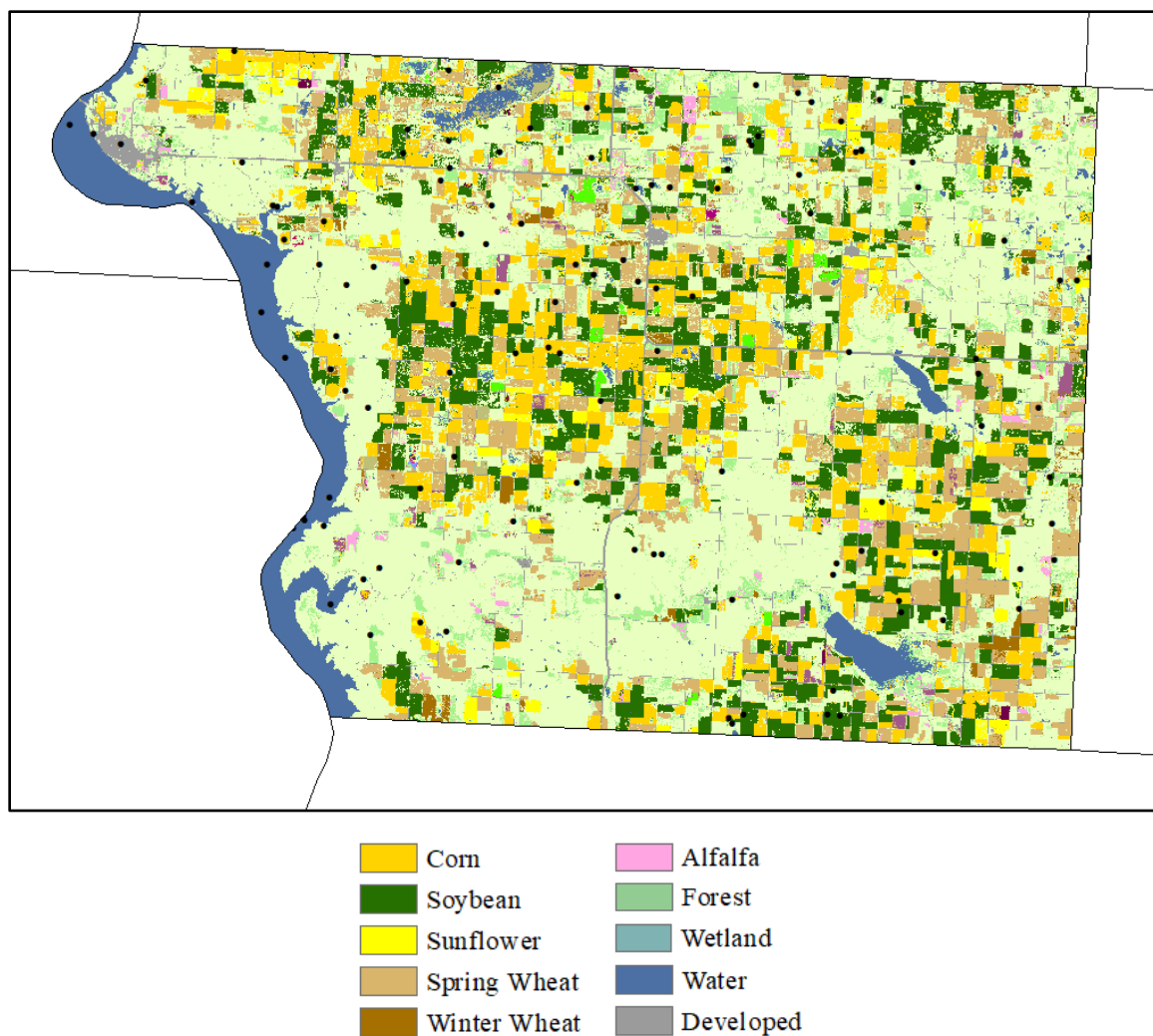
Figure S2.3. Walworth County 2014 CDL and reference points (black circles).
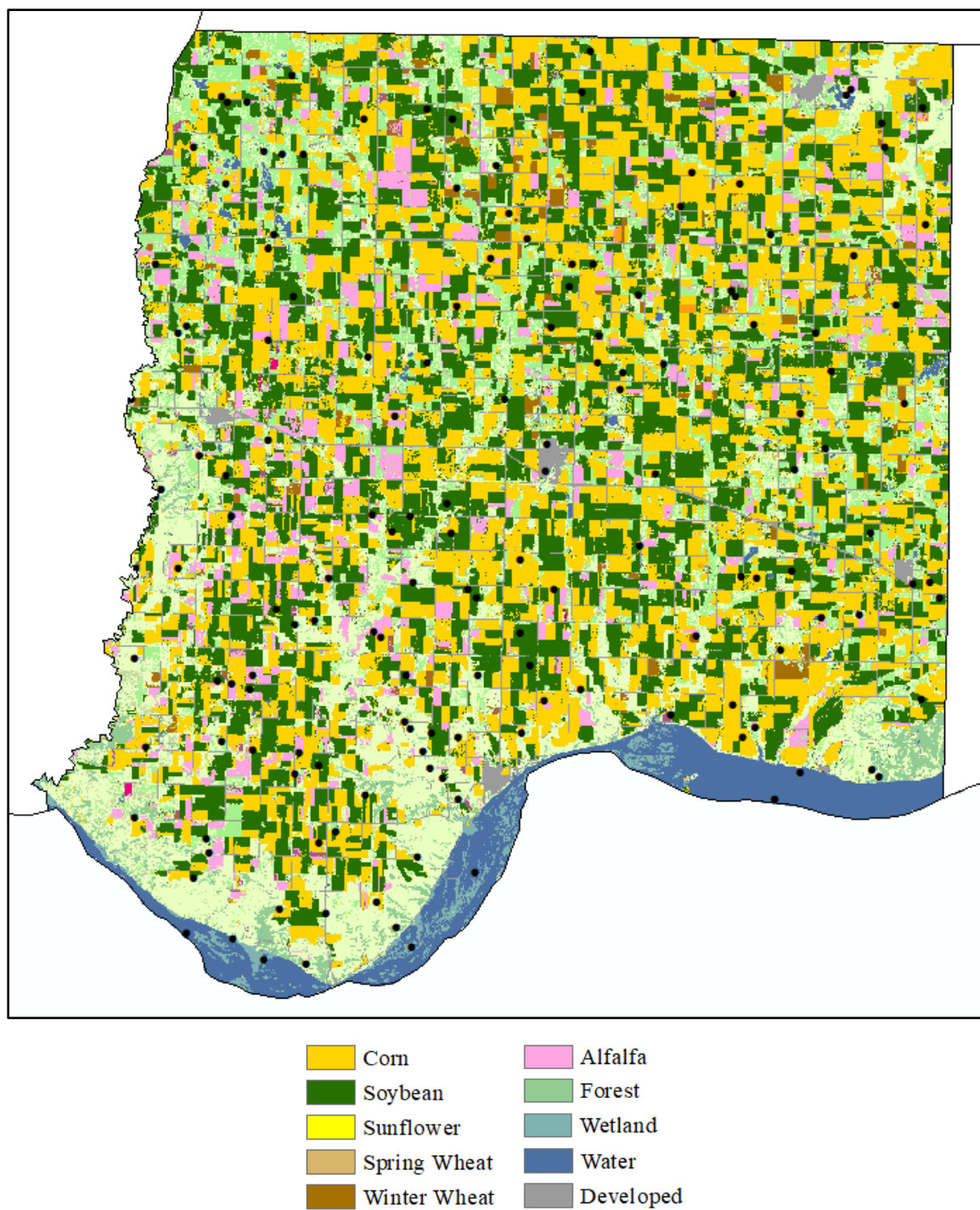
Figure S2.4. Bon Homme County 2014 CDL and reference points (black circles).
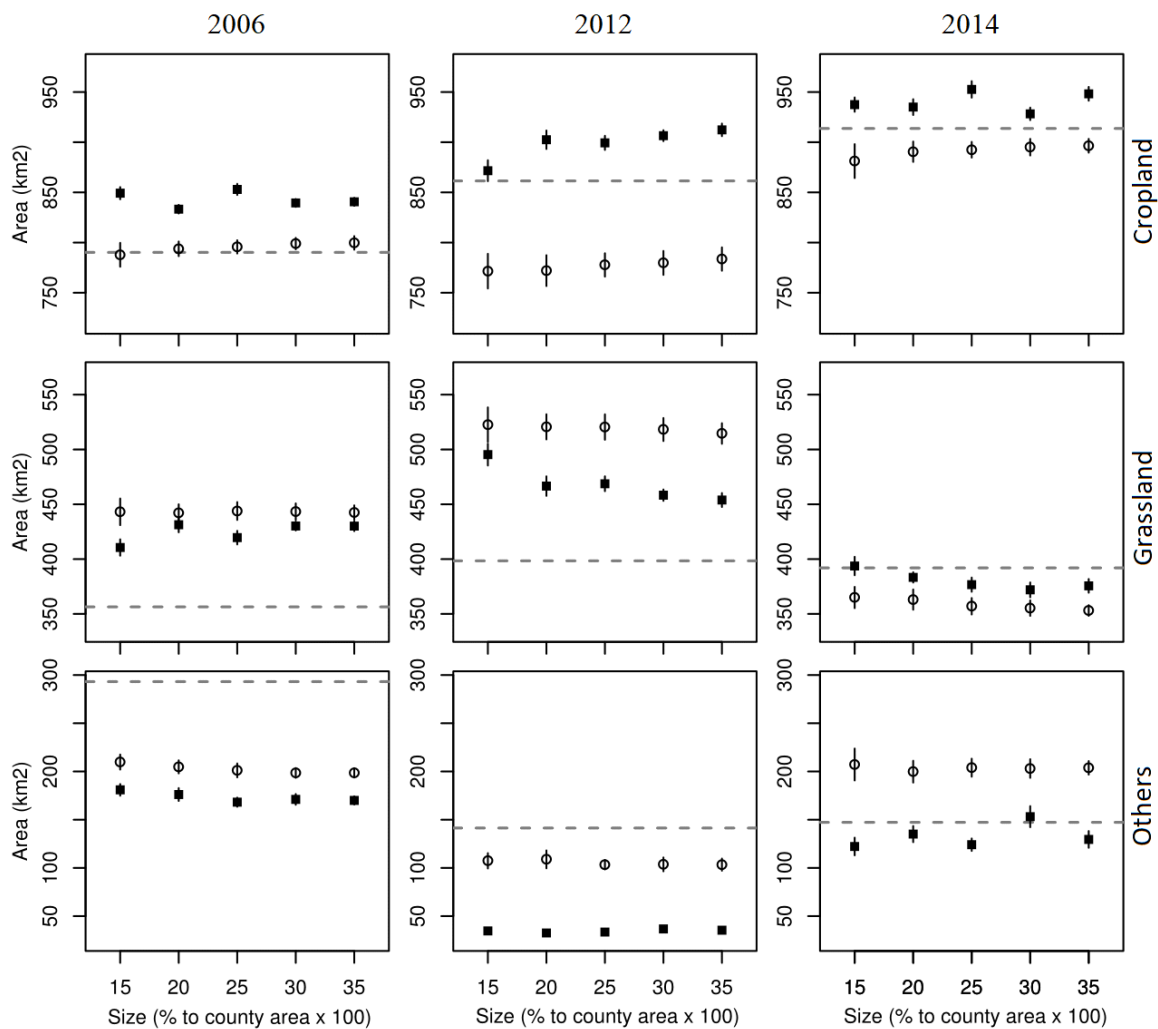
Figure S2.5. Area of each land cover type (mean ± 1σ) estimated by 'Same Distribution' (square) and 'Same Size' (circle) RFC models for Bon Homme County. Dashed lines present areas of corresponding land cover types estimated by re-classified CDL.
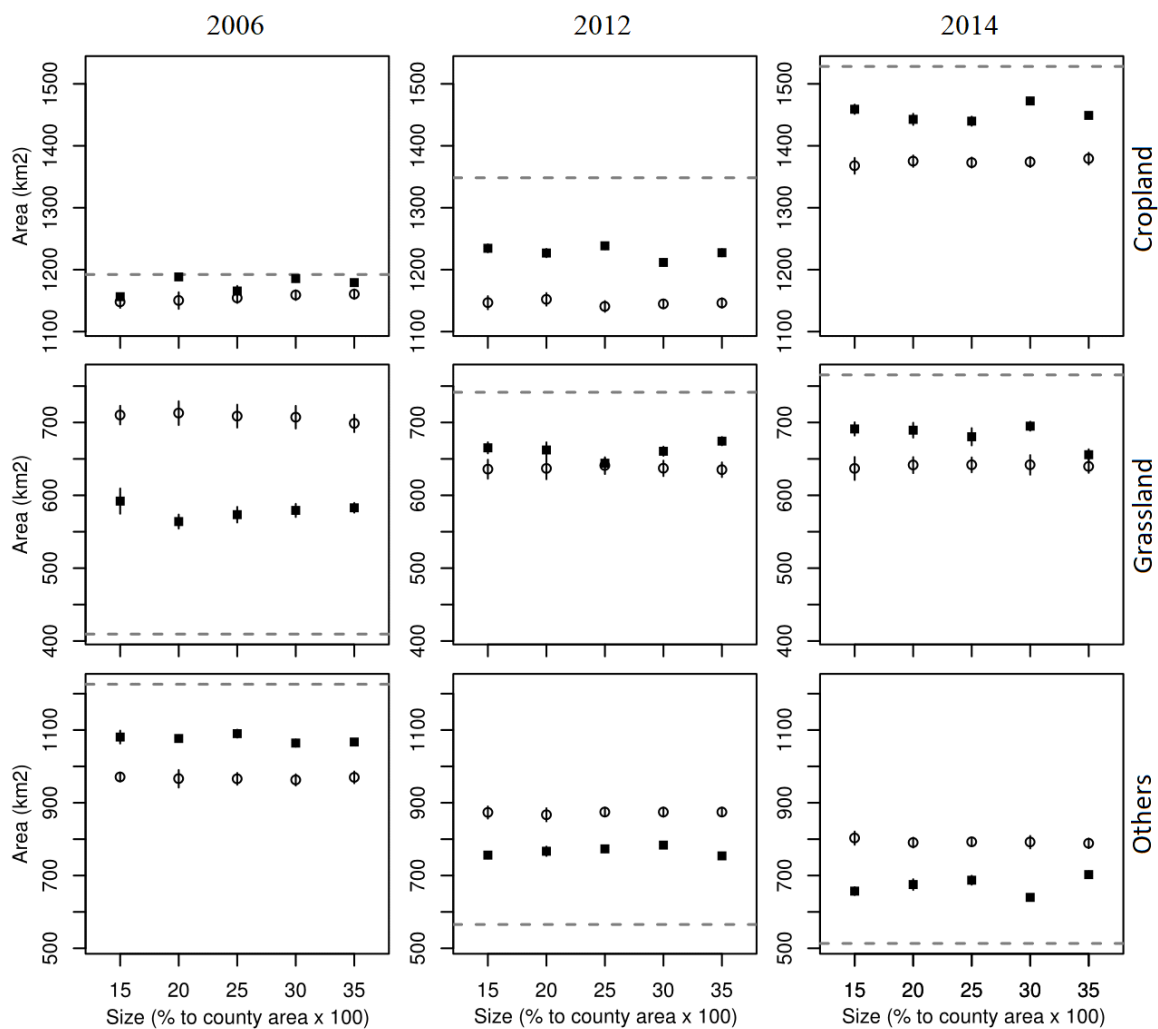
Figure S2.6. Area of each land cover type (mean ± 1σ) estimated by 'Same Distribution' (square) and 'Same Size' (circle) RFC models for Roberts County. Dashed lines present areas of corresponding land cover types estimated by re-classified CDL.
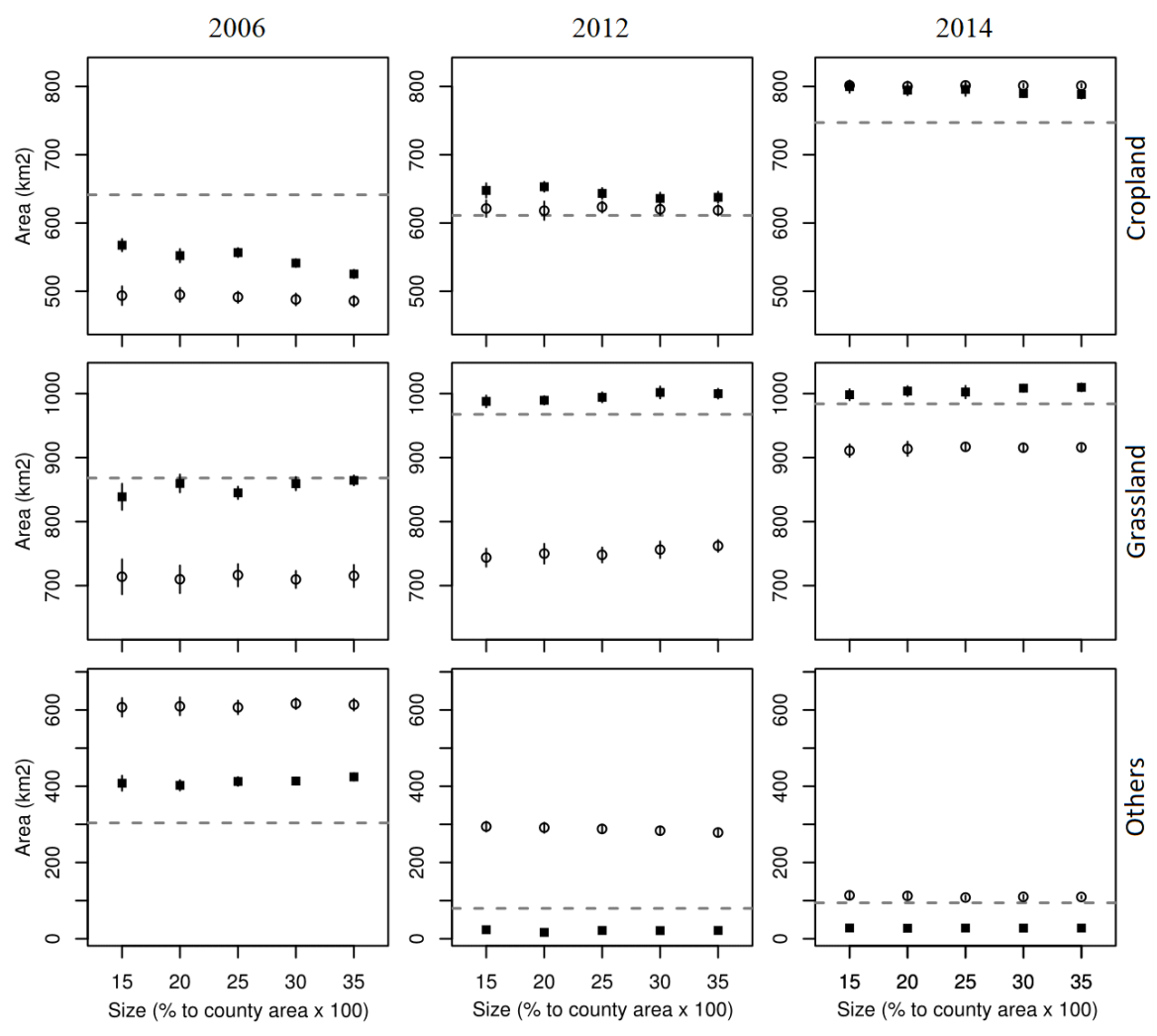
Figure S2.7. Area of each land cover type (mean ± 1σ) estimated by 'Same Distribution' (square) and 'Same Size' (circle) RFC models for Walworth County. Dashed lines present areas of corresponding land cover types estimated by re-classified CDL.
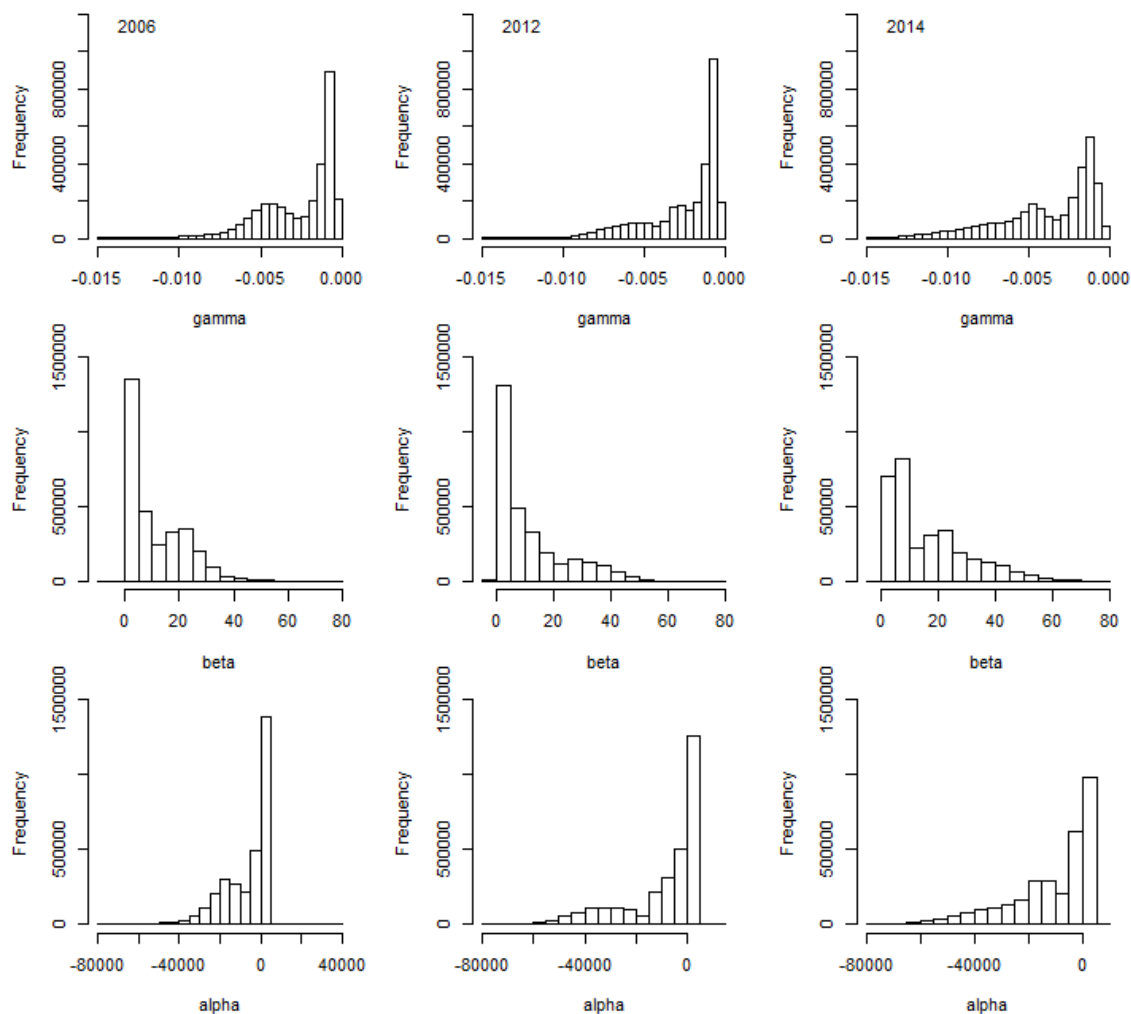
Figure S2.8. Histograms of fitted parameter coefficients (scaled by 10,000) in Roberts County.
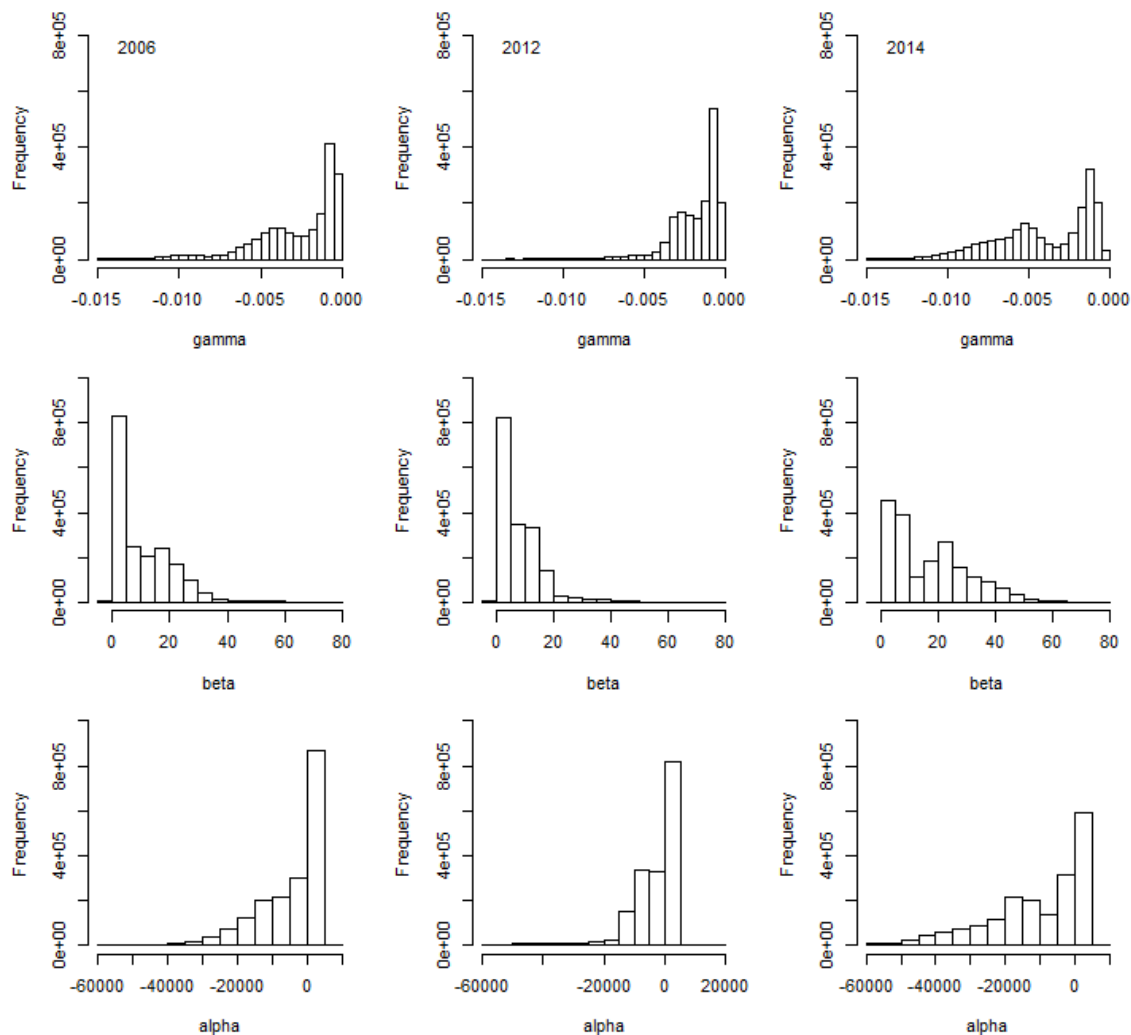
Figure S2.9. Histograms of fitted parameter coefficients (scaled by 10,000) in Codington County.
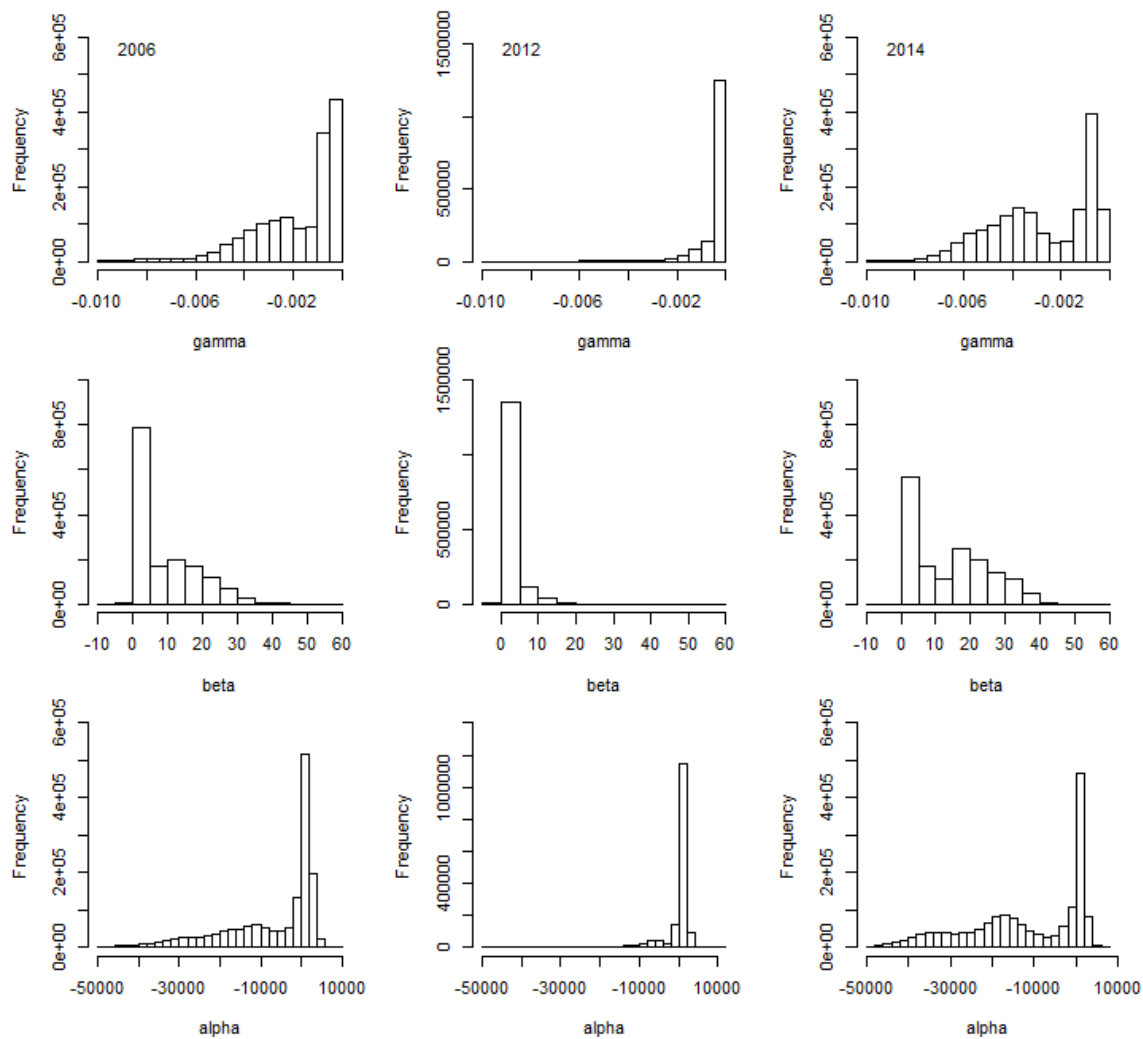
Figure S2.10. Histograms of fitted parameter coefficients (scaled by 10,000) in Bon Homme County.
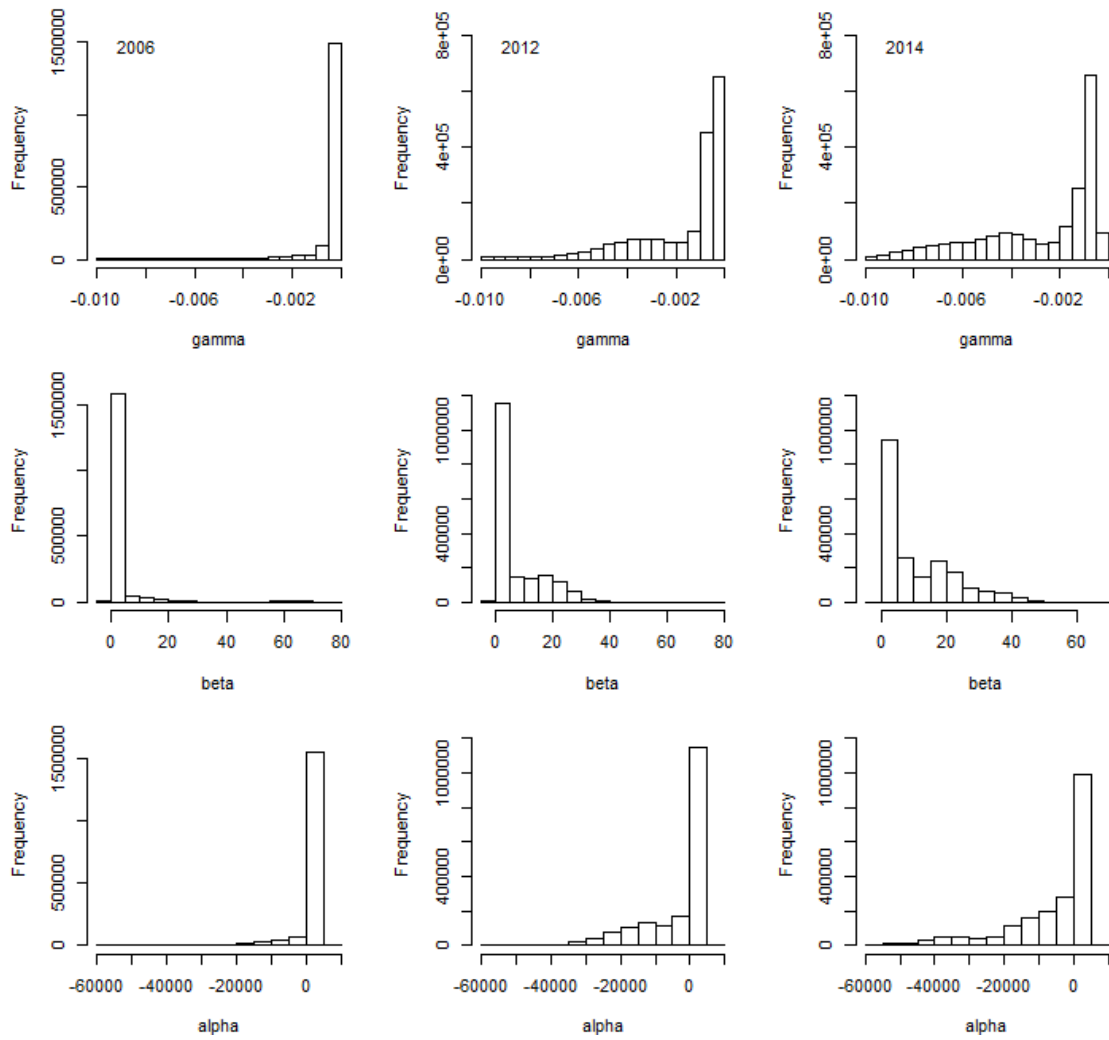
Figure S2.11. Histograms of fitted parameter coefficients (scaled by 10,000) in Walworth County.
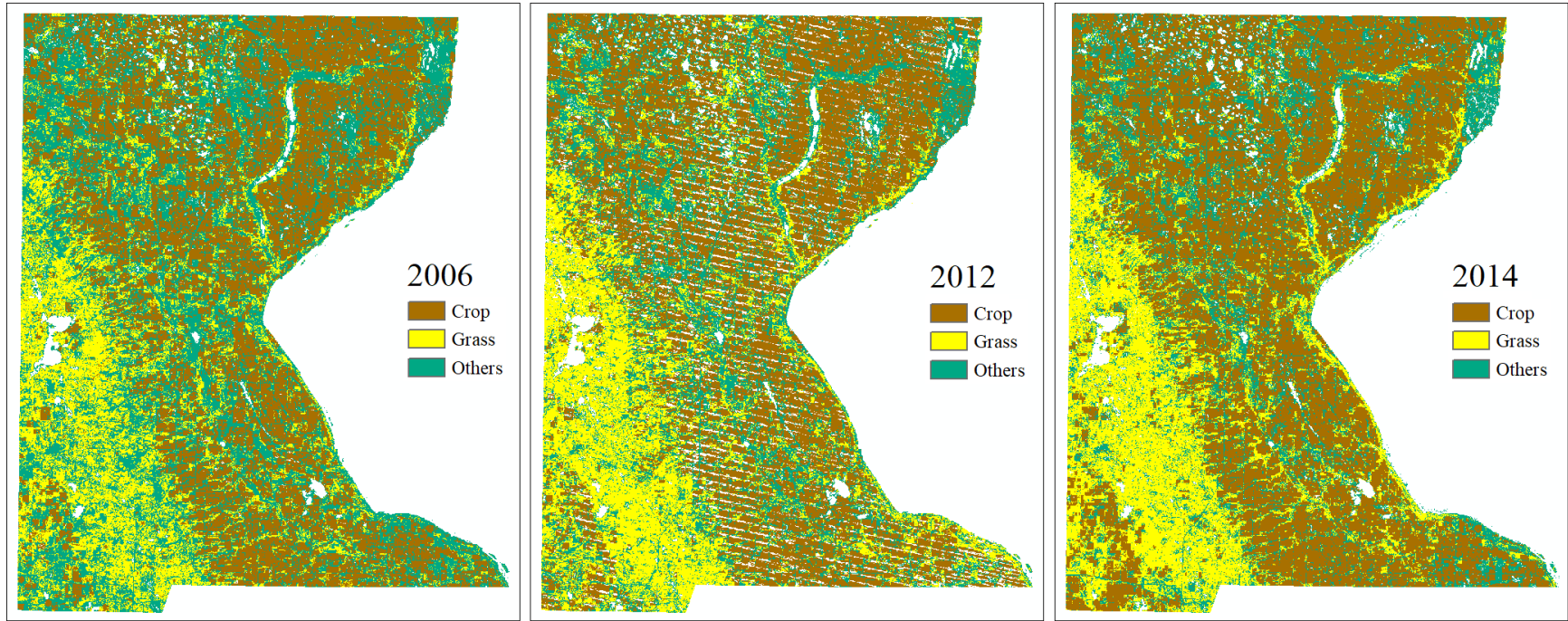
Figure S2.12. Sample output for Roberts. RFC Model: same distribution, reduced variable, sample dataset covers 0.25% of county area.
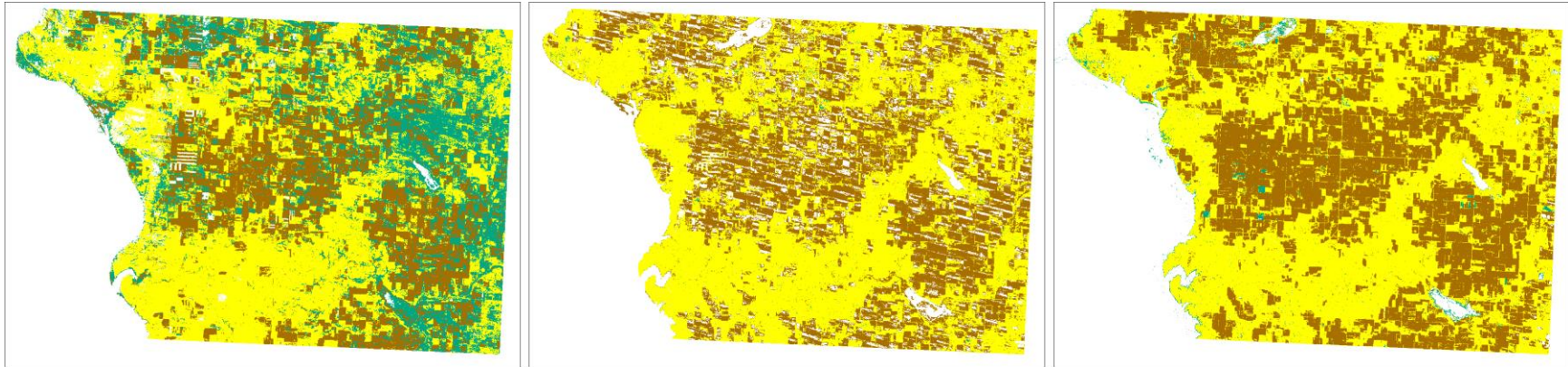
Figure S2.13. Sample output for Walworth. RFC Model: same distribution, reduced variable, sample dataset covers 0.25% of county area.
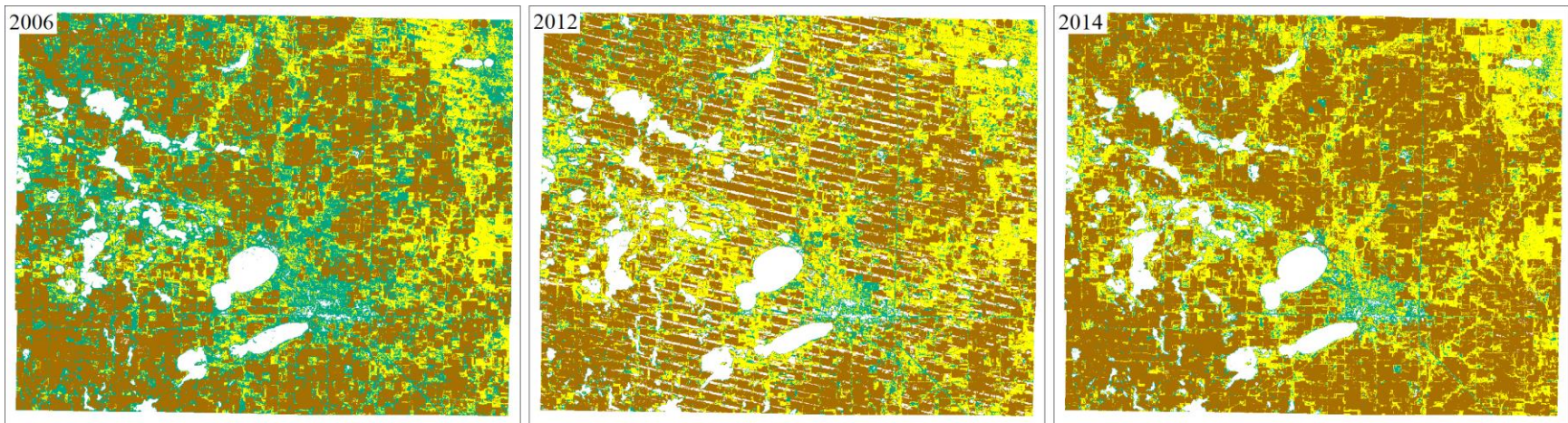
Figure S2.14. Sample output for Codington. RFC Model: same distribution, reduced variable, sample dataset covers 0.25% of county area.
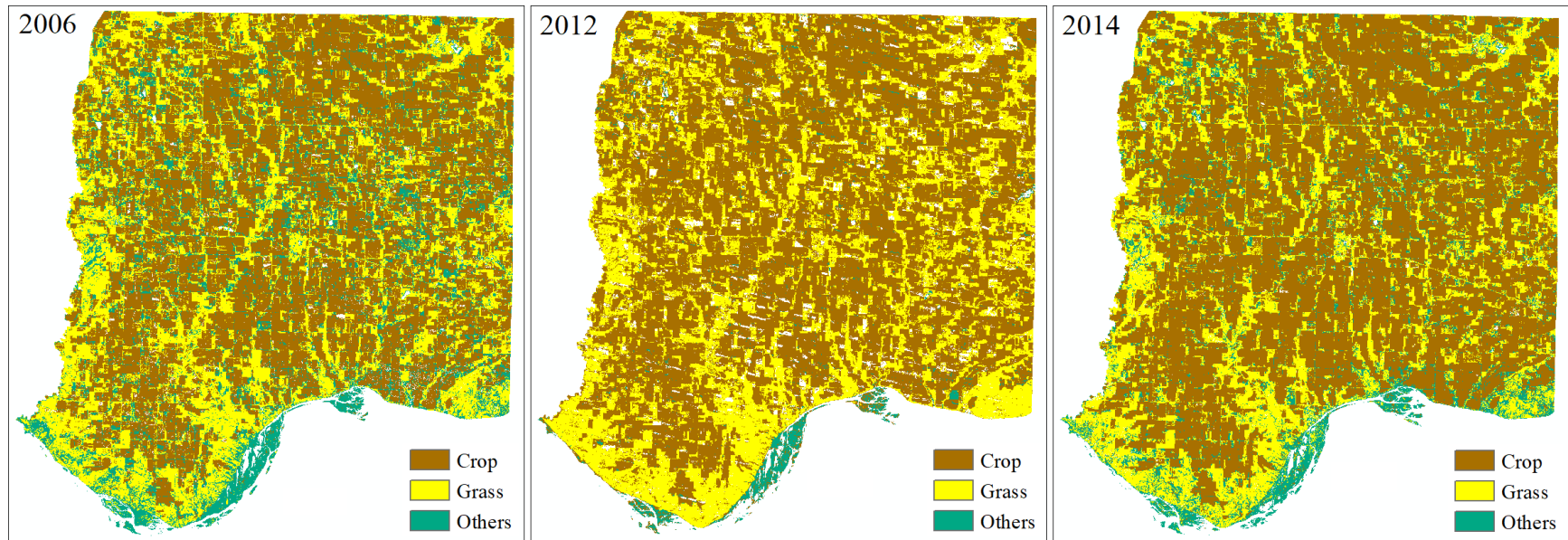
Figure S15. Sample output for Bon Homme. RFC Model: same distribution, reduced variable, sample dataset covers 0.25% of county area.
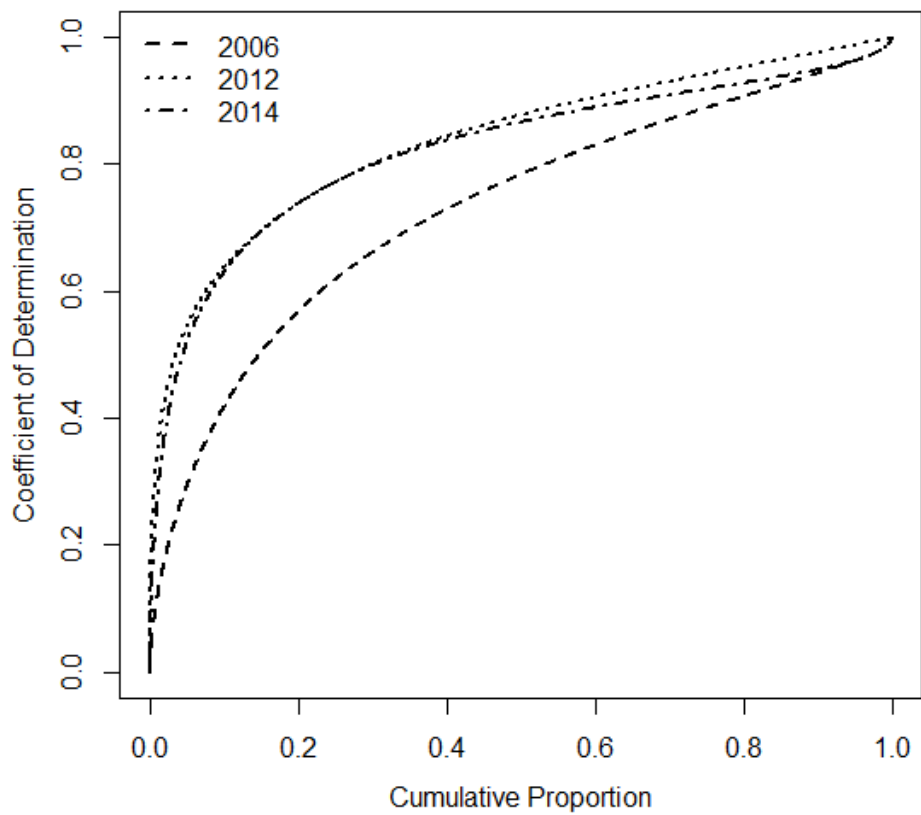
Figure S2.16. Cumulative proportion of r^2 of fitted models for Roberts County.

CHAPTER 3

IMPROVED CHANGE DETECTION WITH TRAJECTORY-BASED APPROACH:

APPLICATION TO QUANTIFY CROPLAND EXPANSION IN SOUTH DAKOTA

3.0   Abstract

The growing demand for biofuel production increased agricultural activities in South Dakota, leading to the conversion of grassland to cropland. Although a few land change studies have been conducted in this area, they lacked spatial details and were based on the traditional bi-temporal change detection that may return incorrect rates of conversion. This study aimed to provide a more complete view of land conversion in South Dakota using a trajectory-based analysis that considers the entire satellite-based land cover/land use time series to improve change detection. We estimated cropland expansion of 5,447 km^2 (equivalent to 14% of the existing cropland area) between 2007 and 2015, which matches much more closely the reports from the National Agriculture Statistics Service—NASS (5,921 km^2) and the National Resources Inventory—NRI (5,034 km^2) than an estimation from the bi-temporal approach (8,018 km^2). Cropland gains were mostly concentrated in 10 counties in northern and central South Dakota. Urbanizing Lincoln County, part of the Sioux Falls metropolitan area, is the only county with a net loss in cropland area over the study period. An evaluation of land suitability for crops using the

Soil Survey Geographic Database—SSURGO indicated a scarcity in high-quality arable land available for cropland expansion.

3.1   Introduction

Agriculture is the leading industry of South Dakota, contributing approximate $22 billion to the state's economy each year (*Decision Innovation Solutions, 2014*). Both crop production and livestock play important roles, sharing 60.7% and 39.3% of the total agricultural revenue. Over the past decade, the growing demand for biofuel production led to a considerable expansion of cropland area in South Dakota (increase of 1-5% annually; *Wright & Wimberly, 2013*), especially for corn/maize (for corn-based ethanol) and soybean (for biodiesel) (*Wright & Wimberly, 2013; Reitsma et al., 2014*). Corn/maize and soybean are South Dakota's two largest crops by area, accounting for 68% of the harvested field crops in 2017 (43,370 km^2) (*USDA-NASS 2017*). Conversion from grassland or wetland to cropland can alter the landscape (*Wright & Wimberly, 2013; Reitsma et al., 2014*) and hurt the ecosystem as well as the environment (*Vaché et al., 2002; Montgomery, 2007; Stephens et al., 2008; Meehan et al., 2010; Mutter et al., 2015; Wimberly et al., 2018*).

Several efforts have been made to characterize land dynamics in the U.S Northern Great Plains (spans over multiple states, including the entire South Dakota). Most extant studies take a bi-temporal "snapshot" approach (*Wright & Wimberly, 2013; Reitsma et al., 2014; Faber et al., 2012; Decision Innovation Solutions, 2013*) that only compares data between two points in time and disregards data from intermediate years. The bi-temporal approach does not capture the regular rotation of lands into and out of cultivation and, thus, may retrieve an incorrect rate of conversion (*Lark et al., 2015*). There are only two studies that provide more temporal context to land change analysis in South Dakota (*Lark et al.,*

2015; Arora & Wolter, 2018). However, one examined land change only for a brief period (2008-2012) (*Lark et al., 2015*), while the other provided a long-term analysis (1984-2016) but covered only a portion of the state (Landsat path 30, rows 28-29) (*Arora & Wolter, 2018*). Although the extant research has reported different amounts of land change, each study showed considerable losses of grassland with conversion to cropland (mostly corn/maize and soybean) in the Northern Great Plains, especially east of the Missouri river.

Only one study has evaluated land change exclusively in South Dakota (*Reitsma et al., 2014*). However, that study did not provide adequate spatial details about the changes. Since accurate geospatial accounts of land use change are critical in assessing long-term risk of the conversions, this paper aims to provide a more complete view of land conversion in South Dakota by characterizing the dominant land cover transition from 2006 to 2016 and its spatiotemporal patterns. To overcome the limitations of the bi-temporal change detection, we propose a trajectory-based approach that considers the entire land cover/land use time series to determine if there was actual land change at a particular location. First, land cover/land use (LCLU) maps were generated annually from 2006 to 2016 at 30-meter resolution (eleven maps in total) using the phenometric-based classification described in *Nguyen et al. (2019)*. This LCLU dataset allows us to analyze land changes in the study area over the past decade. The results are then compared against various official data sources released by the United States Department of Agriculture (USDA).

3.2   Methods

3.2.1   Generation of land cover/land use dataset

To characterize land changes in the study area, a fine spatiotemporal resolution land cover dataset with just three broad categories ("cropland", "grassland", and "others") was

generated using the phenometric-based classification as described in *Nguyen et al., 2019*. First, at each pixel, the seasonal variation of EVI time series derived from Landsat Collection-1 Surface Reflectance product (courtesy of the U.S. Geological Survey) was modeled as a downward-arching convex quadratic function (*de Beurs & Henebry, 2004; Henebry & de Beurs, 2013*) of accumulated growing degree-days (AGDD) derived from MODIS level-3 V005 global Land Surface Temperature and Emissivity 8-day composite products (*NASA LP DAAC, 2016; Wan, 2008*). Then, from the fitted model parameter coefficients, a suite of 16 metrics (Table S3.1) was derived to serve as input for the land cover/land use classification task using a Random Forest Classifier (RFC).

The USDA NASS Cropland Data Layer (CDL) (*Boryan et al., 2011*) was used to select training and testing pixels (the sample dataset) for the RFC models—despite known issues with the CDL (*Reitsma et al., 2016; Lark et al., 2017*)—due to a scarcity of ground observations. To increase the accuracy of sample dataset, only pixels that were surrounded by eight pixels of the same cover type and repeated cover type at least once in the study period were selected. Because the generated land cover maps contain just three broad categories, the CDL layers were regrouped into those classes to be used in the sample selection.

LCLU classifications were performed annually on the county basis to improve accuracy (*Zhang & Roy, 2017*). From each county-year sample pool, multiple stratified random sample datasets were selected, each covering 0.25% of the county area (*Nguyen et al., 2019*). For each county-year, twenty RFC models were performed and ensembled to retrieve a final LCLU map. The accuracies of the generated land cover/ land use maps were examined using a rich reference dataset that contains 14,400 points in each of just three

years: 2006, 2012, 2014 (*Reitsma et al., 2016*). In addition, predicted land cover areas from this study were compared with those from existing studies and databases.

Due to cloud/snow contamination or band gaps caused by failure of the Scan Line Corrector on Landsat 7 ETM+ (*Loveland et al., 2008*), there were not enough observations (EVI data points) over a year to fit the convex quadratic function at certain pixels at image edges. The phenometric-based classification was not able to assign cover types for those pixels, thereby generating gaps in the cover maps (*Nguyen et al., 2019*). In order to create seamless maps for change detection, the gaps of a current year were filled by land cover information from the two adjacent years, favoring the previous year's information. For example, missing land cover information from 2008 was first filled using the 2007 data. If there was also a data gap in 2007, the 2009 land cover information was used to fill gaps in the 2008 cover map.

3.2.2   Characterizing cropland expansion using the trajectory-based change detection

Phenometric-based classification can generate very accurate cropland maps but may not work as well for other land covers (*Nguyen et al., 2019*). Therefore, this analysis focused only on changes in cropland. We developed the trajectory-based approach that considered land cover information from the entire time series to determine whether land cover at a pixel shifted cover types. Since the analysis focused on characterizing cropland expansion, all land cover maps were regrouped into "cropland" and "non-cropland" ("grassland" and "others") for the change detection.

The trajectory-based change analysis of each crop pixel is presented in Figure 3.1. First, land cover maps from 2006 to 2016 were stacked to create a time series. Then the map series was converted into a binary time series with "0" for "non-cropland" and "1" for

"cropland". A binary array at each pixel indicates land cover and changes over time. To counter misclassifications in the land cover dataset, we assumed that a single year of different land cover in the entire 11-year time series was "noise".
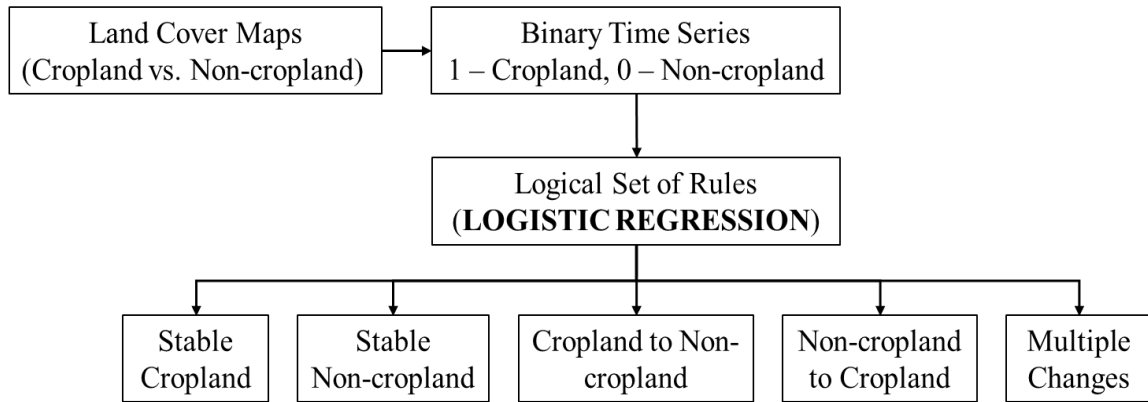
Figure 3.1. The trajectory-based change detection for cropland.

First, from the binary time series, locations were identified with very strong temporal signal of being "cropland" or "non-cropland". Over eleven years, if a number of times (counts) that a pixel appeared as "cropland" is less than two, the pixel was assigned as "stable non-cropland". Similarly, a pixel with counts of greater than nine was considered as a "stable cropland" pixel.

After masking out pixels with less than two or more than nine appearances as "cropland", all other possible "cropland" pixels ($2 \leq$ "counts" ≤ 9) were subjected to further change analysis. To incorporate land formation from all eleven years of data in change detection, the logistic regression was performed on each binary time series using calendar year as the only predictor variable:

$$\ln \left(\frac{p}{1-p} \right) = \text{intercept} + \text{slope} \times \text{year} \qquad \text{(Equation 3.1)}$$

where p is the probability of the pixel appearing as "cropland" over time and (1-p) is the probability of the same pixel appearing as "non-cropland" over the same interval.

From the fitted parameters (intercept and slope), we computed the p values for all years and used those values to help detect changes. If p value of the first year (2006) was less than 0.2 and p values of the last two years (2015 and 2016) were greater than 0.8 and 0.9 respectively, a pixel is subjected as "changed" from "non-cropland" to "cropland". On the complementary side, if p values of 2006 and 2007 were greater than 0.8 and 0.9, respectively, and p value of 2016 was less than 0.2, a pixel was subjected as "changed" from "cropland" to "non-cropland". All other pixels that did not meet these two conditions were further evaluated using the average probability of crop events over 11 years (pmean). If pmean was less than 0.2, a pixel was considered as a "stable non-cropland" pixel. Pixels with pmean of above 0.8 were assigned as "stable cropland". Pixels with pmean ranging between 0.2 and 0.8 were tagged as "multiple changes". The dominant land cover type and trend of "multiple changes" pixels are described by pmean (greater than 0.5: "cropland", less than 0.5: "non-cropland") and slope (greater than 0: toward "cropland", less than "0": toward "non-cropland"). Only changes detected from the analysis of probability were considered as "true" land cover changes. The 3x3 majority filter was applied to remove spatial noise—primarily singletons and doublets--in the land change layer.

Recall that we assumed a single mismatch in the entire land cover time series (eleven years) as "noise" due to misclassification rather than an actual change. Therefore, the proposed approach did not detect changes just happened in either 2006 or 2016. Instead, these results account for changes only between 2007 and 2015.

The suitability for crop production on stable and converted croplands was determined using SSURGO data (*USDA-NRCS, 2018*). Crop pixels were overlaid on the SSURGO Non-Irrigated Capability Class-Dominant Condition layer to extract land

capability classes (LCC), the broadest category, which are coded from 1 to 8 indicating progressively greater limitations and narrower choices for practical use. Classes 1 to 4 are considered as arable lands, and classes 5 to 8 are suitable mainly as pasture or rangeland. We regrouped and labeled LCC 1-2 as "prime" land for cultivation, LCC 3-4 as "fragile", and LCC 5-8 as "unsuitable" (*Lark et al., 2015*).

3.3   Results

3.3.1   Accuracy assessment of cropland maps of South Dakota

Phenometric-based classification using the sample dataset generated from the CDL performed very well: the RFC models consistently retrieved overall accuracies of greater than 90% across counties-years (Figure S3.1). Consequently, general patterns of land cover/land use from the RFC models and the reclassified CDL are similar, especially for "cropland" (Figure S3.2, Table S3.2). The similarity in cropland pattern between the CDL and RFC is also shown in Figure 3.2 as data points (county-year cropland areas) of the scatterplot are distributed closely around the 1:1 line. From 2006 to 2009, our dataset showed slightly larger cropland areas in South Dakota (Figure 3.2) compared to the CDL values. After 2009, estimated cropland areas from the two datasets showed difference of less than 1% (Figure 3.3). The predicted land cover maps from this study and the reclassified CDL both underestimated cropland areas reported by field-based statistics from the USDA. However, these satellite-based datasets (CDL and this study) and the field-based NASS estimates showed similar temporal patterns and strong correlation to each other (Figure 3.3, Table S3.3). However, the area in "grassland" is overestimated by RFC, while the area in "others" is underestimated (Figure 3.2, 3.3). The differences in "grassland" and "others" between the CDL and this study seem to be systematic (Figure

3.3), possibly due to misclassification of grass-like covers in "others" (e.g., urban lawns, playing fields). While the CDL adopted the "developed" classes from the National Land Cover Database (NLCD), many "developed" pixels were classified as "grassland" or "cropland" pixels in RFC models. Those pixel classifications are not fundamentally incorrect from a land surface phenology perspective but require supplementary contextual information to shift the pixels from "grassland" (pasture/grazing use) to "others" (recreational use).
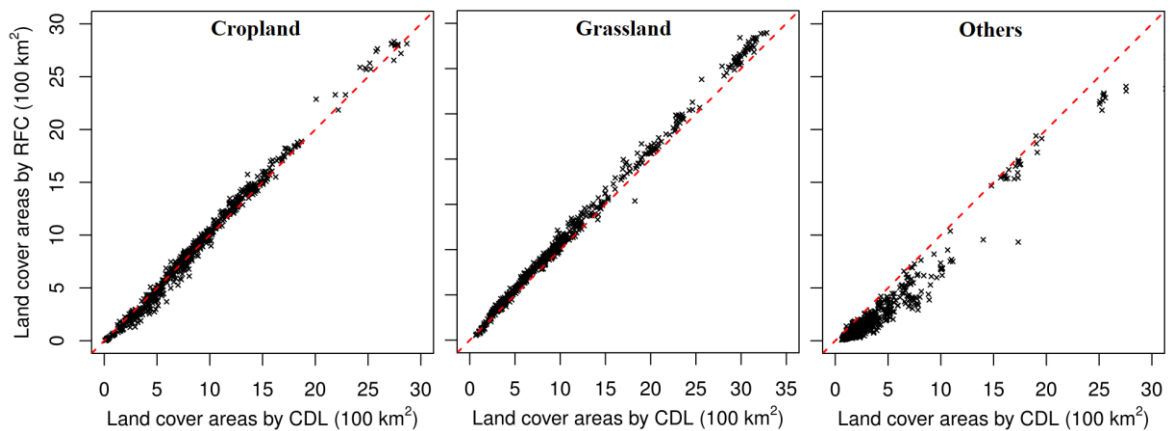
Figure 3.2. County-level comparison between estimated land cover area from this study and the CDL. The total of 726 data points (66 counties x 11 years) were plotted.
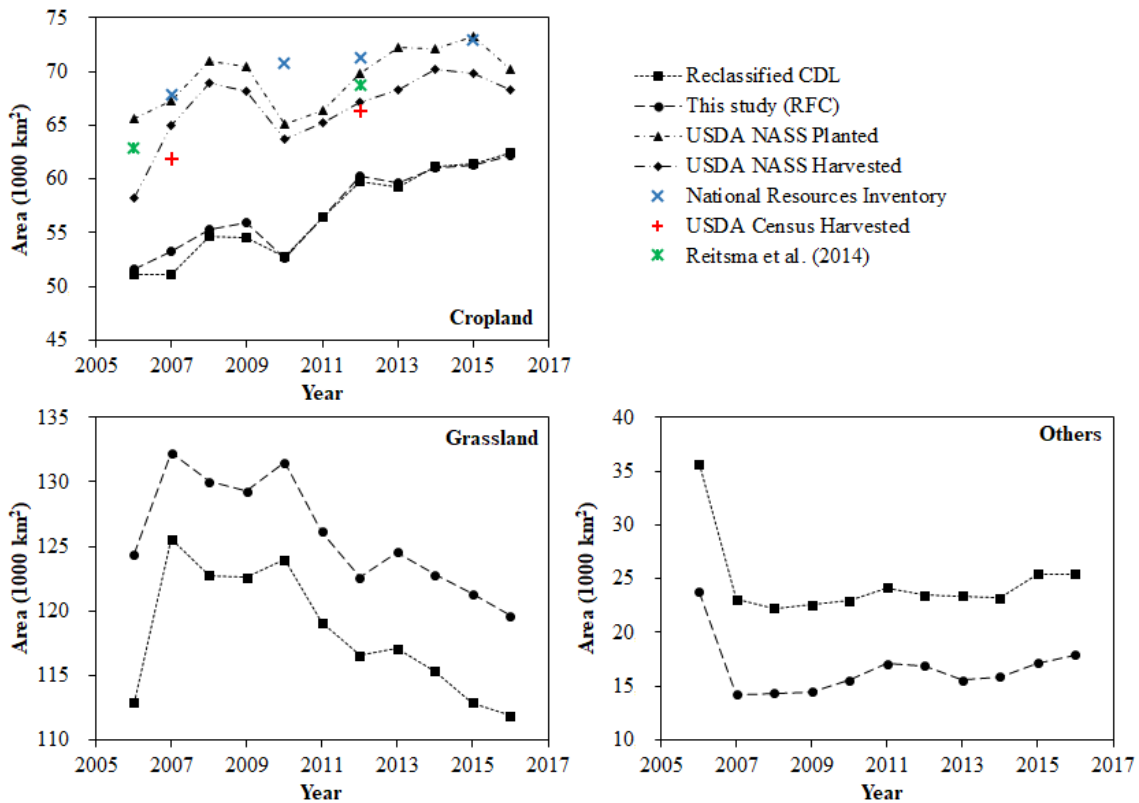
Figure 3.3. State-level comparison between estimated land cover areas from this study (RFC) and other data sources. Both USDA NASS Planted/ Harvested and Census of Agriculture for South Dakota data were retrieved from the NASS website (*USDA, 2017; USDA-NASS 2009,2014*) with no accuracy information. The 95% confident interval (CI) for 2007, 2010, 2012, and 2015 NRI data are $\pm$ 2138 km^2, $\pm$ 1823 km^2, $\pm$ 1525 km^2, and $\pm$ 1708 km^2 respectively (*USDA. 2009, 2013, 2015, 2018*). Cropland areas found in (*Reitsma et al., 2014*) also did not have accuracy information.

To evaluate the newly generated LCLU dataset further, a state-level accuracy assessment of the predicted land cover maps (RFC) and the reclassified CDL was performed using the independent reference point dataset (Table 3.1). Although the RFC has lower overall accuracies and kappa statistics compared to the reclassified CDL (Table 3.1), relative differences are minor: less than 5% and 2% for kappa and overall accuracy, respectively. The slightly lower accuracy metrics of the RFC are most likely due to

differences in "grassland" and "others" land covers (Figure 3.2). User and producer accuracies for cropland are also slightly better for the CDL for all three years. Considering the good accuracy of CDL crops and the use of CDL as the training dataset for our classification, these results are expected. The RFC tends to show lower user accuracy but higher producer accuracy for "grassland" compared to the reclassified CDL, indicating more "cropland" or "others" pixels were classified incorrectly as "grassland", but fewer "grassland" pixels were classified incorrectly as "cropland" or "others" (Figure 3.2, 3.3). It is possibly due to pixels with failed crops (in "cropland"), urban vegetation, or dry wetlands (in "others") having similar phenological patterns to those of the "grassland" class. On the other hand, the RFC shows higher user accuracy but lower producer accuracy in "others", indicating fewer "cropland" or "grassland" pixels were classified incorrectly as "others", but more "others" pixels were classified incorrectly as "cropland" or "grassland" (Figure 3.2). It is possibly due to misclassification of urban vegetation as "grassland" (mostly lawns, and playing fields) or, less common, as "cropland" (gardens, residential trees).

Figure 3.4. (a) County-level stable cropland coverage in %, and (b) cropland net expansion in km². The maps were overlaid with NASS reporting districts. Numbers on (b) present net cropland gains summarized by NASS districts.

Stable cropland areas were concentrated mostly in the eastern South Dakota where lands are more suitable for cultivation (Figure 3.4a), especially in the East-Central and Southeast NASS districts. Less than 10% of the western region was covered by stable croplands (Figure 3.4a). Between 2007 and 2015, cropland gains were mostly concentrated either in the northern or central South Dakota. Among the nine NASS reporting districts, North-Central, Central, and Northeast are the top three regions for cropland expansion with the net gains of 1,307 km², 977 km², and 926 km², respectively (Figure 3.4b). There was less available cropland in the eastern South Dakota during the study period; thus, croplands expanded westward to neighboring counties, especially in the Northeast and South-Central districts (Figure 3.4b). Few gains in cropland were observed in the Southeast district, due to very limited land resources available for new cropland expansion (Figure 3.4a). There was also not much cropland expansion in Harding County (Figure 3.4b), located at the northwestern corner of the state, due to its dry and cold climate that does not support crop growth. The western side of the West Central and South West districts also showed little

cropland expansion. These areas are mostly covered by lands unsuitable for cultivation (*e.g.*, Black Hills National Forest, Badlands National Park, Buffalo Gap National Grassland). Of South Dakota's 66 counties, Lincoln County in eastern South Dakota was the only one with a net loss in cropland area due to urban and suburban expansion of the Sioux Falls metropolitan area in northern Lincoln County (Figure 3.4b). Total population of the Sioux Falls metropolitan area grew from 132,358 to 171,530 (29.6%) between 2005 and 2015 (*ACS, 2018*).

In South Dakota, previous croplands are mostly concentrated on "prime" farmland, accounting for 75% of the total stable cropland area (Figure 3.5a). For more than 44,000 km^2 of the stable cropland, only 3% cropland areas were in lands "unsuitable" for cultivation. However, between 2007 and 2015, farmers have been expanding croplands into more "fragile" and "unsuitable" lands to meet growing demand (48% of new expansion; Figure 3.5b). This trend indicates increasing scarcity of farm land: less area of prime cropland is available for cultivation.
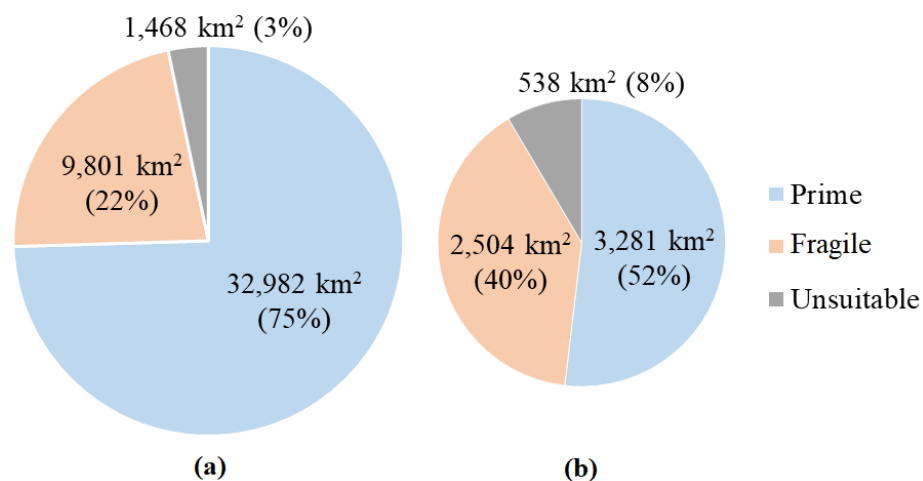
Figure 3.5. Areas (km^2) in (a) stable cropland and (b) converted cropland by Land Capability Class from SSURGO.

3.4   Discussion

3.4.1   Uncertainties in land cover/land use dataset

A new land cover dataset for South Dakota was generated to characterize recent land cover/land use changes. In contrast to the CDL that went through several changes in methodology and input data, our product was generated more consistently to enable better change detection. One limitation of the new dataset is that the CDL was used as training dataset in a supervised classification process. Although the major commodity crops are mapped quite well in the CDL (generally, producer/use accuracies of above 95% for corn/soybean and approximate 90% for spring/winter wheat), accuracies of other classes can be low (*Boryan et al., 2011*). Thus, multiple steps were applied to reduce errors and uncertainties in the sample data. By using the CDL as the training dataset, we may magnify misclassifications found in CDL dataset. However, validation using the independent reference points showed comparable accuracy between LCLU maps from this study and the reclassified CDL (Table 3.1). Furthermore, our estimated cropland areas from our land cover maps are similar with those from the CDL and even closer to NASS statistics in the 2006-2009 period (Figure 3.3).

3.4.2   Characterizing land changes using the trajectory-based approach

Field-based estimations from the USDA are the most reliable data sources to quantify changes, but they offer limited spatial and/or temporal details. Therefore, researchers often turn into satellite-derived land cover/land use products with much higher spatiotemporal detail that allow depiction of both timing and location of land change. However, all satellite-based products suffer from misclassification errors that can falsely accentuate or attenuate changes between classes. To overcome this limitation, we

developed a trajectory-based analysis that considers the entire land cover series to better determine land change at a particular location. The proposed approach stems from idea of examining a temporal array of "cropland" and "non-cropland" to determine land use status (*Lark et al., 2015*). However, instead of looking at all possible combinations ($2^{11} = 2048$ options) as demonstrated in *Lark et al. (2015)*, we modeled "cropland" and "non-cropland" time series as a binary array and described land use trajectory using logistic regression. This approach not only allows us to examine long time series but also to describe temporal land dynamics as a sigmoidal curve for a simple and consistent determination of land change. The results showed that cropland expansion estimated by the trajectory-based approach is much closer to the field-based reports from the USDA compared to the traditional bi-temporal change detection.

A key limitation of this study is low accuracy of the "non-cropland" classes (cf. section 3.1); therefore, the analysis was focused only on cropland expansion since "cropland" is the most reliable class in our dataset (Table 3.1). Although, transitions in "cropland" can translate into changes in other classes, especially "grassland", as published studies have shown that most of grassland losses are due to cropland expansion (*Wright & Wimberly, 2013; Reitsma et al., 2014*), the lower accuracy of "grassland" and "others" prevents exact quantification of their losses due to cropland expansion. In addition, our new dataset does not offer crop-specific categories like the CDL, preventing identification of the crops were planted on the newly cultivated lands.

3.5   Conclusions

Crop production plays critical role in South Dakota's economy, especially commodity crops like corn/maize and soybean. However, losses of grassland and wetland

area due to crop expansion may pose long-term risks to ecosystem services and biodiversity. Assessment of those risks requires accurate geospatial accounts of land change. This paper presented a comprehensive picture of LCLU transition in South Dakota between 2007 and 2015 using the fine spatiotemporal land cover dataset and the trajectory-based change detection approach. We found a net increase in cropland area of approximate 14% (5,447 km^2) in South Dakota. This result confirms substantial cropland expansion into grassland reported by previous studies (*Decision Innovation Solutions, 2013; Lark et al., 2015; Arora & Wolter, 2018*). Scarcity of land suitable for further cropland expansion was identified, pointing to the need for careful joint design and implementation of agriculture and energy policies to allow enable bioenergy demands to be met while protecting remaining wetlands and grasslands.

3.6   Acknowledgments:

3.7   References

ACS — U.S. Census Bureau, American Community Survey. (2018). ACS 5-Year estimates of Total Population for Sioux Falls city, SD; generated by Lan Nguyen; using American FactFinder; <http://factfinder.census.gov>; (15 December 2018).

Arora, G.; Wolter, P.T. Tracking land cover change along the western edge of the US Corn Belt from 1984 through 2016 using satellite sensor data: observed trends and contributing factors. J. Land Use Sci. 2018, 1-22.

Boryan, C.; Yang, Z.; Mueller, R.; Craig, M. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer program. Geocarto Int. 2011, 26, 341-358.

de Beurs, K.M.; Henebry, G.M. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. Remote Sens. Environ. 2004, 89, 497–509.

Decision Innovation Solutions. 2013 Multi-State Land Use Study: Estimated Land Use Changes 2007-2012. Urbandale, IA 50322: Decision Innovation Solutions. 2013. Available online: http://www.decision-innovation.com/spatial-time-series-analysis.

Decision Innovation Solutions. 2014 South Dakota Agricultural Economic Contribution Study. Decision Innovation Solutions: Urbandale, IA 50322. 2014. Available online: https://sdda.sd.gov/office-of-the-secretary/publications/.

Faber, S.; Rundquist, S.; Male, T. Plowed under: how crop subsidies contribute to massive habitat losses. Environmental Working Group: Washington, DC 20009. 2012. Available online: https://www.ewg.org/research/plowed-under

Henebry, G.M.; de Beurs, K.M. Remote Sensing of Land Surface Phenology: A Prospectus. In Phenology: An Integrative Environmental Science 2nd ed.; M.D. Schwartz, Ed.; Springer: New York, 2013; pp. 385-411.

Lark, T.J.; Mueller, R.M.; Johnson, D.M.; Gibbs, H.K. Measuring land-use and land-cover change using the US department of agriculture's cropland data layer: Cautions and recommendations. Int. J. Appl. Earth Obs. Geoinf. 2017, 62, 224-235.

Lark, T.J.; Salmon, J.M.; Gibbs, H.K. Cropland expansion outpaces agricultural and biofuel policies in the United States. Environ. Res. Lett. 2015, 10, 044003.

Loveland, T.R.; Cochrane, M.A.; Henebry, G.M. Landsat still contributing to environmental research. Trends Ecol. Evol. 2008, 23,182-183.

Meehan, T.D.; Hurlbert, A.H.; Gratton, C. Bird communities in future bioenergy landscapes of the Upper Midwest. PNAS. 2010, 107, 18533-18538.

Montgomery, D.R. Soil erosion and agricultural sustainability. PNAS. 2007, 104, 13268-13272.

Mutter, M.; Pavlacky, D.C.; Lanen, N.J.; Grenyer, R. Evaluating the impact of gas extraction infrastructure on the occupancy of sagebrush-obligate songbirds. Ecol. Appl. 2015, 25, 1175-1186.

NASA LP DAAC, 2006, MODIS Terra/Aqua Land Surface Temperature/Emissivity 8-Day L3 Global 1 km SIN Grid. Version 5. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov), accessed January 15, 2017.

Nguyen, L.H.; Joshi, D.R.; Clay, D.E; Henebry, G.M. Characterizing land cover/land use from multiple years of Landsat and MODIS time series: a novel approach using land

surface phenology modeling and random forest classifiers. Remote Sens. Environ. 2019. https://doi.org/10.1016/j.rse.2018.12.016

Reitsma, K.D.; Clay, D.E.; Carlson, C.G.; Dunn, B.H.; Smart, A.J.; Wright, D.L.; Clay, S.A. Estimated South Dakota land use change from 2006 to 2012. iGrow Agronomy. 2014.

Reitsma, K.D.; Clay, D.E.; Clay, S.A.; Dunn, B.H.; Reese, C. Does the US cropland data layer provide an accurate benchmark for land-use change estimates? Agron. J. 2016, 108, 266-272.

Stephens, S.E.; Walker, J.A.; Blunck, D.R.; Jayaraman, A.; Naugle, D.E.; Ringelman, J.K.; Smith, A.J. Predicting risk of habitat conversion in native temperate grasslands. Conserv. Biol. 2008, 22, 1320-1330.

USDA — U.S. Department of Agriculture, National Agricultural Statistics Service. 2009. 2007 Census of Agriculture: South Dakota state and county data. Available from: https://www.nass.usda.gov/AgCensus/

USDA — U.S. Department of Agriculture, National Agricultural Statistics Service. 2014. 2012 Census of Agriculture: South Dakota state and county data. Available from: https://www.nass.usda.gov/AgCensus/

USDA — U.S. Department of Agriculture. Summary Report: 2007 National Resources Inventory, Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. 2009. Available from: http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb11 67354.pdf

USDA — U.S. Department of Agriculture. *Summary Report: 2010 National Resources Inventory*; Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. 2013. Available from: http://www.nrcs.usda.gov/technical/NRI/2007/2007_NRI_Summary.pdf. (accessed on 10 March 2019).

USDA — U.S. Department of Agriculture. *Summary Report: 2012 National Resources Inventory*; Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. 2015. Available from: http://www.nrcs.usda.gov/technical/nri/15summary. (accessed on 10 March 2019).

USDA — U.S. Department of Agriculture. *Summary Report: 2015 National Resources Inventory*; Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa. 2018. Available from: http://www.nrcs.usda.gov/technical/nri/15summary. (accessed on 10 March 2019).

USDA — United States Department of Agriculture (USDA) (2017) Quick Stats 2.0. U.S. Department of Agriculture, National Agricultural Statistics Service, Washington DC. https://quickstats.nass.usda.gov/ (21 December 2018).

USDA-NASS — U.S. Department of Agriculture-National Agricultural Statistics Service. 2017 State Agriculture Overview: South Dakota. Available online: https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=SOUTH%20DAKOTA. (accessed on 21 December 2018)

USDA-NRCS — U.S. Department of Agriculture-Natural Resources Conservation Service. (2018). Soil Survey Geographic (SSURGO) Database. Available online: https://nrcs.app.box.com/v/soils. (accessed on 5 October 2018)

Vaché, K.B.; Eilers, J.M.; Santelmann, M.V. Water quality modeling of alternative agricultural scenarios in the US Corn Belt. J. Am. Water Resour. Assoc. 2002, 38, 773-87.

Wan, Z. New refinements and validation of the MODIS land-surface temperature /emissivity products. Remote Sens. Environ. 2008, 112, 59-74.

Wimberly, M.C.; Narem, D.M.; Bauman, P.J.; Carlson, B.T.; Ahlering, M.A. Grassland connectivity in fragmented agricultural landscapes of the north-central United States. Biol. Conserv. 2018, 217, 121-130.

Wright, C.K.; Wimberly, M.C. Recent land use change in the Western Corn Belt threatens grasslands and wetlands. PNAS. 2013, 110, 4134-4139.

Zhang, H.K.; Roy, D.P. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. Remote Sens. Environ. 2017, 197, 15-34.

3.8   Supplementary

Table S3.1. Fitted parameter coefficients, derived metrics from the Convex Quadratic (CxQ) model for land surface phenology.

Parameters	Meaning
α, β, γ	Fitted parameter coefficients of CxQ model (Equation 1)
TTP	Thermal time to peak (AGDD at the max fitted EVI) (TTP = -β/2×γ)
PH	Peak height EVI (max fitted EVI) (PH = α − β²/4×γ)
HTV	Value of EVI at half-TTP (HTV = α + β×TTP/2 + γ×TTP²/4)
ymax	Highest observed EVIs
r²	Coefficient of determination of the fitted model
lpos, rpos	Observation index of start and end of the fitting window
o_fit	Number of observations used to fit the CxQ model
o_per	Ratio of "o_fit" to the total number of observations
minx, maxx	AGDD at left and right ends of the fitted curve in the first quadrant
peaks	Number of high EVI values (≥0.8*ymax) outside the fitting window
jumps	Number of times that ΔEVI ≥ 0.2

Table S3.2. Pixel-wise comparison between predicted land cover maps (RFC) and the reclassified CDL for (a) 2006 and (b) 2012. PA is producer's accuracy; UA is user's accuracy. Overall accuracy appears in **bold**.

(a)		CDL (km^2)				
	2006	Cropland	Grassland	Others	Total	UA
RFC	Cropland	42,265	5,053	4,236	51,554	82.0%
	Grassland	7,388	103,232	13,775	124,395	83.0%
(km^2)	Others	1,503	4,613	17,662	23,778	74.3%
	Total	51,156	112,898	35,673	199,727	
	PA	82.6%	91.4%	49.5%		**81.7%**

(b)		CDL (km^2)				
	2012	Cropland	Grassland	Others	Total	UA
RFC	Cropland	52,258	5,018	2,940	60,216	86.8%
	Grassland	6,587	108,440	7,577	122,604	88.4%
(km^2)	Others	906	3,066	12,935	16,907	76.5%
	Total	59,751	116,524	23,452	199,727	
	PA	87.5%	93.1%	55.2%		**86.9%**
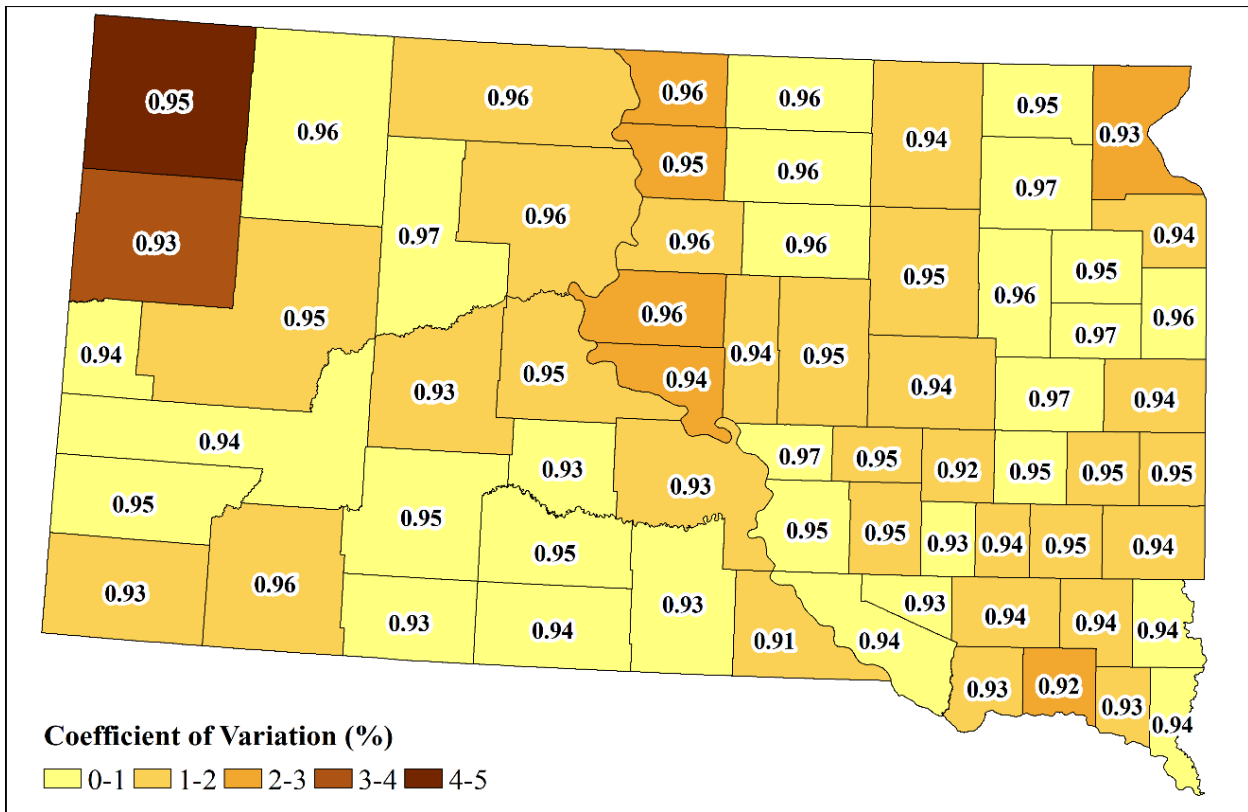
Figure S3.1. Accuracy of county-level Random Forest models. Numbers are mean overall accuracies of 11 years, and color-coded background shows temporal coefficient of variation.

Table S3.3. Correlation between estimated crop areas by this study (RFC), the reclassified CDL and the NASS statistics. Indication of significance: *, **, and *** for p-values less than 0.05, 0.01, and 0.001, respectively.

	Pearson's r	Spearman's r
CDL ~ NASS_Planted	0.74**	0.62*
CDL ~ NASS_Harvested	0.73*	0.76**
RFC ~ NASS_Planted	0.79**	0.65*
RFC ~ NASS_Harvested	0.78**	0.76**
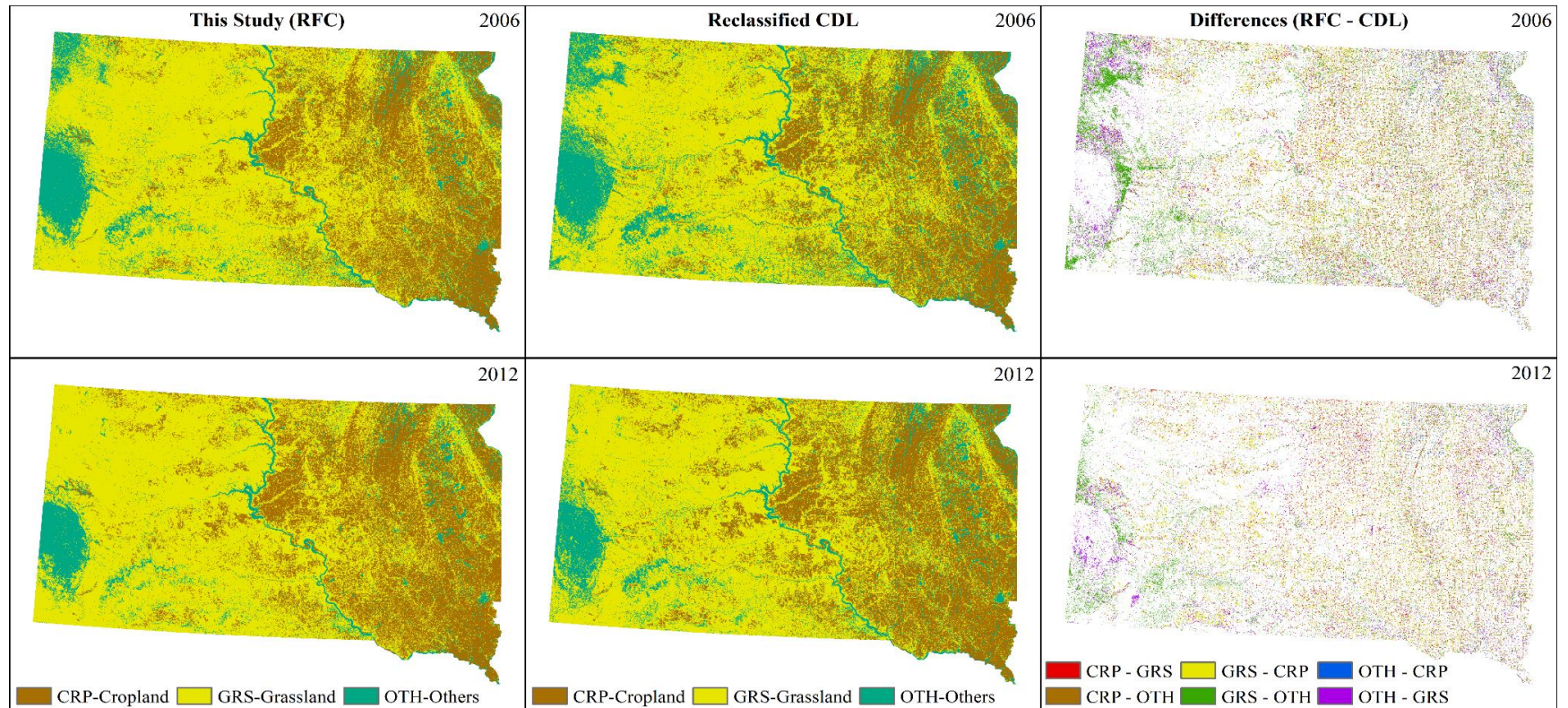
Figure S3.2. Comparison between our land cover maps and the reclassified CDL.

References: Nguyen, L.H.; Joshi, D.R.; Clay, D.E; Henebry, G.M. Characterizing land cover/land use from multiple years

of Landsat and MODIS time series: a novel approach using land surface phenology modeling and random forest classifiers.

Remote Sens. Environ. 2019. https://doi.org/10.1016/j.rse.2018.12.016

CHAPTER 4

CHARACTERIZING LAND COVER/LAND USE USING MULTI-SENSOR

TIME SERIES FROM THE PERSPECTIVE OF LAND SURFACE

PHENOLOGY

4.0   Abstract

Due to a rapid increase of accessible Earth Observation data coupled with high computing and storage capabilities, multiple efforts over the past few years have aimed to map land cover/land use using image time series with promising outcomes. Here we evaluate the comparative performance of alternative land cover classifications generated by using only (1) phenological metrics derived from either of two land surface phenology models, or (2) a suite of spectral variables, or (3) the combination of phenological metrics and spectral variables. First, several annual time series of remotely sensed data were assembled: accumulated growing degree-days (AGDD) from the MODIS 8-day land surface temperature products, 2-band Enhanced Vegetation Index (EVI2), and the spectral variables from the Harmonized Landsat Sentinel-2 as well as from the USGS Landsat Analysis Ready Data surface reflectance products. Then, at each pixel, EVI2 time series were fitted using two different land surface phenology models: the Convex Quadratic model (CxQ), in which EVI2=f(AGDD); and the Hybrid Piecewise Logistic Model (HPLM), in which EVI2=f(Day of Year). Phenometrics and spectral variables were

submitted separately and together to Random Forest Classifiers (RFC) to depict land cover/land use in Roberts County, South Dakota. HPLM RFC models showed slightly better accuracy than CxQ RFC models (about 1% relative higher in overall accuracy), mostly due to more accurate location of land cover. Compared to phenometrically-based RFC models, spectrally-based RFC models yielded more accurate land cover maps, especially for non-crop cover types. However, the RFC models built from spectral variables could not classify accurately the wheat class, which contained mostly spring wheat with some fields in durum or winter varieties. The most accurate RFC models were obtained when using both phenometrics and spectral variables as input. The combined-variables RFC models overcame weaknesses of both phenometrically-based classification (low accuracy for non-vegetated covers) and spectrally-based classification (low accuracy for wheat). The analysis of important variables indicated that land cover classification for this study area was strongly driven by variables related to the initial green-up phase of seasonal growth and maximum fitted EVI2. For a deeper evaluation of RFC performance, RFC classifications were also executed with several alternative sampling scenarios, including different spatiotemporal filters to improve accuracy of sample pools and different sample sizes. Results indicated that a sample pool with less filtering yielded the most accurate predicted land cover map, and a stratified-random sample dataset covering approximately 0.25% or more of the study area was required to achieve an accurate land cover map. In case of data scarcity, a smaller dataset might be acceptable, but should not smaller than 0.05% of the study area.

4.1   Introduction

Knowledge about land cover/land use (LCLU) is fundamental for natural resource management, agricultural policy making, as well as for regional and urban planning. Data sources for LCLU information include the periodic surveys from governmental agencies, e.g., the National Resource Inventory and the National Agricultural Statistics Service (NASS), both in the United State Department of Agriculture (USDA) (*Goebel, 1998; Miller et al., 2009*). However, those datasets often lack spatial and temporal details, which prevents comprehensive analysis of land change. Remote sensing technology can complement field observations and surveys. Conventional classification approaches, such as those applied in the National Land Cover Dataset (NLCD) (*Homer et al., 2001; Xian et al, 2009; Homer et al., 2015*) and Cropland Data Layer (CDL) (*Boryan, 2018*), were developed in an era of data scarcity and limited computational power and data storage. Thus, they have focused on mapping annual land cover from multispectral data from one or just a few image dates. However, in areas with frequent morning cloud cover, collecting even a few cloud-free scenes over a year can be challenging. The recent rapid increase of accessible Earth observation data coupled with improved computing and storage capabilities is leading to the emergence of methods for mapping land cover using multi-date imagery and dense image time series (*Gómez et al, 2016*). Compared to the traditional approach, the use of image time series often improves classification accuracy by incorporating both spectral and temporal profiles (*Key et al., 2001; Mitchell et al., 2013; Franklin et al., 2015*).

Land surface phenology (LSP) has been a useful approach to characterize seasonal vegetation dynamics on vegetation index time series (*Henebry & de Beurs, 2013*). Over the past few years, several efforts have been made to map LCLU using phenological

metrics derived from satellite image time series with promising outcomes (*Zhong et al., 2011; Jia et al., 2014; Xue et al., 2014; Yan et al., 2015; Kong et al., 2016; Qader et al., 2016; Nguyen et al., 2019*). Due to the relatively low return interval of orbital sensors with spatial resolutions finer than 50 m, many studies—with notable exceptions *Jia et al., 2014*, *Kong et al., 2016*, and *Nguyen et al., 2019*—have relied on MODIS time series to capture phenological characteristics of land surfaces, and thus, often produced cover maps at spatial resolutions (e.g., 250-1000 m) that is coarse relative to human land uses such as agriculture and settlements. To overcome limited temporal coverage of Landsat-like data and map land covers at finer spatial resolutions, *Jia et al. (2014)* and *Kong et al. (2016)* fused MODIS NDVI (Normalized Different Vegetation Index) (*Tucker, 1979*) with Landsat and GF-1 NDVI time series, respectively. Although each produced land cover maps at finer spatial resolution (at 30 m for Landsat and 16 m for GF-1), neither *Jia et al. (2014)* nor *Kong et al. (2016)* were able to map more than Level-1 NLCD Land Cover Classification System, except for coniferous and broadleaf forest in *Kong et al., 2016* (Level-2 NLCD).

In 2016, the United States Geological Survey reorganized the Landsat archive into a tiered collection, namely the Landsat Collections, to facilitate time series analysis and data stacking (*USGS, 2018a*). Taking advantage of the Landsat Collections data, *Nguyen et al. (2019)* performed a phenometrically-based classification for sample areas in South Dakota using all available Tier-1 (highest quality) images from Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI). At each pixel, EVI time series calculated from Landsat Collections data was simulated as a convex quadratic function of accumulated growing

degree days (AGDD). Results showed that classification using only phenometrics generated from the fitted model could map accurately broad thematic land cover classes (water, developed, grassland) as well as commodity crops (corn/maize, soybean, wheat) in Codington and Roberts counties in South Dakota for two years (2012 and 2014). However, they also pointed out some challenges of phenometrically-based classification. First, the classification accuracy varied since the form of the chosen LSP model might be more suitable for some certain vegetation types than others. Second, the phenometrically-based classification performed well only for vegetated classes, particularly crops. Third, many cloud/snow/shadow-free observations were needed at each pixel over a year to fit the LSP model well and to avoid data gaps in the predicted land cover map. Regarding the last point, they also showed that an adequate number of observations could be gathered by combining data from comparable sensors, especially in sidelap zones of Landsat swaths. Finally, in addition to pointing out the challenges of classification based on phenometrics, *Nguyen et al. (2019) also* discussed the potential opportunity to improve classification accuracy by incorporating both phenological and spectral variables.

Here we explore challenges of phenometrically-based classification and potential way to improve classification accuracy as demonstrated in *Nguyen et al. (2019)*. This study focuses on evaluating the performance of alternative land cover classifications using either (1) only phenological metrics derived from either of different land surface phenology (LSP) models: the Convex Quadratic Model, in which EVI2=f(AGDD) *(Henebry & de Beurs, 2013; de Beurs & Henebry, 2004)* and the Hybrid Piecewise Logistic Model, in which EVI2=f(Day of Year) *(Zhang, 2015)*, or (2) a suite of spectral variables, or (3) both phenological metrics and spectral variables. In our evaluation, we address three research

questions. The first question is whether the maps from the phenometrics are more accurate than maps from spectral variables alone. As land surface phenology have been a useful tool to characterize the dynamics of the vegetated land surface *(Henebry & de Beurs, 2013)*, we hypothesized that land cover classifications using only phenometrics could be more accurate for vegetated land covers, especially for commodity crops, than those using only spectral variables. The second question asks which set of phenometrics—derived either from the Convex Quadratic Model (CxQ) or from the Hybrid Piecewise Logistic Model (HPLM)—performs better. In the temperate ecosystem, plant development is sensitive to variation in temperature. We hypothesized, therefore, that the Convex Quadratic model, which links vegetation growth with the progression of thermal time, would be better suited to land cover classification of our study area in northeastern South Dakota. The third question asks whether combining the phenometrics and spectral variables would result in superior performance. Studies have indicated that classification accuracies were improved by incorporating phenological features *(Jia et al., 2014; Kong et al., 2016)*. Thus, we hypothesized that classification using the combination of spectral variables and phenometrics would be consistently more accurate than those using only phenometrics or spectral variables. To build a more complete picture of classification performance, we ran RFCs with different sampling scenarios and sets of input variables.

First, three annual time series of remotely sensed data were constructed, including accumulated growing degree-days from the MODIS 8-day composites of land surface temperatures and 2-band Enhanced Vegetation Index (*Jiang et al., 2008*) and spectral variables from surface reflectance products from (1) Landsat Analysis Ready Data (ARD) and (2) Harmonized Landsat Sentinel-2 (HLS) data, separately. At each pixel, EVI2 time

series were then fitted to the LSP models: CxQ or HPLM. Phenometrics derived from the fitted LSP models as well as spectral variables were submitted individually and in combination to Random Forest Classifiers (RFC) to map land cover/ land use of the study area. Accuracy assessments for both RFC models and predicted land cover maps were reported using both conventional accuracy metrics (overall, producer's, and user's accuracies) (*Congalton & Green, 2008*) and alternatives for kappa (*Pontius & Millones, 2011*).

Our assessment of classification performance is twofold. First, RFC model performance was evaluated by submitting different input datasets randomly generated from the CDL. Accuracy comparisons between classification scenarios were tested by both Mann-Whitney U *(Nachar, 2008)* and equivalence tests *(Foody, 2009; Lakens, 2017)*. These two tests are based on opposite but complementary evaluation perspectives. The nonparametric U test indicates whether the two sets of accuracy metrics are statistically different, regardless how difference magnitude. The equivalence test, on the other hand, examines whether differences fall within a certain user-defined threshold and, thus, deemed equivalent or are large enough to be deemed not equivalent. The second step was to compare the predicted land cover maps with the CDL.

4.2   Data and Study Area

4.2.1. Study Area

The proposed classification exercise was demonstrated for Roberts County, SD in two years, 2016 and 2017. Roberts County is at the northeastern corner of South Dakota with a total area of 2,940 km2 and a current population of approximately 10,000. According to 2016 Cropland Data Layer (Figure 1), cropland is a dominant land cover in Roberts

County, accounting for approximately 53.3% of the county area. Other cover types in the County include grassland (25.9%), wetland (8.9%), water (5.2%), barren/developed (4.4%), and forest (2.2%). The County falls within overlap zones of Landsat paths, which allows retrieval of more cloud-free observations.
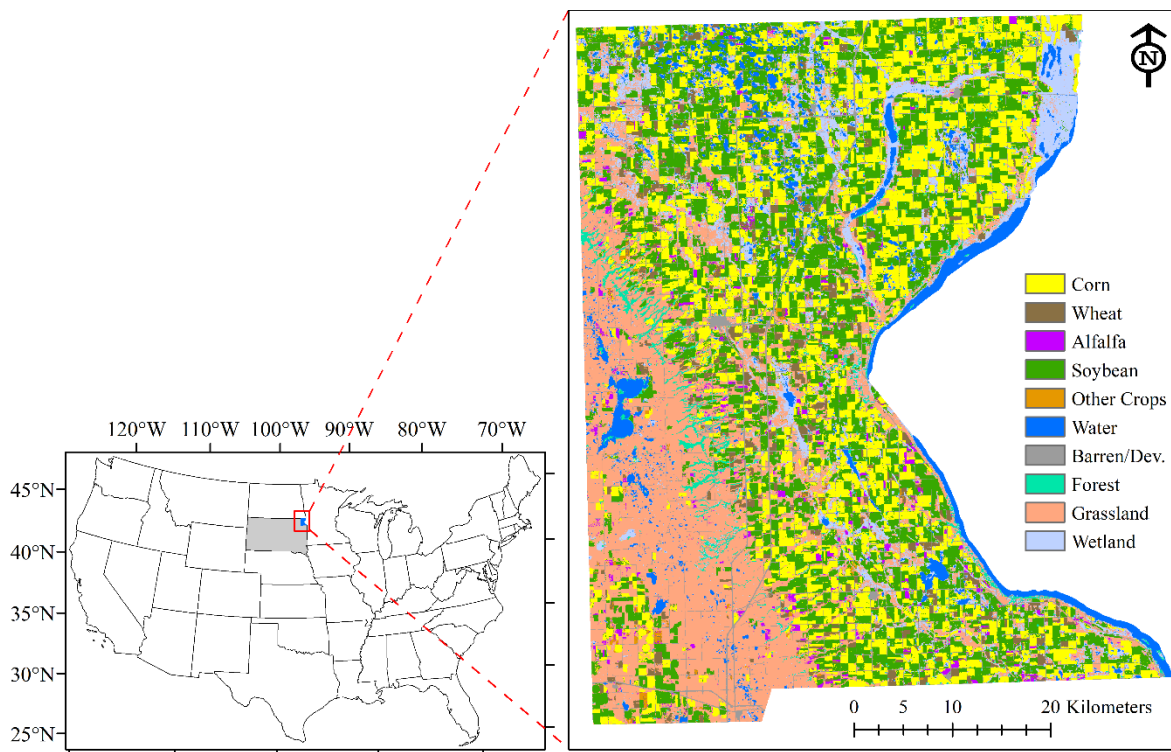
Figure 4.1. The 2016 reclassified Cropland Data Layer for Roberts County, South Dakota.

4.2.2. Input Data

4.2.2.1. U.S. Landsat Analysis Ready Data

The Landsat Analysis Ready Data products from the U.S. Geological Survey are designed to reduce the amount of data preparation for scientists and to facilitate time series analysis by generating data at the highest scientific standards required for direct use in applications (*USGS, 2018b*). Landsat Collection 1 Level-1 scenes serve as the input for generating all ARD products. The ARD dataset is defined in the Albers Equal Area (AEA)

projection and World Geodetic System 1984 datum (WGS84). The products are distributed in 150 × 150 km tiles instead of the traditional Landsat swaths in the WRS-2 path-row coordinate system. We used the ARD surface reflectance (SR) product.

4.2.2.2. Harmonized Landsat Sentinel-2

The Harmonized Landsat and Sentinel-2 product suite is a combined surface reflectance dataset consisting observations from both the Landsat 8 Operational Land Imager and Sentinel-2 Multi-Spectral Instrument (MSI) (*Claverie et al., 2018*). We used two products from HLS version 1.4: (1) S30 - SR derived from Sentinel-2 MSI L1C data and resampled to 30m; and (2) L30 - 30m SR derived from Landsat-8 OLI L1T data. Both S30 and L30 products provide nadir BRDF-adjusted reflectance (NBAR) data gridded with the Sentinel-2 tiling system in Universal Transverse Mercator (UTM) projection and World Geodetic System 1984 datum (WGS84). The Sentinel-2 MSI radiometry is adjusted to mimic the spectral bandpasses of Landsat 8 OLI for visible, near infrared, and shortwave infrared bands.

4.2.2.3. MODIS Land Surface Temperate

We used the Collection 6 MODIS level-3 land surface temperature (LST) 8-day composites at 1000 m spatial resolution from both Aqua (MYD11A2) and Terra (MOD11A2) satellites (*Wan et al., 2015a&b*). The MODIS LST data are provided in a sinusoidal grid format and display the mean clear-sky LST in Kelvin observed during the 8-day compositing period. All MODIS data were reprojected and resampled to 30 m using bilinear interpolation into UTM zone 14N to work with the HLS data and into AEA projection to work with the ARD. The LST time series were converted from Kelvin to degrees Celsius for calculation of thermal time used in the LSP modeling.

4.2.2.4. Cropland Data Layer

The USDA Cropland Data Layer (CDL) is a crop-specific land cover raster created annually for the continental United States by the NASS using moderate resolution satellite imagery and extensive agricultural ground observations (*Boryan et al., 2011*). It is distributed in AEA projection and North American 1983 datum (NAD83). The CDL was first produced in 1997 for North Dakota but has covered the contiguous US yearly only since 2008. The product has approximately 130 classes and a spatial resolution of 30 m at best. We regrouped the CDL layers into ten classes (Table S4.1), and then used this reclassified data to generate sample datasets for input to the RFCs. The reclassified CDL layer also provided a reference against which to evaluate the predicted land cover maps. To work with HLS data, the CDL data were reprojected into UTM zone 14N. Due to differences in the original projections and datums, the reclassified CDL, ARD, and HLS pixels are not perfectly co-aligned. While offsets between the CDL and ARD pixels are only about 3 meters in both latitude and longitude direction, offsets between the CDL and HLS pixels are 15 meters (half pixel) in each direction. We did not resample these data into a common grid, as this step would introduce another source of uncertainty into the analysis.

4.3   Methodology

4.3.1. Land surface phenology modeling

(a) EVI2 time series from ARD and HLS surface reflectance

The two-band Enhanced Vegetation Index (EVI2) was calculated from ARD and HLS surface reflectance products using Equation 4.1 (*Jiang et al., 2008*). Poor-quality observations—snow, high confidence cloud, or cloud shadow pixels—were masked out using quality control layers delivered with the ARD and HLS products. EVI2 values

outside the valid range (from 0 to 1) were also excluded. The remaining "good" EVI2

values at each pixel were then stacked in chronological order from the first day of the year

(DOY=1) to the final day of the year (DOY=365 or 366 in leap years).

$$EVI2 = 2.5 \ \frac{(NIR-R)}{(NIR+2.4R+1)} \qquad (4.1)$$

(b) AGDD time series from MODIS LST

From MODIS LST, we calculated the accumulated growing degree-days (AGDD)

as follow:

$$GDD_t = \max\left\{\frac{T_{max,t}+T_{min,t}}{2}, 0\right\} \qquad (4.2)$$

$$AGDD_t = AGDD_{t-1} + 8 \times GDD_t \qquad (4.3)$$

where GDDt are the growing degree-days for compositing period (t), Tmax,t and Tmin,t

are the highest and lowest LST values from available MODIS observations from both Aqua

and Terra during the compositing period. Since the compositing period is 8 days, we

multiplied the GDD by 8 to achieve a proportional accumulation of GDD for each of 46

composites per year.

(c) Convex Quadratic Model

We fitted the EVI2 time series as a quadratic function of AGDD (Equation 4) using

the fitting process described in *Nguyen et al., 2019*:

$$EVI2 = \alpha + \beta \times AGDD - \gamma \times AGDD^2 \qquad (4.4)$$

where α, β, γ (alpha, beta, and gamma, respectively) are the parameter coefficients to be

fitted. Alpha—a constant component—regulate directly the peak EVI2 value over the

growing season as changing value of alpha solely would move the fitted curve up or down

along the EVI2-axis. Beta—a linear component—affects the position of the peak on the

thermal time axis (timing of peak growth) as changing value of beta solely would move

the fitted curve in an upward quadratic pattern. Changing value of gamma—a quadratic component—would make the fitted quadratic curve become thinner or fatter (how fast values on the two sides depart from the peak). The negative sign on gamma in Equation 4 indicates that we accepted only a fitted curve that is downward arching, since the EVI2 values will rise, peak, and then decrease over the growing season. From each fitted model, we derived a suite of 17 variables to be used the LCLU classification, including fitted parameter coefficients, derived phenological metrics (phenometrics), and model fit statistics (Table 4.1).

Table 4.1. Variables derived from the Convex Quadratic Model

Parameters	Meaning
α, β, γ	Fitted parameter coefficients of CxQ model (Equation 4.4)
TTP_{CxQ}	Thermal time to peak (AGDD at the max fitted EVI2) (TTP = $-\beta/2\times\gamma$)
PH_{CxQ}	Peak height EVI2 (max fitted EVI2) (PH = $\alpha - \beta^2/4\times\gamma$)
HTV	Value of EVI2 at half-TTP (HTV = $\alpha + \beta\times TTP/2 + \gamma\times TTP^2/4$)
ymax	Highest observed EVI2
R^2	Coefficient of determination of the fitted model
lpos, rpos	Observation index of start and end of the fitting window
o_all	The total number of "good" observations
o_fit	Number of observations used to fit the CxQ model
o_per	Ratio between "o_fit" and "o_all"
minx, maxx	AGDD at left and right ends of the fitted curve in the first quadrant
peaks	Number of high EVI2 values (≥0.8*ymax) outside the fitting window
jumps	Number of times that $\Delta EVI2 \geq 0.2$

(d) Hybrid Piecewise Logistic Model

The Hybrid Piecewise Logistic Model (HPLM) (*Zhang, 2015*) is an improvement of the widely-used logistic model that formed the basis for the MODIS Land Cover Dynamics product (MCD12Q2) before Collection 6 (*Zhang et al., 2003*). During the growing season, plant can suffer from water stress or other impacts leading to a different greenness trajectory compared to one under favorable weather condition. A key advance in the HPLM was incorporation of alternative conditions for vegetation growth: favorable or stressed. To determine whether the plant is under favorable or stressed conditions, the two functions of Equation 4.5 were fitted to the EVI2 time series and the function with a higher agreement index was chosen.

$$\text{EVI2} = \begin{cases} \frac{c_1}{1+e^{a_1+b_1 t}} + \text{EVI2}_b \\ \frac{c_2+dt}{1+e^{a_2+b_2 t}} + \text{EVI2}_b \end{cases} \quad (4.5)$$

where t is time in the day of year (DOY), a is related to the vegetation growth time, b is associated with the rate of plant leaf development, c is the amplitude of EVI2 variation, d is a vegetation stress factor, EVI2b is the background EVI2 value, and the subscripts 1 and 2 refer to parameters for favorable and stressed conditions, respectively. From each fitted model, we derived a suite of 14 variables to be used in LCLU classification, including timings of vegetation growth and corresponding EVI2 values (Table 4.2). We note that fitted parameter coefficients from the HPLM were not used directly in classification (as with the CxQ) because the EVI2 time series at each pixel were fitted with multiple logistic curves.

Table 4.2. Variable derived from the Hybrid Piecewise Logistic Model.

Parameters	Meaning
gri, vi_gri	DOY and EVI2 of green-up start
gre, vi_gre	DOY and EVI 2 of green-up end
grMD, vi_grMD	Middle of "gri" and "gre" and its corresponded EVI2
sei, vi_sei	DOY and EVI 2 of senescence start
see, vi_see	DOY and EVI 2 of senescence end
se_MD, vi_seMD	Middle of "sei" and "see" and its corresponded EVI2
DP_{HPLM}, PH_{HPLM}	DOY with the highest fitted EVI2 and its EVI2

4.3.2. Spectral variables

From the ARD and HLS surface reflectance, we generated three sets of annual spectral variables, including the 20th, 50th and 80th percentiles of blue (B), green (G), red (R), NIR (band 8A in HLS), SWIR 1 (S1) and SWIR 2 (S2). For each set of percentiles, twelve normalized band ratios were computed, including: (G-R)/(G+R); (NIR–R)/(NIR+R); (NIR–B)/(NIR+B); (NIR–G)/(NIR+G); (S1–R)/(S1+R); (S1–B)/(S1+B); (S1–G)/(S1+G); (S1–NIR)/(S1+NIR); (S2–R)/(S2+R); (S2–B)/(S2+B); (S2–G)/(S2+G); and (S2–NIR)/(S2+NIR). The 20th and 80th percentiles were used, to reduce sensitivity to shadows and residual cloud and atmospheric contamination effects. Similar variables were used previously to produce the NLCD-like land cover map for North America using WELD data (*Zhang & Roy, 2017*). In total, 42 spectral variables were generated for each 30 m pixel location. Those variables were named using the following convention: "percentile_band(normalized ratio)" (e.g., "P50_S2R" is a normalized ratio between the 50th percentile of SWIR-2 and Red bands).

4.3.3. Land cover/ land use classification using Random Forest Classifier

The Random Forest Classifier (RFC) (*Breiman, 2001*) is an ensemble of decision trees—each created with a random subset of training samples and variables—and allows them to vote for the most popular class. By growing "random forest" of multiple trees N, RFC creates a set of classification rules with high variance but low bias. The size and design of sample data have been found to affect RFC models (*Colditz, 2015; Millard & Richardson, 2015*). To better understand those influences in our study, we performed land cover classifications using different scenarios: (1) sample pools—different ways to build sample pools from the CDL; and (2) sample sizes—different sizes of sample datasets selecting from the pool. In addition, we examined RFC models arising from various sets of input variables. We generated 12,800 RFC models in total—50 trials $\times$ 2 years $\times$ 2 input data source $\times$ 4 sample pools $\times$ 4 sample sizes $\times$ 4 sets of input variables— using the "scikit-learn" library in Python (*Pedregosa et al., 2011*). For each trial, a new sample dataset was randomly selected from the CDL data pool. All sample datasets were class-balanced (same proportional distribution of cover types to the CDL) and divided half for training and half for testing.

(a) Sample pool scenarios

Although RFC is not very sensitive to mislabeled pixels in the sample dataset (*Gislason et al., 2006; Rodriguez-Galiano et al., 2012*), it is still critical to improve land cover accuracy in our sample data as they contain considerable error. First, the overall accuracy of the agriculture class for the 2016 and 2017 CDL are only 89.3% and 81.7%, respectively, and it is likely worse for non-agricultural classes. In addition, CDL pixels are not perfectly co-aligned with ARD and HLS pixels due to differences in their original

datums and projections and, thus, may lead to incorrect land cover information when selecting the sample dataset. To improve the accuracy of the land cover information, we used sample selection by selecting only core pixels from the CDL, i.e., pixels surrounded by pixels of the same type, to avoid misclassification that can occur more frequently at the edge and off-sets between CDL and ARD/ HLS pixels. Another way to increase accuracy of the sample data is to compare land cover types at a same pixel between different years (here 2016 and 2017): a pixel presenting the same cover type for two or more years is more likely to be classified correctly. Improvement in land cover accuracy of the sample dataset may reduce the predictive power of RFC models (despite their good accuracy metrics) since complex spatial characteristics of particular cover types may be excluded through this selection process. In addition, selecting only core pixels may lead to a higher degree of spatial autocorrelation in the sample dataset, thereby inflating accuracy metrics (*Mannel et al., 2011*). To find a good balance between accuracy and representativeness of the sample dataset, we examined land cover classifications arising from four sample pool scenarios as described in Table 4.3.

Table 4.3. Sample pool scenarios.

Name	Practice
C1S	Only keep pixels surround by 8 same neighbors (C1: 1 pixel away from the focal pixel), S: land cover of a single year.
C1M	C1 and matched (M) land cover in 2016 and 2017
C2S	Only keep pixels surround by 24 same neighbors (C2: 2 pixels away from the focal pixel), S: land cover of a single year.
C2M	C2 and matched (M) land cover in 2016 and 2017

(b) Sample size scenarios

Random Forest Classifier performs better with larger sample datasets (*Deng & Wu, 2013; Du et al., 2015*). Tradeoffs for better performance include higher cost in data collection and longer computational time. Although, previous studies suggested that the sample dataset should represent about 0.25% of the total study area (*Nguyen et al., 2019; Colditz, 2015*), it remains unclear how smaller sample datasets might affect classifications. To explore this issue, we examined performance of RFC models using sample datasets at four different sizes of the total county area: 0.01% (P01), 0.05% (P05), 0.15% (P15), and 0.25% (P25).

(c) Input set scenarios:

We examined performance of RFC model using four sets of input variables (Table 4.4) to understand how well phenometrically-based and spectrally-based variables can be used in land cover classification individually and in combination.

Table 4.4. Input variables for RFC modeling.

Name	Practice
CxQ	Use only the 17 variables from the Convex Quadratic Model
HPLM	Use only the 14 variables from the Hybrid Piecewise Logistic Model
SPL	Use only the 42 spectrally-based variables
CMB	Use the combination of 73 variables from CxQ, HPLM and SPL

4.3.4. Accuracy assessment and feature importance of Random Forest Classifier

We evaluated RFC model accuracy assessment (model AA) using multiple metrics, including producer's accuracy (PA), user's accuracy (UA), overall accuracy (OA) (*Zhang, 2015*), and two alternatives to Cohen's kappa: kappa for location (k_L) and kappa for

quantity (k_Q) (*Congalton & Green, 2008*). Given fixed sizes for all cover classes (or fixed proportional distribution), higher k_L indicates larger areas of matched land covers (or larger overlap between the predicted map and the reference). Given a fixed matched land cover area, higher k_Q indicates smaller area of unmatched cover types between the predicted map and the reference. All accuracy metrics are reported for each tested scenario as average values of multiple RFC models. In addition to the mean accuracy metrics, nonparametric Mann–Whitney U test and equivalence test using two one-sided procedure (TOST) were performed to support cross-comparison of RFC performance under different scenarios. For the TOST test, we chose an indifference zone, measured by Cohen's d, of (–0.35, 0.35). The chosen effect size lies between Cohen's suggested values for a small effect size of 0.2 and a medium effect size of 0.5 (*Fritz et al, 2012*). To understand the contribution of each variable to the classification, the sum of Gini Importance (GI) was computed for each variable from 12,800 RFC models. A higher summation value of GI indicates a more important variable.

4.3.5. Ensemble land cover maps from multiple RFC models

A total of 12,800 RFC predicted land cover maps were generated and divided into fourteen major groups for comparison, including four types of sample pools, four types sample sizes, and six types of input variable sets. Each major group was also separated by year (2016 or 2017) and source of input data (ARD or HLS), resulting in 56 smaller groups. In each smaller group, a number of times a particular cover type appeared at each pixel was counted (referred to as Count). Next, an ensemble land cover map was generated for each group by assigning land cover at a particular pixel with cover type that have the highest Count. We then compared those ensemble land cover maps with the CDL.

4.3.6. Cross comparison between predicted maps and the CDL

In addition to the accuracy assessment of the RFC output, we compared the predicted land cover maps with the reclassified CDL. The cross-comparison was reported as the map accuracy assessment (map AA). Although the CDL's accuracy ranged from higher for commodity crops to lower for non-agricultural classes, the CDL remains one of the more reliable land cover datasets for the US. Thus, cross-comparison between our predicted maps and the reclassified CDL should provide a good indicator of the accuracy the ensemble land cover maps generated by the RFC. Note that ARD, HLS predicted land cover maps and the reclassified CDL are in different projections and/or datums. To allow cross-comparison, pixels from those datasets were co-registered to match perfectly to each other. Because off-sets between the CDL and ARD in latitude and longitude directions are small, co-registration between the two layers was just simple pixel snapping—moved ARD pixels to match the CDL pixels on the nearest direction. For cross-comparison between the CDL and HLS data, we examined four different adjustments to HLS pixels: moving the raster half pixel in up-right, down-right, up-left, down-left directions. The up-right adjustment, which yielded the highest number of matched pixels between the CDL and HLS, was reported here.

4.4.   Results

4.4.1. Accuracy assessment of RFC models

Overall accuracy, kappa indices for location and quantity of 2016 RFC models are summarized by sample pools and sizes in Table 4.5. The pairwise comparison of accuracy metrics using the Mann–Whitney U and the TOST equivalence tests appear in Table S4.3 and S4.4, respectively. Generally, RFC models using C2 sample pools (2 pixels away from

the evaluated pixel: C2S, C2M) had significantly higher accuracy metrics than those using C1 sample pools (C1S, C1M) for all combinations of year and data source. Sample pools that matched 2016 and 2017 land covers (M) yielded more accurate RFC models than those based on land cover from only a single year (S). For all combinations of year and data source, RFC models using larger sample size had significant higher accuracy metrics. We observed largest improvements in accuracy metrics from P01 to P05 RFC models with relative increases of 4.1%, 1.7%, and 6.4% for OA, k_L and k_Q, respectively. Larger increases in k_Q compared to k_L indicated that improvement in model accuracy was mostly due to better quantity agreement of P05 compared to P01 RFC models. In other words, proportional distributions of land cover classes in P05 RFC models in are generally closer to the CDL than those of P01 RFC models. P05 samples are five times larger than P01 samples, which enabled better description for all classes, especially minor cover types. Accuracy improvement from P05 to P15 RFC models was moderate with relative increases of 1.7%, 1.1%, and 1.8% for OA, k_L and k_Q, respectively. Relative differences in accuracies of P15 and P25 RFC models were minor, less than 0.6% for all three metrics. Among RFC models using different input datasets, models using phenometrics (CxQ and HPLM) had the lowest accuracy metrics. There was no obvious choice between the 2016 RFC models using CxQ versus HPLM: the HPLM RFC models performed better on ARD data and the CxQ RFC models are better on HLS data. For 2017 data, HPLM RFC models slightly edged CxQ RFC models with less than 1% higher OA (Table S4.2). Although differences in 2017 OA between CxQ and HPLM RFC models were statistically significant in the Mann-Whitney U tests, the TOST equivalence tests indicated that the differences are within a user-defined indifference zone (i.e., the two models are equivalent). Spectrally-

based RFC models (SPL) were more accurate than the phenometrically-based RFC models, with approximately 3% higher OA relatively. Unlike the sample size scenarios, improvement in SPL RFC models mostly came from better location of pixels (small change in proportional distribution) shown by higher relative increase in k_L (approximately 5%) compared to k_Q (approximately 1.7% and -3.2% in 2016 and 2017, respectively), indicating that locations of land covers were described more accurately using spectral information. More importantly, RFC models with combined variables (CMB) consistently outperformed RFC models using solely spectral variables (SPL) or phenometrics (CxQ or HPLM). Similar results were found for the 2017 data (Table S4.2).

Table 4.5. Overall accuracy (in percent), kappa indices for location and quantity of 2016 RFC models summarized by sample pools, sample sizes, and input variables. A particular scenario (current row) was compared to a scenario right above it (above row) using the nonparametric Mann–Whitney U test and the TOST equivalence test. The null hypothesis of the U test is that a random accuracy metric of the first scenario (above row) will be less than a random accuracy metric of the second scenario (current row). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are highlighted in light blue for "not equivalent" and light yellow for "equivalent". Full pairwise comparisons are provided in Tables S4.3 and S4.4.

	Scenario	ARD OA		k_L		k_Q		HLS OA		k_L		k_Q	
Sample Pool	C1S	88.8		0.904		0.917		86.8		0.884		0.906	
	C1M	90.7	***	0.923	***	0.926	***	88.7	***	0.904	***	0.914	***
	C2S	90.4	NS	0.921	NS	0.924	NS	89.4	***	0.909	**	0.921	***
	C2M	91.8	***	0.935	***	0.932	***	90.7	***	0.922	***	0.927	***
Sample Size	P01	87.1		0.906		0.877		84.8		0.883		0.862	
	P05	90.5	***	0.919	***	0.929	***	89.0	***	0.903	***	0.922	***
	P15	91.8	***	0.927	***	0.944	***	90.6	***	0.914	***	0.940	***
	P25	92.3	***	0.931	*	0.949	***	91.2	***	0.919	**	0.945	***
Input Set	CxQ	86.8		0.879		0.914		86.0		0.871		0.911	
	HPLM	88.6	***	0.900	***	0.919	NS	85.2	NS	0.867	NS	0.898	NS
	SPL	92.2	***	0.943	***	0.927	***	91.0	***	0.928	***	0.925	***
	CMB	94.1	***	0.961	***	0.938	***	93.4	***	0.953	***	0.936	***

Tables 4.6 and S4.5 show producer's and user's accuracies for the RFC models using C1S and C2M sample pools (the worst and the best sample pool scenarios based on

results in Tables 4.5 and Table S4.2). Between C1S and C2M RFC models, relative differences in both producer's and user's accuracies were less than 2.5% for corn, soybean, and water classes. Those three classes were also higher accuracy classes. C2M RFC models had relatively higher producer's and user's accuracies than C1S RFC models in all other classes, including wheat (4.6%-12.5%), alfalfa (11.4%-16.3%), barren/developed (6.7%-18.4%), wetland (16.6%-20.4%), and other crops (90% - 356%). Compared to corn, soybean, and water, the other cover types have more complicated aggregates of phenological and spectral characteristics that make it more difficult to map those classes. For example, barren/developed includes both vegetated (lawn, garden) and non-vegetated (barren, impervious surface) land covers. In addition, minor crops and non-agriculture classes are likely to have lower accuracy in the CDL compared to corn and soybean (commodity crops) or to open water (distinct spectral characteristics), resulting in lower accuracy in the training and validating data. Nevertheless, all improvements from C1S and C2M RFC models were statistically significant (Table S4.6). However, differences in PA/UA of corn, soybean, water and barren/developed were generally within the indifference zones (or no obvious improvements for those classes).

Between the least accurate and the most accurate sample size scenarios (P01 versus P25 RFC models), relative differences in water was less than 2.3% (Tables 4.7, S4.7), likely due to very distinct spectral responses of water compared to other covers. Relative improvements in PA/UA of major classes (corn, soybean, and grassland) were also minor (less than 5%) as there were already many training pixels in each class even with the smallest sample size. However, increases in both PA and UA of those classes were statistically significant (Table S4.6). Both producer's and user's accuracies improved

significantly for minor crops and non-agricultural cover types (10%-70% relative higher PA/UA), including wheat, alfalfa, other crops, wetland, and barren/developed. Considering that minor cover types have mixed spectral and phenological characteristics, larger sample sizes would allow those classes to be described more thoroughly in the training, thereby improving accuracies.

Table 4.6. Producer's and user's accuracies (in percent) of 2016 RFC models using C1S and C2M sample pools. Significance level of the U test (C1S < C2M) across rows are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence tests across rows are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

| | Producer's Accuracy (%) | | | | | | User's Accuracy (%) | | | | | |
| | ARD | | | HLS | | | ARD | | | HLS | | |
Land Cover	C1S	C2M		C1S	C2M		C1S	C2M		C1S	C2M	
Corn	94.6	95.7	***	91.2	92.6	***	94.8	95.9	***	89.8	91.5	***
Wheat	75.4	78.6	***	70.2	74.4	***	84.4	90.2	***	82.0	89.1	***
Alfalfa	73.5	82.6	***	69.3	79.9	***	83.7	91.2	***	79.8	89.6	***
Soybean	95.4	96.6	***	90.3	92.1	***	93.6	95.2	***	90.4	92.2	***
Other Crops	11.2	50.9	***	15.3	55.0	***	35.8	70.2	***	46.6	71.0	***
Water	97.4	98.3	***	96.2	98.2	***	97.5	97.6	***	96.3	96.9	***
Barren/Dev.	55.1	60.9	***	47.4	56.1	***	77.0	79.0	***	72.4	75.9	***
Forest	87.7	94.0	***	84.3	94.5	***	86.7	91.1	***	84.8	92.4	***
Grassland	93.0	94.9	***	93.6	96.1	***	84.9	88.6	***	85.6	90.4	***
Wetland	66.8	79.2	***	73.9	87.7	***	74.2	84.4	***	76.3	87.5	***

Table 4.7. Producer's and user's accuracies (in percent) of 2016 RFC models using P01 and P25 sample sizes. Significance level of the U test (P01 < P25) across rows are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence tests across rows are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

	Producer's Accuracy (%)								User's Accuracy (%)							
	ARD			HLS					ARD			HLS			
Land Cover	P01	P25		P01	P25				P01	P25		P01	P25	
Corn	94.1	96.2	***	89.8	93.6	***			93.0	96.7	***	87.4	92.6	***
Wheat	58.1	87.4	***	49.6	85.0	***			84.5	90.2	NS	80.6	89.0	NS
Alfalfa	61.9	87.9	***	55.0	86.5	***			75.2	92.9	NS	69.9	92.0	NS
Soybean	94.6	96.8	***	88.5	92.8	***			92.2	96.1	***	87.9	93.6	***
Other Crops	7.1	48.3	***	9.1	55.9	***			7.0	84.7	***	9.0	89.5	***
Water	97.5	98.5	NS	97.0	97.9	NS			97.4	97.8	NS	96.1	96.8	NS
Barren/Dev.	49.6	64.2	***	38.8	58.9	***			72.5	81.7	**	66.0	78.9	***
Forest	86.6	93.4	NS	84.5	92.8	NS			85.3	91.6	NS	84.7	91.6	NS
Grassland	92.6	94.7	***	93.7	95.7	***			83.7	88.5	***	84.8	90.1	***
Wetland	64.8	77.6	***	74.5	84.5	***			72.8	83.5	***	74.6	86.3	***

Tables 4.8 and S4.8 show PA/UA of RFC models using different sets of input variables. SPL RFC models generally performed better than CxQ and HPLM RFC models in most classes, including the three dominant cover types: corn, soybean, and grassland. However, only increases in non-crop types were significant (Table S4.9). Compared to phenometrically-based models, SPL RFC models were much more accurate in barren/developed, forest and wetland. On the other hand, CxQ and HPLM RFC models yielded higher PA/UA values for wheat and other crops. Among all scenarios, RFC models using combined set of variables consistently had the highest accuracy metrics. The CMB RFC models overcame weaknesses of both SPL RFC models (wheat and other crops) and phenometrically-based RFC models (barren/developed) (Table 4.8, S4.8, S4.10, S4.11).

Table 4.8. Producer's and user's accuracies in percent (%) of 2016 RFC models summarized by sets of input variables. A particular scenario (current column) was compared to a scenario on the left (left column) using nonparametric Mann–Whitney U and TOST equivalence tests. The null hypothesis of the U test is that a random accuracy metric of the first scenario (left column) will be less than a random accuracy metric of the second scenario (current column). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of TOST equivalence tests are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

Metrics	ARD							HLS						
	CxQ	HPLM		SPL		CMB		CxQ	HPLM		SPL		CMB	
PA_Corn	93.0	93.8	***	96.0	***	98.0	***	89.5	90.0	***	93.2	***	95.6	***
PA_Wheat	79.3	75.5	NS	66.1	NS	89.1	***	75.6	71.8	NS	60.0	NS	83.2	***
PA_Alfalfa	63.6	87.0	***	80.8	NS	84.0	***	61.1	85.0	***	77.6	NS	79.9	***
PA_Soybean	93.2	95.2	***	97.3	***	98.5	***	87.0	88.4	***	94.2	***	95.7	***
PA_Other.Crops	36.2	34.2	NS	14.1	NS	32.6	***	34.9	36.7	NS	31.3	NS	36.6	**
PA_Water	94.5	98.8	***	99.6	***	99.5	NS	93.7	97.3	***	99.6	***	99.5	NS
PA_Barren/Dev.	40.2	48.3	***	73.2	***	72.6	NS	37.3	40.6	***	64.6	***	62.5	NS
PA_Forest	82.8	89.2	***	95.9	***	96.4	*	80.3	88.7	***	95.2	***	95.5	NS
PA_Grassland	91.1	92.5	***	95.8	***	96.3	***	93.7	92.6	NS	96.1	***	97.4	***
PA_Wetland	68.4	67.7	NS	77.5	***	78.9	***	84.0	64.6	NS	86.0	***	89.4	***
UA_Corn	92.0	94.3	***	97.1	***	98.1	***	86.0	87.2	***	94.4	***	95.6	***
UA_Wheat	88.4	88.5	*	82.1	NS	92.6	***	88.5	84.7	NS	77.6	NS	92.3	***
UA_Alfalfa	79.5	88.8	NS	89.4	***	93.7	***	78.3	87.8	NS	87.5	***	89.5	***
UA_Soybean	92.4	95.4	***	93.1	NS	97.2	***	89.3	91.4	***	90.4	NS	94.8	***
UA_Other Crops	52.9	59.3	***	42.4	NS	62.7	***	56.2	58.4	NS	59.7	***	64.3	*
UA_Water	93.3	97.7	***	99.8	***	99.7	NS	92.3	94.3	***	99.7	***	99.4	NS
UA_Barren/Dev.	61.1	67.1	***	93.0	***	92.1	NS	56.3	62.1	***	89.9	***	88.2	NS
UA_Forest	83.6	85.1	***	93.5	***	94.5	**	83.0	85.2	***	92.9	***	94.2	***
UA_Grassland	83.6	84.4	***	89.5	***	89.5	NS	88.1	83.1	NS	90.2	***	91.2	***
UA_Wetland	73.2	73.2	NS	84.4	***	87.5	***	79.6	72.6	NS	86.5	***	90.1	***

4.4.2. Variable Importance

Table 4.9 presents the top ten important variables for the phenometrically-based RFC models. For CxQ RFC models, the three fitted parameter coefficients, HTV, and minx

were consistently in top six most important variables. All three CxQ parameter coefficients (α, β, γ) contributed significantly to the classification as the entire EVI2 pattern of the growing season can be described with those three values. The three fitted parameter coefficients of CxQ model followed the same rank order (decreasing importance: $\alpha \rightarrow \beta \rightarrow \gamma$) for all four combinations of year and data source, indicating that phenological characteristics did not contribute equally to the classification. Alpha (α) was consistently ranked as the most important variable among three parameter coefficients as well as all other CxQ variable indicating that peak fitted EVI2 is a main driver to the classification. Note that α, PHCxQ and ymax—constant component of the quadratic curve, max fitted EVI2, max observed EVI2—are correlated to each other as those variables all refer to the highest EVI2 value over the growing season. Thus, contribution of one variable to classification will lower contributions of other variables. Nevertheless, PHCxQ and ymax also consistently appeared in the sixth and seventh places. The second-ranked important phenological property was the rate of green-up controlled by beta (β) fitted parameter coefficient value. Both HTV and minx—EVI2 value at half TTP and AGDD value at the left end of the fitted curve in the first quadrant—were measurements on the first half of the growing season indicating strong influence of variables related to the initial green-up phase of seasonal growth.

For HPLM RFC models, PHHPL, gri, vi_gri, giMD, gre—highest fitted EVI2, DOY of green-up start, EVI2 at gri, DOY in the middle of green-up start and end, DOY of green-up end—were consistently in top six important variables. Similar to the CxQ RFC models, the modeled peak EVI2 value is an important variable for HPLM RFC models. Moreover, gri, vi_gri, giMD and gre were also timings in the first half of the growing

season, confirming the strong influence of variables related to the initial green-up phase found with the CxQ RFC models.

In the spectrally-based RFC models, contribution of SWIR-2 (S2) was striking: it appeared twelve times (out of 20) in top five important variables (Table 4.10). Despite that SPL RFC models tend to perform slightly better than CxQ and HPLM RFC models (Tables 4.5, 4.8), more phenometrically-based variables were considered important for classification in the CMB RFC models (Table 4.10). The LSP related variables from CxQ (α, β, γ, HTV, minx) and HPLM (gri, giMD) consistently appeared in top ten for important variables in the CMB models. Together, phenometrics from CxQ and HPLM appeared thirty-one times (out of 40) in the top ten important variables. Consistent appearances of HTV, minx, gri, giMD as highly ranked important variables in the CxQ, HPLM as well as in the CMB RFC models indicated that the classification is driven by variables related to the initial green-up phase of seasonal growth.

Table 4.9. Top 10 important variables of CxQ and HPLM RFC models. Variables highlighted in light yellow are those consistently appeared in top 6 most important variables (at least three times out of four year-data combinations). **Bolded** items are variables related to the initial green-up phase of the seasonal growth.

#	CxQ RFC models				HPLM RFC models			
	2016		2017		2016		2017	
	ARD	HLS	ARD	HLS	ARD	HLS	ARD	HLS
1	**α**	**α**	**α**	**α**	**gri**	**gri**	**giMD**	**giMD**
2	**β**	**β**	**HTV**	**HTV**	**giMD**	PH_{HPLM}	**gre**	**gri**
3	**HTV**	ymax	r^2	**minx**	**gre**	vi_sei	**gri**	**gre**
4	**minx**	**minx**	**minx**	**β**	PH_{HPLM}	**giMD**	PH_{HPLM}	see
5	ymax	**HTV**	**β**	r^2	vi_gre	vi_gre	**vi_gri**	PH_{HPLM}
6	**γ**	PH_{CxQ}	**γ**	**γ**	vi_sei	**vi_gri**	see	**vi_gri**
7	PH_{CxQ}	o_fit	ymax	PH_{CxQ}	see	gre	DP_{HPLM}	vi_sei
8	r^2	**γ**	PH_{CxQ}	ymax	vi_gri	see	vi_sei	DP_{HPLM}
9	TTP_{CxQ}	r^2	o_per	TTP_{CxQ}	seMD	seMD	vi_see	vi_gre
10	o_fit	o_all	TTP_{CxQ}	o_per	DP_{HPLM}	vi_seMD	vi_gre	seMD

Table 4.10. Top 10 important variables of SPL and CMB RFC models. In the SPL RFC models, spectral variables in top 5 that involved SWIR-2 band are highlighted in light yellow. In the CMB RFC models, spectral variables are highlighted in light blue, phenometrics from the CxQ and HPLM are highlighted in light orange and light green, respectively. **Bolded** items are variables related to the initial green-up phase of the seasonal growth.

	SPL RFC models				CMB RFC models			
	2016		**2017**		**2016**		**2017**	
#	**ARD**	**HLS**	**ARD**	**HLS**	**ARD**	**HLS**	**ARD**	**HLS**
1	P20_S1	P20_B	P20_S1	P20_B	P20_S1	α	P20_S1	α
2	P20_S2	P80_S2	P80_S2	P20_S2N	giMD	P20_B	α	giMD
3	P80_S2	P80_N	P20_S2	P80_S2	α	β	P20_S2	HTV
4	P20_S2R	P80_S2G	P80_S2G	P20_R	gri	minx	HTV	minx
5	P80_S2G	P20_G	P20_N	P80_S2G	P20_S2	γ	minx	gre
6	P80_N	P20_R	P50_S2N	P20_N	β	o_fit	giMD	β
7	P20_N	P50_NR	P20_R	P20_G	minx	HTV	gre	P20_B
8	P20_S2G	P50_NB	P20_S2N	P80_S2R	HTV	gri	β	gri
9	P80_S2R	P80_S2R	P50_S2	P20_S2R	γ	ymax	γ	P20_R
10	P50_NB	P20_S2R	P20_S2G	P50_NR	P20_S2R	P20_G	gri	γ

4.4.3. Cross-comparison between predicted land cover maps and the CDL

Table 4.11 and S4.12-S4.15 showed results of pixel-based comparison between predicted land cover maps and the CDL for different sample pools and sizes. Unlike model AA, pixel-based comparison between predicted land cover maps and the CDL (map AA) revealed that the C1S RFC map agreed more with the CDL than the C2M RFC map. RFC models trained with larger sample size did yield better land cover prediction. Although improvement of OA in the map AA (about 2% to 3% for ARD and HLS data, respectively) were not as large as those in the model AA (about 6%), this result was expected because many more predictions were made to create land cover map (the entire study area) than in model AA (up to 0.25% of study area). Even a small increase of the map OA (e.g., 0.1%) translates into a large area. Higher k_Q compared to k_L in all scenarios indicated that the

proportional distribution of predicted maps was quite close to the CDL and the majority of classification errors came from misallocation of pixels.

Table 4.11. Pixel-based comparison between 2016 predicted land cover maps and CDL summarized by sample pools and sample sizes.

	Scenario	ARD			HLS		
		OA%	k_L	k_Q	OA%	k_L	k_Q
Sample Pool	C1S	80.8	0.820	0.878	77.9	0.782	0.875
	C1M	80.4	0.819	0.868	77.4	0.777	0.871
	C2S	80.4	0.819	0.867	77.4	0.779	0.866
	C2M	80.1	0.817	0.863	77.0	0.780	0.857
Sample Size	P01	78.9	0.808	0.853	75.5	0.776	0.828
	P05	80.1	0.817	0.866	77.2	0.776	0.868
	P15	80.7	0.821	0.872	78.0	0.783	0.874
	P25	81.0	0.823	0.876	78.3	0.786	0.876

Generally, SPL land cover maps had higher agreement with the CDL than the phenometrically-based maps (Tables 4.12, S4.16). Similar to the model AA, the SPL RFC map clearly improved accuracy of barren/developed but was not as accurate in wheat as CxQ and HPLM RFC maps. Land cover map created from CMB RFC models had higher agreement to the CDL than those created with only spectrally-based or phenometrically-based RFC models. Among all ensemble maps, CMB RFC maps were consistently the most accurate. Compared to phenometrically-based and SPL, CMB maps improved PA and UA for both wheat and barren/developed classes. It is also important to note that the barren/developed areas estimated from the phenometrically-based maps were closer to

CDL compared to values from the spectrally-based maps. However, both PA and UA of barren/developed from the phenometrically-based maps were lower, and thus, resulted in fewer correctly assigned pixels (Table 4.12, Figure 4.2).

Overall accuracies of ARD maps were greater than those of HLS maps by 1% to 3% (Tables 4.11, 4.12). A possible reason leaded to lower agreement between HLS and CDL is the difference in projection of the two datasets. While HLS data was produced using the UTM projection for zone 14N, both the CDL and ARD data were produced in the AEA projection. Although CDL and ARD pixels were not perfectly aligned, offsets were only about 3 meters in both latitude and longitude direction. On the other hand, offsets between HLS and CDL data were 15 meters in each direction. Re-projection and pixel co-registration to allow pixel-based comparison would negatively affect cross-comparison between the CDL and HLS-based maps than cross-comparison between the CDL and ARD-based maps. The large offset between HLS and CDL pixels was also observed in kappa indices for location and quantity. Compared to ARD-based maps, HLS-based maps had similar k_Q but lower k_L values, especially in the more accurate maps: C1S, P25, and CMB (cf. Tables 4.11, 4.12), indicating that proportional distribution of the two maps are similar, but HLS-based maps have less accurate pixel allocation

Figure 4.2. 2016 ARD-based land cover maps of area around Sisseton, the county seat of Roberts County, South Dakota**.** Note the barren/developed class in phenometrics-based RFC maps (CxQ and HPLM) are not as accurate as in SPL and CMB RFC maps.

Table 4.12. Pixel-wise comparison between 2016 predicted land cover maps and CDL summarized by input variables. Land cover area is in km^2.

Land Cover	Info.	CDL	ARD CxQ	ARD HPLM	ARD CMP	ARD CMB	HLS CxQ	HLS HPLM	HLS CMP	HLS CMB
Corn	Area	662	674	636	613	641	701	699	615	644
	UA		83.8	89.0	93.3	93.0	76.9	78.6	88.3	88.2
	PA		85.4	85.5	86.5	90.1	81.6	83.1	82.0	85.8
Wheat	Area	103	86	93	135	121	81	92	132	107
	UA		81.9	80.0	48.9	69.1	79.6	75.0	45.4	71.9
	PA		68.7	72.5	64.2	81.2	62.6	67.6	58.3	74.7
Alfalfa	Area	45	29	36	33	31	28	37	34	33
	UA		73.5	67.2	66.8	74.3	67.2	61.4	59.2	65.9
	PA		47.5	54.5	48.3	51.9	41.7	49.9	45.0	47.9
Soybean	Area	746	693	687	780	731	663	646	769	729
	UA		87.2	91.1	85.0	90.7	83.3	87.0	81.2	86.2
	PA		81.0	83.9	88.8	88.9	74.1	75.3	83.7	84.2
Other Crops	Area	13	1	1	0	0	0	1	0	0
	UA		0.0	61.2	0.0	0.0	0.0	60.8	0.0	0.0
	PA		0.0	4.3	0.0	0.0	0.0	6.9	0.0	0.0
Water	Area	154	122	120	126	121	124	125	131	126
	UA		92.8	94.8	97.0	97.5	89.8	90.6	93.4	94.5
	PA		73.2	74.1	79.5	76.9	72.5	73.6	79.6	77.4
Barren/Dev.	Area	131	69	66	59	60	81	56	53	56
	UA		25.5	29.2	52.9	50.6	19.9	23.4	43.5	39.8
	PA		13.5	14.6	23.9	23.1	12.4	10.0	17.5	17.1
Forest	Area	64	49	65	70	68	41	66	74	68
	UA		57.2	56.4	64.8	65.6	58.7	54.3	56.4	60.2
	PA		43.2	57.3	70.7	69.6	36.9	55.5	65.1	63.2
Grassland	Area	761	872	903	886	916	899	931	917	940
	UA		74.5	73.8	75.7	74.5	72.3	70.5	72.0	72.0
	PA		85.5	87.6	88.2	89.7	85.5	86.4	86.7	88.9
Wetland	Area	262	347	333	239	251	322	286	216	239
	UA		35.9	35.3	56.1	53.1	40.9	36.9	58.1	54.5
	PA		47.5	44.8	51.1	50.8	50.2	40.2	47.9	49.7
OA			74.6	76.4	79.1	80.8	71.7	72.4	75.5	77.7
k_loc			0.748	0.775	0.806	0.821	0.718	0.726	0.763	0.782
k_quan			0.857	0.850	0.862	0.875	0.837	0.837	0.849	0.870

4.5.   Discussion

Convex quadratic versus hybrid piecewise logistic modeling of LSP for land cover classification

The hybrid piecewise logistic model (HPLM) was first designed to detect vegetation phenology from MODIS time series (*Zhang et al., 2003* The HPLM has a well-refined fitting algorithm, but strict requirements in a number and temporal distribution of observations (*Zhang et al., 2003; Zhang, 2015*). On the other hand, the convex quadratic model (CxQ) has been recently to characterize seasonal patterns of vegetated surfaces, by incorporating the use of MODIS LST 8-d composites at 1 km spatial resolution (*Henebry & de Beurs, 2013; Krehbiel et al., 2016, 2017*). The fitting algorithm and data requirements for the CxQ are more flexible than for the HPLM, due to fewer parameter coefficients to estimate (*Nguyen et al., 2019*). When data requirements were satisfied, the HPLM could fit the observed EVI2 pattern more precisely than the CxQ, leading to higher classification accuracy. However, when fewer observations were available, *e.g.*, outside the Landsat sidelap zones, the CxQ model could serve as a back-up algorithm in this temperate climate where temperatures constrain the initiation and tempo of spring growth. The fundamental challenge for both CxQ and HPLM is dealing with gaps in observations during the growing season arising from few good observations available in some years and/or over some areas. Although many observations were available for the study area in both years (Figure S4), a lower minimum number of observations (at least ten) was required for the fitting the HPLM in this study (compared to the fitting for MODIS data in *Zhang et al., 2003* and for AVHRR data in *Zhang, 2015*) to generate a map without gaps. Even within the Landsat sidelap zones, we were not always able to retrieve sufficient observations to fit the LSP model

(*Nguyen et al., 2019*). However, the temporal density of observations could be increased by bringing together complementary sensors. For example, our results show that Sentinel-2 data can be used with Landsat ARD in phenometrically-based classification. An alternative feasible solution may be to leverage very high spatiotemporal but low spectral resolution data from a small satellite constellation to infill gaps (*Houborg & McCabe, 2018*).

Phenometrically-based versus spectrally-based classification

Our hypothesis was that phenometrically-based RFC models would be more accurate than the spectrally-based RFC models in vegetated cover types, at least for crops. However, the results showed that the spectrally-based classification yielded slightly higher accuracy metrics (compared to the phenometrically-based classifications) for most classes, including corn and soybean (Table 4.8). One possible reason for this result is that both spectral and phenometric variables have their own strengths in classification. Compared to spectrally-based RFC models, phenometrically-based RFC models have an advantage of containing the seasonal information or timing of vegetation growth, thus mapping wheat more accurately. However, the phenometrically-based RFC relied on vegetation growth as represented by EVI2 time series calculated from the red and NIR bands. On the other hand, spectral variables in spectrally-based RFC models contain far more spectral information from multiple bands and normalized ratios that give the spectrally-based classification an edge in feature separation. The rich information from input variables could help spectrally-based models to perform better for most classes, including some vegetated covers (e.g., corn, soybean, wetland, and grass) (Tables 4.8, S4.8). Analysis of important variables (Tables 4.9, 4.10) also helps to explain the better performances of spectrally-based models

compared to phenometrically-based models. Both CxQ and HPLM RFC classifications were strongly driven by the maximum fitted EVI2 as showed by consistent appearance of α as the most important variable in the CxQ RFC models (Table 4.9) and appearance of PHHPLM in the top five HPLM RFC models (Table 4.10). In the study area, all vegetated covers can be classified effectively without seasonal information, except for wheat. A comparison between phenometrically-based and spectrally-based classifications in an area with a more complicated cropping patterns (e.g., the wheat-fallow system used in the western Great Plains) might demonstrate better the relative strengths and weakness of these complementary approaches. Nevertheless, the combination of spectral and phenometric variables yielded the most accurate land cover/land use map.

Impact of sample size on classification accuracy

Our results confirmed previous findings that larger sample sizes would lead to better classification, and a sample size covering 0.25% of the study area would be adequate for classification study (*Colditz, 2015; Nguyen et al., 2019*). However, in case of data scarcity, smaller sample sizes, covering 0.15% and at least 0.05% of the study area, might provide acceptable results (cf. Tables 4.5, S4.7, S4.14, S4.15; *Rodriguez-Galiano et al., 2012*). Note that RFC models' accuracy metrics and predicted land cover maps in this study were an ensemble of multiple RFC models. In the case of having a single sample dataset for training and testing, classification may have quite higher or lower accuracy metrics than the expected value (Figures S4.1-S4.3).

Trade-offs between randomness and accuracy in sample dataset

Accuracies of sample pool scenarios in the model AA and the map AA are in opposite orders (Tables 4.5, 4.11). In the model AA, C1S and C2M RFC models performed

the worst and the best, respectively. On the other hand, C1S RFC models displayed higher accuracy metrics than C2M RFC models in the map AA. To some extent, this result is reasonable. C2M RFC models have higher accuracy metrics in the model AA due to more accurate land cover information as well as higher spatial autocorrelation, but they may have lower predictive power (lower OA in the map AA) as some actual characteristics were excluded at edges. Although it might be necessary to improve accuracy of the sample datasets to compensate for the low accuracy of some CDL classes and the spatial offsets between HLS, ARD, and CDL data, our results suggest that only minimum corrections–C1S sample pool–are needed since there was no improvement in accuracy of the predicted land cover map using sample pools with higher level of correction.

4.6   Conclusions

The main focus of this study was to evaluate classification accuracy using different sets of input variables derived from either Landsat ARD or HLS time series, including phenometrics generated from two land surface phenology models (CxQ and HPLM), spectral variables, and the combined set of phenometrics and spectral variables. Between the two phenometrically-based classifications, HPLM RFC models exhibited slightly better accuracy but absolute differences in OA are minor (<1%), mostly due to more precise pixel allocation of land cover. Compared to the phenometrically-based RFC models, the spectrally-based RFC models yielded more accurate land cover maps, especially for non-crop cover types. However, the spectrally-based RFC models could not classify wheat accurately. As hypothesized, the most accurate RFC models were retrieved when using both phenometrics and spectral variables as inputs. The combined-variables RFC models overcame weaknesses of both phenometrically-based classifications (low accuracies for

non-vegetated covers) and spectrally-based classifications (low accuracies for wheat). The analysis of important variables indicated that classifications of the study area were strongly driven by variables related to the initial green-up phase of seasonal growth and highest EVI2 over the growing season.

We have explored land cover/ land use classification under different sample pool and sample size scenarios. First, to improve land cover accuracy of sample data, both spatial and temporal filters were applied to compensate classification errors of the CDL and offsets between input datasets. The results indicated that a sample pool with a minimum correction of land cover information yielded the most accurate predicted map. Next, land cover classification was also tested with different sample sizes. Although previous findings suggested that a sample size should cover at least 0.25% of the study area to achieve an accurate (OA≥0.90) land cover map, smaller datasets would be acceptable for classification, but should not smaller than 0.05% of the study area, since classification accuracy would decrease rapidly below that threshold.

Land surface phenology modeling requires a substantial number of good quality observations over a year (*Zhang et al., 2017*); thus, it may be less suitable for areas with persistent cloud cover if only optical data are available to characterize the LSP. However, the prospect of using phenometrics to enhance land cover/land use classification is very promising. First, our results proved that the use of phenometrics and spectral variables together yielded the most accurate classification and overcame limitations of both phenometrically-based and spectrally-based classifications. Second, seasonality information from all spectral band and ratio time series could be extracted to enhance classification accuracy (e.g., *Zhu & Woodcock, 2014*). Finally, the temporal resolution of

satellite data can be improved by using comparable sensor datastreams, e.g., Landsat and Sentinel-2, but substantial pre-processing is required to achieve compatibility.

4.7 Acknowledgments

4.8 References

Boryan, C. (2018). The USDA NASS Cropland Data Layer Program Transition from Research to Operations (2006-2009). White Paper. Available from: https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer program. Geocarto Int. 26 (5), 341–358.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., ... & Justice, C. 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145-161.

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 Sens. 7 (8), 9655–9681.

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

de Beurs, K.M., Henebry, G.M., 2004. Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. Remote Sens. Environ. 89 (4), 497–509.

Deng, C., Wu, C., 2013. The use of single-date MODIS imagery for estimating large-scale urban impervious surface fraction with spectral mixture analysis and machine learning techniques. ISPRS J. Photogramm. Remote Sens. 86, 100–110.

Du, P., Samat, A., Waske, B., Liu, S., Li, Z., 2015. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS J. Photogramm. Remote Sens. 105, 38–53.

Foody, G.M. 2009. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence, and non-inferiority. Remote Sens. Environ. 113, 1658–1663.

Franklin, S.E., Ahmed, O.S., Wulder, M.A., White, J.C., Hermosilla, T., Coops, N.C., 2015. Large area mapping of annual land cover dynamics using multi-temporal change detection and classification of Landsat time-series data. Can. J. Remote. Sens. 41, 293–314.

Fritz, C. O., Morris, P. E., & Richler, J. J. 2012. Effect size estimates: current use, calculations, and interpretation. J. Exp. Psychol. Gen. 141(1), 2.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. 2006. Random forests for land cover classification. Pattern Recognit Lett. 27(4), 294-300.

Goebel, J.J. 1998. The National Resources Inventory and its role in US agriculture. International Statistical Institute.

Gómez, C., White, J.C., & Wulder, M.A. 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS J. Photogramm. Remote Sens. 116, 55-72.

Henebry, G.M., de Beurs, K.M., 2013. Remote sensing of land surface phenology: a prospectus. In: Schwartz, M.D. (Ed.), Phenology: An Integrative Environmental Science, 2e. Springer, pp. 385–411.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J. and Megown, K. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information. ISPRS J. Photogramm. Remote Sens. 81(5), 345-354.

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. 2004. Development of a 2001 national land-cover database for the United States. ISPRS J. Photogramm. Remote Sens. 70(7), 829-840.

Houborg, R., McCabe, M.F., 2018. A cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sens. Environ. 209, 211–226.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., & Wang, X. 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. 6(11), 11518-11532.

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sens. Environ. 112(10), 3833-3845.

Key, T., Warner, T.A., McGraw, J.B., & Fajvan, M.A. 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. Remote Sens. Environ. 75(1), 100-112.

Kong, F., Li, X., Wang, H., Xie, D., Li, X., & Bai, Y. 2016. Land cover classification based on fused data from GF-1 and MODIS NDVI time series. Remote Sens. 8(9), 741.

Krehbiel, C., Zhang, X., Henebry, G.M., 2017. Impacts of thermal time on land surface phenology in urban areas. Remote Sens. 9 (5), 499.

Krehbiel, C.P., Jackson, T., Henebry, G.M., 2016. Web-enabled Landsat data time series for monitoring urban heat island impacts on land surface phenology. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 9 (5), 2043–2050.

Lakens, D. 2017. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc. Psychol. Personal. Sci. 8(4), 355-362.

Mannel, S., Price, M., & Hua, D. (2011). Impact of reference datasets and autocorrelation on classification accuracy. Int. J. Remote Sens. 32(19), 5321-5330.

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 Sens. 7 (7), 8489–8515.

Miller, D., McCarthy, J., & Zakzeski, A. 2009. A fresh approach to agricultural statistics: data mining and remote sensing. In Proceedings of the Joint Statistical Meetings, 1-6.

Mitchell, J., Shrestha, R., Moore-Ellison, C., & Glenn, N. 2013. Single and multi-date Landsat classifications of basalt to support soil survey efforts. Remote Sens. 5(10), 4857-4876.

Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. Tutorials in quantitative Methods for Psychology, 4(1), 13-20.

Nguyen, L.H.; Joshi, D.R.; Clay, D.E; Henebry, G.M. 2019. Characterizing land cover/land use from multiple years of Landsat and MODIS time series: a novel approach using land surface phenology modeling and random forest classifiers. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.12.016.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12(Oct), 2825-2830.

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

Qader, S.H., Dash, J., Atkinson, P.M., Rodriguez-Galiano, V., 2016. Classification of vegetation type in Iraq using satellite-based phenological parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 9 (1), 414–424.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93-104.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8(2), 127-150.

USGS (U.S. Geological Survey). 2018a. Landsat collections: U.S. Geological Survey Fact Sheet 2018–3049. https://doi.org/10.3133/fs20183049.

USGS (U.S. Geological Survey). 2018b. U.S. Landsat Analysis Ready Data: U.S. Geological Survey Fact Sheet 2018–3053. https://doi.org/10.3133/fs20183053.

Wan, Z., Hook, S., Hulley, G. 2015a. MOD11A2 MODIS/Terra Land Surface Temperature/ Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD11A2.006

Wan, Z., Hook, S., Hulley, G. 2015b. MYD11A2 MODIS/Aqua Land Surface Temperature/ Emissivity 8-Day L3 Global 1km SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MYD11A2.006

Xian, G., Homer, C., & Fry, J. 2009. Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods, Remote Sens. Environ. 113(6), 1133-1147.

Xue, Z., Du, P., & Feng, L. 2014. Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7(4), 1142-1156.

Yan, E., Wang, G., Lin, H., Xia, C., & Sun, H. 2015. Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. Int. J. Remote Sens. 36(2), 489-512.

Zhang, H.K., & Roy, D.P. 2017. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. Remote Sens. Environ. 197, 15-34.

Zhang, X. 2015. Reconstruction of a complete global time series of daily vegetation index trajectory from long-term AVHRR data. Remote Sens. Environ. 156, 457-472.

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C., Gao, F., ... Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84 (3), 471–475.

Zhang, X., Wang, J., Gao, F., Liu, Y., Schaaf, C., Friedl, M., Yu, Y., Jayavelu, S., Gray, J., Liu, L., Yan, D., Henebry, G.M. 2017. Exploration of scaling effects on coarse resolution land surface phenology. Remote Sens. Environ. 190, 318-330.

Zhong, L., Hawkins, T., Biging, G., Gong, P., 2011. A phenology-based approach to map crop types in the San Joaquin Valley, California. Int. J. Remote Sens. 32 (22), 7777–7804.

Zhu, Z., & Woodcock, C.E. 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152-171.

4.9 Supplementary

Table S4.1. Reclassification of CDL land use/land cover classes for Robert County, SD.

CDL Code	CDL Cover	New Cover	CDL Code	CDL Cover	New Cover
0	Background	NoData	53	Peas	Other Crops
1	Corn	Corn	59	Sod/Grass Seed	Grassland
4	Sorghum	Other Crops	61	Fallow/Idle Cropland	Grassland
5	Soybeans	Soybean	111	Open Water	Water
6	Sunflower	Other Crops	121	Developed/Open Space	Barren/Dev.
21	Barley	Other Crops	122	Developed/Low Intensity	Barren/Dev.
22	Durum Wheat	Wheat	123	Developed/Med Intensity	Barren/Dev.
23	Spring Wheat	Wheat	124	Developed/High Intensity	Barren/Dev.
24	Winter Wheat	Wheat	131	Barren	Barren/Dev.
27	Rye	Other Crops	141	Deciduous Forest	Forest
28	Oats	Other Crops	142	Evergreen Forest	Forest
29	Millet	Other Crops	143	Mixed Forest	Forest
36	Alfalfa	Alfalfa	176	Grassland/Pasture	Grassland
37	Other Hay/Non Alfalfa	Grassland	190	Woody Wetlands	Wetland
39	Buckwheat	Other Crops	195	Herbaceous Wetlands	Wetland
41	Sugarbeets	Other Crops	205	Triticale	Wheat
42	Dry Beans	Other Crops	241	Dbl Crop Corn/Soybeans	Other Crops
44	Other Crops	Other Crops			

Table S4.2. Overall accuracy (in percent), kappa for location and quantity of 2017 RFC models summarized by sample pools, sample sizes, and input sets. A particular scenario (current row) was compared to a scenario right above it (row above) using the nonparametric Mann–Whitney U test and the TOST equivalence test. The null hypothesis of the U test is that a random accuracy metric of the first scenario (row above) will be less than a random accuracy metric of the second scenario (current row). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are highlighted in light blue for "not equivalent" and light yellow for "equivalent.

	Scenario	ARD OA	ARD k_L	ARD k_Q	HLS OA	HLS k_L	HLS k_Q
Sample Pool	C1S	89.0	0.909	0.913	87.9	0.898	0.908
	C1M	90.2 ***	0.921 ***	0.920 ***	89.6 ***	0.915 ***	0.917 ***
	C2S	90.5 **	0.923 NS	0.924 **	90.3 ***	0.922 ***	0.920 **
	C2M	91.6 ***	0.933 ***	0.929 ***	91.3 ***	0.932 ***	0.927 ***
Sample Size	P01	87.1	0.907	0.871	86.3	0.899	0.866
	P05	90.2 ***	0.918 ***	0.926 ***	89.7 ***	0.914 ***	0.923 ***
	P15	91.7 ***	0.928 ***	0.942 ***	91.3 ***	0.924 ***	0.939 ***
	P25	92.2 ***	0.932 **	0.947 ***	91.9 ***	0.929 ***	0.944 ***
Input Set	CxQ	88.2	0.892	0.923	87.9	0.887	0.924
	HPLM	89.1 ***	0.901 ***	0.926 NS	88.8 ***	0.901 ***	0.921 NS
	SPL	89.9 ***	0.931 ***	0.899 NS	88.6 NS	0.919 ***	0.890 NS
	CMB	94.1 ***	0.961 ***	0.938 ***	93.8 ***	0.958 ***	0.936 ***
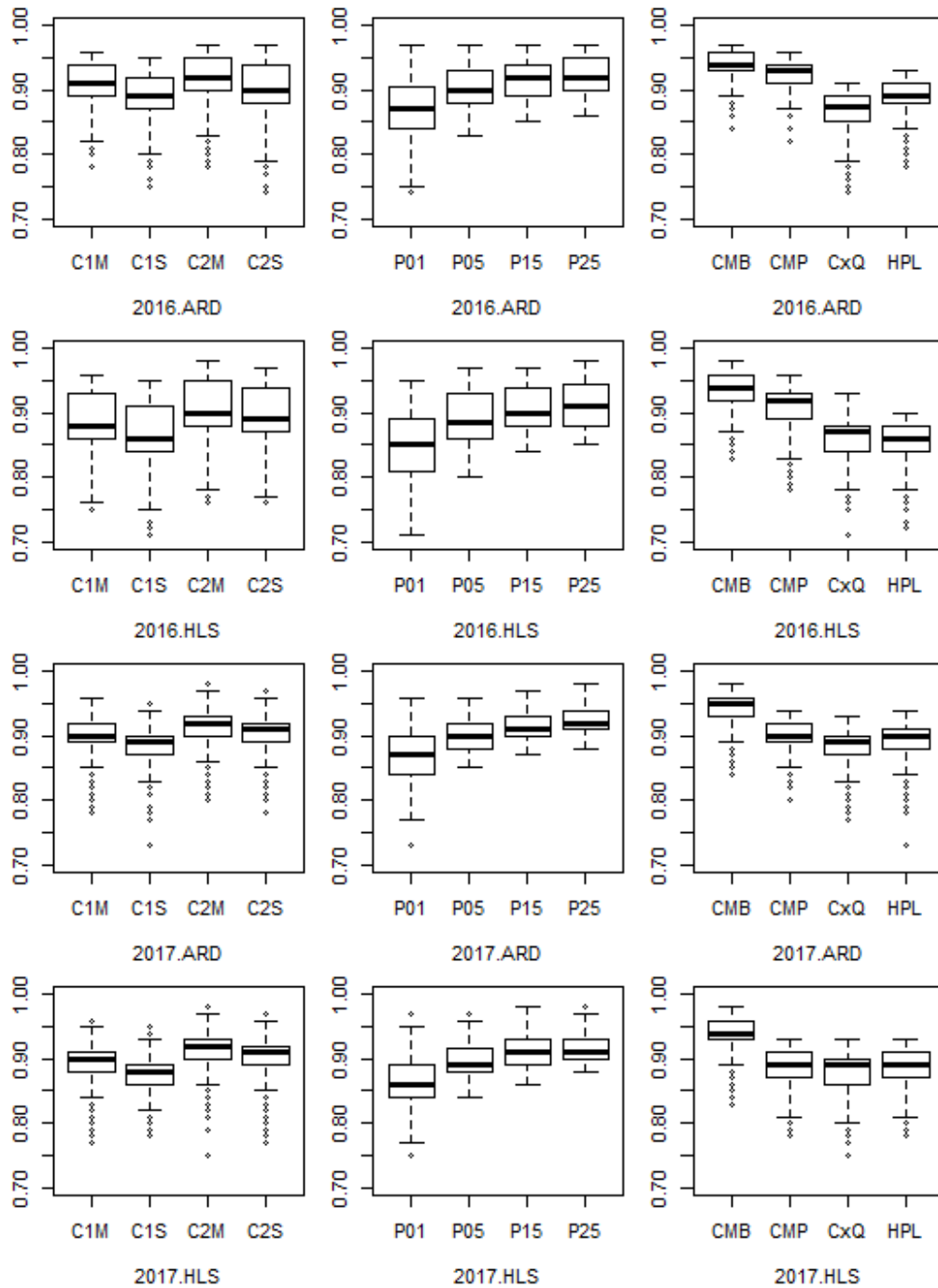
Figure S4.1. Boxplots of overall accuracy (OA) metrics from different scenarios for each combination of year and data source.
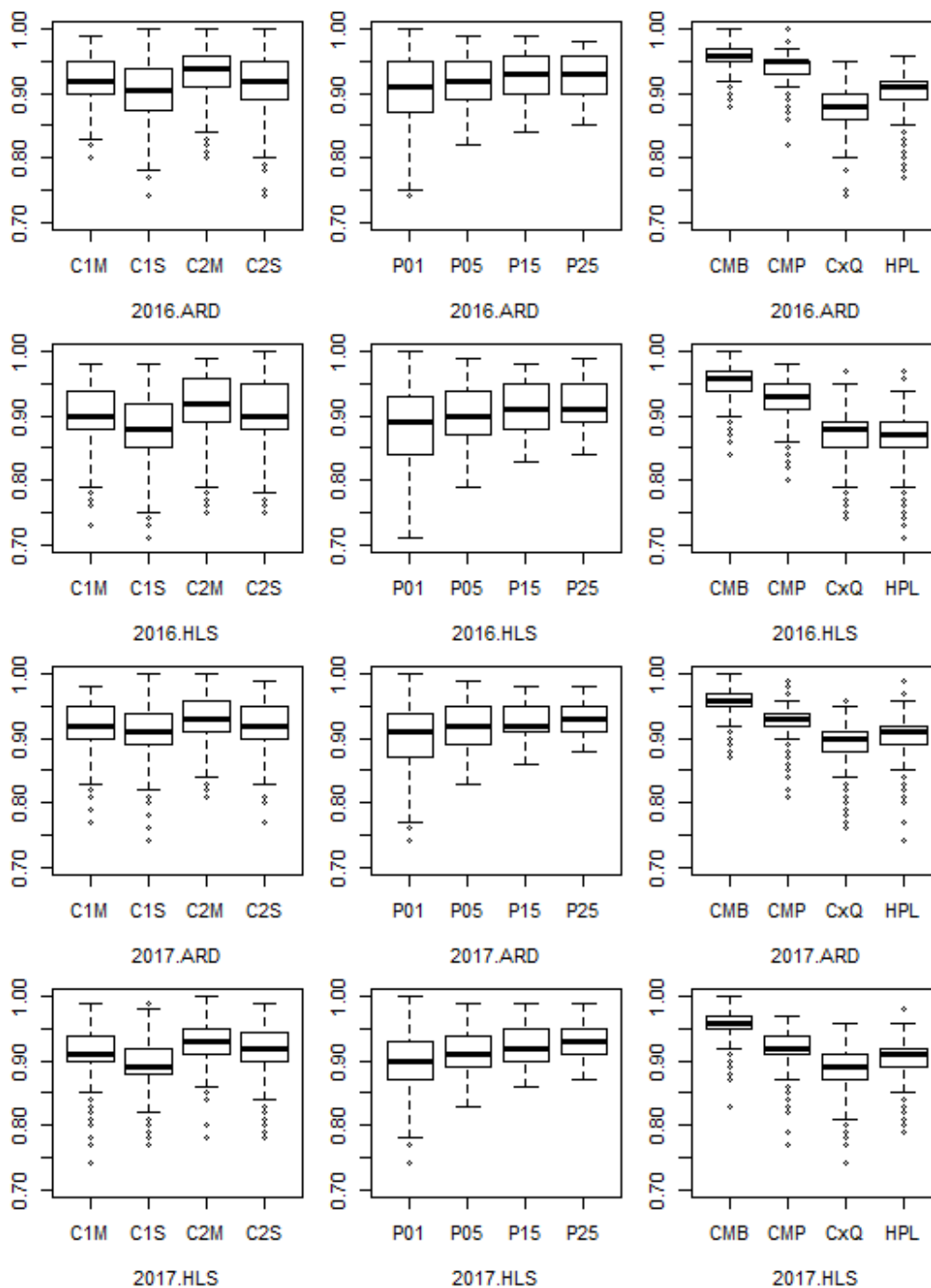
Figure S4.2. Boxplots of kappa_Location (k_L) metrics from different scenarios for each combination of year and data source.
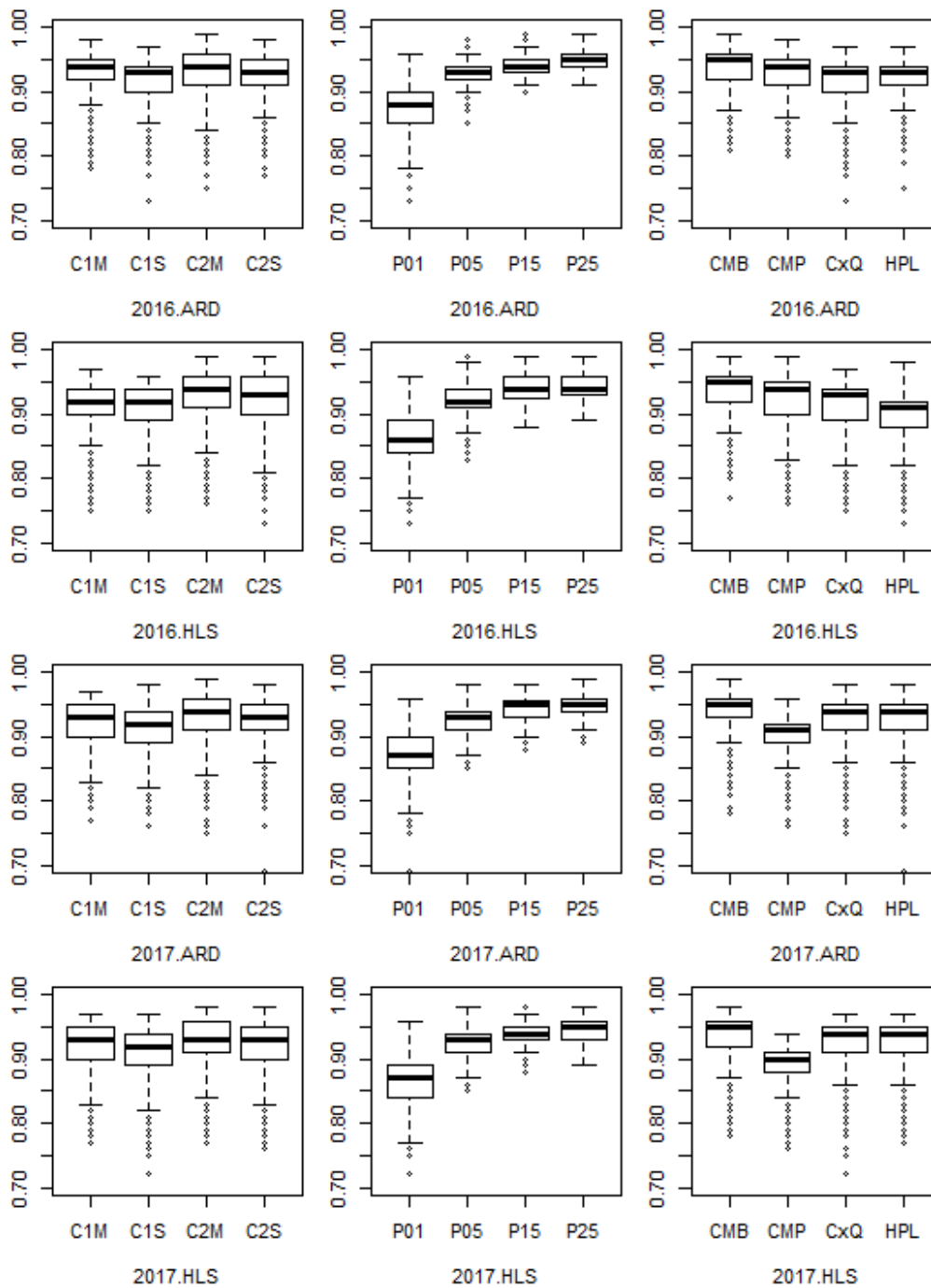
Figure S4.3. Boxplots of kappa_Quantity (k_Q) metrics from different scenarios for each combination of year and data source.

Table S4.3. Significance level of the nonparametric Mann–Whitney U test for pairwise OA, k_L and k_Q comparison. ***, ** and * indicate p-values of less than 0.001, 0.01, 0.05, respectively. NS means "not significant". Area colored in light yellow (upper right) show results computed on ARD-RFC models, and light blue (lower left) show results from HLS-RFC models. The null hypothesis is that a random accuracy metric of the first scenario will be less than a random accuracy metric of the second scenario (or the first population has smaller values than the second population). For the test on ARD-RFC models, the first scenario is selected from the table's row, and second scenarios is selected from the column. Scenario selection for the test on HLS-RFC models is in the opposite direction (column → row). For example, the significance level for comparison of k_L between C1M and C2M are "NS" and "**" for ARD- and HLS-RFC models, respectively.

		Overall Accuracy				Kappa for Location				Kappa for Quantity			
		C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M
2016	C1S	-	***	***	***	-	***	***	***	-	***	***	***
	C1M	***	-	NS	***	***	-	NS	***	***	-	NS	***
	C2S	***	***	-	***	***	**	-	***	***	***	-	***
	C2M	***	***	***	-	***	***	***	-	***	***	***	-
2017	C1S	-	***	***	***	-	***	***	***	-	***	***	***
	C1M	***	-	**	***	***	-	NS	***	***	-	**	***
	C2S	***	***	-	***	***	***	-	***	***	**	-	***
	C2M	***	***	***	-	***	***	***	-	***	***	***	-
		P01	P05	P15	P25	P01	P05	P15	P25	P01	P05	P15	P25
2016	P01	-	***	***	***	-	***	***	***	-	***	***	***
	P05	***	-	***	***	***	-	***	***	***	-	***	***
	P15	***	***	-	***	***	***	-	*	***	***	-	***
	P25	***	***	***	-	***	***	**	-	***	***	***	-
2017	P01	-	***	***	***	-	***	***	***	-	***	***	***
	P05	***	-	***	***	***	-	***	***	***	-	***	***
	P15	***	***	-	***	***	***	-	**	***	***	-	***
	P25	***	***	***	-	***	***	***	-	***	***	***	-
		CxQ	HPLM	SPL	CMB	CxQ	HPLM	SPL	CMB	CxQ	HPLM	SPL	CMB
2016	CxQ	-	***	***	***	-	***	***	***	-	NS	***	***
	HPLM	NS	-	***	***	NS	-	***	***	NS	-	***	***
	SPL	***	***	-	***	***	***	-	***	***	***	-	***
	CMB	***	***	***	-	***	***	***	-	***	***	***	-
2017	CxQ	-	***	***	***	-	***	***	***	-	NS	NS	***
	HPLM	***	-	***	***	***	-	***	***	NS	-	NS	***
	SPL	***	NS	-	***	***	***	-	***	NS	NS	-	***
	CMB	***	***	***	-	***	***	***	-	***	***	***	-

Table S4. TOST equivalence tests. "E" indicates that the two sets of accuracy metrics are equivalent or differences between those are within a user-defined range of Cohen's d. "×" that the two sets of accuracy metric are not equivalent. Area colored in light yellow (upper right) and light blue (lower left) show results computed on ARD- and HLS-RFC models, respectively. For the test on ARD-RFC models, the first scenario is selected from the table's row, and second scenarios is selected from the column. Scenario selection for the test on HLS-RFC models is on the opposite direction (column → row). For example, the results of equivalence test between C1S and C2S are "×" and "E" for ARD- and HLS-RFC models, respectively.

		Overall Accuracy				Kappa for Allocation				Kappa for Quantity			
		C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M
2016	C1S	-	×	×	×	-	×	×	×	-	E	E	×
	C1M	×	-	E	E	×	-	E	E	E	-	E	E
	C2S	×	E	-	×	×	E	-	E	×	E	-	E
	C2M	×	×	E	-	×	×	E	-	×	E	E	-
2017	C1S	-	×	×	×	-	E	×	×	-	E	E	×
	C1M	×	-	E	×	×	-	E	×	E	-	E	E
	C2S	×	E	-	E	×	E	-	E	E	E	-	E
	C2M	×	×	E	-	×	×	E	-	×	E	E	-
		P01	P05	P15	P25	P01	P05	P15	P25	P01	P05	P15	P25
2016	P01	-	×	×	×	-	E	×	×	-	×	×	×
	P05	×	-	×	×	×	-	E	E	×	-	×	×
	P15	×	×	-	E	×	E	-	E	×	×	-	×
	P25	×	×	E	-	×	×	E	-	×	×	E	-
2017	P01	-	×	×	×	-	E	×	×	-	×	×	×
	P05	×	-	×	×	×	-	E	×	×	-	×	×
	P15	×	×	-	E	×	E	-	E	×	×	-	E
	P25	×	×	E	-	×	×	E	-	×	×	E	-
		CxQ	HPLM	SPL	CMP	CxQ	HPLM	SPL	CMP	CxQ	HPLM	SPL	CMP
2016	CxQ	-	×	×	×	-	×	×	×	-	E	E	×
	HPLM	E	-	×	×	E	-	×	×	E	-	E	×
	SPL	×	×	-	×	×	×	-	×	E	×	-	E
	CMP	×	×	×	-	×	×	×	-	×	×	E	-
2017	CxQ	-	E	×	×	-	×	×	×	-	E	×	×
	HPLM	E	-	E	×	×	-	×	×	E	-	×	E
	SPL	E	E	-	×	×	×	-	×	×	×	-	×
	CMP	×	×	×	-	×	×	×	-	E	×	×	-

Table S4.5. Accuracy assessment of RFC models summarized by sample pool scenarios.

| | 2016 | | | | | | | | 2017 | | | | | | | |
| | ARD | | | | HLS | | | | ARD | | | | HLS | | | |
Metrics	C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M
PA_Corn	94.6	95.3	95.3	95.7	91.2	92.1	92.3	92.6	93.9	94.5	94.7	95.4	91.8	93.2	93.4	94.0
PA_Wheat	75.4	77.9	78.3	78.6	70.2	72.9	73.1	74.4	59.0	63.4	63.8	66.3	57.0	62.5	61.3	63.3
PA_Alfalfa	73.5	80.3	79.1	82.6	69.3	77.1	77.2	79.9	74.2	80.2	78.2	82.7	71.2	80.1	76.4	82.8
PA_Soybean	95.4	96.1	96.1	96.6	90.3	91.1	91.8	92.1	95.3	95.7	95.7	96.2	93.6	94.6	94.6	95.1
PA_Other Crops	11.2	35.4	19.7	50.9	15.3	43.9	25.2	55.0	15.6	21.2	22.3	29.7	15.6	25.1	21.2	33.2
PA_Water	97.4	98.4	98.2	98.3	96.2	97.9	97.8	98.2	96.4	97.1	97.6	97.5	95.4	96.8	97.3	97.3
PA_Barren/Dev.	55.1	58.6	59.7	60.9	47.4	47.2	54.3	56.1	54.3	55.6	57.6	58.5	48.6	44.6	50.8	51.9
PA_Forest	87.7	89.5	93.1	94.0	84.3	87.8	93.1	94.5	83.3	84.2	87.5	88.3	81.6	84.4	88.9	90.8
PA_Grassland	93.0	94.1	93.8	94.9	93.6	95.0	95.1	96.1	93.4	94.2	94.2	94.9	94.0	94.7	95.2	95.8
PA_Wetland	66.8	74.4	72.1	79.2	73.9	80.4	82.0	87.7	67.7	72.8	74.6	78.9	69.9	76.7	79.8	84.2
UA_Corn	94.8	95.5	95.4	95.9	89.8	90.6	91.3	91.5	94.7	95.3	95.3	95.9	91.8	93.0	93.1	93.4
UA_Wheat	84.4	89.3	87.6	90.2	82.0	86.5	85.5	89.1	80.6	84.2	83.1	86.1	80.4	82.8	83.3	84.3
UA_Alfalfa	83.7	88.8	87.7	91.2	79.8	87.2	86.6	89.6	87.9	91.0	90.9	92.9	85.0	91.5	90.0	93.6
UA_Soybean	93.6	94.7	94.7	95.2	90.4	91.6	91.7	92.2	92.7	93.1	93.4	93.9	91.5	92.6	92.9	93.4
UA_Other Crops	35.8	61.3	49.9	70.2	46.6	67.0	54.0	71.0	41.0	43.8	50.3	52.8	40.4	47.6	45.0	55.1
UA_Water	97.5	97.6	97.7	97.6	96.3	96.2	96.4	96.9	96.4	96.0	96.2	96.5	95.9	96.0	95.8	96.3
UA_Barren/Dev.	77.0	80.2	77.1	79.0	72.4	74.1	74.0	75.9	77.0	78.5	77.7	78.6	75.7	73.4	76.2	77.1
UA_Forest	86.7	88.5	90.4	91.1	84.8	87.5	90.6	92.4	80.7	82.9	85.4	86.6	82.4	84.3	88.3	89.7
UA_Grassland	84.9	87.2	86.4	88.6	85.6	87.6	88.9	90.4	85.5	87.4	88.0	89.3	85.3	87.5	88.6	90.0
UA_Wetland	74.2	79.5	80.1	84.4	76.3	81.6	83.4	87.5	73.9	77.7	78.8	82.1	75.1	79.5	81.6	84.9
OA	88.8	90.7	90.4	91.8	86.8	88.7	89.4	90.7	89.0	90.2	90.5	91.6	87.9	89.6	90.3	91.3
k_L	0.904	0.923	0.921	0.935	0.884	0.904	0.909	0.922	0.909	0.921	0.923	0.933	0.898	0.915	0.922	0.932
k_Q	0.917	0.926	0.924	0.932	0.906	0.914	0.921	0.927	0.913	0.920	0.924	0.929	0.908	0.917	0.920	0.927

Table S4.6. The nonparametric Mann–Whitney U test and the TOST equivalence test for the (P01 versus P25 RFC models) and (C1S versus C2M RFC models) comparison. The null hypothesis is that a random accuracy metric of the first scenario will be less than a random accuracy metric of the second scenario (or the first population has smaller values than the second population). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

Metrics	P01 versus P25				C1S versus C2M			
	2016		2017		2016		2017	
	ARD	HLS	ARD	HLS	ARD	HLS	ARD	HLS
PA_Corn	***	***	***	***	***	***	***	***
PA_Wheat	***	***	***	***	***	***	***	***
PA_Alfalfa	***	***	***	***	***	***	***	***
PA_Soybean	***	***	***	***	***	***	***	***
PA_Other Crops	***	***	***	***	***	***	***	***
PA_Water	NS	NS	NS	NS	***	***	***	***
PA_Barren/Dev.	***	***	***	***	***	***	***	**
PA_Forest	NS	NS	NS	NS	***	***	***	***
PA_Grassland	***	***	***	***	***	***	***	***
PA_Wetland	***	***	***	***	***	***	***	***
UA_Corn	***	***	***	***	***	***	***	***
UA_Wheat	NS	NS	NS	NS	***	***	***	***
UA_Alfalfa	NS	NS	NS	NS	***	***	***	***
UA_Soybean	***	***	***	***	***	***	***	***
UA_Other Crops	***	***	***	***	***	***	***	***
UA_Water	NS	NS	NS	NS	***	***	*	***
UA_Barren/Dev.	**	***	*	**	***	***	**	**
UA_Forest	NS	NS	NS	NS	***	***	***	***
UA_Grassland	***	***	***	***	***	***	***	***
UA_Wetland	***	***	***	***	***	***	***	***
OA	***	***	***	***	***	***	***	***
k_L	***	***	***	***	***	***	***	***
k_Q	***	***	***	***	***	***	***	***

Table S4.7. Accuracy assessment of RFC models summarized by sample size scenarios.

| | 2016 | | | | | | | | 2017 | | | | | | | |
| | ARD | | | | HLS | | | | ARD | | | | HLS | | | |
Metrics	P01	P05	P15	P25	P01	P05	P15	P25	P01	P05	P15	P25	P01	P05	P15	P25
PA_Corn	94.1	94.8	95.8	96.2	89.8	91.8	93.1	93.6	93.3	94.4	95.2	95.6	90.7	93.2	94.1	94.5
PA_Wheat	58.1	79.2	85.3	87.4	49.6	73.7	82.4	85.0	45.8	64.0	70.3	72.5	44.4	62.4	67.4	69.9
PA_Alfalfa	61.9	79.9	85.7	87.9	55.0	77.9	84.2	86.5	56.8	81.4	87.5	89.7	55.0	80.0	87.1	88.5
PA_Soybean	94.6	96.2	96.6	96.8	88.5	91.6	92.5	92.8	93.3	95.8	96.7	97.1	92.1	94.4	95.5	96.0
PA_Other Crops	7.1	21.8	40.0	48.3	9.1	27.2	47.3	55.9	0.0	16.6	31.6	40.6	0.0	18.8	33.2	43.0
PA_Water	97.5	98.1	98.3	98.5	97.0	97.4	97.9	97.9	96.3	96.9	97.6	97.8	96.0	96.3	97.1	97.4
PA_Barren/Dev.	49.6	58.2	62.4	64.2	38.8	50.7	56.5	58.9	47.0	55.8	60.8	62.5	36.1	48.0	54.8	57.0
PA_Forest	86.6	91.4	92.8	93.4	84.5	90.3	92.0	92.8	77.9	86.7	89.0	89.7	79.1	87.0	89.4	90.3
PA_Grassland	92.6	94.0	94.5	94.7	93.7	94.9	95.5	95.7	93.1	94.1	94.5	94.9	93.9	94.8	95.4	95.6
PA_Wetland	64.8	73.7	76.4	77.6	74.5	81.3	83.6	84.5	64.4	73.8	77.3	78.6	69.2	78.3	81.0	82.2
UA_Corn	93.0	95.6	96.4	96.7	87.4	91.1	92.2	92.6	92.8	95.4	96.4	96.7	89.9	92.8	94.0	94.6
UA_Wheat	84.5	87.5	89.4	90.2	80.6	85.6	88.0	89.0	70.5	84.9	88.7	89.8	67.6	82.0	89.9	91.3
UA_Alfalfa	75.2	91.2	92.1	92.9	69.9	90.1	91.1	92.0	81.3	92.8	94.0	94.6	79.1	92.4	93.9	94.7
UA_Soybean	92.2	94.3	95.6	96.1	87.9	91.3	93.0	93.6	91.0	93.0	94.3	94.8	89.8	92.6	93.8	94.3
UA_Other.Crops	7.0	48.0	77.6	84.7	9.0	55.8	84.3	89.5	0.0	27.1	73.9	86.9	0.0	29.4	73.2	85.5
UA_Water	97.4	97.6	97.7	97.8	96.1	96.2	96.6	96.8	94.9	96.2	96.9	97.1	95.0	96.0	96.4	96.5
UA_Barren/Dev.	72.5	78.3	80.7	81.7	66.0	73.7	77.9	78.9	70.5	78.1	81.0	82.2	67.1	75.7	79.1	80.4
UA_Forest	85.3	89.0	90.8	91.6	84.7	88.2	90.8	91.6	80.0	83.4	85.6	86.7	82.1	85.8	87.9	88.9
UA_Grassland	83.7	86.9	87.9	88.5	84.8	88.1	89.5	90.1	84.5	87.6	88.8	89.3	84.8	87.8	89.2	89.7
UA_Wetland	72.8	79.6	82.4	83.5	74.6	82.7	85.2	86.3	72.2	77.7	80.7	81.9	73.8	79.9	83.1	84.2
OA	87.1	90.5	91.8	92.3	84.8	89.0	90.6	91.2	87.1	90.2	91.7	92.2	86.3	89.7	91.3	91.9
k_L	0.906	0.919	0.927	0.931	0.883	0.903	0.914	0.919	0.907	0.918	0.928	0.932	0.899	0.914	0.924	0.929
k_Q	0.877	0.929	0.944	0.949	0.862	0.922	0.940	0.945	0.871	0.926	0.942	0.947	0.866	0.923	0.939	0.944

Table S4.8. Producer's and user's accuracies in percent (%) of 2017 RFC models summarized by sets of input variables. A particular scenario (current column) was compared to a scenario on the left (left column) using the nonparametric Mann–Whitney U test and the TOST equivalence tests. The null hypothesis of the U test is that a random accuracy metric of the first scenario (left column) will be less than a random accuracy metric of the second scenario (current column). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

Metrics	ARD							HLS						
	CxQ	HPLM		SPL		CMB		CxQ	HPLM		SPL		CMB	
PA_Corn	92.2	95.2	***	94.0	NS	97.1	***	91.5	94.2	***	90.9	NS	95.9	***
PA_Wheat	81.2	77.7	NS	14.7	NS	79.0	***	74.9	81.1	***	8.1	NS	80.0	***
PA_Alfalfa	71.5	86.2	***	75.9	NS	81.9	***	70.4	88.8	***	69.9	NS	81.4	***
PA_Soybean	94.5	92.7	NS	97.0	***	98.8	***	92.5	93.0	**	94.6	***	97.8	***
PA_Other Crops	34.6	25.4	NS	5.9	NS	22.9	***	31.7	29.3	NS	4.6	NS	29.4	***
PA_Water	93.8	96.1	***	99.4	***	99.4	NS	93.1	94.7	***	99.5	***	99.3	NS
PA_Barren/Dev.	43.4	43.5	NS	69.5	***	69.7	NS	40.1	35.1	NS	59.3	***	61.4	**
PA_Forest	73.7	78.7	***	94.8	***	96.1	***	75.2	80.7	***	94.4	***	95.4	***
PA_Grassland	92.1	93.2	***	95.0	***	96.3	***	93.0	93.2	*	96.2	***	97.3	***
PA_Wetland	71.0	71.4	NS	72.7	***	78.9	***	77.7	72.2	NS	77.5	***	83.2	***
UA_Corn	95.1	94.7	NS	93.3	NS	98.2	***	92.8	94.4	***	87.2	NS	97.0	***
UA_Wheat	89.8	90.7	NS	61.4	NS	92.1	***	88.7	91.1	***	57.5	NS	93.6	***
UA_Alfalfa	87.3	91.4	NS	89.7	*	94.4	***	86.7	93.3	NS	86.0	NS	94.1	***
UA_Soybean	92.9	94.2	***	89.4	NS	96.6	***	92.2	93.7	***	88.3	NS	96.1	***
UA_Other Crops	55.0	51.5	NS	28.4	NS	52.9	***	52.8	53.7	NS	24.7	NS	56.9	***
UA_Water	92.3	93.2	***	99.9	***	99.8	NS	91.2	93.3	***	99.8	***	99.6	NS
UA_Barren/Dev.	62.2	63.8	***	93.3	***	92.4	NS	59.2	63.0	***	91.8	***	88.3	NS
UA_Forest	72.0	77.1	***	93.1	***	93.4	NS	76.7	82.8	***	91.3	***	94.0	***
UA_Grassland	86.2	86.4	NS	88.1	***	89.6	***	87.6	85.2	NS	88.8	***	89.9	***
UA_Wetland	70.7	74.5	***	81.0	***	86.3	***	73.0	74.5	***	85.5	***	87.9	***

9 **Table S4.9.** The nonparametric Mann–Whitney U test and the TOST equivalence test for the comparison

10 between phenometrically-based and combined RFC models. The null hypothesis of the U test is that a

11 random accuracy metric of the first scenario (left column) will be less than a random accuracy metric of the

12 second scenario (current column). Significance level of the U test are indicated by ***, ** and * for p-

13 values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are

14 highlighted in light blue for "not equivalent" and light yellow for "equivalent".

Metrics	CxQ versus SPL				HPLM versus SPL			
	2016		2017		2016		2017	
	ARD	HLS	ARD	HLS	ARD	HLS	ARD	HLS
PA_Corn	***	***	***	NS	***	***	NS	NS
PA_Wheat	NS	NS	NS	NS	NS	NS	NS	NS
PA_Alfalfa	***	***	NS	NS	NS	NS	NS	NS
PA_Soybean	***	***	***	***	***	***	***	***
PA_Other Crops	NS	NS	NS	NS	NS	NS	NS	NS
PA_Water	***	***	***	***	***	***	***	***
PA_Barren/Dev.	***	***	***	***	***	***	***	***
PA_Forest	***	***	***	***	***	***	***	***
PA_Grassland	***	***	***	***	***	***	***	***
PA_Wetland	***	***	***	NS	***	***	***	***
UA_Corn	***	***	NS	NS	***	***	NS	NS
UA_Wheat	NS	NS	NS	NS	NS	NS	NS	NS
UA_Alfalfa	***	**	NS	NS	***	***	*	NS
UA_Soybean	***	***	NS	NS	NS	NS	NS	NS
UA_Other Crops	NS	***	NS	NS	NS	***	NS	NS
UA_Water	***	***	***	***	***	***	***	***
UA_Barren/Dev.	***	***	***	***	***	***	***	***
UA_Forest	***	***	***	***	***	***	***	***
UA_Grassland	***	***	***	***	***	***	***	***
UA_Wetland	***	***	***	***	***	***	***	***
OA	***	***	***	***	***	***	***	NS
k_L	***	***	***	***	***	***	***	***
k_Q	***	***	NS	NS	***	***	NS	NS

15

Table S4.10. The nonparametric Mann–Whitney U test and the TOST equivalence test for the comparison between phenometrically-based and combined RFC models. The null hypothesis of the U test is that a random accuracy metric of the first scenario (left column) will be less than a random accuracy metric of the second scenario (current column). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

Metrics	CxQ versus HPLM				SPL versus CMB			
	2016		2017		2016		2017	
	ARD	HLS	ARD	HLS	ARD	HLS	ARD	HLS
PA_Corn	***	***	***	***	***	***	***	***
PA_Wheat	NS	NS	NS	***	***	***	***	***
PA_Alfalfa	***	***	***	***	***	***	***	***
PA_Soybean	***	***	NS	**	***	***	***	***
PA_Other Crops	NS	NS	NS	NS	***	**	***	***
PA_Water	***	***	***	***	NS	NS	NS	NS
PA_Barren/Dev.	***	***	NS	NS	NS	NS	NS	**
PA_Forest	***	***	***	***	*	NS	***	***
PA_Grassland	***	NS	***	*	***	***	***	***
PA_Wetland	NS	NS	NS	NS	***	***	***	***
UA_Corn	***	***	NS	***	***	***	***	***
UA_Wheat	*	NS	NS	***	***	***	***	***
UA_Alfalfa	NS	NS	NS	NS	***	***	***	***
UA_Soybean	***	***	***	***	***	***	***	***
UA_Other Crops	***	NS	NS	NS	***	*	***	***
UA_Water	***	***	***	***	NS	NS	NS	NS
UA_Barren/Dev.	***	***	***	***	NS	NS	NS	NS
UA_Forest	***	***	***	***	**	***	NS	***
UA_Grassland	***	NS	NS	NS	NS	***	***	***
UA_Wetland	NS	NS	***	***	***	***	***	***
OA	***	NS	***	***	***	***	***	***
k_L	***	NS	***	***	***	***	***	***
k_Q	NS	NS	NS	NS	***	***	***	***

16
17
18
19
20
21
22

Table S4.11. The nonparametric Mann–Whitney U test and the TOST equivalence test for the comparison between phenometrically-based and combined RFC models. The null hypothesis of the U test is that a random accuracy metric of the first scenario (left column) will be less than a random accuracy metric of the second scenario (current column). Significance level of the U test are indicated by ***, ** and * for p-values of less than 0.001, 0.01, 0.05 and NS for "not significant". Results of the TOST equivalence test are highlighted in light blue for "not equivalent" and light yellow for "equivalent".

Metrics	CxQ versus CMB				HPLM versus CMB			
	2016		2017		2016		2017	
	ARD	HLS	ARD	HLS	ARD	HLS	ARD	HLS
PA_Corn	***	***	***	***	***	***	***	***
PA_Wheat	***	***	**	***	***	***	***	**
PA_Alfalfa	***	***	***	***	NS	NS	NS	NS
PA_Soybean	***	***	***	***	***	***	***	***
PA_Other Crops	NS	NS	NS	NS	NS	NS	NS	NS
PA_Water	***	***	***	***	***	***	***	***
PA_Barren/Dev.	***	***	***	***	***	***	***	***
PA_Forest	***	***	***	***	***	***	***	***
PA_Grassland	***	***	***	***	***	***	***	***
PA_Wetland	***	***	***	***	***	***	***	***
UA_Corn	***	***	***	***	***	***	***	***
UA_Wheat	***	***	***	***	***	***	***	***
UA_Alfalfa	***	***	***	***	***	***	***	***
UA_Soybean	***	***	***	***	***	***	***	***
UA_Other Crops	***	***	NS	***	***	***	*	***
UA_Water	***	***	***	***	***	***	***	***
UA_Barren/Dev.	***	***	***	***	***	***	***	***
UA_Forest	***	***	***	***	***	***	***	***
UA_Grassland	***	***	***	***	***	***	***	***
UA_Wetland	***	***	***	***	***	***	***	***
OA	***	***	***	***	***	***	***	***
k_L	***	***	***	***	***	***	***	***
k_Q	***	***	***	***	***	***	***	***

Table S4.12. Pixel-based comparison between 2016 predicted land cover maps and the CDL summarized by sample pools. Areal units are km^2.

Land Cover	Info.	CDL	ARD				HLS			
			C1S	C1M	C2S	C2M	C1S	C1M	C2S	C2M
Corn	Area	662	643	639	643	639	668	666	670	670
	UA		92.7	92.9	92.6	92.7	86.2	86.0	85.9	85.8
	PA		90.2	89.6	90.0	89.5	87.0	86.6	87.0	86.8
Wheat	Area	103	100	104	103	105	97	101	102	105
	UA		82.5	80.2	80.9	79.7	79.5	77.1	76.1	74.1
	PA		80.6	81.5	81.2	81.6	75.1	75.6	75.7	76.1
Alfalfa	Area	45	30	32	31	33	31	33	31	35
	UA		76.5	74.0	75.1	72.6	69.5	66.1	68.3	64.4
	PA		51.0	52.5	52.1	53.7	47.2	48.8	47.6	49.5
Soybean	Area	746	730	721	721	714	712	696	696	688
	UA		90.9	91.4	91.4	91.8	87.3	88.2	88.2	88.7
	PA		88.9	88.4	88.3	87.9	83.3	82.2	82.3	81.7
Other Crops	Area	13	0	0	0	0	0	1	1	1
	UA		0.0	0.0	0.0	0.0	0.0	0.0	89.1	0.0
	PA		0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
Water	Area	154	127	123	121	119	131	126	125	125
	UA		96.7	97.2	97.4	97.5	93.0	94.1	94.4	94.5
	PA		79.7	77.7	76.3	75.6	79.2	77.1	76.6	76.5
Barren/Dev.	Area	131	51	50	47	51	47	53	48	58
	UA		51.0	48.6	49.5	46.6	41.6	36.3	38.7	33.1
	PA		20.1	18.6	17.8	18.4	15.0	14.8	14.2	14.7
Forest	Area	64	62	64	64	67	63	65	63	65
	UA		68.3	67.1	66.7	64.8	62.9	61.6	62.2	61.0
	PA		65.7	66.2	66.1	67.2	61.1	61.8	60.6	61.3
Grassland	Area	761	923	918	924	916	933	939	944	955
	UA		74.6	74.6	74.5	74.7	72.9	72.5	71.9	71.2
	PA		90.5	90.1	90.4	90.0	89.5	89.5	89.2	89.5
Wetland	Area	262	274	290	288	295	259	262	262	240
	UA		48.0	45.5	45.3	43.9	51.9	50.3	50.4	51.2
	PA		50.0	50.3	49.7	49.5	51.3	50.2	50.3	46.9
OA			80.8	80.4	80.4	80.1	77.9	77.4	77.4	77.0
k_loc			0.820	0.819	0.819	0.817	0.782	0.777	0.779	0.780
k_quan			0.878	0.868	0.867	0.863	0.875	0.871	0.866	0.857

Table S4.13. Pixel-based comparison between 2017 predicted land cover maps and the CDL summarized by sample pools. Areal units are km^2.

Land Cover	Info.	CDL	ARD C1S	ARD C1M	ARD C2S	ARD C2M	HLS C1S	HLS C1M	HLS C2S	HLS C2M
					ARD				**HLS**	
Corn	Area	685	607	597	601	595	616	602	606	602
	UA		95.4	95.8	95.8	95.9	92.0	92.9	92.6	92.7
	PA		84.6	83.6	84.1	83.3	82.8	81.6	81.9	81.5
Wheat	Area	87	63	65	65	65	63	63	63	63
	UA		88.0	87.7	87.6	87.7	87.6	87.5	87.5	87.4
	PA		64.3	65.6	65.4	65.8	63.8	63.9	63.9	63.8
Alfalfa	Area	50	37	36	39	38	36	36	38	38
	UA		83.8	82.2	81.0	80.2	80.1	79.1	78.4	76.8
	PA		61.4	60.1	64.0	61.1	57.9	56.6	59.6	57.8
Soybean	Area	803	817	806	811	803	820	811	812	808
	UA		89.7	90.3	90.1	90.5	87.1	87.6	87.5	87.7
	PA		91.3	90.7	90.9	90.4	89.0	88.5	88.5	88.2
Other Crops	Area	8	0	0	1	0	0	0	0	0
	UA		0.0	0.0	99.0	0.0	0.0	0.0	0.0	0.0
	PA		0.0	0.0	8.3	0.0	0.0	0.0	0.0	0.0
Water	Area	153	121	120	116	115	125	122	119	119
	UA		97.0	97.2	97.8	97.8	93.7	94.2	95.1	95.1
	PA		76.7	75.9	73.8	73.7	76.2	75.2	73.8	73.6
Barren/Dev.	Area	111	46	50	48	53	44	49	46	51
	UA		52.4	50.0	51.3	48.7	46.2	43.4	43.3	40.7
	PA		21.6	22.5	22.5	23.5	18.3	19.1	18.1	18.8
Forest	Area	63	66	69	72	73	60	63	62	61
	UA		63.6	62.4	60.7	59.5	63.2	61.6	61.8	62.0
	PA		67.0	68.2	69.2	69.4	60.3	61.6	60.6	60.0
Grassland	Area	744	920	914	911	905	938	937	946	949
	UA		73.6	73.7	73.7	73.8	72.0	71.8	71.3	70.9
	PA		91.1	90.6	90.3	89.8	90.7	90.4	90.6	90.5
Wetland	Area	237	263	283	277	293	239	258	248	250
	UA		48.7	45.6	46.1	43.8	50.5	47.0	47.5	46.3
	PA		54.0	54.5	54.0	54.3	50.9	51.3	49.9	48.9
OA			81.2	80.8	80.8	80.4	79.5	79.0	78.9	78.6
k_loc			0.841	0.838	0.838	0.833	0.814	0.811	0.810	0.806
k_quan			0.852	0.846	0.848	0.843	0.858	0.851	0.851	0.851

Table S4.14. Pixel-based comparison between 2016 predicted land cover maps and the CDL summarized by sample sizes. Areal units are km^2.

Land Cover	Info.	CDL	ARD				HLS			
			P01	P05	P15	P25	P01	P05	P15	P25
Corn	Area	662	651	638	640	642	690	661	661	662
	UA		91.0	92.8	93.0	93.1	82.9	86.3	87.0	87.1
	PA		89.5	89.5	90.0	90.3	86.5	86.3	86.9	87.2
Wheat	Area	103	75	103	111	112	66	100	111	113
	UA		87.6	79.0	77.4	77.4	86.3	75.0	72.8	73.1
	PA		64.4	79.6	83.7	84.6	55.7	73.0	79.1	80.6
Alfalfa	Area	45	24	31	34	34	23	32	35	35
	UA		82.2	74.1	72.9	72.9	75.8	66.3	64.6	64.9
	PA		43.3	51.9	54.4	55.8	39.5	47.8	50.0	50.9
Soybean	Area	746	724	722	721	722	683	702	704	706
	UA		90.0	91.2	91.6	91.8	86.7	87.6	88.2	88.5
	PA		87.3	88.2	88.6	88.8	79.4	82.4	83.3	83.7
Other Crops	Area	13	0	0	0	0	0	0	1	1
	UA		0.0	0.0	0.0	0.0	0.0	0.0	87.9	83.4
	PA		0.0	0.0	0.0	0.0	0.0	0.0	4.2	6.4
Water	Area	154	119	121	123	125	125	126	127	128
	UA		97.3	97.4	97.3	97.1	93.9	94.3	94.2	94.1
	PA		75.6	76.6	77.6	78.6	76.2	76.9	77.7	78.0
Barren/Dev.	Area	131	46	49	53	55	44	50	52	54
	UA		44.8	49.8	50.7	51.1	32.8	37.2	40.2	40.7
	PA		15.7	18.7	20.7	21.6	11.0	14.2	16.0	16.7
Forest	Area	64	64	65	64	64	62	65	64	64
	UA		65.8	66.0	66.9	67.3	61.8	61.0	62.0	62.3
	PA		65.1	66.6	66.8	66.8	59.0	61.8	61.9	61.9
Grassland	Area	761	966	917	903	897	988	949	930	925
	UA		71.9	74.6	75.5	76.0	69.6	71.7	72.9	73.3
	PA		91.3	90.0	89.7	89.7	90.4	89.5	89.2	89.2
Wetland	Area	262	272	294	291	289	260	256	255	253
	UA		43.5	44.7	46.4	47.3	47.8	50.5	51.9	52.7
	PA		45.0	50.0	51.5	52.1	47.4	49.2	50.4	50.8
OA			78.9	80.1	80.7	81.0	75.5	77.2	78.0	78.3
k_loc			0.808	0.817	0.821	0.823	0.776	0.776	0.783	0.786
k_quan			0.853	0.866	0.872	0.876	0.828	0.868	0.874	0.876

Table S4.15. Pixel-based comparison between 2017 predicted land cover maps and the CDL summarized by sample sizes. Areal units are km².

Land Cover	Info.	CDL	ARD				HLS			
			P01	P05	P15	P25	P01	P05	P15	P25
Corn	Area	685	598	596	604	608	603	604	609	612
	UA		94.4	95.6	95.8	95.8	91.2	92.5	92.7	92.8
	PA		82.4	83.3	84.5	85.0	80.3	81.5	82.5	82.9
Wheat	Area	87	41	62	69	70	44	62	67	69
	UA		92.3	87.3	86.8	86.6	91.4	87.3	86.2	85.9
	PA		43.6	62.7	68.7	70.4	46.0	62.2	66.9	68.3
Alfalfa	Area	50	24	37	41	41	25	37	39	40
	UA		90.1	80.1	80.1	80.5	86.6	76.5	77.0	77.1
	PA		42.6	59.7	65.3	66.6	43.4	56.8	59.8	61.3
Soybean	Area	803	825	810	806	806	823	810	812	813
	UA		88.0	89.9	90.7	90.8	85.6	87.5	87.8	88.0
	PA		90.4	90.6	90.9	91.1	87.7	88.2	88.8	89.0
Other Crops	Area	8	0	0	0	1	0	0	0	0
	UA		0.0	0.0	0.0	99.4	0.0	0.0	0.0	0.0
	PA		0.0	0.0	0.0	7.2	0.0	0.0	0.0	0.0
Water	Area	153	115	117	119	120	119	120	121	122
	UA		97.7	97.6	97.4	97.2	94.8	94.7	94.5	94.3
	PA		73.5	74.4	75.4	76.0	73.4	74.2	74.9	75.4
Barren/Dev.	Area	111	43	51	53	54	35	49	53	54
	UA		49.4	49.9	51.5	51.7	44.0	42.6	43.6	44.2
	PA		19.2	22.8	24.8	25.5	14.1	18.8	20.8	21.5
Forest	Area	63	67	72	71	71	58	64	63	62
	UA		61.7	60.5	61.4	62.0	62.8	60.9	62.0	62.8
	PA		65.6	69.0	69.6	69.7	58.1	61.5	61.7	61.4
Grassland	Area	744	964	909	896	894	989	937	928	923
	UA		70.8	73.8	74.7	74.9	69.1	71.6	72.3	72.6
	PA		91.7	90.1	89.9	90.0	91.9	90.2	90.1	90.1
Wetland	Area	237	264	287	282	276	245	259	249	247
	UA		43.2	44.5	46.7	48.1	44.1	46.2	49.1	50.2
	PA		48.3	54.0	55.7	56.0	45.6	50.5	51.6	52.2
OA			79.0	80.4	81.3	81.6	77.3	78.7	79.4	79.7
k_loc			0.830	0.834	0.838	0.839	0.805	0.806	0.810	0.812
k_quan			0.821	0.842	0.857	0.863	0.823	0.851	0.862	0.866

Table S4.16. Pixel-based comparison between 2017 predicted land cover maps and the CDL summarized by sets of input variables. Areal units are km^2.

| Land Cover	Info.	CDL	ARD				HLS			
			CxQ	HPLM	SPL	CMB	CxQ	HPLM	SPL	CMB
Corn	Area	685	597	625	602	604	613	627	641	608
	UA		91.6	90.7	90.9	95.1	88.3	88.4	82.7	92.1
	PA		79.9	82.8	80.0	83.9	79.1	81.0	77.4	81.8
Wheat	Area	87	64	66	14	70	59	70	2	67
	UA		81.7	84.5	55.2	83.4	80.9	81.6	76.7	84.7
	PA		60.0	64.5	8.7	67.1	55.4	66.2	1.8	65.8
Alfalfa	Area	50	38	42	39	36	34	43	35	36
	UA		81.2	73.6	69.0	82.1	80.2	70.6	67.9	78.8
	PA		62.3	61.3	53.7	59.0	54.9	61.0	47.7	56.9
Soybean	Area	803	765	755	912	816	769	762	910	812
	UA		88.2	89.5	80.2	89.6	85.7	87.1	77.7	87.4
	PA		84.1	84.0	91.1	91.0	82.1	82.6	88.1	88.4
Other Crops	Area	8	0	0	0	0	0	0	0	0
	UA		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	PA		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Water	Area	153	118	121	119	116	119	115	124	120
	UA		91.6	91.7	97.5	97.8	89.1	92.3	94.1	94.9
	PA		70.5	72.1	75.5	74.2	69.5	69.0	76.3	74.5
Barren/Dev.	Area	111	71	64	53	57	65	46	46	61
	UA		25.1	31.1	55.5	52.9	24.8	28.4	50.6	43.6
	PA		16.1	17.9	26.6	27.1	14.7	11.9	21.1	24.1
Forest	Area	63	56	73	76	74	37	65	73	67
	UA		46.2	48.1	60.5	61.0	55.7	49.7	55.0	60.1
	PA		40.9	55.7	73.0	72.1	32.4	51.1	63.9	64.1
Grassland	Area	744	833	886	906	915	866	927	931	938
	UA		76.1	73.8	72.5	73.3	73.9	70.4	70.4	71.5
	PA		85.2	87.9	88.2	90.1	86.0	87.6	88.1	90.1
Wetland	Area	237	399	310	220	253	377	286	178	231
	UA		33.7	38.5	51.0	49.9	35.4	38.8	56.1	49.8
	PA		56.7	50.4	47.4	53.3	56.4	46.9	42.1	48.6
OA			75.7	77.1	77.3	80.9	74.6	75.5	74.8	78.9
k_loc			0.778	0.787	0.810	0.834	0.766	0.767	0.780	0.805
k_quan			0.831	0.845	0.818	0.857	0.826	0.842	0.811	0.861
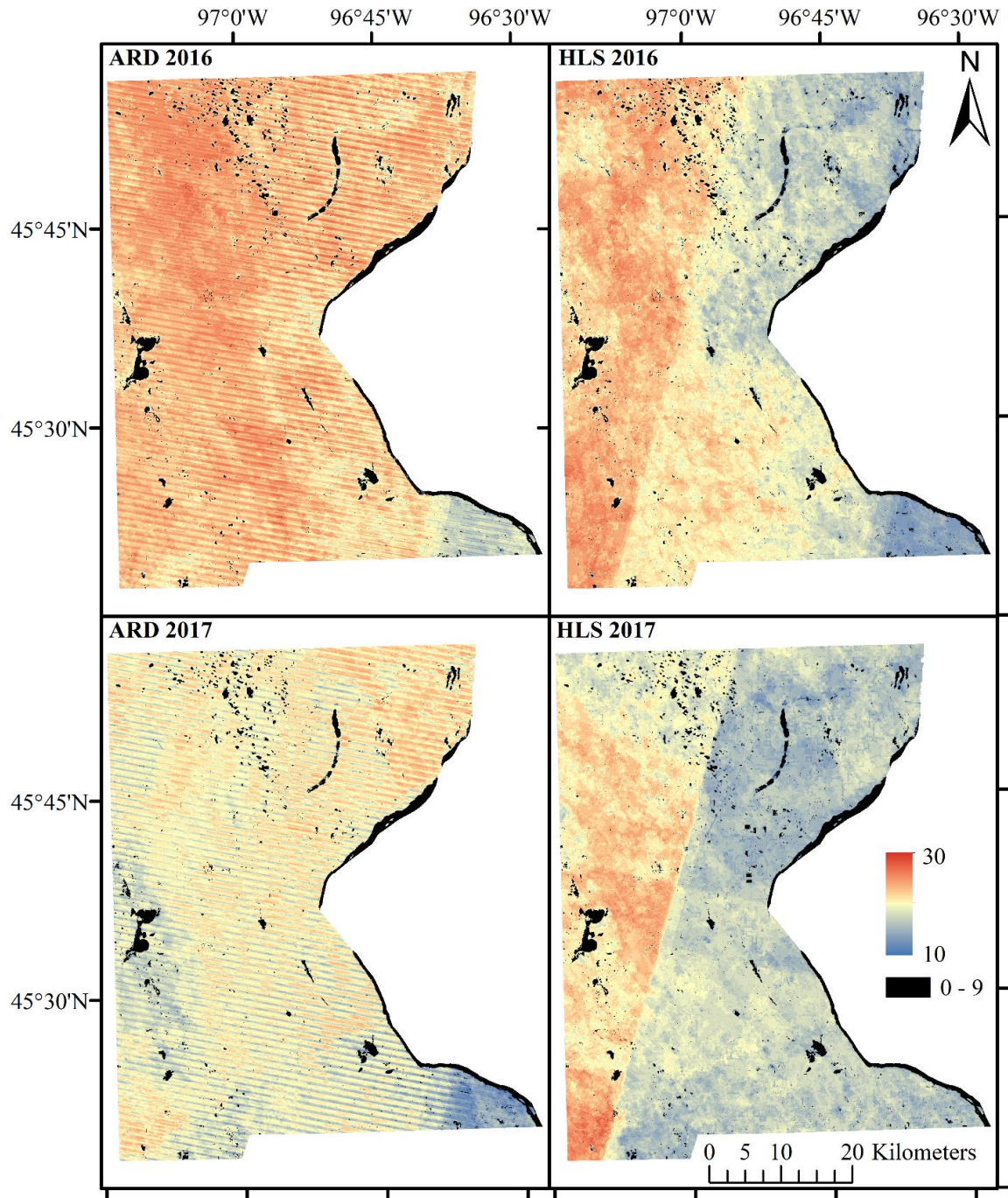
Figure S4.4. Number of valid observations (cloud/snow/shadow-free, EVI2>0) over the study area for each combination of year and data source.

CHAPTER 5

RESEARCH SUMMARY AND RECOMMENDATIONS

5.1   Summary and Key Findings

Understanding rapid land change in the U.S. Northern Great Plains region is not only critical for management and conservation of prairie habitats and ecosystem services, but also for projecting production of crops and biofuels and the impacts of land conversion on water quality and rural transportation infrastructure. Hence, it raises the need for an LCLU dataset with good spatiotemporal coverage as well as consistent accuracy through time to enable change analysis. This dissertation aimed (1) to develop a novel classification method from the perspective of land surface phenology, which utilizes time series images from comparable sensors, and (2) to apply the land cover/land use dataset generated from the phenometrically-based classification approach to quantify crop expansion in South Dakota. My dissertation research advances the researcher's toolkit for land cover mapping and change analysis as well as shines fresh light on what has been a controversial issue since 2013: the conversion of mixed-grass prairie to commodity crops, particularly to corn and soybean, due to the increasing demand for biofuels and animal feed. The primary results of this research are summarized below

Chapter 2: The main focus of Chapter 2 was to evaluate "*how well does land cover mapping perform if phenological metrics alone are used as input to the classification algorithm*" (research question #1). In addition, responses of RFC models to different sample sizes and sampling designs were also examined to identify which sampling scenario yielded the most accurate classification.

The classification based only on phenometrics derived from the Convex Quadratic model could accurately differentiate major commodity crops with PA/UA of above 0.7 and 0.9 for 2012 and 2014 RFC models, respectively. However, accuracy of non-vegetated

classes, especially for developed, are limited. Among sampling designs, the "same distribution" RFC models (proportional distribution of the sample is like proportional distribution of the population) tend to yield best land cover prediction. Without control on sample dataset, larger sample sizes did not necessary lead to better RFC models. The "same distribution" sample dataset covering 0.25% of the study area seems to be adequate to achieve accurate classification.

Chapter 3: The objective of Chapter 3 was to *quantity the rate of cropland expansion and its spatial pattern in South Dakota over the past decade* (research question #2). I proposed a trajectory-based approach that considers the entire land cover/land use time series to determine if there was actual land change at a particular location, to overcome the limitations of the bi-temporal change detection.

Between 2007 and 2015, the trajectory-based change detection approach estimated a cropland expansion of 5,447 km^2 in South Dakota (equivalent to 14% of the existing cropland area), which matches much more closely the reports from the National Agriculture Statistics Service (5,921 km^2) and the National Resources Inventory (5,034 km^2) than an estimation from the bi-temporal approach (8,018 km^2). Cropland gains were mostly concentrated in 10 counties in northern and central South Dakota. Urbanizing Lincoln County, part of the Sioux Falls metropolitan area, is the only county in South Dakota with a net loss in cropland area over the study period. An evaluation of land suitability for crops using the Soil Survey Geographic Database indicated a scarcity in high-quality arable land available for cropland expansion in South Dakota.

Chapter 4: The goal of Chapter 4 was further exploration of the phenometrically-based classification approach presented in Chapter 2 by addressing two questions: (1) *how*

the selection of different LSP models used to fit annual time series impacts classification accuracy (research question #3) and (2) *how LSP-based classification can be improved by incorporating information from spectral variables* (research question #4). Similar to Chapter 2, land cover classifications in Chapter 4 were also evaluated with alternative sampling scenarios to examine which sampling scenario yielded the most accurate classification. The assessment was conducted for two years 2016 and 2017 using two different sources for the imagery: Landsat ARD and HLS data.

There was no obvious choice between the 2016 RFC models using CxQ versus HPLM: the HPLM RFC models performed better with the ARD and the CxQ RFC models were better with the HLS data. For 2017 data, HPLM RFC models slightly edge CxQ with about 1% higher overall accuracy, mostly due to more precise allocation of land cover. Indeed, the TOST test of equivalence indicated that overall accuracies of the 2017 HPLM and CxQ RFC models exhibited equivalent performance. The spectrally-based RFC models were more accurate than the phenometrically-based RFC models, especially for non-crop cover types. However, the spectrally-based RFC models could not classify the wheat class accurately. The combined spectral-phenological variables RFC models consistently overcame weaknesses of both phenometrically-based classification (low accuracy for non-vegetated covers) and spectrally-based classification (low accuracy for wheat).

A sample pool with a minimum correction of land cover information yielded the most accurate predicted map despite its lowest RFC models' accuracy. A random stratified sample dataset should cover at least 0.25% of the study area to achieve accurate land cover

map. In case of data scarcity, smaller datasets would yield acceptable for classification but ought not be smaller than 0.05% of the study area.

5.2   Multi-temporal land cover classification and change detection: a synthesis

A rapid increase of available Earth observations and advances in computer science led to an emergence of land cover classification (see *Gomez et al, 2016*; Chapters 2 & 4) based on time series data, starting in 2010s. Compared to the conventional land cover classification approaches, the phenology-based method exploits the rich temporal information from time series to map land cover more accurately (*Mitchell et al., 2013; Franklin et al., 2015*). The phenology-based method is also fully compatible for an operational process, which is a critical advantage for generating land cover maps at very large scale. Beside those advantages, phenology-based classification (or classification using time series data, in general) are still facing several challenges that need to be addressed by future studies.

A substantial number of good quality observations required for the fitting of land surface phenology model poses the greatest challenges in application of the phenometrically-based classification (*Zhong et al., 2011; Jia et al., 2014; Kong et al., 2016;* Chapter 2). Thus, most extant studies focused on using MODIS NDVI or EVI time series, which are available with higher temporal coverage but at a coarser spatial resolution (*Zhong et al., 2011; Xue et al., 2014; Yan et al., 2015; Qader et al., 2016*). My research has shown that land cover can be mapped accurately at finer spatial resolution using data from Landsat and Sentinel satellites. However, data gaps might be created during the fitting process arising from model failure due to low number of good observations in some years and over some areas. Because the gaps were produced in different places across the years,

a direct comparison between yearly RFC outputs was not appropriate, in most cases, without spatiotemporal interpolation. Even within the Landsat sidelap zone, it was not always able to retrieve a sufficient number of good observations to fit the land surface phenology model.

The temporal density of observations, however, could be increased by bringing together complementary sensors. For example, results from my research demonstrated that Sentinel-2 data can be used with Landsat ARD in the phenometrically-based classification. Another feasible solution is to leverage very high spatiotemporal but low spectral resolution data from a small satellite constellation to infill gaps (*Houborg & McCabe, 2018*). The temporal resolution of Landsat-like data can also be improved for phenometrically-based classification by fusing with data that has lower spatial but substantial higher temporal resolution, such as MODIS (*Jia et al., 2014; Kong et al., 2016*). Nevertheless, it is important to note that the use of multi-sensor time series would require a huge preprocessing effort. In addition, in areas with persistent cloud cover, such as in the moist tropics, it may be impractical to collect a sufficient number of good observations spanning the growing season, regardless how many optical sensors are observing. In those areas, land cover classification would benefit from leveraging Synthetic Aperture Radar (SAR) data (*Waske & Braun,* 2009; *Qi et al, 2012*)

Extant phenology-based land cover classification studies, including my dissertation research, only focus on the annual pattern of vegetation index time series. It is possibly due to the long use of NDVI (or other vegetation indices) in land cover classification study (*Tucker et al., 1985; DeFries & Townshend, 1994*) and availability of well-developed land surface phenology models (see *Henebry & de Beurs, 2013*). Phenometrics derived from

time series can complement the spectral information to improve classification accuracy (Chapter 4; *Jia et al., 2014; Kong et al., 2016*). However, the use of spectral data in my dissertation research, *Jia et al. (2014)* and *Kong et al. (2016)* was limited to the generation of spectral variables, partly due to a lack of a well-studied tool to simulate annual pattern of spectral data (excepts for vegetation indices). Temporal trajectory of spectral bands and their related variables proved to be useful in land cover classification (*Zhu & Woodcock 2014*) and should be utilized more in the future classification study. However, a robust land surface seasonality model is needed to simulate annual pattern of different spectral variable time series (bands and ratios) and to extract seasonality information from the fitted curves.

Number of variables in phenometrically-based classification can increase significantly when combining features from different land surface phenology models or incorporating additional information from other spectral bands and ratios. Although RFC models are not sensitive to correlated or noise variables (*Biau, 2012*), a larger number of input variables would increase the computational complexity. In addition, some input variables may be linked to the same biogeophysical property of the land surface. Including all of those variables in land cover/land use classification may weaken a contribution of the underlying biogeophysical process to the classification. Thus, future land cover classification studies with high dimensional data, such as those using phenology-based approach, would consider applying dimension reduction techniques (*Fodor, 2002*) or variable selection methods (*Degenhardt et al., 2013*) to achieve efficient computation and clear understanding of variable importance.

Land cover/land use classifications in my dissertation study were limited to a few simple cover classes. Cropland mapping in this study was constrained to a few commodity

crops (partly due to the homogeneous crop planting in the study area) with a single growing seasonal per year, including corn, soybean, wheat, sunflower, alfalfa. In addition, forest and grassland were not broken into smaller classes (partly due to the focus on cropland of the study). To demonstrate fully the capability and advantage of the phenology-based classification, future studies should be conducted in areas with more complex land cover/land use, *e.g.*, multiple cropping in California's Central Valley or different types of forest in the eastern US region.

Compared to the traditional bi-temporal approach, the land cover change detection exploring time series data with the trajectory-based approach (Chapter 3) yielded a closer estimation of cropland expansion in South Dakota to the USDA's reports. However, my study only quantified changes in cropland area, partly due to limited accuracy of non-cropland classes in the input land cover datasets. Thus, exact quantification of grassland or wetland losses due to cropland expansion, another information of interest for conservation purposes, was not available. In addition, the land cover dataset for South Dakota generated from my study does not offer crop-specific categories like the CDL (only mapped three broad categories: cropland, grassland and others), preventing identification of the crops were planted on the newly cultivated lands. Applying the trajectory-based change detection approach on a land cover dataset with more detailed classes, such as CDL, could yield valuable information about changes in the study area.

5.4   Future Research

Examine capability of land cover/land use classification using time series

My next research would further explore capability of LSP-based classification by performing classification in areas with complex land cover such as: multiple cropping in

California's Central Valley or different types of forest in the eastern US region. I would use surface reflectance from Landsat 7, 8 and Sentinel-2 together to obtain more good quality observations. As comparable classification accuracies were achieved from ARD and HLS data, the combination of Landsat 7, 8 and Sentinel-2 for land cover/land use classification is feasible. Evaluating classification performance (based on similar sets of input variables) for those areas would further confirm robustness of the proposed approach.

High spatiotemporal dataset from Planet for land cover classification

In another future study, I would want to explore a private dataset from PlanetScope (*Planet, 2018*)—a constellation of approximate 130 small satellites—available at 3 meter spatial resolution and up to daily coverage since 2017. Although the PlanetScope dataset has several limitations, such as a maximum of 10,000 km^2 of free data each month per user, only offers four spectral bands—RGB and NIR and observed at different times (some satellites have a sun-synchronous orbit and others have an international space station orbit), a very high spatiotemporal resolution of the archive make it a great data source for land cover/land use classification, especially when a high spatial resolution land cover map is needed at local scale. Using a very dense time series of PlanetScope could also allow to examine how well existing land surface phenology models work on other spectral bands and ratios. It would be a good start for developing a new generation of land surface seasonality models.

5.3   References

Biau, G. 2012. Analysis of a random forests model. J. Mach. Learn. Res. 13, 1063-1095.

DeFries, R.S., & Townshend, J.R.G. 1994. NDVI-derived land cover classifications at a global scale. Int. J. Remote Sens. 15(17), 3567-3586.

Degenhardt, F., Seifert, S., & Szymczak, S. 2017. Evaluation of variable selection methods for random forests and omics data sets. Brief. Bioinform. 20(2), 492-503.

Fodor, I. K. 2002. A survey of dimension reduction techniques (No. UCRL-ID-148494). Lawrence Livermore National Lab., CA (US).

Franklin, S.E., Ahmed, O.S., Wulder, M.A., White, J.C., Hermosilla, T., Coops, N.C., 2015. Large area mapping of annual land cover dynamics using multi-temporal change detection and classification of Landsat time-series data. Can. J. Remote. Sens. 41, 293–314.

Gómez, C., White, J.C., & Wulder, M.A. 2016. Optical remotely sensed time series data for land cover classification: A review. ISPRS J. Photogramm. Remote Sens. 116, 55-72.

Henebry, G.M., de Beurs, K.M., 2013. Remote sensing of land surface phenology: a prospectus. In: Schwartz, M.D. (Ed.), Phenology: An Integrative Environmental Science, 2e. Springer, pp. 385–411.

Houborg, R., McCabe, M.F., 2018. A cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing Planet, Landsat and MODIS data. Remote Sens. Environ. 209, 211–226.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., & Wang, X. 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens. 6(11), 11518-11532.

Key, T., Warner, T.A., McGraw, J.B., & Fajvan, M.A. 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification

of individual tree species in a temperate hardwood forest. Remote Sens. Environ. 75(1), 100-112.

Kong, F., Li, X., Wang, H., Xie, D., Li, X., & Bai, Y. 2016. Land cover classification based on fused data from GF-1 and MODIS NDVI time series. Remote Sens. 8(9), 741.

Mitchell, J., Shrestha, R., Moore-Ellison, C., & Glenn, N. 2013. Single and multi-date Landsat classifications of basalt to support soil survey efforts. Remote Sens. 5(10), 4857-4876.

Planet. 2018. Planet Imagery Product Specifications. Available online: https://assets.planet .com/docs/Combined-Imagery-Product-Spec-Dec-2018.pdf

Qader, S.H., Dash, J., Atkinson, P.M., Rodriguez-Galiano, V., 2016. Classification of vegetation type in Iraq using satellite-based phenological parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 9 (1), 414–424.

Qi, Z., Yeh, A.G.O., Li, X., & Lin, Z. 2012. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. Remote Sens. Environ. 118, 21-39.

Tucker, C.J., Townshend, J.R., & Goff, T.E. 1985. African land-cover classification using satellite data. Science, 227(4685), 369-375.

Waske, B., & Braun, M. 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. ISPRS J. Photogramm. Remote Sens. 64(5), 450-457.

Xue, Z., Du, P., & Feng, L. 2014. Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7(4), 1142-1156.

Yan, E., Wang, G., Lin, H., Xia, C., & Sun, H. 2015. Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. Int. J. Remote Sens. 36(2), 489-512.

Zhong, L., Hawkins, T., Biging, G., Gong, P., 2011. A phenology-based approach to map crop types in the San Joaquin Valley, California. Int. J. Remote Sens. 32 (22), 7777–7804.

Zhu, Z., and Woodcock, C.E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152-171.