CHARACTERIZING MAJOR CROP CHANGE TRENDS IN FLORIDA FROM MULTIPLE
SOURCES: A COMPARISON OF USDA CROPLAND DATA LAYER AND STATEWIDE
RANDOM FOREST CLASSIFICATIONS OF THE TOP FIVE CROPS

By

TYLER A SCHAPER

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2021

I dedicate this to all the people who helped me along the way, and especially my loved ones.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

CHARACTERIZING MAJOR CROP CHANGE TRENDS IN FLORIDA FROM MULTIPLE
SOURCES: A COMPARISON OF USDA CROPLAND DATA LAYER AND STATEWIDE
RANDOM FOREST CLASSIFICATIONS OF THE TOP FIVE CROPS

By

Tyler A Schaper

May 2021

Chair: Jane Southworth
Cochair: Gregory Glass
Major: Geography

Monitoring crop coverage and crop change in agricultural areas is of paramount importance
for the understanding and management of food production, statewide economies, environmental
impacts and global environmental change. The USDAs Cropland Data Layer (CDL) is the most
commonly used database of crop coverage produced annually for the conterminous US. This
research compared the output of the published CDL products to a statewide classification focused
on the top five agricultural crops in the state of Florida. Given the prominence of agriculture to
the statewide economy and the extensive area of the state in crop production, the consequence of
accurate understanding of crop coverage and crop changes is key. The research utilized a
multi-date, within season, image stack to produce a five-crop classification using a Random
Forest (RF) classifier. 250 testing sites were used to compare both the RF and CDL products, for
both 2008 and 2016, and results showed the CDL to have lower than 40% overall accuracy,
compared to over 85% overall accuracy for the RF classification, for these five crops. Change in
crop coverage statewide decreased about 4% between 2008 and 2016, and for individual crop
coverages, over 12% decrease occurred in citrus and over 4% decrease in peanuts. Cotton and
strawberry coverages increased significantly, although both are much less significant crops in
terms of area statewide. Sugarcane remained stable in area over time. Changes in agricultural
production, especially given the position of Florida as the top citrus and sugarcane producing

state, and their importance to the statewide economy, are of key concerns to agricultural and land managers alike.

CHAPTER 1
INTRODUCTION

Geography as a study has long sought to understand and quantify the interactions of humans and their home Earth in the broadest sense [32]. One of the most prominent and direct ways that humans have impacted the Earth's surface is through agriculture. Agriculture can be attributed to one of the most important steps and practices in human history, and its role has not diminished in relation to humans as time has gone on [17]. Agriculture not only played a role in allowing humans to settle and sustain them, but also plays a large economic role. Agriculture, the land put into cropping and other agricultural practices are ubiquitous across the Earth's surface. Their impacts are not limited to the area of agricultural production, but have far reaching consequences such as environmental (runoff, pesticides, fertilizers, etc.), economic (crops for local, national, and international markets) and social (crop type, job opportunities, development, etc.) issues. Other effects and impacts include habitat loss leading to endemic species extinction, eutrophication, and other natural environment disruptions stemming from human systems[31]. Urban areas and croplands cover approximately 10 to 15% of the surface of Earth. Additionally, 6 to 8% of the Earth's land area has been converted into pasture [37]. Considering this level of human-environment interaction, understanding agriculture through the lens of geography is a critical analysis of increasing import as conversion continues. The magnitude of these impacts and roles can be examined and analyzed at different scales, from global through to local all with differing results depending on the scope of analysis. Given agriculture's importance to both humans and their environment, not to mention the ample ways to quantify their interaction, it is of paramount importance to accurately measure agricultural activity. Remote sensing, one of the most important data sources and techniques for quantification human-environment interactions is employed in this study to gain a greater understanding of the extent of agricultural cover and its change over time.

The monitoring of agriculture via remote sensing lends itself to the discipline of land change science. Land change science combines social and biophysical sciences with geographical information sciences [34] to answer the question of where change happened, how much changed, why change took place, and when change occurred. Scientific entities such as the National

Aeronautics and Space Administration (NASA) and the United States Department of Agriculture (USDA) utilize programs like Land-Cover/Land-Use Change (LCLCUC) and the Cropland Data Layer (CDL) to combine those social, biophysical, and geographical information sciences. These programs leverage satellite-based studies, representative field-based process studies, and modeling, to monitor these land-cover/land-use changes [8]. With the archive of satellite records spanning the last half century, it is possible to monitor changes in land-use and land-cover at a coarse temporal scale.. However, looking towards the future, improvements and advancements are being made to provide both finer temporal and spatial resolutions.

Remote sensing is now critical for mapping and quantifying change [23] and one of the most frequently applied techniques for land-cover/land-use is the classification of satellite images into land cover maps. A plethora of classification techniques used to classify images fall into one of two categories; unsupervised or supervised classification. Recently, machine-learning algorithms such as Random Forests (RF) have emerged prominently. RF generally has a higher accuracy than other traditional classification approaches. This is because it creates an ensemble of trees and then utilizes them to vote for the most popular class [5]. With this technique, a wide variety of variables are able to be used and Random Forest will ignore insignificant covariates. Random Forest has the great advantage of not over-fitting data. This is due to the large number of trees Random Forest "grows" for analysis. In order to grow these large number of trees, random prediction variables are used, and no pruning is used. This large number of trees results in a low generalization error [5]. This makes Random Forest an ideal classifier for classifying highly complex and heterogeneous landscapes.

While these concepts of quantifying human and environment interaction through land-cover/land-use change via remotely sensed image classification can be applied anywhere on earth and at a multi-scale level; for example, a global, regional, local, or other scale; This research focuses on a state scale, emphasising the major agricultural crops in Florida, leading to the creation of two specific classified images (2008 and 2016) utilizing the RF method. Images are compared to federally published data to compare results and their relative accuracy. The research

objective is to develop an integrated high resolution, Landsat-based, remotely sensed analysis to understand the changing spatial and temporal patterns of the five most important crops in Florida. Specific aims are to: (1) develop a RF-based approach which integrates both spatial and temporal data to estimate agricultural crop areas in Florida, (2) compare these results to the federal data and field data to determine which method is better, and (3) map the areas in citrus, cotton, peanut, sugarcane and strawberry and determine changes between 2008 and 2016. The results will assist agricultural planners and land managers alike, by furnishing a method of determining location, quantity, and agriculture type with greater accuracy.

CHAPTER 2
METHODS, RESULTS, AND DISCUSSION

Humans have altered the earth's landscape for many reasons, but none as much as agriculture. Agriculture represents the largest anthropogenic use of land on earth's surface [7, 25]. At the global scale, the production of agriculture has increased faster than population in recent years [12]. While agriculture production has increased as a whole, world prices for agricultural products have generally seen a decline. This has been possible through the use of technological advancements to increase crop yields, most notably since the 1970's. Advancements such as improved seeds, irrigation, fertilizers, and pesticides [12]. In recent years, increased agricultural production has been achieved through crop yield rather than crop area. By switching from expansion of crop area to intensive input use as the primary production growth method has reduced land conversion by 1 billion hectares globally since the 1960's [18]. Despite the lessened need for cropping area expansion to produce more food, agriculture still stands as the top category for land use globally. Since remote sensing is used to capture earth's surface through time, it is one of the most useful tools to determine land cover and land use across the globe. With the recent dramatic increase in the availability of satellite-based products for the large-scale study of various land covers, it is now much easier to monitor agricultural cover at a finer temporal and spatial scale. In addition, we can utilize these geospatial data to provide extensive and more detailed information regarding agricultural productivity [16]. Climate and soil, land classification and crop inventories, productivity, as well as crop stress and disease are a few areas remote sensing has direct applications to in agriculture [30].

It is increasingly important to have accurate and current estimates of crop area and amounts, including crop type, spatial arrangement, and location, for the analysis of food production, biodiversity, agricultural pests and diseases and their related inputs, agricultural policy and land management. Indeed, it is the mission of the National Agricultural Statistics Services (NASS) which is part of the US Department of Agriculture (USDA) to produce such data at an annual timestep for just these reasons [14]. NASS produces the Cropland Data Layer (CDL) product which is a raster based, georeferenced, land cover map of crop type. The data is created based on moderate resolution satellite imagery, USDA collected ground control data and the National Land

Cover Dataset along with other ancillary products (see Boryan et al. 2011, for detailed description of NASS CRL products). Published analysis of the 2009 products describe accuracies of 85-95% for the major crop categories. Despite these published accuracies, and especially given the importance of the crop type information at the field level, many other researchers have found significant discrepancies between the actual crop covers and the CDL products for their regions of interest, with some areas experiencing over or underestimations of 50% [16]. As such, it is deemed essential to review the CDL products for specific regions of interest prior to their use within any research studies. Florida has a total area of 170,310 km^2 and hosts a rich agricultural diversity. According to the USDA and the NASS, oranges are so important to the state of Florida that in 2016 they generated $905,138,000 in the value of production. The next most profitable crops are grapefruit ($136,495,000), and hay ($129,600,000) [35]. In 2015, Florida produced 60% and 58% of the total value in the United States for oranges and grapefruit, respectively, according to the Florida Department of Agriculture and Consumer Services [6] and as such, citrus is a key crop across the state [2]. Florida also produces 54% of the US value in sugarcane for sugar and seed production, valuing at $515 million, as well as 13% of strawberries, valuing at $291 million [35]. Other prominent crops in the state are peanuts and cotton. As such, Florida is the top producer nationally of oranges and sugarcane and economically speaking, in 2014, almost 14% of all jobs in the state were linked to agriculture, with $155 billion from direct industry output and $127.3 billion in value-added contributions. In 2017 Florida had 47,000 farm operations covering 9.45 million acres of farmland which was another year of decline, from its peak of 48,000 farms covering 9.55 million acres in 2013 [36]. Accuracy in the location, amount and type of crops, as well as changes in production over time, is therefore of key interest to land managers and the state overall. This research looks to improve upon current methods used to identify and estimate areas in different agricultural crops and to better determine changes in agricultural crops over time in the state. The research objective is to develop an integrated high resolution and Landsat based remotely sensed analysis to understand the spatial and temporal pattern of the top five crops in Florida. Specifically this research aims to: (1) Develop a random

forest based approach which integrates both spatial and temporal data to estimate agricultural crop areas in Florida, (2) To compare these results with the federally published data and field data to determine which method is better, and (3) To map the areas in citrus, cotton, peanut, sugarcane and strawberry and determine area and location of change between 2008 and 2016.

2.1 Methods

Determining area and extent of crop coverage across the state is of key importance for economic and environmental concerns. The top five crops in Florida are Citrus, Sugarcane, Strawberry, Peanut and Cotton. Various researchers have highlighted limitations of existing agricultural data sources [15, 24, 3, 26, 9], specifically the use of the USDA/NASS CDL product, which is also widely recognized as the best available national level data. As such, given our interest is in location, extent and change in the top five crops between 2008 to 2016, we created two land cover classifications based on field data. We then compared these results and the CDL product to determine if our product outperforms the existing nationally and most widely used crop product available. Using field sites, visited by our team, we determined the accuracy of both products and determined changes in the top five crops for the state of Florida between the time periods of interest.

2.1.1 Satellite Image Classification

The remotely sensed imagery used in this study was obtained from the Landsat-ETM+ (Enhanced Thematic Mapper Plus) sensor as well as the Landsat-OLI/TIRS (Operational Land Imager/Thermal Infrared Sensor). These sensors obtain imagery at a 30 m spatial resolution for most bands, save for the thermal band on the ETM+ sensor which has a spatial resolution of 60 m and the thermal infrared bands on the OLI/TIRS sensors that have a spatial resolution of 100 m. Both satellite platforms have a temporal resolution of 16 days.

When creating classifications of agricultural landscapes, it is essential to account for the crop calendar for each crop being grown. Specifically, this helps identify ideal times to be able to extract land cover information which helps to differentiate crop types, determine grassland from agricultural fields and highlight citrus versus seasonal crops. The full suite of available imagery

was first analyzed, and 5 dates selected for each crop that most clearly differentiated the different crop types from their surrounding landscapes. Based on differences in crop types and their related growing patterns, the selected dates differed by crop (Figure 2-1). The resultant crop calendar graphs by spectral signature for the imagery used in this study, highlights times of differentiation and also shows variation in signatures throughout the growing seasons. This information was compiled across all five crop types and the best image dates were selected to differentiate crop landscapes and types. The growing seasons and distinguishable development points for each of the five key crops spanned the entire calendar year. Development points for the crops included stages such as planting, emerging, budding, flowering, fruiting, and harvesting for example. These stages were not identical between the five crops and so the need for multiple image dates, rather than only one or two, to best represent the range of growing conditions for all crops studied. This resulted in the creation of two composites of the different growing periods, for 2008 and 2016. In these composites, all five different dates throughout the growing season were used to create a single image stack of each year's composite (Table 2-1, Figure 2-1 and 2-2). There was an attempt to keep month and days consistent between the two years, but due to the temporal resolution of the satellite, as well as image quality when considering variables such as cloud coverage and growing days, slight deviations were made when selecting exact dates for each of the year's composites.
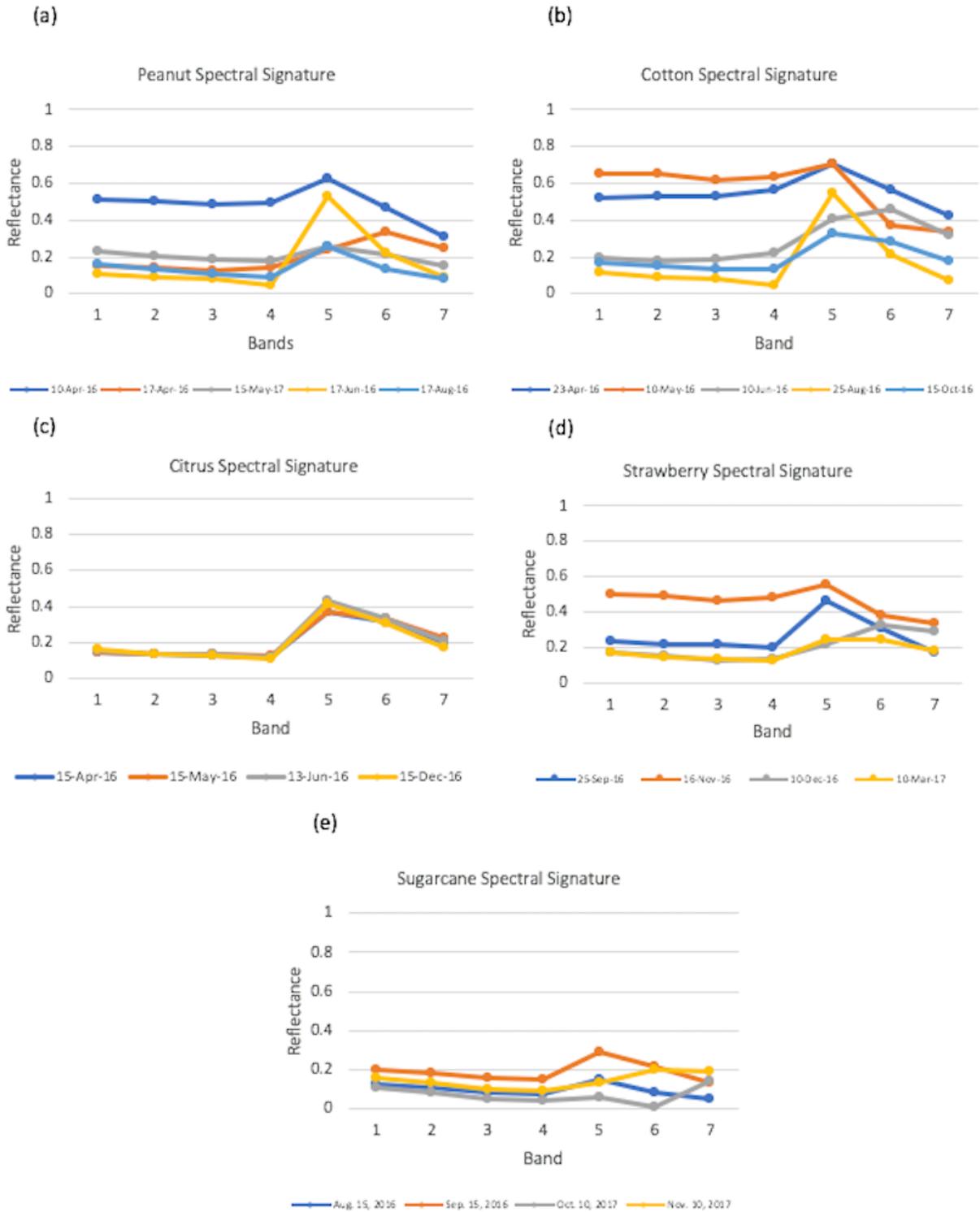
Figure 2-1. Spectral signatures of (**a**) peanut, (**b**) cotton, (**c**) citrus, (**d**) strawberry, and (**e**) sugarcane for Landsat at different development points within the growing season.
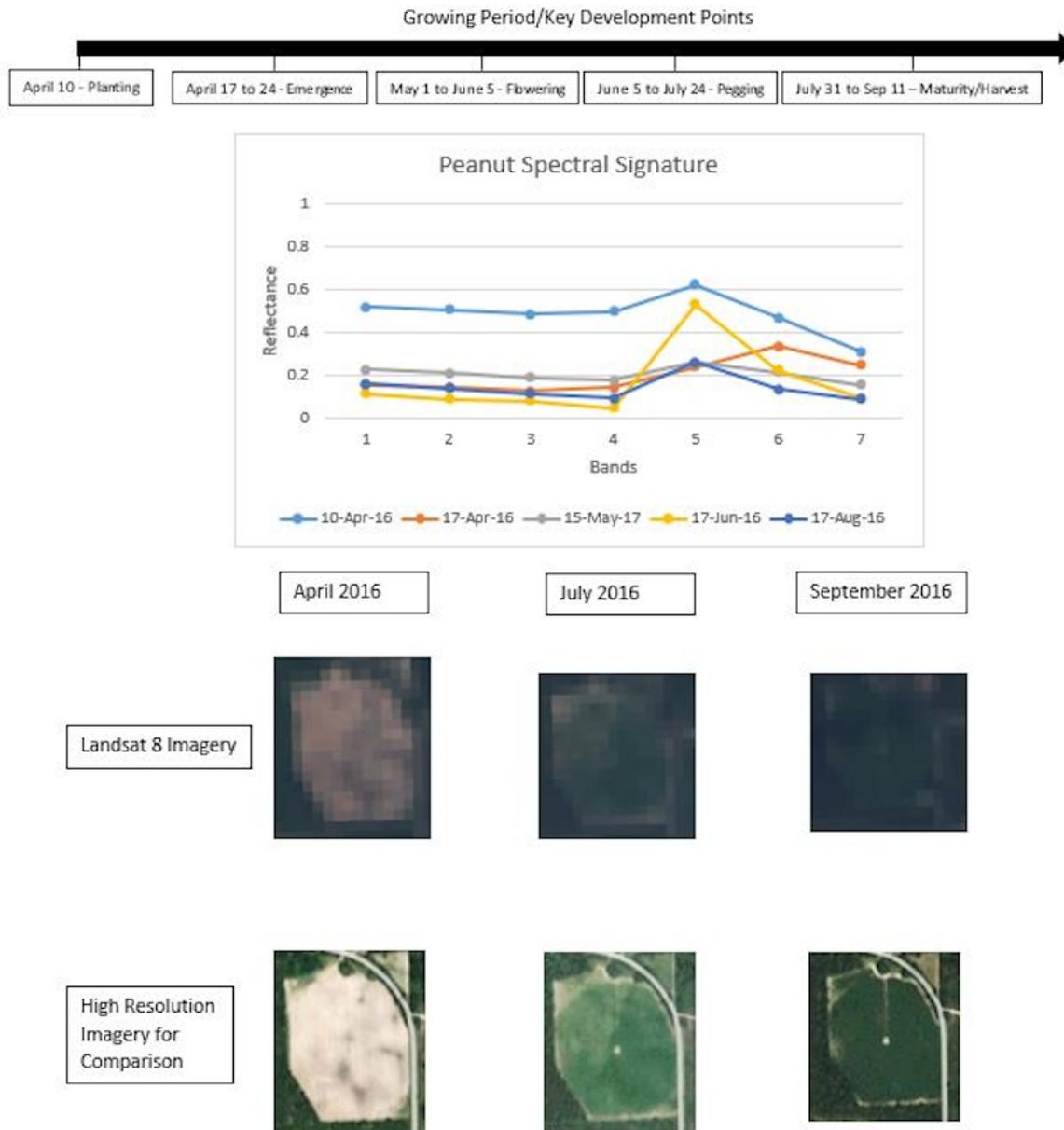
Figure 2-2. Crop calendar information for Peanuts, with a spectral signature plot for Landsat 8 data, and their high resolution (RapidEye) and Landsat image views for a peanut site.

In addition to the individual Landsat image bands a series of spectral indices were created to highlight differences in vegetated surfaces. These included the Normalized Difference Vegetation Index (NDVI), Normalized Difference Wetness Index (NDWI), and Enhanced

Vegetation Index (EVI). NDVI and EVI are commonly used indices to highlight vegetation growth, vigor and health and in association with the NDWI are excellent at highlighting slight changes in vegetation cover, growth and health [29, 33, 1]. As such, they can separate crop types across the growing season and also to differentiate crops from other natural land cover types within the state. Therefore, the input data for the classifications included the multiple image dates for bands with the created additional indices. The classification was performed in two stages. In the first stage we analyzed the full suite of land cover data for the state. This scheme included 12 classes. The classes consisted of the five key crops of cotton, citrus, peanuts, strawberries, and sugarcane and seven additional classes of urban, forest, grassland, woody wetland, herbaceous wetland, shrubland, and fallow. These latter seven classes were selected because they represented at least 1% of Florida's land cover, according to the CDL. To create the reference data for the classifications, random samples were collected during both years of interest. For 2016, a total of 833 reference data points were generated for the 12 classes. In 2008 a total of 725 reference date points were created. Given most agricultural land in the state has been in its designated cover for some time, most changes were in cropping type or rotation within a given year or abandonment of areas formerly in agriculture. As such, the 12 classes were reduced to only those classes of interest: the class of agricultural land, and specifically to the 5 main crops of interest. This allowed for improved accuracy of classification, and reduced confusion of classes not part of the agricultural landscape. As such, the classification only addressed changes in crop type within known agricultural lands within the state, from 2008 to 2016. An ensemble learning method was used for the supervised classification. Random Forest (RF) is an advanced machine-learning algorithm that does not require parametric assumptions to be made [5]. This method is widely-used because it tends not to over-fit data and has been shown to work at a high degree of accuracy even in complex systems [13]. Generally, RF performs well compared to other supervised classification techniques [11]. The reason random forest and tree voting works is through the use of the Strong Law of Large Numbers so the classifications converge and over-fitting is reduced as a problem [5]. This technique uses the variables in many statistical trees

in which the variables at different branching nodes randomly mutate and each tree 'votes' for the potentially most popular class [5]. 80% of reference points were used for training the RF classifier and the remaining 20% were used for accuracy assessment.

Table 2-1. Landsat image dates used for the random forest classification of crop type

2008 Landsat 7 Imagery		2016 Landsat 8 Imagery	
Image Dates	Image Bands Used	Image Dates	Image Bands Used
2008-02-11	Raw Bands 1-7	2016-02-09	Raw Bands 1-7
2008-05-01	NDVI	2016-02-09	NDVI
2008-06-02	NDWI	2016-06-16	NDWI
2008-11-09		2016-10-06	

2.1.2   USDA's Cropland Data Layer (CDL)

The most readily available crop coverage estimates were the national crop coverage generated by the USDA. The USDA's CDL contains information on land cover across all states, focusing specifically on the most dominant crop types nationally [21]. The CDL is a raster, geo-referenced, crop-specific land cover data layer created annually for the continental United States using moderate resolution satellite imagery and extensive agricultural ground truth [20]. Originally released at a 56 m resolution, the 2008 CDL was reprocessed using Landsat 5 Thematic Mapper data and released at a 30 m resolution. The 2016 CDL utilized imagery from Deimos-1, UK-DMC 2, and Landsat 8 to produce a 30 m resolution product for the continental United States [20]. The CDL has been used for many real-world applications, ranging from disaster assessments to agricultural policy decisions [27]. As such, the CDL products for 2008 and 2016 were downloaded and mapped to compare our own satellite based classified products too. In addition, the USDA NASS's Quick Stats portal also gives crop coverage in acres at multiple geographic scales including national, state-level, and county-level [22]. The Quick Stats coverage data is not geo-referenced data, but rather is a numerical product in which some level of accuracy analysis has been undertaken and corrections attempted to improve the CDL product. As such, many researchers use the NASS estimates as a more realistic validation of the actual acreage in each crop at a state level. Given this, we also extracted the NASS data for each crop for the state of Florida, for comparison, numerically, with our own satellite derived land cover

classification product.

2.2.1   RF Classification of 5 crop types

To assess the accuracy of the classification, a total of 250 sites, 50 sites for each of the five crops, were randomly selected to be ground truthed. For 2016, of the 250 ground truthing sites, a subset of 80 sites were physically visited. The remaining 170 sites were verified using high-resolution satellite imagery. Using 20% validation data, from the created reference data points for each year, accuracy assessments were completed, with confusion matrixes, for both the 2016 and 2008 random forest crop classifications. For both years, the overall accuracy of the random forest classifier was over 85%. In the 2008 classification, the accuracy for the validation data is 89.9% (Table 2-2). In the 2016 classification, the validation accuracy is 86.5% (Table 2-3). The resultant crop coverage maps for Florida for 2008 and 2016 indicated that crop locations are spatially distinct across the state (Figure 2-3) with set zones of different agricultural growth regions, reflecting the climatic and biotic conditions across the state.

Table 2-2. Validation Confusion Matrix for the Random Forest Classification for 2008 imagery

		Cotton	Citrus	Reference Strawberries	Peanuts	Sugarcane	Total	Producer Accuracy	User Accuracy
Classified	Cotton	7	1	0	3	0	11	0.778	0.636
	Citrus	0	28	0	0	0	28	0.875	1.0
	Strawberries	0	0	11	1	0	13	0.917	0.846
	Peanuts	2	0	1	19	0	23	0.826	0.826
	Sugarcane	0	0	0	0	5	5	1.0	1.0
	Total	9	32	12	23	5	167	Overall Accuracy	86.5%

Table 2-3. Validation Confusion Matrix for the Random Forest Classification for 2016 imagery

		Cotton	Citrus	Reference Strawberries	Peanuts	Sugarcane	Total	Producer Accuracy	User Accuracy
Classified	Cotton	1	0	0	1	1	3	1.0	0.333
	Citrus	0	22	0	1	1	24	0.917	0.917
	Strawberries	0	1	6	1	0	9	0.857	0.667
	Peanuts	0	0	0	12	0	12	0.8	1.0
	Sugarcane	0	0	0	0	33	34	0.868	0.971
	Total	1	24	7	15	38	145	Overall Accuracy	89.9%

When looking at change in crop area over the two dates of analysis the overall, total acreage for the five crops has been stable, with only a relatively small decrease (4.17%), from 5,571.88

km^2 to 5,339.71 km^2, between 2008 and 2016. Citrus, sugarcane, and peanut are the most common crops for the both periods, while cotton and strawberry crops accounted for a small portion of the total acreage. The acreage change for cotton, citrus, and strawberry were substantial. Citrus acreage decreased by 12.43% from 3,016.51 km^2 to 2,641.69 km^2. Cotton acreage increased by 50.19% from 236.12 km^2 to 354.63 km^2. The largest percent of acreage increase was observed for strawberry, from 1.45 km^2 to 23.72 km^2. Sugarcane and peanut crops experienced small acreage changes. Sugarcane acreage increased by 2.15%, from 1,615.03 km^2 to 1,649.79 km^2, while peanut acreage decreased by 4.68%, from 702.74 km^2 to 669.87 km^2.

Table 2-4. Change in crop area from 2008 to 2016 across Florida in km^2

	Cotton	Citrus	Strawberry	Peanut	Sugarcane	Totals
2008	236.12	3016.51	1.45	702.74	1615.03	5571.88
2016	354.63	2641.69	23.72	669.87	1649.79	5339.71
Change (%)	50.19	-12.43	1535.86	-4.68	2.15	-4.17
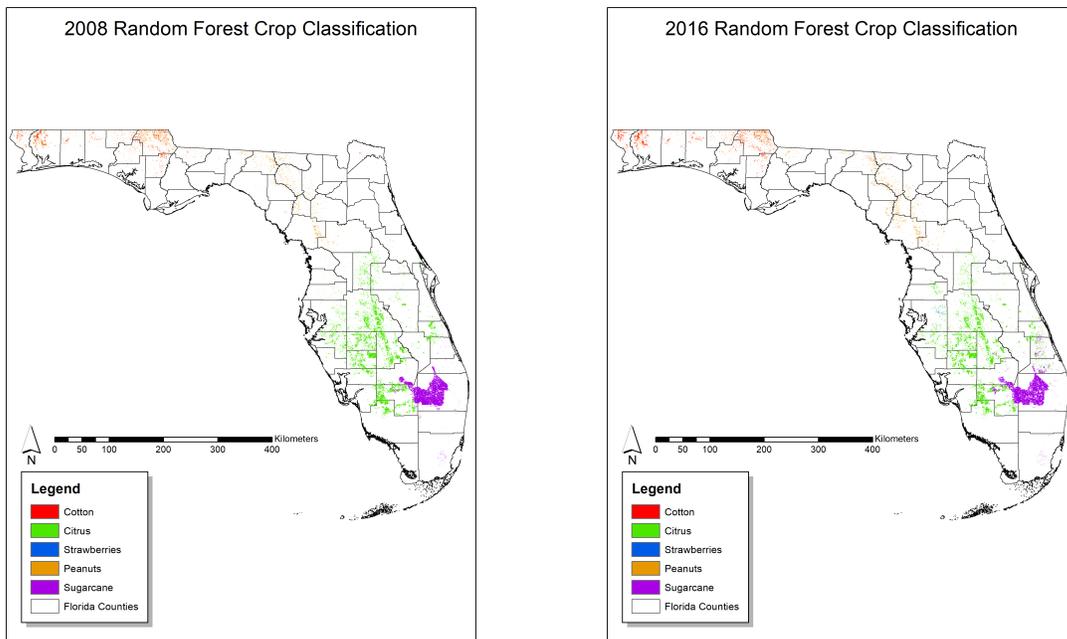
Figure 2-3. Crop Classification from the random forest classification for Florida for 2008 and 2016

2.2.2 USDA's Cropland Data Layer Assessment

To assess the accuracy of the USDA's Cropland Data layer with our classification, we utilized the same 250 sites, 50 sites for each of the five crops, to test the actual crop type based on the field visits and high-resolution imagery, to the CDL product. From the 250 sites, a total of 87 sites were accurately classified by the USDA CDL for a 34.8% overall accuracy. The same 250 sites were used for 2008 and verified through imagery. The accuracy for 2008 was 32.4%.

2008 Cropland Data Layer

Legend

Florida Counties

Cropland DataLayer 2008

Class Name

- Cotton
- Citrus
- Peanuts
- Strawberries
- Sugarcane

2016 Cropland Data Layer

Legend

Florida Counties

Cropland Data Layer 2016

Class Name

- Cotton
- Citrus
- Peanuts
- Strawberries
- Sugarcane
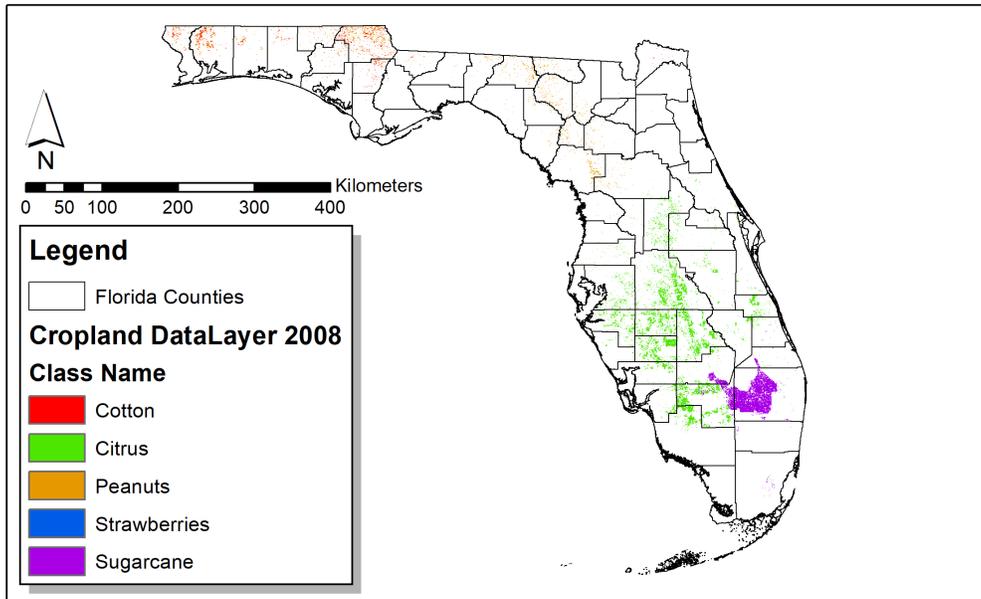
Figure 2-4. The USDA Cropland Data Layer classification for the five key crops in Florida for 2008 and 2016

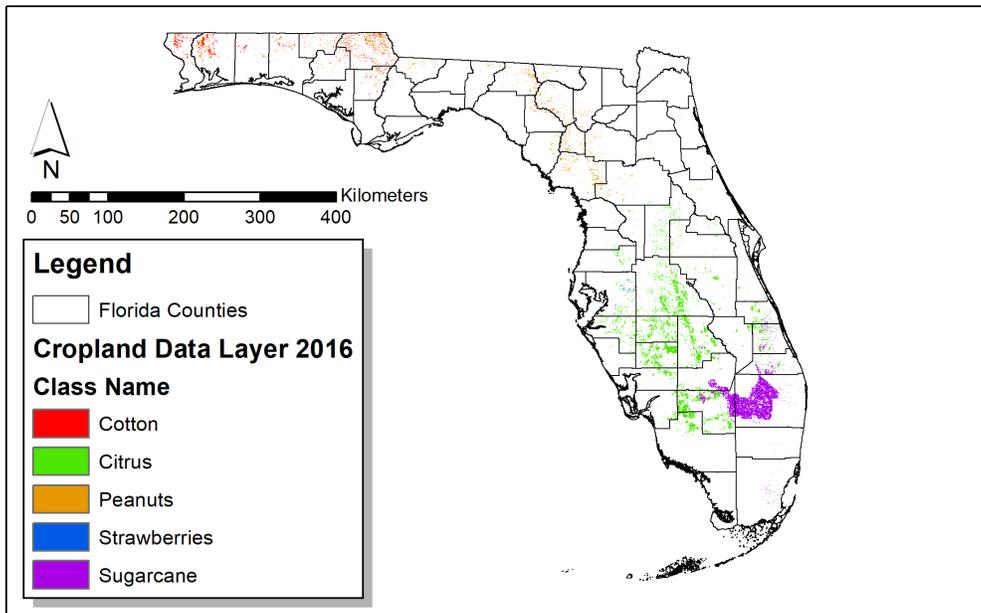### 2.2.3 Comparison of RF classification, CDL and NASS Statistics

We compared the two raster based spatial products, along with the USDA NASS data which is the adjusted CDL estimate. There were differences between the CDL and official NASS estimates for acreage statistics because CDL estimates depend upon pixel counting. Pixel counting is usually downward biased when compared to the official estimates. The bias is corrected via regression methods that are applied and generates the official NASS data. As such, in this analysis we are comparing the CDL raw data, the NASS official statistics and the RF classification produced in our research. In this way we hope to account for any possible error in the CDL product by also including the regression corrected NASS data. Crop acreage coverages from the random forest classifications are compared with both the USDA's CDL classification, and the NASS' official acreage counts for the specified crop types in Florida (Table 2-5). When viewing the data (Table 2-5), with the exception of strawberries, which are massively underestimated by both CDL and the RF classifiers in 2008, the RF classifier produces values much closer to the official NASS statistics than the CDL. Given the NASS data are reportedly based on some regression adjustment of the CDL data this result seems most interesting. Our classifier significantly outperforms the CDL both in terms of the NASS data and in terms of validation. If we select a couple of crop examples to highlight the differences in performance of the CDL versus the RF classifier (Figure 2-5) we can clearly see a higher quality and more spatially refined product from the RF classifier than the CDL product.

Table 2-5. Acreage counts for official NASS estimates, Cropland Data Layer classification estimates, and random forest classification estimates in km^2.

2008 Acreage	Cotton	Citrus	Strawberry	Peanut	Sugarcane	Totals
NASS	263.04	1877.33	27.92	566.55	1622.78	4357.65
CDL	271.74	3809.28	2.05	904.37	2027.90	7015.36
RF Classifier	236.12	3016.51	1.45	702.74	1615.03	5571.88
2016 Acreage	Cotton	Citrus	Strawberry	Peanut	Sugarcane	Totals
NASS	335.88	1566.13	44.11	728.43	1655.16	4329.73
CDL	396.97	3640.95	37.10	893.54	1936.95	6905.54
RF Classifier	354.63	2641.69	23.72	669.87	1649.79	5339.71

Citrus

CDL 2016   New Classification   High Resolution Imagery

Sugarcane

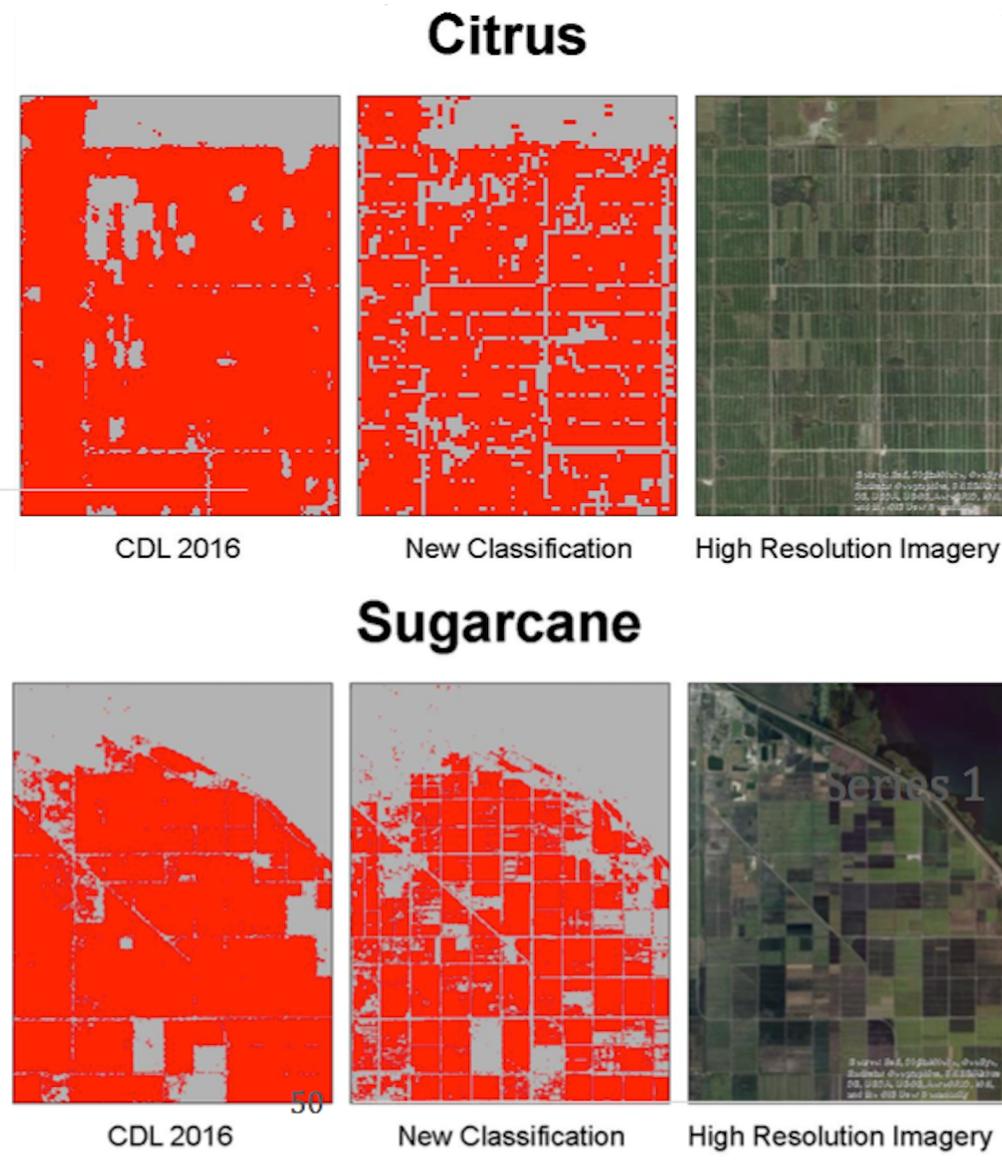CDL 2016   New Classification   High Resolution Imagery

Figure 2-5. Comparison of the CDL and RF classification products for two known field sites of cotton and sugarcane.

2.2.4 Trajctories of crop change from 2008 to 2016 from the RF classifier

Spatially, there were limited areas of crop changes across the state, totaling less than 4% overall in terms of loss of crop coverages. When we view at a pixel level though we can also identify areas that stayed cropped but had a different crop grown or a rotation in crop type. While there may have been a 5% loss in crop area between 2008 and 2016, there was also a conversion of crop types. For all crops, the majority of area was stable. The largest changes were between peanuts and cotton with 123.3 km^2 of change from peanuts to cotton and 68.8 km^2 of change from

cotton to peanuts. The next large changes were observed for change from citrus to sugarcane, for 23.9 km^2, and strawberry, for 6.3 km^2. The land cover trajectories for the state are given in Table 2-6, with all areas of conversion, totaling at least 1.0 km^2 highlighted. Examples highlighting these conversion areas for three sample sites are illustrated visually in Figure 2-6.

Table 2-6. Land cover trajectory classes for the five main crops in Florida from 2008 to 2016. Only crop conversions are shown, and for areas of 1 km^2 or above.

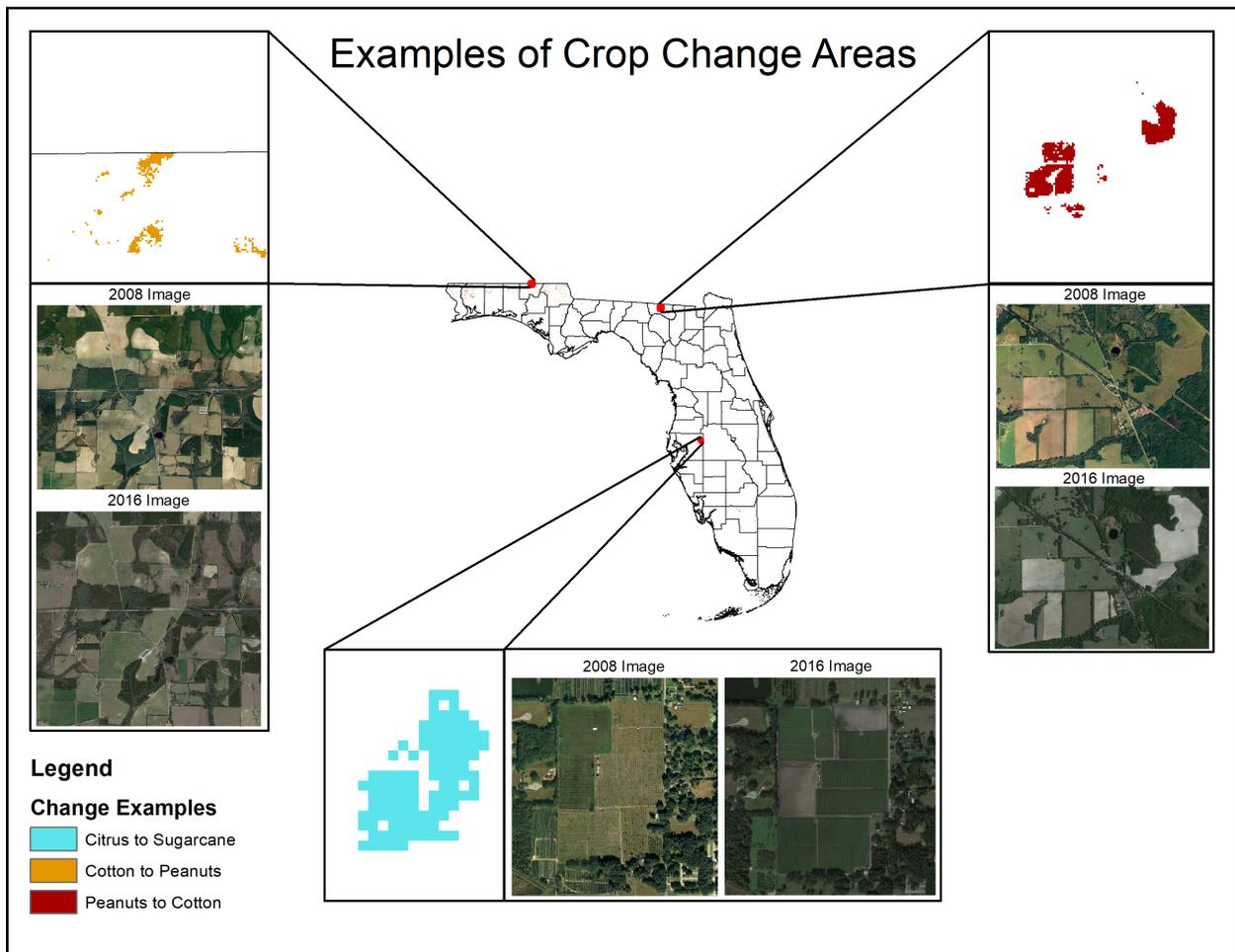| Crop conversion from 2008 to 2016	Area (km^2)
Peanuts to Cotton	123.35
Cotton to Peanuts	68.77
Citrus to Sugarcane	23.86
Citrus to Strawberries	6.31
Sugarcane to Citrus	3.07

Figure 2-6. Examples of crop change trajectories identified by the RF Classifier for three different sites between 2008 to 2016.

2.3 Discussion

The extent and spatial organization of crops and crop types is of paramount importance in terms of global food production and related areas such as biodiversity, management of pests and diseases within agricultural landscapes and also linked to human health and water quality issues [16]. Geospatial data, especially satellite images, holds the greatest promise for the provision of detailed information on agricultural extent, crop type, and health. Such in-season crop identification is essential for management-based decisions [24]. This study found low accuracy of preexisting national level agricultural data, the USDA CDL product, such as these pixel-level data. The results, when coupled with earlier studies indicate a need for caution given their suite of limitations or possible inaccuracies[15]. The random forest classifications yielded 85% overall accuracy for both 2008 and 2016 years, and outperformed the USDA Cropland Data Layer classification with its accuracy of less than 40%. Crop coverage across the state still shows citrus to be the number one crop in terms of area, followed by sugarcane, peanut, cotton and then strawberry. Across the state approximately 4% of the area in agriculture was lost across the study period and the greatest losses in overall crop area were found for citrus (-12%) and peanuts (-5%). Strawberry increased significantly as a percentage, in part due to the, initial, very low area in this crop, and cotton also increased significantly in area. The RF classification was found to be a strong performer in spatial accuracy when compared to actual field visits, and also was closer to the USDA NASS adjusted numbers statewide, than the CDL product on which the NASS numbers are based. Such low total accuracy for the CDL (30-50%) makes this product unsuitable for use for looking at these crops within our study region, especially in terms of addressing changes in these crop areas over time. The very low levels of accuracy for this product make it unusable in any further studies which needs accuracy and spatial location information and make the creation of real crop classification essential for this state. As such, this new RF classification of the top 5 crops for the state of Florida, was found to be a superior product for the study of crop change and also for use in additional studies based on the location and quantity of crops in Florida. While the CDL holds value when considering agriculture crop coverage at a national scale, especially when

28

considering its high temporal frequency, in that it is completed annually for every state [19, 10]. In this example, it does not hold up at the state scale when considering performance of classifiers and accuracy. The USDA CDL classifications do have a series of known errors but the very low accuracy levels indicated here are well below those reported by other researchers. Reitsma et al (2016) investigated CDL cropland accuracy across South Dakota and found highly variable results over space, with overall accuracy levels between 89.2 and 42.6%. These ranges are variable, but the lower end of accuracy does match our findings within Florida for the five selected crops. Similarly, Larsen et al (2015) found that accuracy was highly variable with over or underestimates of +/- 50% possible. Lark et al (2017) cautioned in their research, the use of CDL data for tracking crop changes over time. Their research found that special considerations were necessary when using the data in this way, specifically they recommended a need to account for crop area underestimation within this dataset. Particularly, the known underestimation of crop area if pixel counts are used from the CDL data directly [4], when compared with data such as the NASS, is such that there is a linear regression model built to correct for this error [3]. While the CDL is clearly a useful and unique dataset, available at an annual timestep, and highly beneficial to many research studies, it can also be problematic and should always be validated with in-situ field data prior to being used for land change studies [24]. Other researchers addressing issues of crop change and comparing to the CDL data set have also found the random forest classifier to be a superior technique for classification accuracies [24] in comparison to the CDL direct output. In addition to the use of random forest classifiers these researchers also compared the use of single date versus multi-date imagery for crop classification and reported multi-date image incorporation significantly improved the classification accuracy. Incorporating multi-date images across the season within this research, in a single stacked image layer for classification, was clearly superior to a single image date, for agricultural classifications with highly variable coverage within a season. For land managers and those researchers interested in modeling crop location and area, care must be taken when using national data products such as the CDL. In addition, when interest is focused on only a few crops, it is evident that site specific classifications require extensive field

29

calibration. Use of the CDL products are on the increase [15, 9] making research on their accuracy and site-specific applicability also of key importance to researchers. The work discussed here highlights the importance of accurate crop classifications for Florida where declines in the citrus industry and areas under citrus coverage are declining, and different crops, such as cotton, are increasing in area. Such changes, especially when Florida represents the number one producer nationally for Citrus, are of great importance economically to the state and region. The implications of changes in crop distribution and amount, have real impacts on natural resources statewide, with such agricultural regions and their related management practices which are crop dependent, impacting our water supply, air quality, climate, and wildlife habitats.

CHAPTER 3
CONCLUSIONS

RF enables the classification of the top agricultural crops within the state of Florida to a higher degree of accuracy and permits the location and quantification of change that occurred between t the times of interest. Quantifying this high degree of accuracy and confidence in the results is paramount to studies that utilize any type of data of this sort. The application of random forest for classification of remotely sensed images can be applied at a multiscale level. While this study looks at agricultural change at a state wide level, it could easily be applied at a county, national, or even global level with similar results. Understanding real and actual change can help professionals in a multitude of fields. Land managers, researchers, economist, modelers, and agronomists all rely on this type of data to make impactful decisions. Identifying and displaying this data in a spatial context, via mapping, helps us further understand the interaction humans are continually having with their environment. Identifying change between the two time periods of our study also helps us understand trends of change that are happening within agriculture, since it is such an important facet of the human and environment interaction that has, and continues to be, the forefront of land use. From the results of this research, the random forest classifier proved to be the superior classifier. This classifier boasted accuracy numbers above 85%. This of course was a significant improvement upon the accuracy numbers we were able to ascertain from the CDL data. Through our random forest classification, we were able to determine specific change numbers for each of the five crops of interest. Between 2008 and 2016 there was a total 4.17% crop area loss, citrus coverage decreased from 3,017 km^2 to 2,642 km^2, cotton coverages increased from 236 km^2 to 355 km^2, strawberry coverage increased from 1 km^2 to 24 km^2, sugarcane coverage increased from 1,615 km^2 to 1,650 km^2, and peanut coverage decreased from 703 km^2 to 670 km^2. We were also able to determine the trajectory of crop change between the two years. The largest change was peanuts to cotton with a total of 123 km^2, followed by change from cotton to peanuts with 69 km^2. The next largest changes were from citrus to sugarcane with 24 km^2, citrus to strawberry with 6 km^2, and sugarcane to citrus with 3 km^2. The rest of the crop changes total less than 1 km^2. These results help us understand agriculture trends within the state of Florida. This spatial data can also easily be fed into modeling algorithms to determine a

number of applications in which accurate land cover maps would improve the precision of the results or estimates from the model. A few limitations this study faced was inaccessibility to the CDL methods and training data. With a greater understanding of the exact methods used in the creation of CDL, it would have been easier to improve upon and test this created product. Since the CDL is an annually produced product, the methods are constantly being updated and changed. However, at the time of this research, we were unable to receive the requested information. With more time, a number of improvements could be made to this study. The inclusion of more crop types and land cover types, with more training data could result in even higher accuracy for the random forest classification. Of course, an additional improvement for this study would be to add more training data for the classification. Supplementing the classification with additional training data would likely improve the classification and could improve the results in terms of accuracy. Adding additional image dates within the growing season years of interest could also result in an improvement. A final possible improvement would be using a different type of classifier, such as an ensemble classification. One of the biggest strengths of an ensemble classification method is the use of a multitude of weighted classifications which vote to create a combined classification that out performs each of the components. Ensemble classifications leverage the different strengths of multiple classifications to overcome the weakness of utilizing a single classification method [28]. Some future directions this study could lead to are, as previously mentioned, the inclusion of this analysis into a suite of different modeling analyses. As an example, this spatial data could be paired with meteorological data and fertilization best practices which are stated by crop type, to estimate total amount of fertilizer use and location across the state of Florida. This data could lead to explaining eutrophication of fresh water springs within the state or help explain and understand chronic algal blooms that have been happening here. The research method of using the random forest classifier can also be applied to a wide variety studies that seek to classify spatial data. This could be any study that seeks to classify any type of land-use/land-cover. Habitat mapping is a common use of image classification, and is not limited to just land bound areas, since habitat classification can also be used on benthic aquatic habitats. The scope of this

study could also be expanded to a national level as well. This expanded scope would build on more land-use/land-cover types, but could definitely give a greater insight to the highly complex and heterogeneous landscape. Finally, this research undertaken here is of importance within the field of Geography, as both the geospatial analysis tools utilized here, and the subject matter addressed, are both keys areas of study within this field. Geographers study the spatial location, patterns and change of different phenomena of importance within human and environmental studies. Agriculture, as previously stated, is one of the most impactful and the longest occurring changes to the surface of the earth, and as such our increased understanding and monitoring of such changes is of real concern within this field. This research analyzed the changes in agricultural coverage across a set time period within the state of Florida, looking at the top 5 crops in terms of economic importance for our state. In addition, we applied remotely sensed analysis techniques to analyze this change and compared our results to those produced nationally and found the national data products to be severely lacking. This research highlights the need for field based geographical research, and the utilization of geospatial analysis tools and advanced remotely sensed methodologies to better understand spatial variation in such important cover types. The tools and approaches geographers can bring to such studies is of paramount importance and such research should be continued and built upon in future work.

REFERENCES

[1] BECK, H. E., MCVICAR, T. R., VAN DIJK, A. I., SCHELLEKENS, J., DE JEU, R. A., AND BRUIJNZEEL, L. A. Global evaluation of four avhrr–ndvi data sets: Intercomparison and assessment against landsat imagery. *Remote Sensing of Environment 115*, 10 (2011), 2547–2563.

[2] BORISS, H. Commodity profile: Citrus. http://www.scribd.com/doc/322244/citrus-2006, 2006. Accessed: 2018-05-26.

[3] BORYAN, C., YANG, Z., MUELLER, R., AND CRAIG, M. Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International 26*, 5 (2011), 341–358.

[4] BORYAN, C. G., YANG, Z., DI, L., AND HUNT, K. A new automatic stratification method for us agricultural area sampling frame construction based on the cropland data layer. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7*, 11 (2014), 4317–4327.

[5] BREIMAN, L. Random forests. *Machine learning 45*, 1 (2001), 5–32.

[6] FLORIDA DEPARTMENT OF AGRICULTURE AND CONSUMER SERVICES. Florida agriculture overview and statistics. https://www.fdacs.gov/Agriculture-Industry/Florida-Agriculture-Overview-and-Statistics, 2017. Accessed: 2018-07-17.

[7] FOLEY, J. A., DEFRIES, R., ASNER, G. P., BARFORD, C., BONAN, G., CARPENTER, S. R., CHAPIN, F. S., COE, M. T., DAILY, G. C., GIBBS, H. K., ET AL. Global consequences of land use. *science 309*, 5734 (2005), 570–574.

[8] GUTMAN, G., JANETOS, A. C., JUSTICE, C. O., MORAN, E. F., MUSTARD, J. F., RINDFUSS, R. R., SKOLE, D., TURNER II, B. L., AND COCHRANE, M. A. *Land change science: observing, monitoring and understanding trajectories of change on the earth's surface*, vol. 6. Springer Science & Business Media, 2004.

[9] HAN, W., YANG, Z., DI, L., AND MUELLER, R. Cropscape: A web service based application for exploring and disseminating us conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture 84* (2012), 111–123.

[10] HAN, W., YANG, Z., DI, L., YAGCI, A. L., AND HAN, S. Making cropland data layer data accessible and actionable in gis education. *Journal of Geography 113*, 3 (2014), 129–138.

[11] HASAN, R. C., IERODIACONOU, D., AND MONK, J. Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sensing 4*, 11 (2012), 3427–3443.

[12] HAZELL, P., AND WOOD, S. Drivers of change in global agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences 363*, 1491 (2008), 495–515.

[13] HERRERO, H. V., SOUTHWORTH, J., AND BUNTING, E. Utilizing multiple lines of evidence to determine landscape degradation within protected area landscapes: A case study of chobe national park, botswana from 1982 to 2011. *Remote Sensing 8*, 8 (2016), 623.

[14] JOHNSON, D. M., MUELLER, R., ET AL. The 2009 cropland data layer. *Photogramm. Eng. Remote Sens 76*, 11 (2010), 1201–1205.

[15] LARK, T. J., MUELLER, R. M., JOHNSON, D. M., AND GIBBS, H. K. Measuring land-use and land-cover change using the us department of agriculture's cropland data layer: Cautions and recommendations. *International journal of applied earth observation and geoinformation 62* (2017), 224–235.

[16] LARSEN, A. E., HENDRICKSON, B. T., DEDEIC, N., AND MACDONALD, A. J. Taken as a given: Evaluating the accuracy of remotely sensed crop data in the usa. *Agricultural Systems 141* (2015), 121–125.

[17] LEV-YADUN, S., GOPHER, A., AND ABBO, S. The cradle of agriculture. *Science 288*, 5471 (2000), 1602–1603.

[18] MILLENNIUM ECOSYSTEM ASSESSMENT, M. *Ecosystems and human well-being*, vol. 5. Island Press Washington, DC, 2005.

[19] MUELLER, R., BORYAN, C., AND SEFFRIN, R. Data partnership synergy: The cropland data layer. In *2009 17th International Conference on Geoinformatics* (2009), IEEE, pp. 1–6.

[20] NATIONAL AGRICULTURAL STATISTICS SERVICE. About cdl. https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php, 2017. Accessed: 2017-03-13.

[21] NATIONAL AGRICULTURAL STATISTICS SERVICE. Cropscape – cropland data layer. https://nassgeodata.gmu.edu/CropScape/, 2017. Accessed: 2017-03-10.

[22] NATIONAL AGRICULTURAL STATISTICS SERVICE. Quickstats page. https://quickstats.nass.usda.gov/, 2017. Accessed: 2017-04-15.

[23] NEMANI, R. R., RUNNING, S. W., PIELKE, R. A., AND CHASE, T. N. Global vegetation cover changes from coarse resolution satellite data. *Journal of Geophysical Research: Atmospheres 101*, D3 (1996), 7157–7162.

[24] RAHMAN, M., DI, L., YU, E., ZHANG, C., MOHIUDDIN, H., ET AL. In-season major crop-type identification for us cropland from landsat images using crop-rotation pattern and progressive data classification. *Agriculture 9*, 1 (2019), 17.

[25] RAMANKUTTY, N., EVAN, A. T., MONFREDA, C., AND FOLEY, J. A. Farming the planet: 1. geographic distribution of global agricultural lands in the year 2000. *Global biogeochemical cycles 22*, 1 (2008).

[26] REITSMA, K. D., CLAY, D. E., CLAY, S. A., DUNN, B. H., AND REESE, C. Does the us cropland data layer provide an accurate benchmark for land-use change estimates? *Agronomy Journal 108*, 1 (2016), 266–272.

[27] REITSMA, K. D., DUNN, B. H., MISHRA, U., CLAY, S. A., DESUTTER, T., AND CLAY, D. E. Land-use change impact on soil sustainability in a climate and vegetation transition zone. *Agronomy Journal 107*, 6 (2015), 2363–2372.

[28] REN, Y., ZHANG, L., AND SUGANTHAN, P. N. Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational intelligence magazine 11*, 1 (2016), 41–53.

[29] ROUSE, J. W., HAAS, R. H., SCHELL, J. A., DEERING, D. W., AND HARLAN, J. C. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. *NASA/GSFC Type III Final Report, Greenbelt, Md 371* (1974).

[30] STEVEN, M., AND CLARK, J. A. *Applications of remote sensing in agriculture*. Elsevier, 2013.

[31] TILMAN, D., FARGIONE, J., WOLFF, B., D'ANTONIO, C., DOBSON, A., HOWARTH, R., SCHINDLER, D., SCHLESINGER, W. H., SIMBERLOFF, D., AND SWACKHAMER, D. Forecasting agriculturally driven global environmental change. *science 292*, 5515 (2001), 281–284.

[32] TUAN, Y.-F. A view of geography. *Geographical review* (1991), 99–107.

[33] TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment 8*, 2 (1979), 127–150.

[34] TURNER, B. L., LAMBIN, E. F., AND REENBERG, A. The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences 104*, 52 (2007), 20666–20671.

[35] UNITED STATES DEPARTMENT OF AGRICULTURE/NATIONAL AGRICULTURAL STATISTICS SERVICE. State agriculture overview. https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=FLORIDA, 2016. Accessed: 2018-05-05.

[36] UNIVERSITY OF FLORIDA IFAS. Floirda agricultural facts. Presentation, 2016. Accessed: 2017-06-12.

[37] VITOUSEK, P. M., MOONEY, H. A., LUBCHENCO, J., AND MELILLO, J. M. Human domination of earth's ecosystems. *Science 277*, 5325 (1997), 494–499.

BIOGRAPHICAL SKETCH

Tyler Schaper was born in North Florida and spent most of his childhood moving overseas with is family as part of the foreign service. He learned his love of geography from his interactions overseas and his return trips back home in Florida. In 2014, he graduated with a Bachelor of Arts degree in Geography from the University of Florida. He earned his Master of Science in Geography in 2021 from the university of Florida. He currently lives in Dallas, Texas and enjoys spending time with his friends, family, and animals.