

University of Arkansas, Fayetteville

ScholarWorks@UARK

Graduate Theses and Dissertations

5-2018

Towards Quantifying Relevant Land Cover Change: A Case Study in the Central Flyway of the Monarch Butterfly

Hanna Lenée Ford

University of Arkansas, Fayetteville

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Animal Studies Commons](#), [Geographic Information Sciences Commons](#), [Natural Resources and Conservation Commons](#), and the [Zoology Commons](#)

Citation

Ford, H. L. (2018). Towards Quantifying Relevant Land Cover Change: A Case Study in the Central Flyway of the Monarch Butterfly. *Graduate Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/2678>

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, uarepos@uark.edu.

Towards Quantifying Relevant Land Cover Change:
A Case Study in the Central Flyway of the Monarch Butterfly

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Geography

by

Hanna L. Ford
University of Arkansas at Little Rock
Bachelor of Arts in Liberal Arts, 2010

May 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Jackson Cothren, Ph.D.
Thesis Chair

W. Fred Limp, Ph.D.
Committee Member

Marty Matlock, Ph.D.
Committee Member

Kusum Naithani, Ph.D.
Committee Member

Abstract

Over the past 20 years notable decreases in monarch butterfly populations have led researchers to begin evaluating the landscape for changes and seeking out opportunities for enacting conservation programs to better support their survival. The monarch butterfly has recently come under consideration for listing under the Endangered Species Act which has created a need for a more informed view of the landscape through which the migrate and breed, the central United States. In this research three spatially-explicit models are explored using the most applicable datasets currently available to address pressing policy and land manager decisions regarding monarch butterfly and pollinator conservation. Using the Cropland Data Layer (CDL) and National Land Cover Data (NLCD) datasets, individually and combined, the ability to to evaluate landscape change, annual and decadal, from 2008 – 2017 is evaluated. The CDL and NLCD both present unique data integration challenges for reliably estimating land use change on an annual basis for all land cover types, and for augmenting additional feature data, such as soil productivity and transportation networks that represent valuable target areas for monarch and pollinator research. The result of these spatially-explicit model trials are a more informed process for quantifying uncertainty and moving toward thoughtful inclusion of CDL data in annual change metrics that identifies land conversion for a broad number of categories, including grassland/pasture. The results of these models begin to identify a more consistent and transferrable process for addressing policy and land manager decisions regarding monarch butterfly and pollinator conservation delivery.

©2018 by Hanna L Ford
All Rights Reserved

Acknowledgments

This material is based upon work supported by BASF Corporation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the BASF Corporation, though funding and support is gratefully appreciated. Some materials are also based on previous work by the USGS Monarch

Conservation Science Partnership. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the USGS Monarch Conservation Science Partnership, though the sharing of models and information is appreciated.

Table of Contents

1.	Introduction.....	1
2.	Background.....	1
2.1	Recent Policy Influencing Research Mandates and Policy Development for the Monarch Butterfly.....	1
2.2	Previous Research involving the Monarch Butterfly and Other Pollinators	4
2.3	Current Spatially-Informed Models Influencing Habitat Objectives and Target Locations for Conservation Action.....	5
3.	Enhanced Spatial Decision Models Supporting Monarch and Pollinator Success	12
4.	Methods.....	15
4.1	Study Area	15
4.2	Defining Monarch Butterfly Habitat	16
4.3	Model Structure and Components	17
4.3.1	Data Preparation Models and Processes.....	19
4.3.2	Land Cover Augmentation and Enhancement Models	23
4.4	Detecting Change across the Landscape Over the Past Decade	28
5.	Results.....	31
6.	Discussion	51
	References.....	55
	Appendix.....	59
1.	Data Layers Specifics Required for Models	59
2.	List of all Categories in CDL, original and enhanced	61
3.	Plots by Category	66
3.1	Model 1 – Z-Score Plot by Year, 2008 -2017.....	66
3.2	Model 2 – Z-Score Plot by Year, 2008 -2017.....	67
3.3	Model 3 – Z-Score Plot by Year, 2008 -2017.....	68
3.4	Model 1 – 2017 Frequency and Confidence Plots by Category	69
3.5	Model 2 – 2017 Frequency and Confidence Plots by Category	70
3.6	Model 3 – 2017 Frequency and Confidence Plots by Category	71
4.	R script for Exploratory Data Analysis.....	72

List of Tables

Table 1:	Comparison of model structure between Koh et al. 2016 and Thogmartin et al. 2017.	7
Table 2:	Comparison of land use change methods used in Decision Innovative Solutions 2013 and Wright and Wimberly 2013. *Grass/Hay is not listed as a category within the CDL. However, may be representative of other categories collapsed by the authors without definition.	10
Table 3:	Comparison of results for conversion of grass-land cover types to corn and/or soy crop cover types in South Dakota.	12
Table 4:	Descriptive text for productivity assigned to NCCPI values by commodity crop type.	19
Table 5:	Mean Standard Deviation, Mean Acres, and Coefficient of Variation for Model 1, 2008 - 2017 points with a standard deviation of greater than 2.5 million as highlighted in red in Figure 9.	33
Table 6:	Mean Standard Deviation, Mean Acres, and Coefficient of Variation for Model 2, 2008 - 2017 points with a standard deviation of greater than 2.5 million as highlighted in red in Figure 9.	34
Table 7:	Mean Standard Deviation, Mean Acres, and Coefficient of Variation for Model 3, 2008 - 2017 points with a standard deviation of greater than 2.5 million as highlighted in red in Figure 9.	35
Table 8:	Comparison of Top 10 (by Area) Categories of Change from 2008 to 2017, by Model	36
Table 9:	Grassland/Pasture gains and losses for 2008 – 2017, model 1 output.	43
Table 10:	Table of all CDL original and enhanced versions of data.	61

List of Figures

Figure 1: Number of publications, by decade and publication type, since 1940 returned in citation index searches.	5
Figure 2: Reorganization of grass-land cover types within CDL cover types.	9
Figure 3: Map of central United States with study area boundary displayed for this research. .	16
Figure 4: General flow diagram of the three versions of the models created for this research. .	18
Figure 5: Distribution of NCCPI values by acreage, nationwide and within the study area of this research, by value range data breaks indicated in Table 4, including separation at the 0.400 breakpoint as identified in Thogmartin et al. 2017 for illustration purposes; however, for the purposes of the model in this research the value ranges of 0.226 – 0.400 and 0.441 – 0.475 are combined.	21
Figure 6: Process for delineating portions of grass-land cover types as protected using CDL Code 176, Grassland/Pasture as example.	25
Figure 7: Process for delineating portions of grass-land cover types integrating NLCD data, and protected areas database layers, illustrated using Grassland/Pasture as example.....	26
Figure 8: Process for delineating commodity crops based on NCCPI values, illustrated using corn as an example.....	28
Figure 9: Model 1, 2008 – 2017 Plot by Category and Coefficient of Variation. Data for highlighted points (red); points with a standard deviation of greater than 2,500,000 are highlighted and displayed in Table 5. Note the difference in scale between Model 1 and Models 2 and 3.	33
Figure 10: Model 2, 2008 – 2017 Plot by Category and Coefficient of Variation. Data for highlighted points (red); points with a standard deviation of greater than 2,500,000 are highlighted and displayed in Table 6. Note the difference in scale between Model 1 and Models 2 and 3.	34
Figure 11: Model 3, 2008 – 2017 Plot by Category and Coefficient of Variation. Data for highlighted points (red); points with a standard deviation of greater than 2,500,000 are highlighted and displayed in Table 7. Note the difference in scale between Model 1 and Models 2 and 3.	35
Figure 12: Annual z-scores for 2008-2017 model 1 grassland/pasture category.	37
Figure 13: Annual z-scores for 2008-2017 model 2 grassland/pasture categories.....	37
Figure 14: Annual z-scores for 2008-2017 model 3 grassland/pasture categories. Note the very abrupt changed in z-scores where notated as NLCD-derived, these mark the change in where years reference 2006 NLCD and 2011 NLCD.....	38
Figure 15: Model 1 distribution of acres of grassland/pasture by confidence and frequency for 2017. Data greater than 4 million acres have been omitted from plots to illustrate variation (4 records omitted).....	39
Figure 16: Model 2 distribution of acres of grassland/pasture by confidence and frequency for 2017. Data greater than 4 million acres have been omitted from plots to illustrate variation (4 records omitted).....	40

Figure 17: Model 3 distribution of acres of grassland/pasture by confidence and frequency for 2017. Data greater than 4 million acres have been omitted from plots to illustrate variation (2 records omitted).....	40
Figure 18: Conversion of lands in grassland/pasture category, 2008 -2017 (model 1).....	42
Figure 19: Grassland Pasture Conversion (Model 2 Output) by Pixel Frequency, 2008 - 2017..	47
Figure 20: Grassland Pasture Conversion (Model 2 Output) by Pixel Confidence, 2008 - 2017.	47
Figure 21: Conversion of grassland/pasture to low and marginal corn and soybean production lands between 2008 and 2017 (model 2).	49
Figure 22: Conversion of grassland/pasture any corn or soybean production lands between 2008 and 2017 (model 2).	50
Figure 23: Conversion of grassland to other categories (loss) between 2008 and 2017 as aggregated to HUC12 that intersect the study area and symbolized by number of acres of grasslands converted to other categories using natural breaks in the data.	54

1. Introduction

Determining the amount of existing, and potential, habitat for any species is a difficult task but is an important and necessary part of policy guidance and decision making in present day conservation delivery. Establishing goals and objectives is often driven by the quantification of land cover or land use types on the landscape in contemporary assessment to create baselines, or benchmarks, of conditions that indicate important thresholds for a given species or ecological community. Using these benchmark assessments, conservation delivery plans are designed using a variety of spatial and non-spatial methods to determine what resources are necessary for obtaining desired future conditions to preserve, conserve, or restore these systems. These decision support models place a great deal of importance on the establishment of benchmark conditions, which can be both a strength and weakness. The objective of the models created in this research is to provide a method for more accurately representing benchmark conditions, while also helping to quantify the impact of uncertainty on outcomes and how uncertainties in benchmark calculation methods can inadvertently misguide land managers. The models and examples in this research are developed in support of monarch butterfly research. However, these processes that have been developed and the products of these models have the potential to be utilized beyond monarch butterfly research, and across multiple disciplines.

2. Background

2.1 Recent Policy Influencing Research Mandates and Policy Development for the Monarch Butterfly

Monarch butterflies (*Danaus plexippus*) and other pollinator species have faced many struggles over the past few decades. The monarch butterfly population, specifically, has been estimated to be declining by approximately 0.89 hectares per year at the winter sites in Mexico

(Oberhauser et al. 2017) with steeper declines being estimated since 2008 (Ries, Taron, and Rendón-Salinas 2015). This is a struggle that most citizens would be oblivious to if it were not for the “charismatic” capacity of the monarch butterfly and the sociological concept developed around this insect. This has made it an icon of conservation and subject of countless research, outreach, and seemingly innocuous mentions of their journeys in countless journals, field notes, and works of art (Gustafsson et al. 2015).

A Presidential Memorandum Creating a Federal Strategy to Promote the Health of Honey Bees and Other Pollinators was issued on June 20, 2014 (President Barack Obama 2014). This memorandum established the Pollinator Health Task Force that includes many of the federal agencies governing policies that could be updated to include pollinator-friendly practices. In addition to establishing the Pollinator Task Force, the agencies were tasked with developing: a) a Pollinator Research Action Plan; b) a public education plan; and c) develop recommendations for public-private partnerships (ibid.).

The Presidential Memorandum also called for strategies to be specifically created for native honey bees, managed honey bees, and the monarch butterfly; but also included pollinators at-large in the general scope of strategy development (ibid.). “Pollinators” include many species other than bees and butterflies, such as flies, birds, beetles, bats, ants, and even unexpected ones such as slugs (USDA U.S. Forest Service 2017). Depending on the species of plant and the location where the plant is being grown, pollinator visitation may be exclusively by bee, non-bee, or some combination thereof (Rader et al. 2016), so one might conclude that preservation of all pollinators is prudent.

In August 2014, two organizations and lifelong monarch scientist, Dr. Lincoln Brower, petitioned the U.S. Fish and Wildlife Service (USFWS) to protect the Monarch Butterfly under

the Endangered Species Act (ESA), the USFWS found the petition to warrant a review and initiated a status review of the monarch (Gustafsson et al. 2015, The Center for Biological Diversity 2016b; U.S. Fish and Wildlife Service Midwest Region 2016).

On January 5, 2016, the Center for Biological Diversity and the Center for Center for Food Safety issued a notice of intent to sue the USFWS for failure to reach a 12-month decision on the monarch status review; a settlement was reached later in 2016 that established a deadline of June 30, 2019 as the date by which a status review must be completed for the monarch (The Center for Biological Diversity 2016b; The Center for Biological Diversity 2016a; Maeckle 2016).

The result of these actions increased recognition and public awareness of monarch butterfly decline. Research communities spanning multiple industries and disciplines have responded by identifying many ongoing and compounding threats, including: illegal logging and deforestation threats to overwintering sites in Mexico (Navarrete, Isabel Ramírez, and Pérez-Salicrup 2011; Vidal and Rendón-Salinas 2014), loss of breeding habitat throughout the central United States (referred to as the “milkweed limitation hypothesis”) (Inamine et al. 2016; Pleasants 2016; Thogmartin et al. 2017), climate change impacts on the extent of breeding and overwintering habitat (Inamine et al. 2016; Gustafsson et al. 2015; Lemoine 2015; Sáenz-Romero et al. 2012), episodic weather event impacts of the population at varying stages of migration, breeding, and overwintering (Inamine et al. 2016), implications of agriculturally-enhanced milkweed production over the past century (Gustafsson et al. 2015) and perceived artificial norms of population dynamics that may be largely as a result of anthropogenic activity, and the myriad of socio-economic-political factors contributing adoption rates of conservation

actions in a time where widespread empathy for conservation is competing with numerous, and ever-growing, “greater-good” efforts and initiatives.

The monarch butterfly is not the best pollinator and could even be described as a pest of its host plant, milkweed. However, the monarch has become the symbol of pollinator conservation (Gustafsson et al. 2015, Agrawal 2017). The promotion of monarch butterfly habitat and programs often includes, “and pollinators,” with the general sentiment that many pollinators will benefit from the conservation actions taken on behalf of the dainty charismatic megafauna of the class Insecta.

2.2 Previous Research involving the Monarch Butterfly and Other Pollinators

Since 1948 there have been approximately 1,152 scholarly publications that include Monarch Butterfly (including plural and scientific alternate text) as part of the title, abstract, or as keyword in citation index searches. Publication data were reviewed for duplication, applicability, and reviewed for errors within Zotero reference management software prior to being summarized. The number of publications has been steadily increasing since this time as shown in Figure 1. Publications between January 2014 and March 2018 account for 25% of the total publications recovered through citation index search. This increase may be as a result of increased awareness, or funding opportunities made available after the announcement of the Presidential Memorandum in 2014; or perhaps the growing number of publications have influenced policy to informed action resulting in the issuance in the Presidential Memorandum.

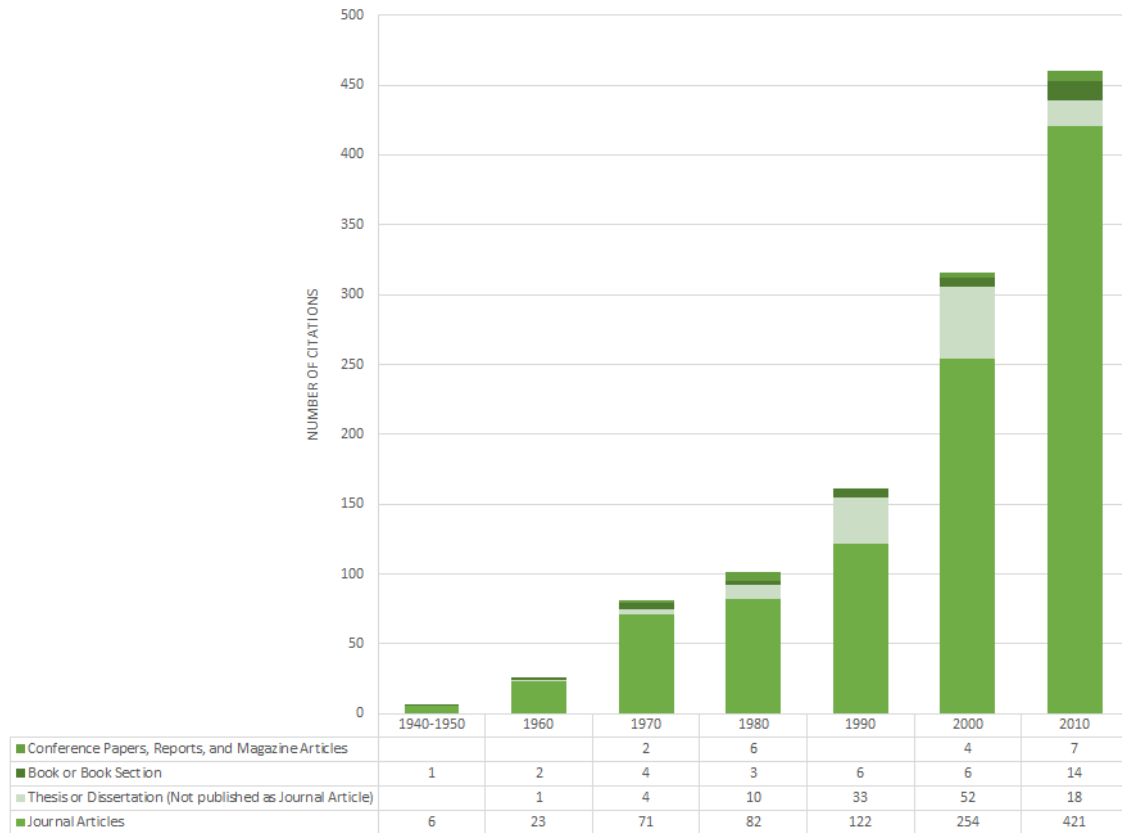


Figure 1: Number of publications, by decade and publication type, since 1940 returned in citation index searches.

2.3 Current Spatially-Informed Models Influencing Habitat Objectives and Target Locations for Conservation Action

The most recent spatially-explicit models to be used for informing conservation delivery with respect to monarchs are those described by Thogmartin et al. 2017. Thogmartin et al. 2017 uses the USDA National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) as the primary base for modeling the extent of lands possibly amenable to milkweed plantings and estimates the possible number of supported stems of milkweeds based off those acreages should plantings be undertaken by landowners. While these estimations are for one year only, there are some calculations over time regarding the amount of acreage possibly affected by more efficient

weed control mechanisms, i.e. glyphosate application, that may have reduced the amount of milkweed available in these areas over the past two decades (Thogmartin et al. 2017).

Another recent model, described by Koh et al. 2016, also uses the CDL as the base for developing the nationwide wild bee estimates of abundance and change from 2008 to 2013. These models found that the conversion of natural habitats to row crops accounted for a 23% decrease in bee abundance within the United States (Koh et. al. 2016).

In both Koh et al. 2016 and Thogmartin et al. 2017 the CDL categories are reclassified from crop cover types to more general categories prior to using data for model input, Table 1 compares model inputs among Koh et al. 2016 and Thogmartin et al. 2017. In Koh et. al. 2016, the crop cover types are reclassified to 32 representative crop categories and 13 non-crop categories, and resampled to 120 x 120 meter pixel resolution; Thogmartin et al. 2017 reclassified crop cover types based on their relevance to milkweed amenability in the upper Midwest region of the monarch butterfly range, resulting in 12 representative crop categories and 30 non-crop categories and retaining a 30 x 30 meter pixel resolution. Thogmartin et al. 2017 further characterized cover type data through introducing and overlaying additional datasets that better described and/or delineated an area of interest regarding milkweed amenability. These additional datasets included: U.S. Geological Survey (USGS) Multi-Resolution Land Characteristics (MRLC) National Land Cover Database (NLCD) 2011; Cropland Reserve Program (CRP) enrollment locations for 2014 (not publicly available); railroads, transmission lines, and road rights of way (as roadways); marginal versus productive farmland (as SSURGO soils data, specifically 2012 National Commodity Cropland Productivity Index); and, characterizations of urban versus exurban environments as determined through generalization processes in ESRI ArcGIS, specifically shrinking and expanding (Thogmartin et al. 2017).

Table 1: Comparison of model structure between Koh et al. 2016 and Thogmartin et al. 2017.

	<i>Koh et al. 2016</i>	<i>Thogmartin et al. 2017</i>
<i>Spatial Resolution</i>	100 x 100 meters	30 x 30 meters
<i>Time Frame</i>	2008 – 2013	2014
<i># Crop Cover Categories</i>	32	12
<i># Non-Crop Cover Categories</i>	13	30
<i>Specifics of Grass-land Categories (i.e. which CDL categories are included within each reclassified grassland category)</i>	Grassland/Pasture Grassland Herbaceous Other Hay/Non Alfalfa Pasture/Hay	Grassland/Pasture Subdivided by overlay with NLCD 2011 categories: Grassland and Pasture/Hay; further subdivided by overlay of Protected Areas Database

In both models the outcome is described as a general guide for directing the future conservation action for their respective concerns, while also recognizing limitations and constraints of existing data. More specifically, that inherent error and accuracy issues that are to be considered when using a national product such as the CDL and can introduce error into land use change calculations over time (Reitsma et al. 2016). Koh et al. 2016 also points out that 1) urban categories are not able to capture the range of pollinator conservation actions that take place within those environments due to spatial scale, and 2) additional classification error and inaccuracy and error in CDL can indicate change where no change is present (Koh et al. 2016, Reitsma et al. 2016). Additionally, in comparing these two datasets the difference in spatial resolution should be noted that can preclude direct comparison. The growing use of CDL data for modeling land use change over time is, in part, because of the temporal frequency of the data. The CDL data is the only nationwide dataset that has an annual release that integrates both crop and non-crop categories. However, the CDL's strength is not found in the ability to model non-crop categories over time. Non-crop categories are spatially informed by the NLCD. For instance, CDL non-crop categories are informed by the most recent version of NLCD released in

2011. Therefore, the CDL data cautions that estimates of change involving non-crop classes are better informed using NLCD directly. However, many recent studies have begun using these data for exactly those purposes (USDA - National Agricultural Statistics Service n.d.). As CDL continues to become an increasingly popular dataset for evaluating land use land cover change for annual or multi-year time frames, in the absence of other annual categorical and spatially appropriate data, methods of how to best utilize these data while minimizing uncertainty and inherent inaccuracy are being developed and vetted with varying and sometimes controversial results.

One popular example of these results are the estimations of grassland conversion. Quantifying grassland conversion is often a difficult task that is made more arduous by similarities, spectrally and categorically, in defining grassland, pastureland, and other grass-land cover types. Within the CDL categories of these varieties that were once separate and now combined in a general non-crop category called “Grass/Pasture” and found under code 176. Figure 2 illustrates how the “Grass/Pasture” category now includes the historical categories: “Pasture/Grass”, “Grassland/Herbaceous”, and “Pasture/Hay.” CDL data were retrofitted and re-processed as far back as 1997, where applicable, to adopt this new categorization scheme for grass-land cover types to reduce redundancy and confusion (USDA - National Agricultural Statistics Service n.d.).

Mueller, R. and M. Harris 2013 reviewed applications of CDL as it was used to between 2006 and 2013, including applications where the primary object was identification of converted lands and/or land change; Reitsma et al. 2016 and Laingen 2015 both highlighted specific examples of how variation in processing, methods, analysis, and interpretation of data can impact results and derivative products when evaluating land use land cover change. Reitsma et al. 2016

used South Dakota as an example in reviewing the CDL-derived spatial data products, along with USDA NASS tabular accuracy products that accompany the release of the annual CDL spatial data product to evaluate producer and user accuracies. The findings of Reitsma et al. 2016 suggest that the temporal variation in cover type accuracies and climatic transition zones may further confound the use of CDL as a surrogate for land change. Laingen 2015 compares change estimates among four popular and readily available datasets: NLCD, CDL, remote sensing of imagery from the National Agriculture Imagery Program, and tabular estimates from the Census of Agriculture. These four datasets evaluated for 2012 South Dakota return values with 6 million acres of difference between the high and low estimates of total cropland. Like Reitsma et al. 2016 and Koh et al. 2016, Laingen 2015 cautions that class selection and data manipulation are important to the outcome of data processing and can lead to unaccounted uncertainty.

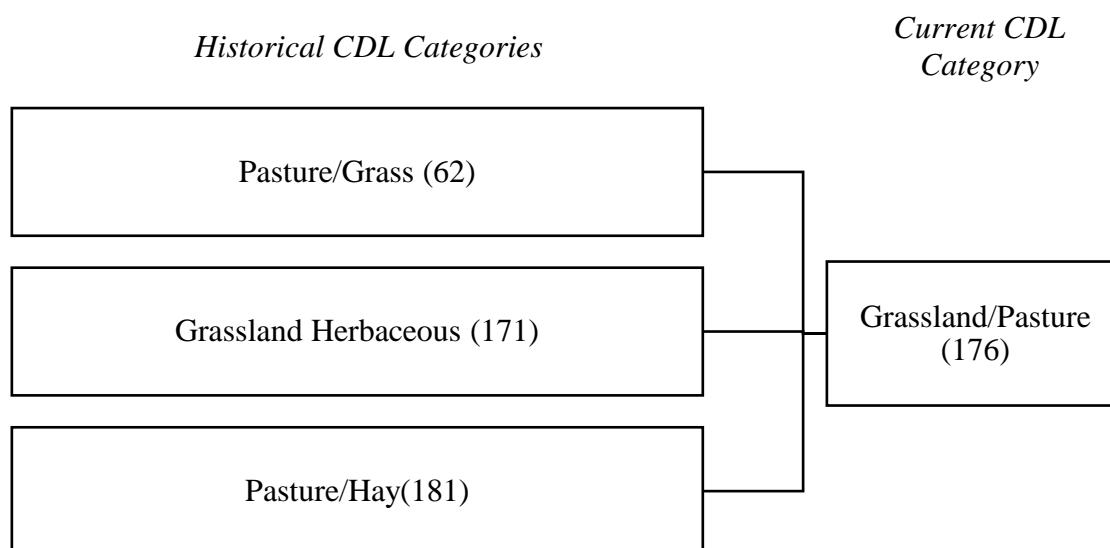


Figure 2: Reorganization of grass-land cover types within CDL cover types.

As an example of how class selection and data manipulation can impact model outcomes, and obscure uncertainty, it is the comparison of the findings of the two publications, Decision

Innovative Solutions 2013 and Wright and Wimberly 2013, that best demonstrate these difficulties. The report from Decision Innovative Solutions 2013 report was completed at the request of the Farm Bureau agencies in seven states in the upper Midwest region of the central United States: These states included: Illinois, Indiana, Iowa, Michigan, Minnesota, Nebraska, and South Dakota. The report focused on land use change from 2007 – 2012 within these states, primarily relying on models that leverage the CDL for annual calculations of land use change. Wright and Wimberly 2013 completed their study as part of the U.S. Department of Energy Sun Grant Initiative’s Regional Biomass Feedstock Partnership with additional funding through a National Science Foundation Grant. The study area in Wright and Wimberly included the following states: Iowa, Minnesota, Nebraska, North Dakota, and South Dakota; and focused on land cover/land use change, specifically grassland conversion, from 2006 – 2011. Both Decision Innovative Solutions 2013 and Wright and Wimberly 2013 began with CDL data, reclassifying subsets of the data for inclusion in their models and resampled data to better accommodate 1) varying resolutions found across the years of CDL coverages and 2) identification of areas of change. Table 2 compares the methods used by Decision Innovative Solutions 2013 and Wright and Wimberly 2013.

*Table 2: Comparison of land use change methods used in Decision Innovative Solutions 2013 and Wright and Wimberly 2013. *Grass/Hay is not listed as a category within the CDL. However, may be representative of other categories collapsed by the authors without definition.*

	<i>Decision Innovative Solutions 2013</i>		<i>Wright and Wimberly 2013</i>
<i>Spatial Resolution</i>	100 x 100 meters		560 x 560 meters
<i>Time Frame</i>	2007 – 2012		2006 – 2011
<i># Crop Cover Categories</i>	6		1
<i># Non-Crop Cover Categories</i>	3		1
<i>Crop Cover Categories List</i>	Alfalfa Other Ag Other Oilseeds Small Grains	Corn Soybeans	Corn-Soybean Combined

Table 2 (Cont.)

	<i>Decision Innovative Solutions 2013</i>	<i>Wright and Wimberly 2013</i>
<i>Non-crop Cover Categories List</i>	All Non-Ag Grassy Habitat Woody Habitat	Grassland
<i>Specifics of Grass-land Categories (i.e. which CDL categories are included within each reclassified grassland category)</i>	Grassland Herbaceous Herbaceous Wetlands Other Hay/Non Alfalfa Pasture/Grass Pasture/Hay Wetlands	Fallow/Idle Cropland Grass/Hay* Grassland Herbaceous Pasture/Grass Pasture/Hay
<i>Methods Summary</i>	USDA NASS CDL 2012 subtracted from CDL 2007 on per pixel basis; resulting data converted to vector, clipped by county, and summarized in SAS.	USDA NASS CDL 2006 compared to CDL 2011 on per pixel basis; generalization by 5x5 majority filter; resample to 560 x 560 with percent change calculation; smoothing by quartic kernel function

Both studies acknowledge inherent difficulties and subsequent inconsistencies that arose from using CDL for land use change; use quite different methods for approaching the calculation of change; and aggregate results to different spatial resolutions. Both are aggregated to spatial resolutions greater than the native grain size which creates uncertainty through the resampling process. The result of the differences in how the same data was used by Decision Innovation Solutions 2013 and Wright and Wimberly 2013 yielded two very different outcomes for South Dakota that did not go un-noticed by industry and academics. The Wright and Wimberly 2013 paper described grassland conversion for 2006 – 2011 as resulting in a net loss of 182,000 hectares (449,732 acres) whereas Decision Innovative Solutions 2013 described 2007 – 2012 grassland conversion as having declined by 879,000 hectares (2,172,056 acres) (Reitsma et al. 2016). However, prudent inspection of the methods reveals that an “apples-to-apples” comparison is not so simple as comparing final outcomes. A more thoughtful comparison would

be to only consider conversion from “Grassy Habitat” to “Corn and/or Soy” from Decision Innovative Solutions 2013 as shown in Table 3 as these categories are more focused on Grassy-Habitat loss specifically to corn and soybean and do not include the other four crop categories for which they evaluated loss. However, even this comparison is still quite misinformed as the Decision Innovative Solutions 2013 “Grassy Habitat” data also includes categories for wetlands as grass-land types. Thus, these data are still considerably different, though improved upon from the general comparisons where class selection is not considered. In summary, no one version can be deemed correct when not comparable in purpose or process. And each iteration of analysis continues to further a more thoughtful inclusion of CDL data into policy and planning as we continue to see extreme differences in summary data as a result of specification problems.

Table 3: A closer look at comparison of results for conversion of grass-land cover types to corn and/or soy crop cover types in South Dakota by only reviewing the data outcomes by the categories, rather than project results.

Decision Innovative Solutions 2013		Wright and Wimberly 2013
<i>Grassy Habitat to Corn (net loss)</i>	<i>Grassy Habitat to Soybean (net loss)</i>	<i>Grassland (net loss)</i>
682,573	414,804	451,000 acres
<i>Grassy Conversion to Corn and/or Soy Total</i>		<i>Grassland Conversion to Corn/Soy Total</i>
1,097,377		451,000

3. Enhanced Spatial Decision Models Supporting Monarch and Pollinator Success

This case study calls for establishing the current extent of monarch butterfly habitat as a baseline for comparison in reviewing 25 years of land change with respect to the monarch butterfly’s central United States flyway. To determine the current extent of monarch butterfly

habitat using CDL data, it would seem logical to first identify what habitats are important to the monarch butterfly and then focus on examination of those data. However, there is no limit to potential habitat for the monarch butterfly, therefore all data are examined across a spatially vast area to determine where change is occurring on the landscape for multiple time periods.

Additionally, efforts are made to use enhanced data to help differentiate among actual change and false change. False change is often present as noise or propagated error, identifying false change present in the data will help to further quantify uncertainty within CDL. Acknowledging these uncertainties will provide the best estimate of baseline information. In this research a 10-year annual land change model is completed that compares land change and conversion between 2008 and 2017. Seeking to develop the most informative version CDL with respect to estimating and identifying monarch butterfly habitat and potential habitats there are two models with augmented datasets tested against the original CDL with no augmentations. The variation in these spatially-explicit models are as follows:

Model 1: CDL with no augmented or enhanced data;

Model 2: CDL with augmented data better characterizing protected areas, soil productivity, transmission lines, and transportation networks.

Model 3: CDL with augmented data better characterizing grass-land cover types and shrub-land cover types (NLCD), as well as protected areas, soil productivity, transmission lines, and transportation networks.

The foundation for model development described in this paper expands and refines the models described in Thogmartin et. al 2017. The enhancements proposed in this research could alter the prescribed habitat objectives passed down to state- and local- level land managers when considering lands that are potentially amenable to pollinator-friendly plantings, including targets

for milkweed stems (genus *Asclepias*), though it is important to note that it will not change the need for a multi-sector approach to monarch and pollinator conservation. Milkweeds are an essential part of the monarch life cycle, they are the plant species on which monarch oviposition takes place, and primary source of food for emerging larva (Agrawal 2017). Increasing the number of milkweed stems planted is the primary conservation action that is addressed by Thogmartin et. al 2017 and is supported and regionally allocated by the models he describes. However, may be argued that this is not the right course of action for the preservation of the monarch butterfly and other pollinators. The models in this research move towards identifying Important Pollinator Areas (IPAs) and establishing High-Priority Pollinator Areas (HPPAs) focusing more on connectivity of the migration corridor and less on specific numbers of milkweeds to be planted. IPAs are identified as those areas that have demonstrated membership by being a vital part of the pollinator connection; and HPPAs being those areas that have been overlooked in creating connective habitat corridors for pollinator species that migrate. The development of the most accurate estimation of monarch butterfly habitat, and potential habitats, using available data, is equally important in addressing the larger research objectives involved in the four-year study funded by the BASF Corporation, of which this research is a part of, and also includes:

1. Identifying where habitat has changed over the past 25 years;
2. Analyzing how habitat has changed by calculating landscape metrics such as patch size, patch-to-patch distance, and rates of land parceling and fragmentation; and,
3. Identifying where conservation practices are taking place and if these efforts are targeting areas where habitat was previously available, and/or can be built upon.

Establish IPAs and HPPAs for targeting conservation efforts and funding, where possible.

Additionally, the spatially-explicit models created in this research provide a more user-friendly, accessible procedure for better identification of land use land change using the CDL that can be applied across multiple disciplines. Both the data and tools are readily available, and easily customizable, to most agencies and institutions contributing to the monarch butterfly and pollinator conservation effort. The resulting data and data models are also products that can be modified in endless ways to benefit any number of projects or initiatives, including derivative models that may eventually improve additional research for monarch butterflies and other pollinators. These models also have the ability to easily expand the extent of the model to include additional data or localities in the future. And to expand the application to other areas of monarch butterfly or conservation research anywhere that similar data may be obtained and processed in place of the study area used in this research.

4. Methods

4.1 Study Area

The study area of this research includes the central flyway of the monarch butterfly. This area includes approximately 20 states within the central United States, as shown in Figure 3. This area was selected because of the overlapping areal extents from three sources on monarch migration and priority geographies: 1) the "Two-way Monarch Migration Map" created for MonarchWatch.org ("Two-Way Monarch Migration Map" 2010); 2) the two central priority geographies designated by the U.S. Fish and Wildlife Service and Monarch Joint Venture (U.S. Fish and Wildlife Service 2015) and, 3) the U.S. Fish and Wildlife Service Monarch Butterfly

Conservation Units in the U.S., specifically inclusive of states encompassing the majority of the North Core and South Core Conservation Units (U.S. Fish and Wildlife Service 2017).

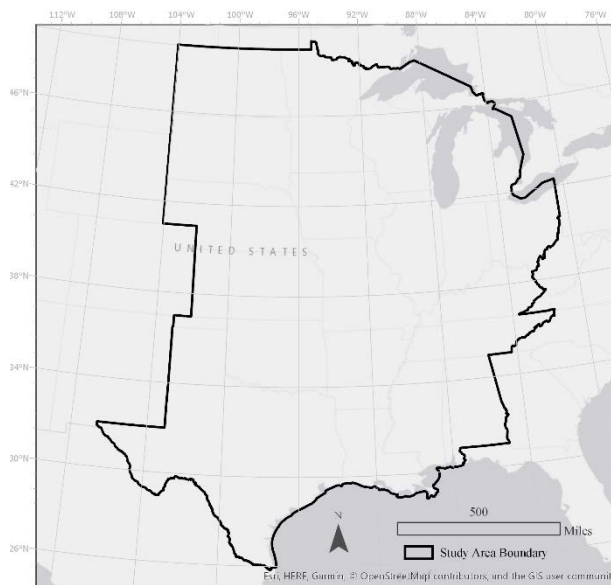


Figure 3: Map of central United States with study area boundary displayed for this research.

The area of this region is approximately 1,449,187 square miles (927,479,680 acres) and is broadly characterized by NLCD 2011 as being 37% in Cultivation, Hay, or Pasture; 25% Grassland-Herbaceous, Shrub, or Scrub; 20% Forest; 6% Developed; 6% Wetlands; 2% Open Water, and 4% Other. The study area includes the states of Alabama, Arkansas, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Nebraska, North Dakota, Ohio, Oklahoma, South Dakota, Tennessee, Texas, and Wisconsin. The study area does not include any portion of New York, Pennsylvania, and West Virginia. These areas could be included in subsequent model runs along with other areas where data permits inclusion.

4.2 Defining Monarch Butterfly Habitat

The resource needs of the monarch butterfly are temporally variable and dependent upon life stage, as well as generation, and can be difficult to characterize among land cover data. This is primarily due to the increase in small habitat enhancements that have been implemented into

many cover types that may have previously been categorically dismissed as having no potential for monarch habitat and typically are too small to be represented by 30 meter pixel. For instance, in the past developed lands may have been excluded as not providing habitat for monarchs and pollinators. However, with the growing adoption of right-of-way management, municipal involvement (for example, National Wildlife Federation Mayor's Monarch Pledge, National Pollinator Garden Network's Million Pollinator Garden Challenge, and the MonarchWatch Monarch Waystation Program), and corporate campus and other private habitat implementations, this is no longer a valid assumption. To better approach defining a monarch butterfly habitat, this research enriches existing land cover types with complementary data that will support better characterization of the landscape with respect to monarch habitat. This approach is opposite from previous studies where CDL categories have been aggregated or collapsed and summarized into more binary categories of habitat vs non-habitat areas. The benefit of expanding the categorization is to more precisely characterize the landscape for targeting conservation practices and programs based on their existing and future potential to support those activities. For example, high productivity corn cover has less potential for implementing a pollinator conservation program than low productivity corn cover. By retaining and creating additional cover types and subtypes for two of the three models a more informed proposal can be made. The output of these models is to serve as a decision-making guide, that combined with local on-the-ground knowledge, will help direct action with more focused and relevant content for the target demographic.

4.3 Model Structure and Components

For each of the three versions of the model, similar processes were completed. The first model reflected solely the CDL without augmentations or enhancements. For model 1, the

preparation and extent steps were matched to the other two models for comparison purposes. For model 2, CDL data was augmented with additional data to better identify key parts of the landscape that have the potential to be targeted more specifically for conservation programs. For model 3, was augmented with additional data and portions of the NLCD. The third model represents the most similar solution to Thogmartin et al. 2017.

The three versions of the model share common data preparation models, each stored as a model constructed within ModelBuilder for ArcGIS Pro. As shown in Figure 4, after data preparation has been completed, two of the three models move on to land cover augmentation and enhancement. In land cover augmentation and enhancement, a model constructed within ModelBuilder for ArcGIS Pro was completed for each year of the CDL where augmentations and enhancements are completed. After all models have run, final computations and data summaries are produced using analysis scripts in R. There are three major components to each completed model run, each component containing several subcomponents: 1) data preparation models; 2) land cover augmentation and enhancement models; and, 3) analysis scripts. Each subcomponent of the model is discussed in more detail in the following text.

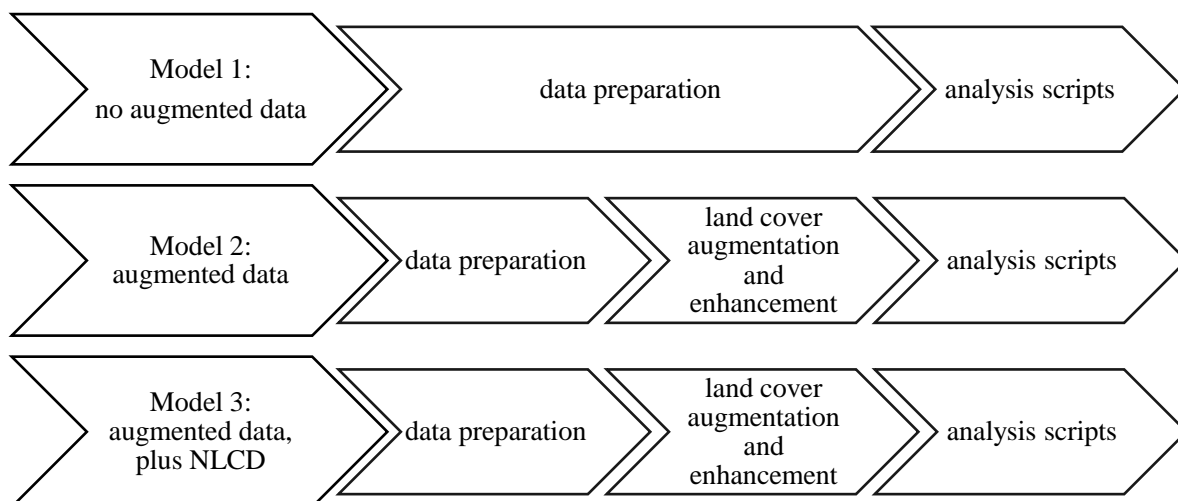


Figure 4: General flow diagram of the three versions of the models created for this research.

4.3.1 Data Preparation Models and Processes

For each year CDL data, 2008 – 2017 the data were clipped to the extent of the study area. The NLCD 2006 and 2011 and gridded SSURGO data were also clipped to the extent of the study area. In all cases, the grid was snapped to a 30 x 30 meter raster map of the study area to ensure complementary alignment and prevent pixel shift during processing. For gridded SSURGO data, the attribute of interest, National Crop Commodity Productivity Index (NCCPI) Version 2.0, was treated separately for each of the commodity crops for which an index is calculated: 1) Corn and soybean; 2) Cotton; and, 3) Small grains. The “USDA NRCS gSSURGO VALU1 Table Outline and Column Descriptions” document indicates that NCCPI values range from 0.01 (low productivity) to .99 (high productivity). Data were clipped to the extent of the study area, sub-categorized using crop-specific data breaks for the study area, and assigned descriptive categories, as shown in Table 4.

Table 4: Descriptive text for productivity assigned to NCCPI values by commodity crop type.

NCCPI Value Range	Corn/ Soybean	Cotton	Small Grains
0.001 – 0.150	Low	Low	Low
0.151 – 0.225	Marginal	Low	Marginal
0.226 – 0.475	Moderate	Marginal	Moderate
0.476 – 0.850	Moderate to high	Moderate to high	Moderate to high
0.851 – 0.991	High	--	High

Figure 5 shows the difference between the NCCPI national and study area specific data breaks, demonstrating the importance of normalizing this data to the study area. These data represent the soil productivity index for all lands as they would relate to the plantings of the

following crops: corn, cotton, small grains, and soybeans. The acreages shown in Figure 5 are not acres of actively cultivated area of the selected crops.

Allowing multiple data breaks, rather than one as indicated in Thogmartin et al. 2017, also provides a more continuous categorization of these data and is more forgiving across the study area where NCCPI values may vary greatly preventing an over-identification of perceived marginal lands in areas where NCCPI value ranges vary. Thogmartin et al. 2017 identified NCCPI value 0.40 as being the break point for determining marginal lands in corn and soybean lands, and only used values for NCCPI for corn and soybean. NCCPI for cotton and small grains were not included in Thogmartin et al. 2017. In this research, several values are identified for classifying NCCPI values, as well as across all categories of NCCPI, including cotton, corn, small grains, and soybeans. Corn and soybeans are combined and represent one NCCPI category within SSURGO but are represented using two datasets in this research in order to retain calculations for soybean and corn separately. The placement of data breaks is important in designation of low and high productivity when data are for a large area. In this data, areas of marginal or moderate productivity corn in the north central United States may be considered high productivity corn in the south. The data breaks are merely a guide for providing conservation opportunity on landscapes where a mutual benefit can be derived among landowners and pollinators. Additionally, NCCPI is designed for use as an indicator of productivity for non-irrigated crops only. Irrigated varieties of selected crops occurring on low and marginal productivity soils may not be recognized at their full potential as supplemental watering regimes are not considered in the NCCPI models.

To include the NCCPI in the models, each of the NCCPI crops was reclassified into a multi-value raster with value ranges aligning to the category numbers in Table 4. Where values

were not present NA was assigned. Raster data sets for corn, small grains, and soybean contained five values aligning to those identified in Table 4, the raster for cotton contains only 3 values, also aligning to those identified in Table 4. The NCCPI raster data sets were referenced in the model as a parameter, making it easier to interchange raster data sets representing NCCPI categories when altering data breaks.

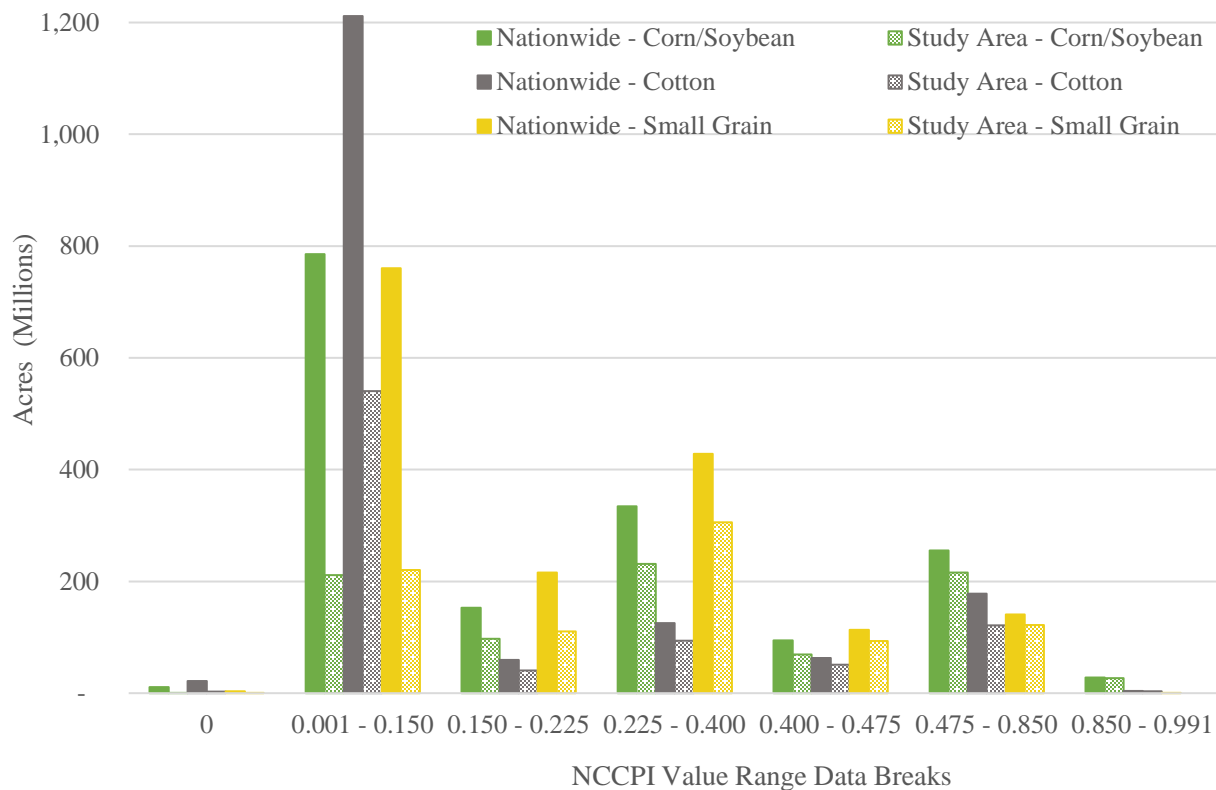


Figure 5: Distribution of NCCPI values by acreage, nationwide and within the study area of this research, by value range data breaks indicated in Table 4, including separation at the 0.400 breakpoint as identified in Thogmartin et al. 2017 for illustration purposes; however, for the purposes of the model in this research the value ranges of 0.226 – 0.400 and 0.441 – 0.475 are combined.

To better characterize the extent of protected areas, transportation networks, and transmission lines (a major focus of right of way activities), each of these input layers were converted from their native format as vector data to a raster data set. Data values were assigned

grid values that were complementary to attributes in those datasets that best characterize the data. Grid values were then used to convert the vector dataset to a 30 x 30 meter grid that was aligned with, and snapped to, the extent of the study area. For protected areas (Protected Areas Database of the United States (PAD-US)), the data were assigned binary grid values prior to conversion to raster. Where data were equal to one the area was defined as “protected areas under the management of local, state, federal or other government management;” where data were equal to zero the area was defined as “un-protected areas in the general public domain.” PAD-US coverage for specific protected land varies based on the data provided by individual states and may also include, in some areas, lands protected under the ownership of non-governmental organizations and/or easements data. Transportation networks are represented by the 2016 TIGER Roads National File Geodatabase, those data were assigned grid values based on the feature class type and description; resulting in four categories: 1) Primary and secondary roads; 2) Local neighborhood roads, rural roads, city streets, ramps, alleys, and service drives; 3) Private roads, vehicular trails, walkways and pedestrian trails; and, 4) Bike paths and trails, and bridle paths; where roads are not present in the raster, NA was assigned. The values that were assigned to the transportation networks are the same as the land use category code that defines those categories within the lookup table. For transmission lines (Electric Power Transmission Lines), the data were assigned as binary grid values prior to conversion to raster; 1 = transmission line present; 0 = transmission line not present. Both transportation networks and transmission lines may be overstated in their estimated reach as the minimum mapping unit pixel is 30 meters. Since both are considered as target areas for conservation activity and as contributors to habitat fragmentation it is imperative to include these areas on the landscape, even if overstated in quantity in some areas.

Other steps completed in data preparation include background removal and mosaicking of the CDL confidence raster data sets. Each confidence raster is generated and provided for download by state through the USDA NASS CDL website. For this model, we mosaic raster data sets of each state using the preparatory models for the CDL confidence raster data sets. Preparatory models were created to ensure consistency in removal of background and mosaicking of data for the four years for which confidence data are provided (2008, 2009, 2016, and 2017). USDA NASS has only recently (2017) added the release of confidence layers to the suite of downloadable products available to accompany CDL products. It is important not to confuse the confidence raster data sets with accuracy assessments also provided through USDA NASS. Accuracy assessments are tabular summaries, as opposed to spatial data sets, providing producer and user accuracies by category and state. The CDL confidence layers provide a value associated with each output pixel, that represents the predicted confidence based upon the rule(s) that were used to classify it. While this information does not provide ground-truth or visually verified information, it does provide important information on the spatial distribution and extent of confidence for classification and uncertainty that may accompany it.

4.3.2 Land Cover Augmentation and Enhancement Models

For Model 1 in Figure 4, the preparation step is the only step required, before moving on to analysis. Model 1 is primarily a test of the CDL with no further enhancements or augmentations of the data. For models 2 and 3, each year of CDL data is processed through a model that augments and enhances the data. The processes are similar, only deviating at the point where further subdivision by NLCD categories is necessary. These models are derivative of processes described in Thogmartin et al. 2017; however, have been extensively modified in hopes of advancing the capacity of the model to better identify areas for conservation action.

Consistent with methodology described by Thogmartin et al. 2017, the initial processes of each model address urban spaces in the CDL. The first process seeks to separate core areas of “Developed/Open Space” from linear areas of the same category by shrinking CDL code 121, “Developed/Open Space” by one pixel; then expanding the output of that process also by one pixel. The output of this process is assigned a new category and code, 120, “Linear Developed/Open Space.” The second set of processes seeks to identify core areas of “Developed/Open Space” and other urban landscapes by shrinking CDL codes 121, 122, 123, and 124 by two pixels, then expanding the output of that process also by two pixels.

The next process introduces the transportation networks data as multi-value raster described in 4.3.1 of this document. Roads data are given preference over urban data in the geoprocessing overlay. Where roads data are not present, and values are zero are present, the previously described urban data are inserted. The dataset to this point includes developed lands and transportation networks and remains consistent in processing for both Model 2 and 3.

The next processing phase begins to see divergence in how Model 2 and 3 address grass-land cover types. For model 2, the process is simple: CDL codes 64, 152, and 176 are isolated from the CDL and subdivided using the binary raster data set for protected areas as shown in Figure 6. The result is the creation of five categories of grass-land cover types. Three categories for areas not contained within a protected area, and two complementary categories for those contained within a protected area; CDL codes 64 and 152 are combined as one protected shrubland category and assigned new code 150.

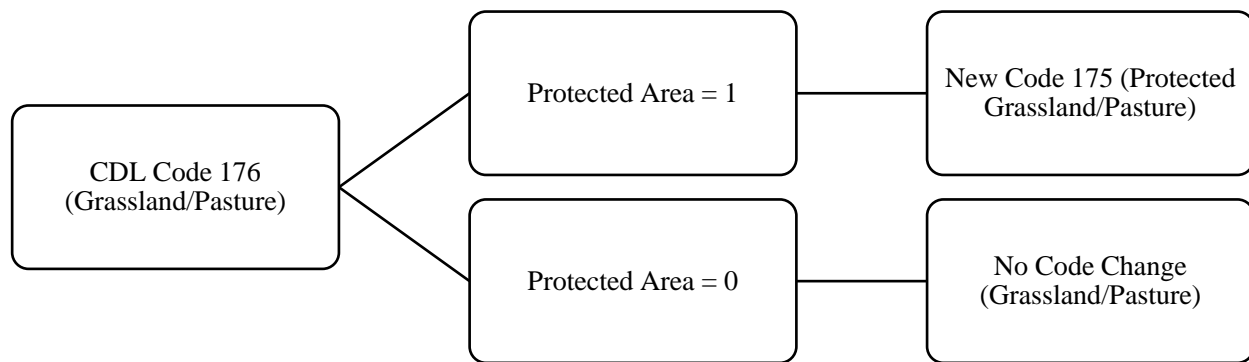


Figure 6: Process for delineating portions of grass-land cover types as protected using CDL Code 176, Grassland/Pasture as example.

For model 3, the process of subdividing grass-land cover type is complicated by the integration of NLCD as a surrogate for more sophisticated means of more clearly the boundaries among herbaceous grasslands and hay or grazing pastures. The annual CDL data used in this research spans the time frame of two NLCD releases and thus, NLCD 2006 and NLCD 2011 have both been used in the combination of CDL to NLCD. For 2008 – 2010, NLCD 2006 is used for grass-land cover delineation and for 2011 – 2017, NLCD 2011 is used. It is important to remember that NLCD is only released every five years, and that the CDL non-crop categories are already quite heavily reliant on NLCD for their categorization. Furthermore, the introduced NLCD categories are only used as a factor of delineation in areas where CDL has already identified the extent of the area as being of the appropriate grass-land cover type as shown in Figure 7. The NLCD categories are not used as a broad replacement for existing categories in CDL (i.e. areas of 2017 CDL crop are not overwritten by areas of 2011 NLCD pasture). For each CDL code, 64, 152, and 176, the category is isolated from the CDL and first divided by the presence (or absence) of NLCD categories within the area, then further subdivided by the portion of the area that is within protected areas. Finally, the segments are pieced back together through a series of conditional statements, the result is a dataset with 11 grass-land cover type categories.

Six for categories not within a protected area, and five for categories within a protected area.

Again, CDL codes 64 and 152 are combined as one protected shrubland category.

In both model 2 and 3, the final output of the grass-land delineation process is combined with the final output of the urban process through a conditional statement. The result of the grass-land delineation process is given priority during conditional statement. More specifically, where no grass-land cover type exists in the final output, the urban final output is added. If a conflict arises where both grass-land cover type and urban cover type exist in the same pixel, the grass-land cover type is selected. However, these conflicts have been minimized by using the urban output as the first input of the grass-land cover type process. In theory, there should be no instances where conflicts of this nature should arise.

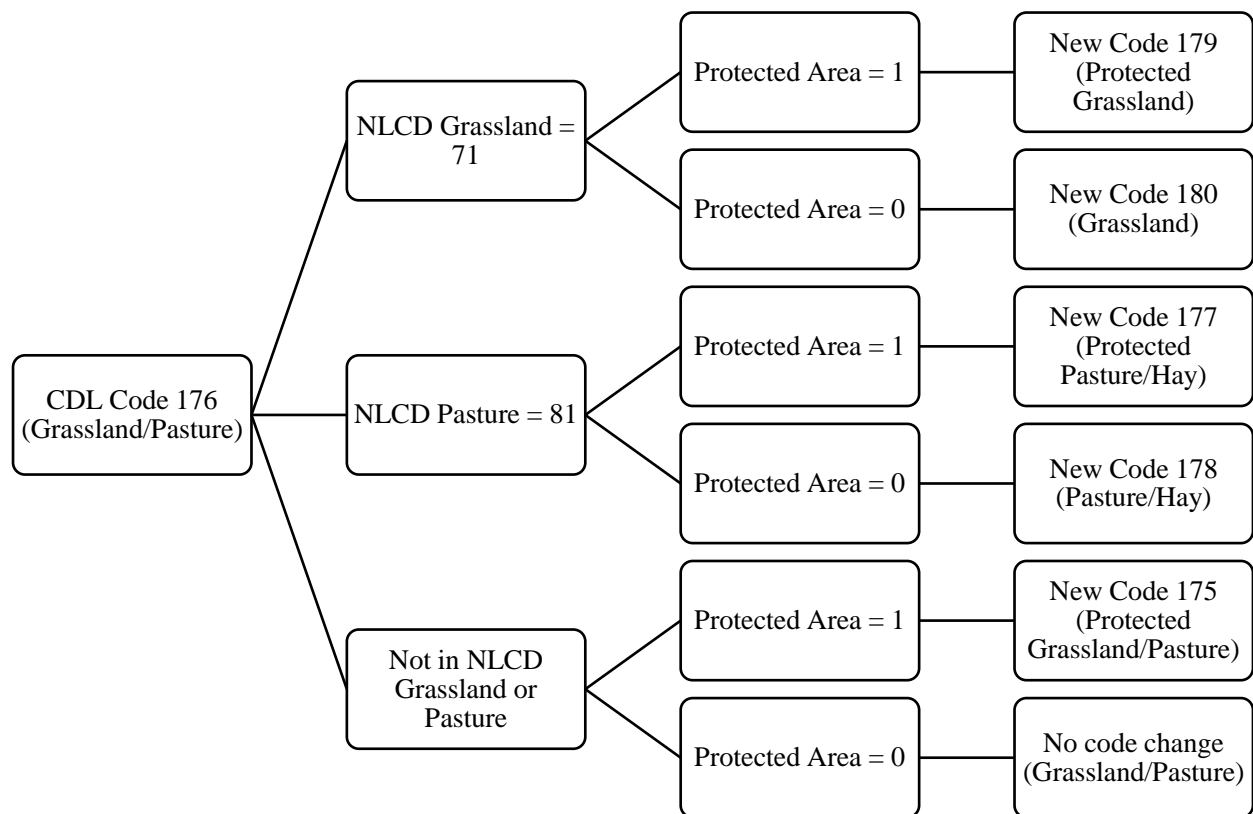


Figure 7: Process for delineating portions of grass-land cover types integrating NLCD data, and protected areas database layers, illustrated using Grassland/Pasture as example.

The next process addresses crop-specific productivity using the 30-meter gridded SSURGO data and the NCCPI data discussed in detail in section 4.3.1 of this document. Models 2 and 3 follow the same structure through this process. Each commodity crop group is isolated from the CDL, aggregated, if needed, and then subdivided based on values in the appropriate NCCPI raster interpretation. Aggregation prior to splitting is required for corn, small grains, and soybeans to accommodate the multiple CDL classes that contribute to the commodity crop class. For example, corn includes CDL codes 1, 12, and 13 as shown in Figure 8, and small grains includes seven different CDL codes: 21, 22, 23, 24, 25, 27, and 28. Cotton is represented by only one CDL code 2, as are soybeans, CDL code 5. After each commodity crop group has been processed they are re-assembled into a crop-specific productivity output with 18 resulting categories and new codes. The output of this process is combined with the previous output from the grass-land cover type delineation process through a conditional statement. No preference is given during the combining process as the outputs are non-overlapping.

The completion of this series of processes and the resulting outputs are finalized by updating code values and descriptions through a look up table where old and new values are cross-walked. Area calculations (acres) are computed, output is stored for each CDL year for models 2 and 3, and summary statistics are generated and written out as table to be referenced in by R scripts in the analysis steps.

The R scripts in Appendix 4 are those used for the processing of the data after all model runs have been completed. These scripts import and join the tables created through the model process create year-to-year comparisons of total acres, mean acres, standard deviation, coefficient of variation, and z-scores for each category. Additionally, year-to-year plots of the

number of acres per category based on confidence and frequency are plotted to better understand how subdivision of the data alters these distributions.

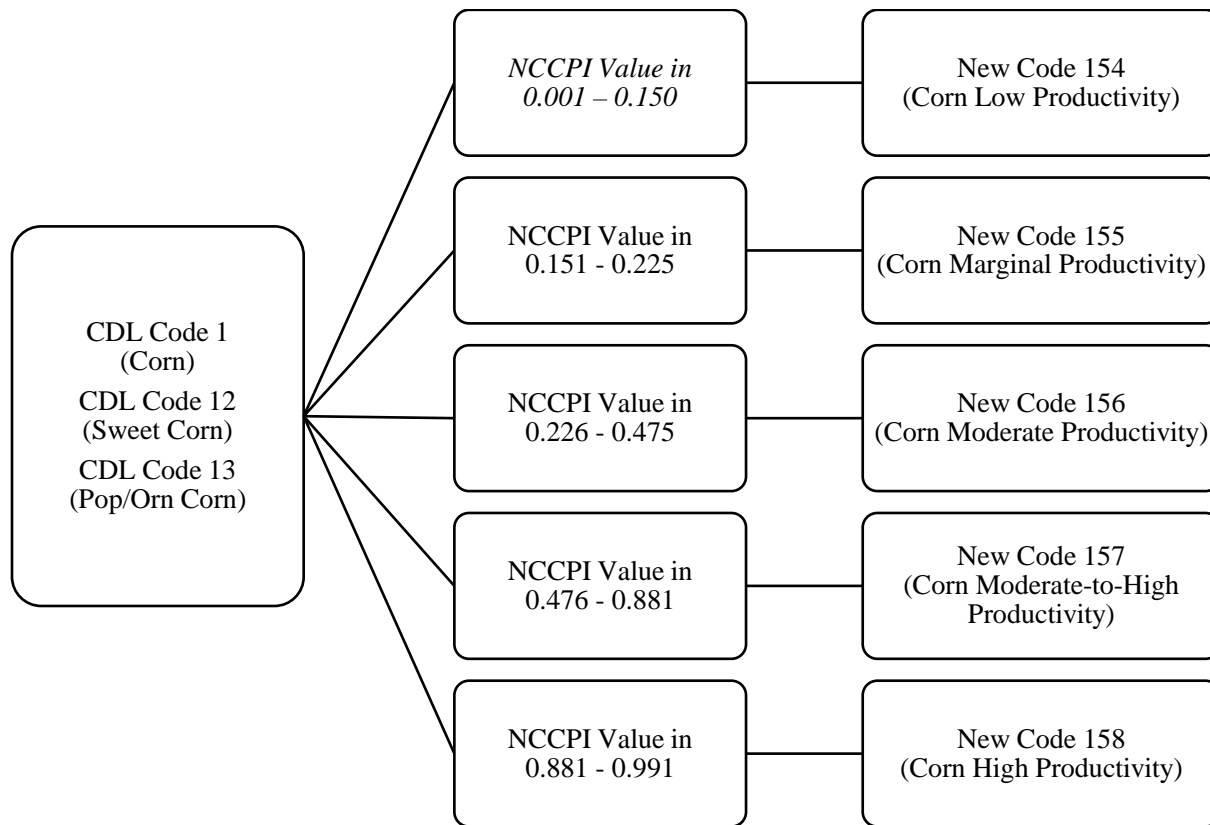


Figure 8: Process for delineating commodity crops based on NCCPI values, illustrated using corn as an example.

4.4 Detecting Change across the Landscape Over the Past Decade

Using the outputs from the land cover augmentation and enhancement models a workflow for generating change matrices over the past decade was developed. Though currently completed as step-by-step procedural process, models for completing these tasks are under development and will be completed during the next research phase.

The process for detecting change across the landscape over the past decade is built on traditional methods for detecting change in raster analysis were. The development of the change matrix for 2008 – 2017, specifically used one of the four concepts from the Multi-Index

Integrated Change Analysis (MIICA) process used for determining spectral change among Landsat pairs (Jin et al. 2013). One process from MIICA is the Change Vector (CV) calculation that can be used in this research to identify categorical change among the pixel pairs. The CV functions similarly for identifying categorical change as it does in the traditional use of identifying spectral change among Landsat pixel pairs. In this case we use the category values of the pixels to look for changes in classification between two-time periods. For each pixel pair in the set of model outputs for 2017 and 2008 $CV = \sum_i (v_{2017} - v_{2008})^2$ is calculated; the absolute value of CV is then inserted into the change code that is assigned to each row. The change code is formed by the following numeric combination: “17” + [2017 Category Value] + “8” + [2008 Category Value] + [CV]. This gives each change of category a unique identifier that also serves as reference that can be deconstructed to determine what change has taken place.

Calculating the CV and change code is the first step in identifying land use conversions among the model outputs and highlights the extent of the change and generates the change matrix. However, the CV is not a stand-alone metric of change in the model outputs. Similar to challenges that are discussed in the NLCD MIICA process for identifying change, the CV identifies all areas of change, including crop rotation, seasonal, and phenology changes (Jin et al. 2013). In addition to considerations for these phenomena, part of the difficulty in using data that has already been processed and categorized is accounting for error propagation. Specifically, edge confusion that may be recorded as change; single pixel or small groupings of pixels that have been misclassified; and, differentiating among types of change. For example, separating naturally occurring change from change as a result of human activity, such as crop rotation, that resembles change on the landscape, but should be quantified separately.

In addition to CV, the frequency and confidence of the 2008 and 2017 model outputs are combined. The resulting layer is used to establish areas of change apart from areas where data support areas of no change. The frequency of each input layer is calculated in ArcGIS Pro 2.1 using the Equal to Frequency geoprocessing tool where all years 2008 – 2016 are compared to 2017. The resulting layer is a pixel-by-pixel assessment of the category value in comparison to 2017, a pixel value of 9 identifies a pixel where the value has been the same for all 10 years of the comparison; similarly, a pixel value of zero identifies a pixel where the value has not been equal to that of the 2017 pixel in any of the preceding 9 years. Where frequency indicates consistent identification for all 10 years, a general assumption about the strength of either the classification of that pixel and/or stable ground conditions is made and is categorized to reflect generally no change. A pixel value between 0 and 9 indicates that at least one time between 2008 and 2016, the pixel value was not equal to the pixel value of 2017, these areas are broadly characterized as being possible change.

The CDL Confidence Layer is also being used to better identify change on the landscape. To include these data, the confidence layer for each state is mosaicked for a seamless representation of the study area. The mosaicking procedure is completed by an ArcGIS model built specifically for the consistent compilation of confidence data as described in section 4.3.1 of this document. The mosaicked confidence layer has a range of values from 0 – 100, where 0 represents little confidence in the pixel categorization through the CDL decision tree; and 100 represents a high confidence in the pixel categorization through the CDL decision tree.

The confidence and frequency of each pixel is used to create a change mask for each model. In each change mask, pixels are classified based on the following: where frequency = 9 AND confidence \geq 75 the pixel should be classified as one; if these conditions are not met, the

pixel should be classified as zero. Areas of pixel values equal to zero indicate areas of potential change and areas of pixel values equal to one indicates areas where change is presumed stable (no change over 10 years). For each model, the change mask is generated for 2017 and 2008; then summed using the plus geoprocessing tool. The resulting datasets divide the study area into three broad classifications: areas where the change mask is equal to zero indicates change in both 2008 and 2017; areas where the change mask is equal to one indicates change in either 2008 or 2017, but not in both years; and, areas where the change mask is equal to two indicates no change in 2008 or 2017. Areas where the change mask agrees in both years are confidently assumed to be areas of no change for the decade. However, to better understand the categorical change, the change code is cross walked to a look up table with all possible combinations of conversion as text descriptors. Using the CV, change code, and raster dataset of combined values from 2008 and 2017 for each model, the change code and text descriptions are reintroduced and are now ready to be visually assessed and quantified across the landscape. The outputs from this model can then be utilized in additional models and data tools to help inform policy and conservation delivery across the landscape.

5. Results

Overall, the model outputs agree regarding general land conversion trends and quantities of land use types. This is expected since most categories are broadly unchanged throughout the models, only further subdivided for more precise consideration. One outcome of this research was to complete development and comparison of the three models and their outputs to identify where augmenting and enhancing the CDL data served to better categorize data for the purposes of targeting conservation actions and best represented the landscape. Additionally, to explore

methods for quantifying the amount of uncertainty that may be masked through the process of subdividing category data due to inherent data issues.

Model 1 and 2 outputs, in general, agree that grassland/pasture category has the greatest quantity of land, the greatest amount of deviation, and downward trending losses over the past 10 years. However, as additional spatial subdivision is completed within the grassland/pasture category these numbers decrease as smaller, more well-defined categories of grassland/pasture increase. This spatial subdivision impacts the overall statistics of the category and can be observed by comparing models 1, 2 and 3 by their coefficient of variation. Model 1 (Figure 9, Table 5) and model 2 (Figure 10, Table 6) are still capturing a great deal of the uncertainty of the grassland/pasture category; while the model 3 (Figure 11, Table 7) processes have masked uncertainty in this category through subdivision of the category to the point that only a small portion of the original grassland/pasture category continues to stand out. Remaining features, not part of the newly enhanced standard features, now have limited standard deviation and coefficient of deviation, that may be masking their need for attention.

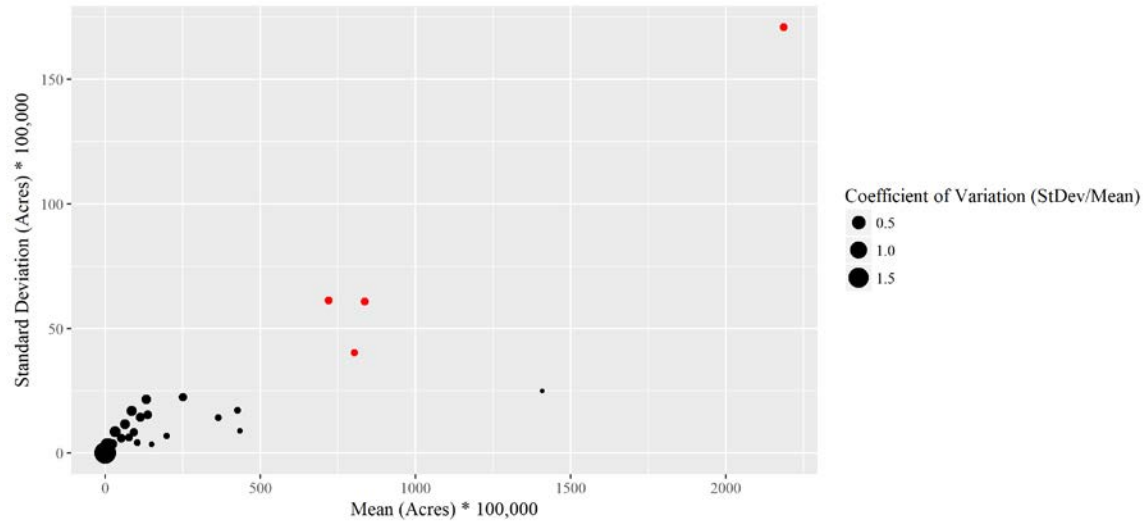


Figure 9: Model 1, 2008 – 2017 Plot by Category and Coefficient of Variation. Data for highlighted points (red); points with a standard deviation of greater than 2,500,000 are highlighted and displayed in Table 5. Note the difference in scale between Model 1 and Models 2 and 3.

Table 5: Mean Standard Deviation, Mean Acres, and Coefficient of Variation for Model 1, 2008 - 2017 points with a standard deviation of greater than 2.5 million as highlighted in red in Figure 9.

Category	Standard Deviation (Acres)	Mean (Acres)	Coefficient of Variation
Corn	4,016,840	80,359,580	0.49985820
Soybeans	6,128,745	72,056,713	0.85054460
Shrubland	6,073,948	83,646,386	0.07261459
Grassland/Pasture	17,083,738	218,704,703	0.07811326

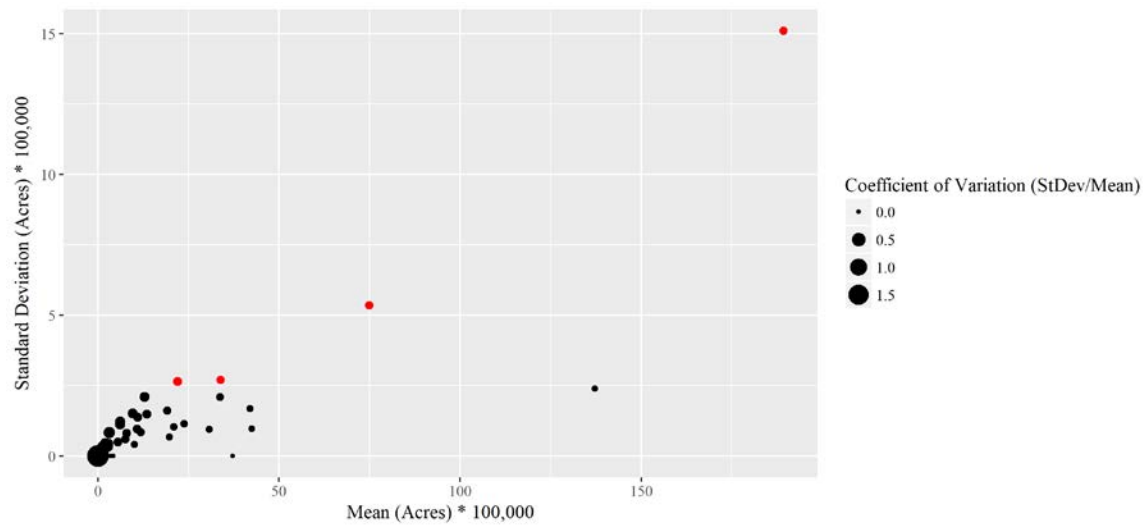


Figure 10: Model 2, 2008 – 2017 Plot by Category and Coefficient of Variation. Data for highlighted points (red); points with a standard deviation of greater than 2,500,000 are highlighted and displayed in Table 6. Note the difference in scale between Model 1 and Models 2 and 3.

Table 6: Mean Standard Deviation, Mean Acres, and Coefficient of Variation for Model 2, 2008 - 2017 points with a standard deviation of greater than 2.5 million as highlighted in red in Figure 9.

Category	Standard Deviation (Acres)	Mean (Acres)	Coefficient of Variation
<i>Deciduous Forest</i>	2,859,511	129,587,995	0.02206617
<i>Shrubland</i>	2,880,063	64,644,903	0.04455205
<i>Grassland/Pasture</i>	9,973,022	40,324,682	0.24731806

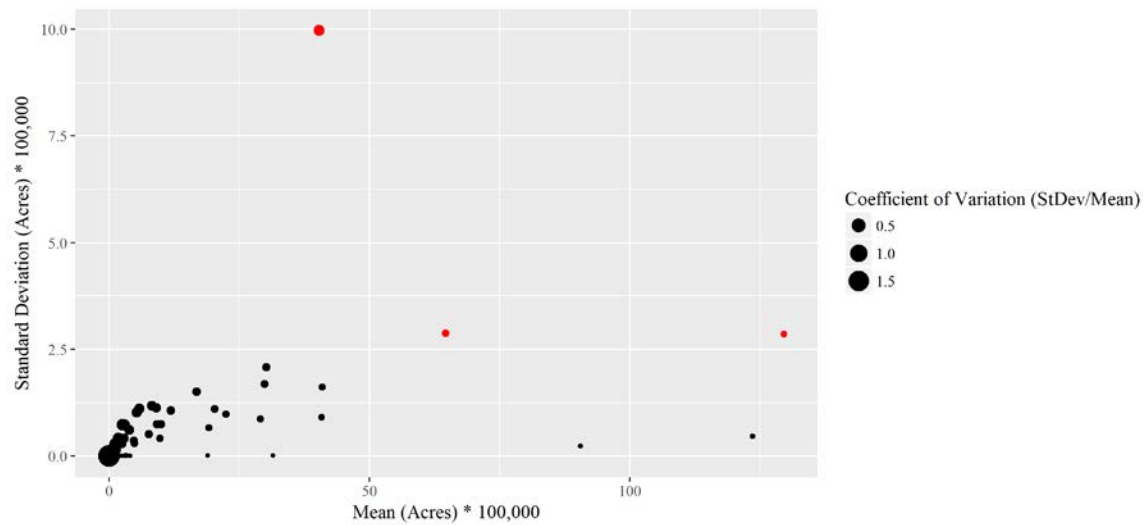


Figure 11: Model 3, 2008 – 2017 Plot by Category and Coefficient of Variation. Data for highlighted points (red); points with a standard deviation of greater than 2,500,000 are highlighted and displayed in Table 7. Note the difference in scale between Model 1 and Models 2 and 3.

Table 7: Mean Standard Deviation, Mean Acres, and Coefficient of Variation for Model 3, 2008 - 2017 points with a standard deviation of greater than 2.5 million as highlighted in red in Figure 9.

Category	Standard Deviation (Acres)	Mean (Acres)	Coefficient of Variation
Core Developed/ Open Space	2,660,228	21,994,547	0.12094943
Soybeans (CDL) Moderate-to-High Productivity (NCPPI)	2,705,582	33,788,474	0.08007412
Grassland/Pasture	10,437,081	264,180,855	0.03950733

This is also apparent in comparing the tabular version of results of change for 2008 to 2017. While the three model outputs are very similar, Table 8 shows the top 10 records out of each change dataset which better demonstrates how the process of augmentation have subdivided large categories into multiple smaller categories that are perceived as having a better fit. However, the extent of the grassland/pasture remains consistent in each model year, only the way in which it is described and subdivided changes throughout each model. This gives a false sense of improvement over the land categories.

Table 8: Comparison of Top 10 (by Area) Categories of Change from 2008 to 2017, by Model

Model 1		Model 2		Model 3	
Change Description from 2017 to 2008	Acres	Change Description from 2017 to 2008	Acres	Change Description from 2017 to 2008	Acres
Corn to Soybeans	35,008,234	Corn (CDL) High Productivity (NCPPI) to Soybeans (CDL) High Productivity (NCPPI)	16,212,194	Corn (CDL) High Productivity (NCPPI) to Soybeans (CDL) High Productivity (NCPPI)	15,687,804
Soybeans to Corn	32,014,247	Soybeans (CDL) High Productivity (NCPPI) to Corn (CDL) High Productivity (NCPPI)	14,958,809	Soybeans (CDL) High Productivity (NCPPI) to Corn (CDL) High Productivity (NCPPI)	14,446,396
Grassland/Pasture (CDL) to Shrubland	13,314,312	Corn (CDL) Moderate-to-High Productivity (NCPPI) to Soybeans (CDL) Moderate-to-High Productivity (NCPPI)	14,569,270	Corn (CDL) Moderate-to-High Productivity (NCPPI) to Soybeans (CDL) Moderate-to-High Productivity (NCPPI)	13,520,694
Grassland/Pasture (CDL) to Other Hay/Non Alfalfa	8,042,320	Soybeans (CDL) Moderate-to-High Productivity (NCPPI) to Corn (CDL) Moderate-to-High Productivity (NCPPI)	13,545,946	Soybeans (CDL) Moderate-to-High Productivity (NCPPI) to Corn (CDL) Moderate-to-High Productivity (NCPPI)	12,585,035
Grassland/Pasture (CDL) to Deciduous Forest	7,418,003	Grassland/Pasture (CDL) to Shrubland	11,673,463	Grassland/Pasture (CDL) to Shrubland	5,956,056
Grassland/Pasture (CDL) to Soybeans	7,395,798	Grassland/Pasture (CDL) to Other Hay/Non Alfalfa	7,291,376	Deciduous Forest to Woody Wetlands	5,029,523
Grassland/Pasture (CDL) to Corn	7,255,516	Grassland/Pasture (CDL) to Deciduous Forest	6,505,381	Woody Wetlands to Deciduous Forest	4,915,450
Other Hay/Non Alfalfa to Grassland/Pasture (CDL)	6,352,080	Other Hay/Non Alfalfa to Grassland/Pasture (CDL)	5,860,616	Deciduous Forest to Evergreen Forest	4,119,724
Deciduous Forest to Woody Wetlands	5,436,658	Deciduous Forest to Woody Wetlands	5,304,163	Grassland/Pasture (CDL) to Deciduous Forest	3,795,819
Woody Wetlands to Deciduous Forest	5,292,635	Woody Wetlands to Deciduous Forest	5,163,344	Evergreen Forest to Deciduous Forest	3,775,034

Exploring the year-to-year data plots and decadal differences begin to highlight interesting trends in the data and areas where models can be extended to further improve performance in estimating extent and change across the landscape.

Comparison among categories of the three model outputs for 2008 – 2017 as year-to-year z-score plots, show important temporal trends that may not be captured in a decadal change analysis, as well as illustrates how the process of subdividing data impacts z-scores (Figure 12, Figure 13, and Figure 14). Plots for grassland/pasture are highlighted here, with similar plots for all other categories in Appendix 2. These extensive charts pose inquiries into the drivers of fluctuations, as well as exposing possible relationships among categories of commonly rotated crops (i.e. corn and soybeans).

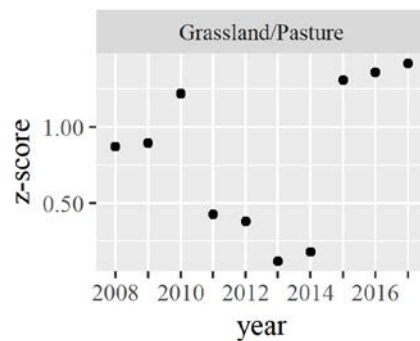


Figure 12: Annual z -scores for 2008-2017 model 1 grassland/pasture category.

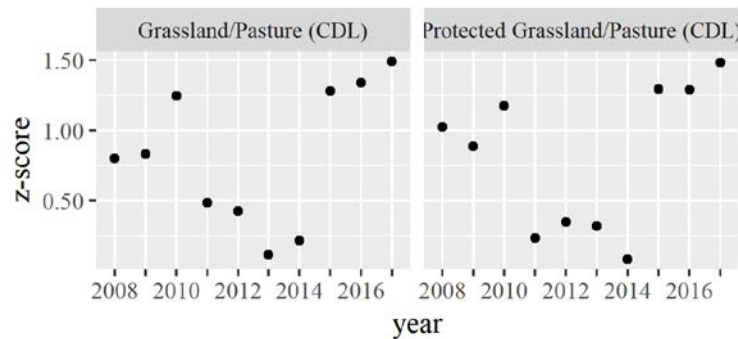


Figure 13: Annual z -scores for 2008-2017 model 2 grassland/pasture categories.

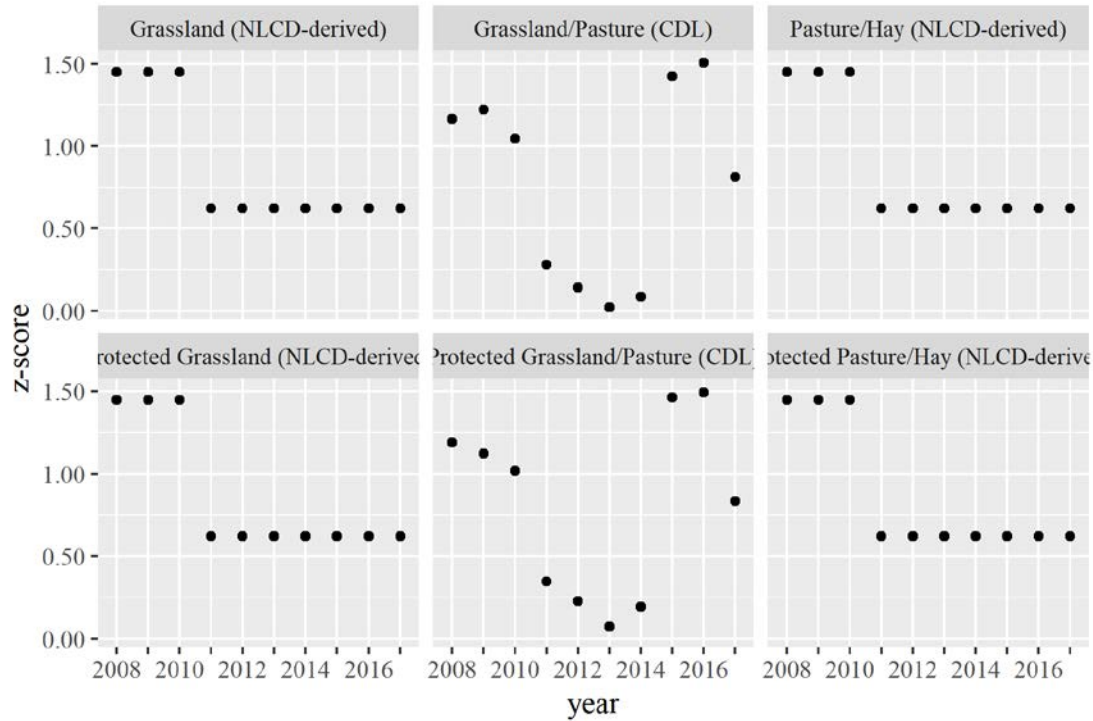


Figure 14: Annual z -scores for 2008-2017 model 3 grassland/pasture categories. Note the very abrupt changed in z -scores where notated as NLCD-derived, these mark the change in where years reference 2006 NLCD and 2011 NLCD.

The weaknesses of the analysis are also more notable in reviewing these charts, particularly those of Model 2 and Model 3. In both, datasets such as transportation and railroads, results are either presented as null or static. This results from the data being unchanging and universally applied to all years of analysis. Ideally, an improved version of this model would introduce temporally appropriate datasets for each year or a subset of years that better captures areas of increasing urbanization and creation of impermeable surfaces. For model 3, the NLCD data also creates interesting patterns in the temporal trends where NLCD is used to improve subdivision of grassland/pasture by further refining the grassland/pasture category into NLCD categories for grassland/herbaceous and pasture/hay, where the CDL category for grassland/pasture already exists on the landscape. For model 3, the 2017 data uses NLCD 2011

for this improved differentiation of grassland/pasture; whereas the 2008 data uses the NLCD 2006. The change in reference NLCD is notable on the year-to-year charts. No attempt to filter the data based on the quality of the pixel categorization has been attempted in the first product of the research. However, the general output of extent and acreage for each output is created that creates a base for future analysis.

An additional series of exploratory data were created to better understand how pixel confidence and frequency could be used to improve estimations of decadal change from 2008 to 2017. The charts in Figure 15, Figure 16, and Figure 17 show the distribution of the pixels for a given category based on confidence and shaded by frequency. As discussed in section 4.4, where the frequency is equal to 9 for a given pixel, all 10 years of data for that pixel were classified as the same category; whereas the confidence is a metric calculated by USDA NASS that provides a value of pixel fit to the CDL model tree. In this research, the measure that identifies lands of no change are those pixels where frequency is equal to 9 and confidence greater than or equal to 75. This may also serve as a preliminary measure for evaluating what areas may be undergoing change in the future (forecasting).

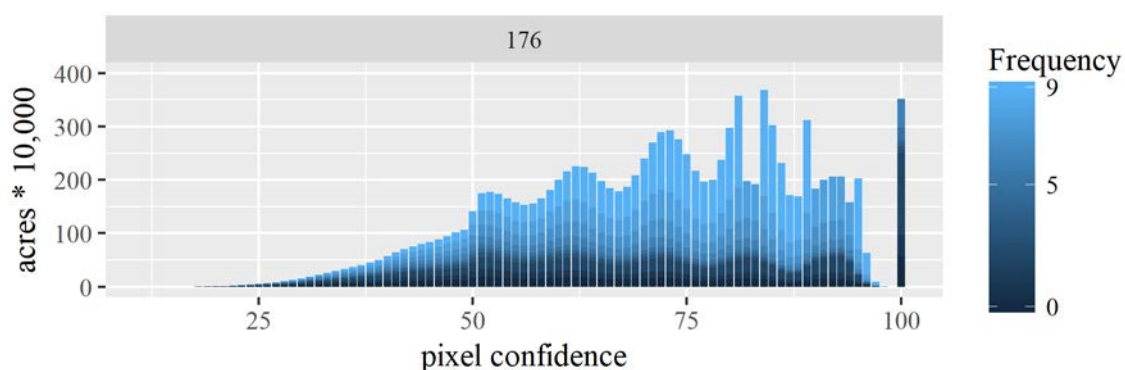


Figure 15: Model 1 distribution of acres of grassland/pasture by confidence and frequency for 2017. Data greater than 4 million acres have been omitted from plots to illustrate variation (4 records omitted).

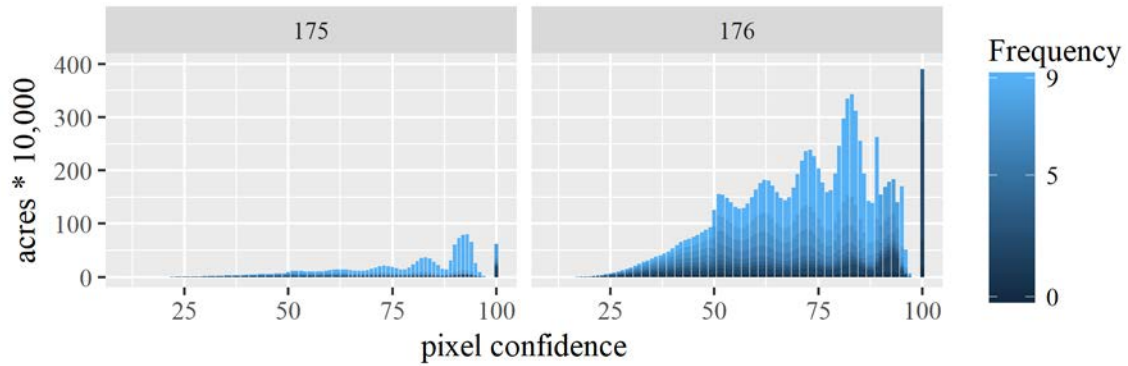


Figure 16: Model 2 distribution of acres of grassland/pasture by confidence and frequency for 2017. Data greater than 4 million acres have been omitted from plots to illustrate variation (4 records omitted).

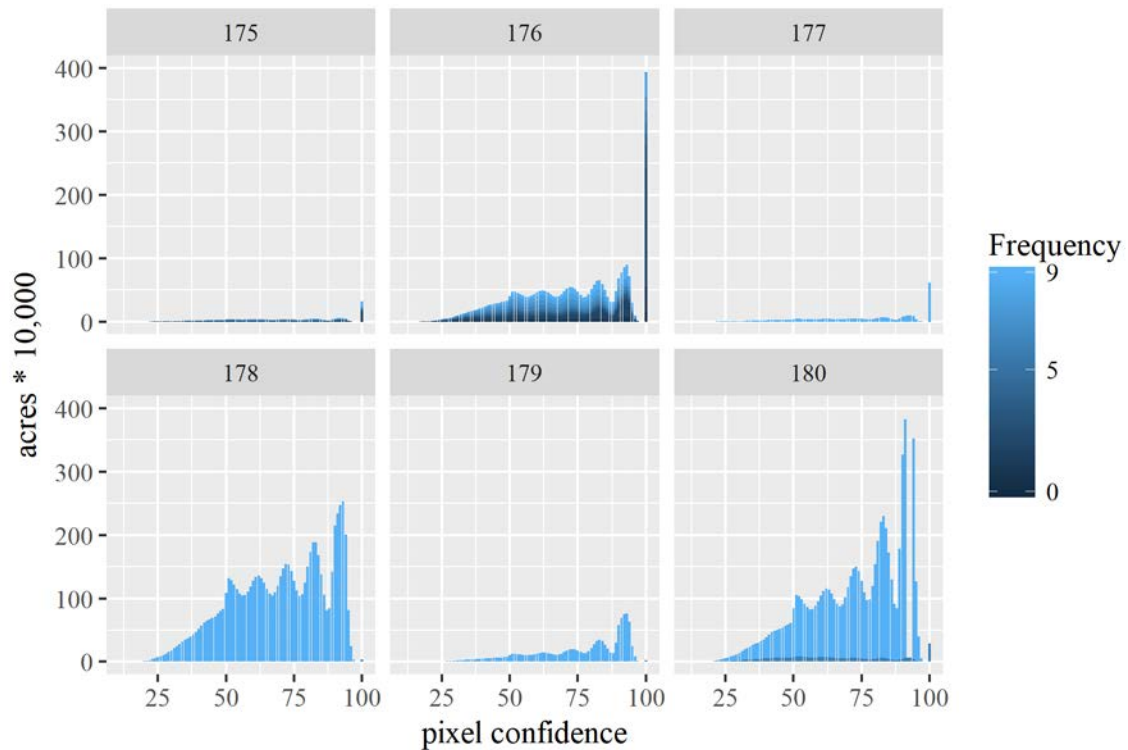


Figure 17: Model 3 distribution of acres of grassland/pasture by confidence and frequency for 2017. Data greater than 4 million acres have been omitted from plots to illustrate variation (2 records omitted).

To demonstrate the value of the model outputs that give a spatial extent to the areas where change has occurred, by category, over the past 10 years (2008 – 2017) the

grassland/pasture category will continue to serve as example. The grassland/pasture category has been initially discussed in section 2.3 of this document. Though grassland/pasture has been selected as example, these data exist for all categories of CDL and all three models. Each category has unique characteristics of uncertainty across that landscape, but also challenges of uncertainty across all categories as there is no way to observe one category without also observing all other categories in transactions in land conversion. The grassland/pasture category represents a non-crop category that is evaluated using CDL despite the lack of strength CDL has in non-crop category identification as discussed earlier. Given this, it is surprising that the primary land category selected to be evaluated for change in recent models using CDL (USDA NASS 2017). One special consideration in the grassland/pasture category is differentiating among actual change and change that has resulted from the purposeful improvement of models to better classify grassland/pasture within the CDL (USDA NASS 2017).

Initial land conversion data for 2008 – 2017 decadal change is a fully expanded matrix of all possible land conversions obtained by first combining the final model outputs of 2017 and 2008, then using the CV calculation to generate the change code described in section 4.4 to arrive at decadal change. For this example, the Extract by Value geoprocessing tool in ArcGIS is used to select and create a new raster of only those records that involve “Grassland/Pasture” from the decadal change raster. The model 1 extract for grassland/pasture is shown in Figure 18 and is symbolized by conversion of grassland/pasture gains (21,711,903 acres), losses (60,892,791 acres), and areas of no change (172,732,548 acres) that are calculated as the difference between 2017 and 2008. Map categories have been grouped based on gain or loss to improve readability. The data in Table 9 shows the individual losses and gains for each category demonstrating how grassland/pasture is lost to a category or gained from a category, or both.

Some of the categories have very small amounts of change in seemingly unlikely categories, as well as some large amounts of complementary change. These are indicators that some of the issues mentioned previously are present in the dataset.

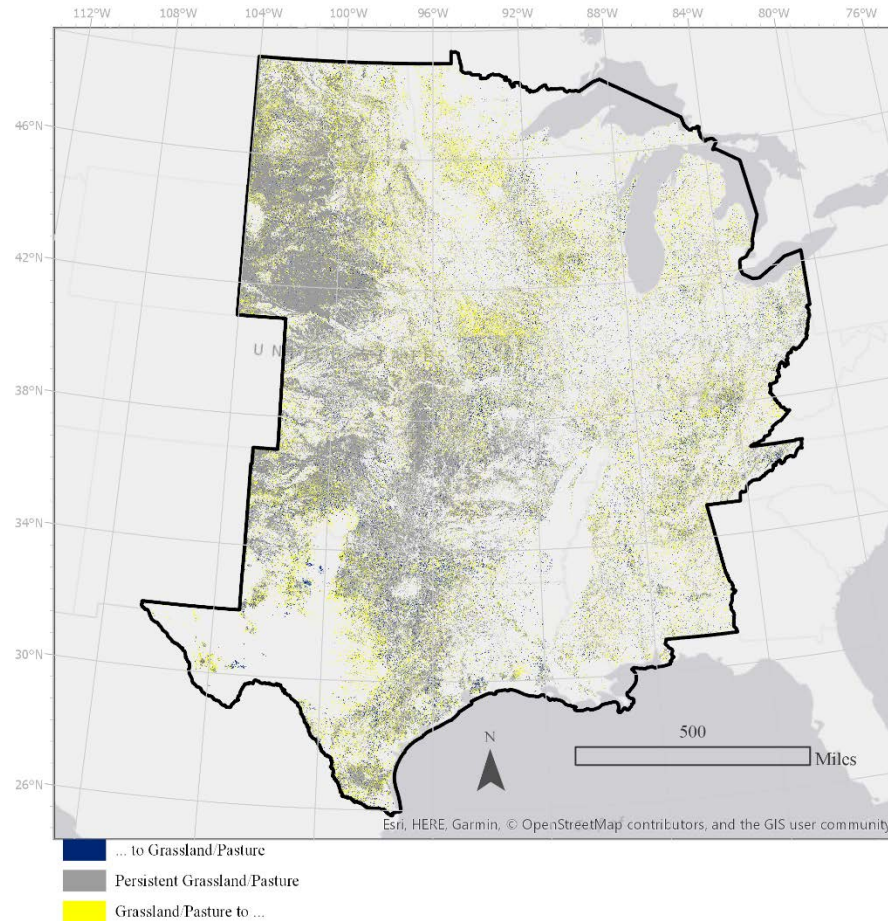


Figure 18: Conversion of lands in grassland/pasture category, 2008 -2017 (model 1).

For areas of that have a very small number of pixels of conversion, noise reduction analysis will need to be completed to possibly remove these areas if they are found to be single pixel areas of grassland. For areas of where a large number of pixels are converted as part of complementary change, we may be seeing error propagation as edge confusion. Edge confusion typically occurs where pixels along the edges of larger, more contiguous pixel patches are

equally likely to be classified as one category or the other. In the CDL edge confusion is prevalent along field edges in cropped areas but can also be found in non-crop areas.

Table 9: Grassland/Pasture gains and losses for 2008 – 2017, model 1 output.

Converted to Grassland/Pasture from Category (Gain)	Category	Grassland/Pasture Converted to Category (Loss)
939,779	Alfalfa	-2,219,165
23,406	Apples	-16,958
6,050	Aquaculture	-24,353
66	Asparagus	-440
34,765	Barley	-54,939
153,089	Barren	-465,394
1,948	Blueberries	-17,128
1	Broccoli	-6
309	Buckwheat	-2,119
258	Cabbage	-620
2	Camelina	0
14,228	Canola	-145,093
28	Cantaloupes	-126
151	Carrots	-919
5	Cauliflower	-131
3	Celery	-414
19,495	Cherries	-20,605
2,072	Christmas Trees	-23,842
1,462	Citrus	-1,286
5,051	Clover/Wildflowers	-13,609
238,636	Core Developed/Open Space	-777,036
1,944,708	Corn	-7,255,516
70,368	Cotton	-464,551
20	Cranberries	-6
518	Cucumbers	-7,055
16	Dbl Crop Barley/Corn	-123
60	Dbl Crop Barley/Soybeans	-489
58	Dbl Crop Corn/Soybeans	-74
61	Dbl Crop Oats/Corn	-364
325	Dbl Crop Soybeans/Oats	-957
3,422	Dbl Crop WinWht/Corn	-7,303
213	Dbl Crop WinWht/Cotton	-7,718
6,860	Dbl Crop WinWht/Sorghum	-9,522
101,859	Dbl Crop WinWht/Soybeans	-239,783
1,157,833	Deciduous Forest	-7,418,003
4,546	Developed/High Intensity	-31,076
35,891	Developed/Low Intensity	-283,838
16,642	Developed/Med Intensity	-144,405

Table 9 (Cont.)

Converted to Grassland/Pasture from Category (Gain)	Category	Grassland/Pasture Converted to Category (Loss)
14,375	Dry Beans	-70,102
77,967	Durum Wheat	-122,230
606,532	Evergreen Forest	-1,176,111
1,277,106	Fallow/Idle Cropland	-1,326,343
13,678	Flaxseed	-17,940
19	Gourds	0
3,969	Grapes	-6,207
550	Greens	-251
551,747	Herbaceous Wetlands	-2,819,634
493	Herbs	-2,850
14	Honeydew Melons	-14
0	Hops	-13
1,001	Lentils	-23,194
1	Lettuce	0
19,239	Millet	-81,199
27	Mint	-1
14	Misc Veggies & Fruits	-30
86,762	Mixed Forest	-209,806
545	Mustard	-528
5	Nectarines	0
106,895	Oats	-308,778
0	Olives	-137
387	Onions	-1,063
286,935	Open Water	-1,087,512
50	Oranges	-308
10,573	Other Crops	-5,304
6,352,080	Other Hay/Non Alfalfa	-8,042,320
36	Other Small Grains	-17
40	Other Tree Crops	-89
537	Peaches	-1,350
3,138	Peanuts	-26,527
9	Pears	-129
14,571	Peas	-52,251
7,602	Pecans	-30,077
109	Peppers	-208
4	Plums	-2
877	Pop or Orn Corn	-9,164
963	Potatoes	-10,759
307	Pumpkins	-2,584
4	Radishes	-193
19	Rape Seed	-26
25,981	Rice	-96,380
30,330	Rye	-102,960
808	Safflower	-3,099
3,274,336	Shrubland	-13,314,312

Table 9 (Cont.)

Converted to Grassland/Pasture from Category (Gain)	Category	Grassland/Pasture Converted to Category (Loss)
103,632	Sod/Grass Seed	-76,490
396,141	Sorghum	-562,992
1,473,358	Soybeans	-7,395,798
280	Speltz	-481
241,259	Spring Wheat	-957,834
317	Squash	-2,753
45	Strawberries	-107
641	Sugarbeets	-12,755
15,766	Sugarcane	-22,986
49,136	Sunflower	-169,441
1,954	Sweet Corn	0
941	Sweet Potatoes	-3,826
803	Switchgrass	-2,433
1,950	Tobacco	-10,959
773	Tomatoes	-604
3,243	Triticale	-27,600
43	Turnips	-75
78	Vetch	-2
15	Walnuts	-626
1,527	Watermelons	-2,792
1,587,036	Winter Wheat	-1,988,208
278,131	Woody Wetlands	-1,047,063
21,711,903	Total	-60,892,791

As example, for identifying the uncertainty in the model result, the frequency and confidence data compiled earlier for model 2 can be used to better categorize the data and the uncertainty that accompanies the data. Using the frequency and confidence mask to further classify the model output, a version of the data is created that can be reviewed by category, conversion, frequency, and confidence. For the model 2 data, of the 185,979,265 acres of lands classified as persistent grassland/pasture and grassland/pasture net gains, approximately 49% (92,549,496 acres) have a frequency = 9 and confidence ≥ 75 . When considering only the confidence of the data, approximately 70% of model 2 persistent grassland/pasture or lands converted to grassland/pasture during the past decade fall within the confidence threshold greater than or equal to 75. The model 2 grassland conversion losses are calculated at 57,675,446 acres

with 24% being attributed to conversion to corn or soybean croplands (of any NCCPI category). The full distribution of the pixel frequency for this data is found in Figure 19, and pixel confidence in Figure 20.

Using the combination of confidence and frequency two observations can be made: 1) areas where change is more likely to have occurred; and 2) areas where change is less likely to be occurring on the landscape. These data can be used to further characterize and select appropriate thresholds at which to consider a pixel not usable for analysis, or by simply acknowledging limitations of use. However, in reviewing these examples of grassland/pasture observations there is a difference in total acres of analyzed lands between model 1 (~133.5 billion acres) and model 2 (~128.3 billion acres). This represents a reduction of total grassland/pasture from model 1 acreages of almost 4%. These differences are due to the enhancement and augmentation process that is undertaken in the model, for example in processes of shrinking and expanding, and in overlaying road networks and transmission lines.

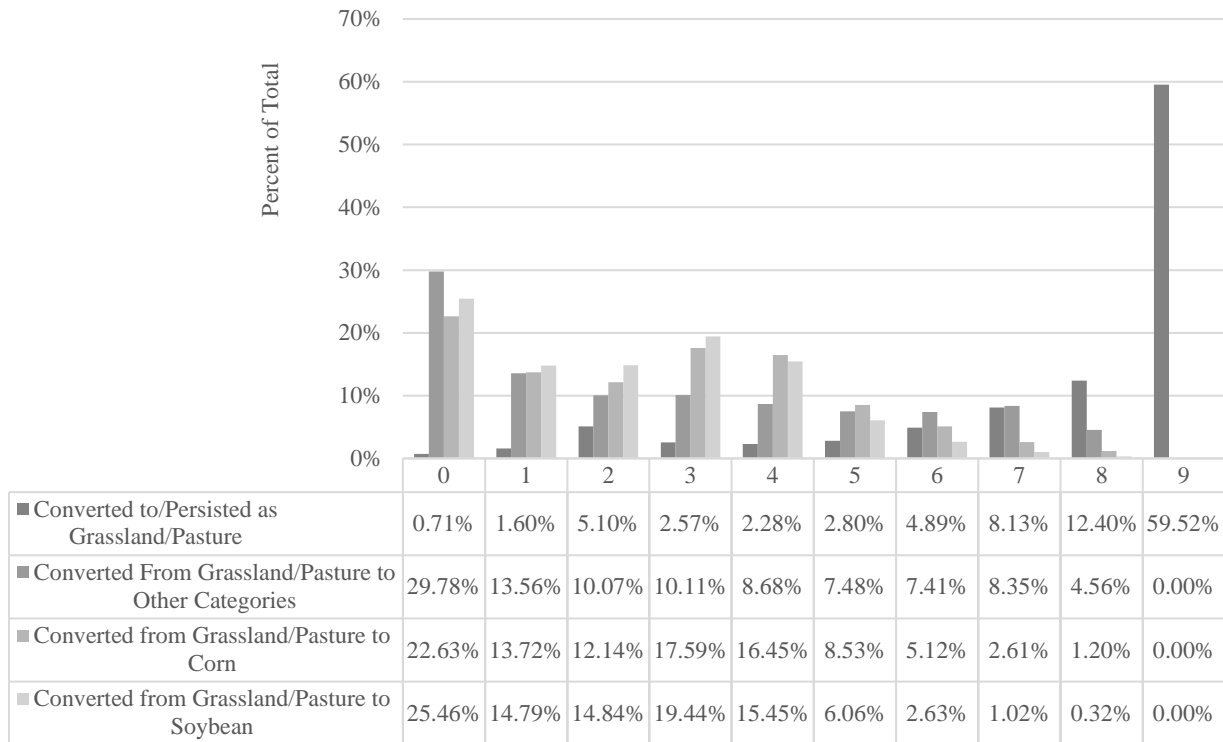


Figure 19: Grassland Pasture Conversion (Model 2 Output) by Pixel Frequency, 2008 - 2017.

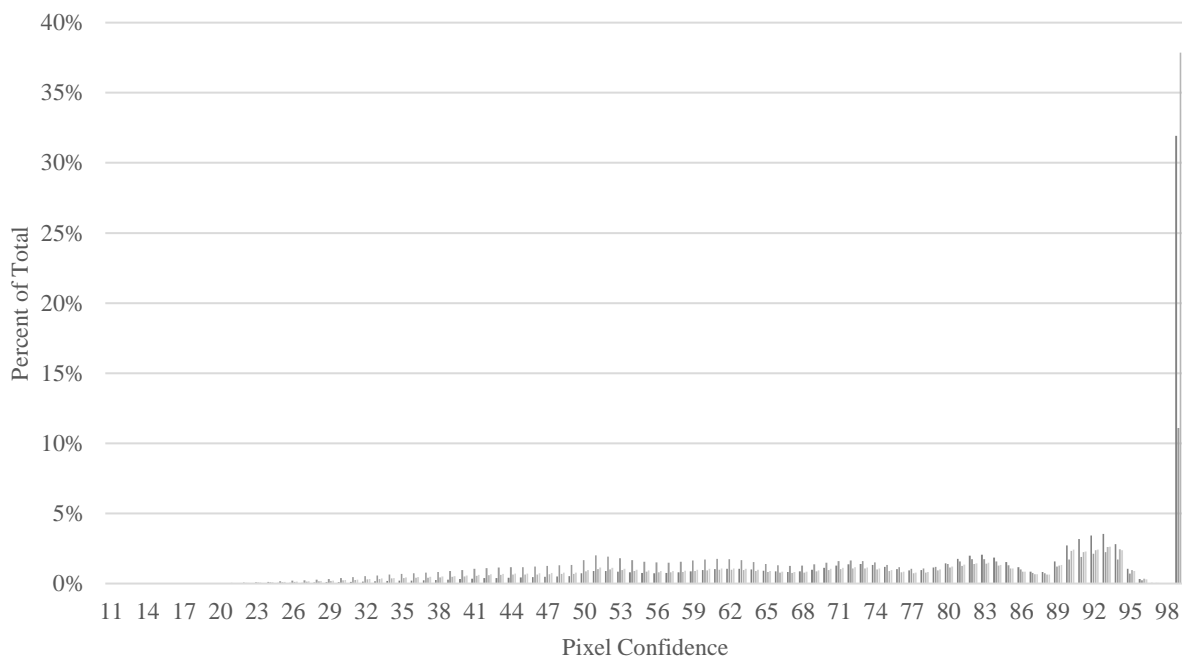


Figure 20: Grassland Pasture Conversion (Model 2 Output) by Pixel Confidence, 2008 - 2017.

Another method explored in this research, and proposed for further research, is applying region group (clump) operations to remove single pixel and small areas of pixels from the dataset. Single pixels are those that share no connectivity to another pixel of the same value by a neighbor connection of eight. Given the 30-meter pixel resolution, it takes approximately 4.5 pixels to be equivalent to one acre. In a trial run, using the model 2 grassland/pasture conversion dataset, pixel clumps composed of 4 or fewer adjacent pixels, were found to be 0.46% (1,131,968 acres), with 0.14% being single pixels. Lands converted to grassland/pasture or persisting as grassland/pasture had a mean pixel clump of 5,484 acres, while those converted from grassland/pasture to another category has a mean pixel clump of 12,603 acres. However, it is still unclear whether these differences point to characterization of grassland/pasture change or change in the data processing that identifies grassland/pasture change. The next step in this research explores the relationship among pixel clumps, confidence, and frequency to evaluate additional methods for quantifying uncertainty in CDL datasets.

The outputs of the model data can also be reviewed to locate specific types of conversion of a category. For example, where are grassland/pasture lands that have been converted to low and marginal corn and soybean production? Where are grassland/pasture lands that have been converted to any corn and soy production, regardless of NCCPI? Using the model 2 grassland/pasture conversion dataset, areas where grassland/pasture has seen conversion to low or marginal corn and soybean production (Figure 21) is identified. The areas identified in Figure 21 could be targeted for incentivized programs to return those areas to fallow while still supporting the economic interests of the farm. Figure 22 shows the extent of converted grassland/pasture to any corn or soybean production, including all levels of NCCPI. However, this data must be used with caution, while we show these lands as conversion to corn or soy and

as low or marginal production these conditions are only valid for non-irrigated cropland. This enhanced data provides a next-level planning approach for land managers seeking to prioritize conservation delivery, but still leaves many gaps that must be filled with local knowledge. One question that may arise from these analyses is the amount of area converted because of declines in CRP enrollment and contracts on existing CRP lands expiring. These models are not capable of making these estimations as CRP data are not part of the augmented data layer set. However, if appropriate data were available this is an area that could be explored in the future.

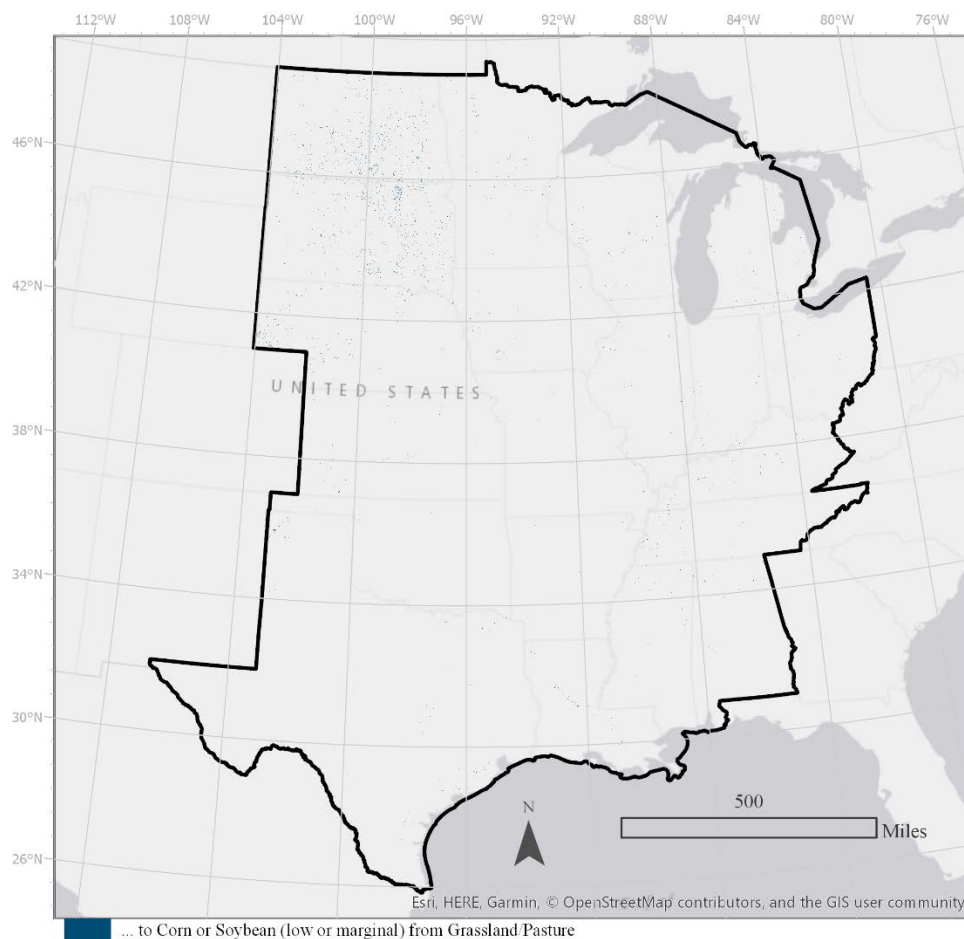


Figure 21: Conversion of grassland/pasture to low and marginal corn and soybean production lands between 2008 and 2017 (model 2).

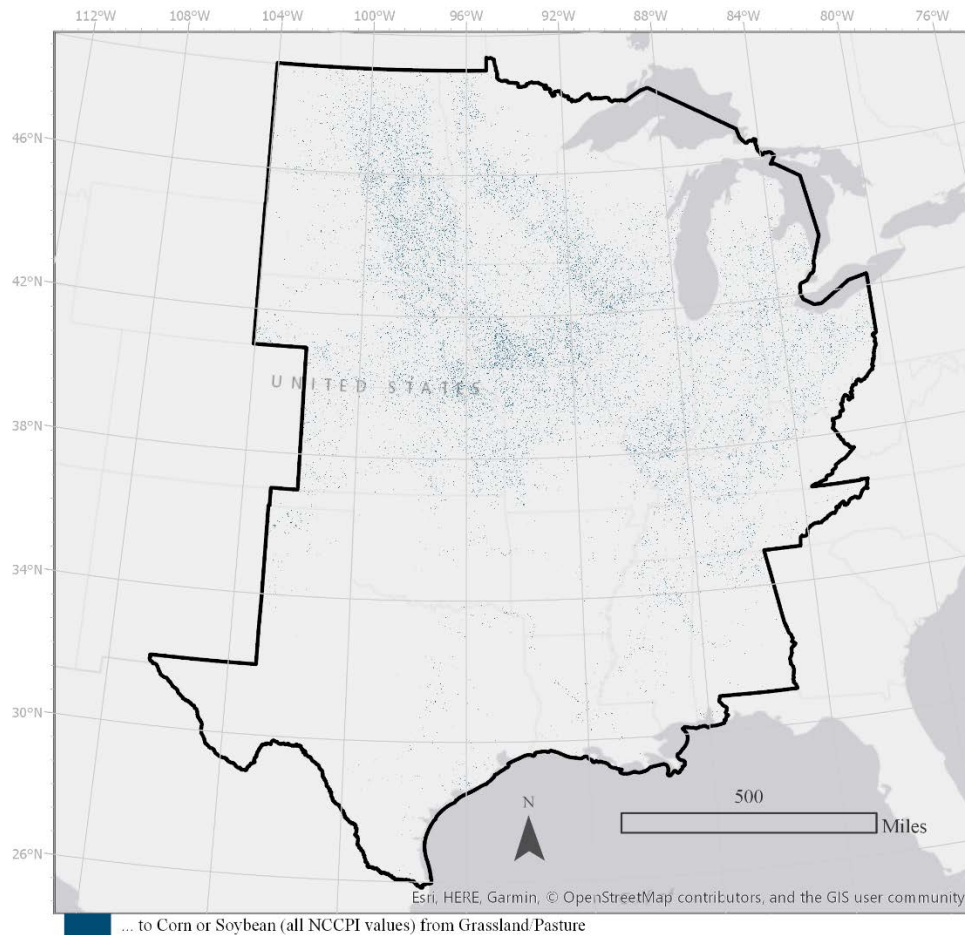


Figure 22: Conversion of grassland/pasture any corn or soybean production lands between 2008 and 2017 (model 2).

The conservation policy and planning impacts of these models would possibly alter the scenarios given in Thogmartin et al 2017, first by enabling an extension of scenarios to a greater area, through re-distributing the conservation delivery within the north central region by re-defining “marginal” croplands, and by expanding the categorization of the landscape.

Thogmartin et al. 2017 discuss conversion of marginal croplands (corn or soybean lands with NCCPI less than 0.40) in corn and soybean lands totaling 50,329 square kilometers (12,436,567 acres) over the study area in their research, which by calculations consistent with those used in this model exceeds 10% of the lands in production of corn and soybean in their study area for

CDL year 2014. Comparatively, the area in this study that would be similarly identified under the conditions of model 2 or 3 would call for possible conversion of low or marginal production lands in corn and soybean with an NCCPI less than or equal to 0.225, equal to 6,730,855 acres in 2014, approximately 4.5% of the total corn and soybean lands in production in the central United States (the study area of this research). Alternatively, if moderate productivity lands (corn or soybean lands with NCCPI less than 0.475, greater than 0.225) are also included as marginal, then the proposed percentage increases to 18.1% of total lands in corn and soybean production within this study area, and to approximately 12% in the study area defined by Thogmartin et al. 2017. In both cases, the higher rate of conversion would undoubtedly be too aggressive even with the best of conservation program incentives.

The NCCPI breaks and other model parameters should be considered as flexible and only serve as a point from which modifications can be made to improve such models to better fit intentions and identification of lands that would most be likely to have a conservation program implemented, which may or may not include low or marginal productivity lands.

6. Discussion

After reviewing the three models, a revision of model 2 is likely to be the best suited to provide the input needed through the remainder of the project objectives. Several improvements could be made that will best provide for the objectives in the next phase of research. These improvements are: 1) correct for land cover type loss due to analysis, such as with the difference of grassland/pasture acres analyzed lands; 2) replace processes, such as shrink and grow, with more controlled selection and replacement through pixel cluster counts and neighbor associations; 3) eliminate model 3 processes and output that includes NLCD lands for further subdividing grassland/pasture data that might best be done at the conclusion of data creation on a

user-by-user basis, rather than broadly applied within the model, especially given the increase in model runtime for accommodating these datasets (model 2: runtime/model = ~ 2.5 hours; model 3: runtime/model = ~ 3.5 hours); and 4) expand transportation network datasets to include multiple years for better temporal characterization of impervious surfaces and opportunities for right-of-way activities.

Additional research to estimate monarch butterfly habitat, arguably the most notable definition absent from these models, is needed to be able to select specific land conversions relevant to monarch research. In researching how to best represent monarch butterfly habitat, it ultimately is not a universal decision that can be applied broadly without respect to life cycle, varying resource needs, and their associated spatial and temporal extents. While additional research is needed to account for many site-specific variables, including the unique needs of each landowner, these models do provide foundational data that can feed larger decision-making matrices in conservation delivery, and provide more easily accessible data in support of identifying incentivized paths to conservation action.

Each model is prepared to support inquiries from a broader-scope of environmental and ecological applications, expanding the impact of the development of these processes and data. The outputs, and many of the methods, of this research could be built into an online interface to 1) help landowners locate programs based on land composition in their area; and 2) help conservation planners identify where resources have been changed on the landscape, resulting in more targeted action, for example by aggregating results to commonly used planning boundaries, such as hydrologic units or counties as in Figure 23.

In addition to the proposed improvements in the models, future research objectives include: 1) creating year-to-year conversion matrices (as opposed to decadal) that provides a

more comprehensive temporal scale from which patterns in land use can be evaluated; 2) converting analysis to take advantage of parallel computing processing environments, where appropriate; 3) perform region group (clump) analysis of each category within the model outputs to improve the characterization of uncertainty; and, 4) development of landscape metrics for each category of land, providing an overall assessment of how the configuration of the landscape has changed over the past 10 years with increasing urbanization and landscape modifications. Data from reviewing data created through these models and in support of the decision-making process can be used in opportunistic determinations of candidate lands that are serviceable by identifying enrollment in existing programs and using those data to identify similar areas. Such programs may be successful in support of further simplifying the process of linking landowners, land managers, and conservation programs.

The challenge in conservation delivery is using best available data and resources to help guide policy and land manager decisions, even when those data are not designed for, or intended for those uses. There is always further research and additional data that can be collected, but similar to discussions in Rudel et al. 2005 on forest transitions and government policy, policymakers often will not wait patiently for results or research to be completed before action must be taken, and in case of the monarch butterfly, time has been given for more thorough investigations; however, a deadline for policy and directives still exists that precludes creation of any new data, or data findings that could be incorporated in these models. As policy-driven research continues to grow we will continue to be faced with challenges of wisely integrating and manipulating existing data to solve emerging problems.

These models, along with the contributions of many other collaborators, citizen scientists, state and federal agencies, private organizations, and non-profits; provide valuable information

for the U.S. Fish and Wildlife Service as part of a summary of ongoing activities supporting monarch and pollinator conservation throughout the United States, and ultimately as a decision support tool to aid in rendering a decision on the status of *Danaus plexippus* by June 30, 2019.

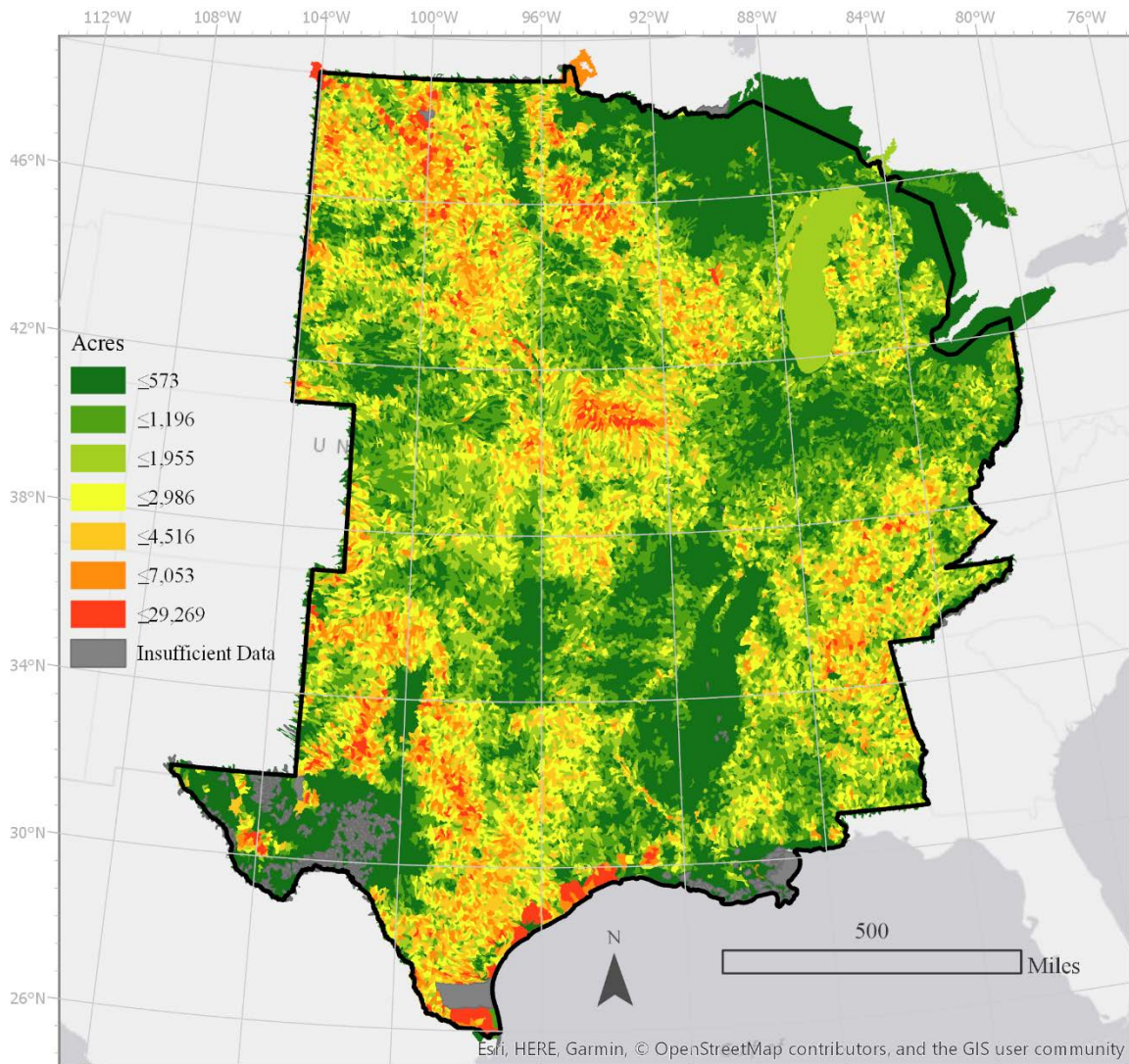


Figure 23: Conversion of grassland to other categories (loss) between 2008 and 2017 as aggregated to HUC12 that intersect the study area and symbolized by number of acres of grasslands converted to other categories using natural breaks in the data.

References

- “Cropland Data Layer.” 2017. Published crop-specific data layer [Online]. Washington, DC.: USDA National Agricultural Statistics Service. Available at <https://nassgeodata.gmu.edu/CropScape/>.
- “Electric Power Transmission Lines.” 2014. Spatial Dataset. Homeland Infrastructure Foundation Level Database Open GP - Energy. U.S. Department of Homeland Security. <https://hifld-geoplatform.opendata.arcgis.com/datasets/electric-power-transmission-lines>.
- “Gridded Soil Survey Geographic (GSSURGO) Database for the Conterminous United States.” 2017. United States Department of Agriculture, Natural Resources Conservation Service. <https://gdg.sc.egov.usda.gov/>.
- Agrawal, Anurag A. 2017. *Monarchs and Milkweed: A Migrating Butterfly, a Poisonous Plant, and Their Remarkable Story of Coevolution*. Princeton: Princeton University Press.
- Decision Innovative Solutions. 2013. “2013 Multi-State Land-Use Study: Estimated Land-Use Changes 2007-2012.” Urbandale, IA 50322: Decision Innovative Solutions. <http://www.decision-innovation.com/webres/File/docs/130715%20Multi-State%20Land%20Use%20Report.pdf>.
- Fry, Joyce, George Z. Xian, Suming Jin, Jon Dewitz, Collin G. Homer, Limin Yang, Christopher A. Barnes, N.D. Herold, and J.D. Wickham. 2011. “Completion of the 2006 National Land Cover Database for the Conterminous United States.” *Photogrammetric Engineering and Remote Sensing* 77 (9): 858–64.
- Gustafsson, Karin M., Anurag A. Agrawal, Bruce V. Lewenstein, and Steven A. Wolf. 2015. “The Monarch Butterfly through Time and Space: The Social Construction of an Icon.” *BIOSCIENCE* 65 (6): 612–22. <https://doi.org/10.1093/biosci/biv045>.
- Homer, Collin G., Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Z. Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. “Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information.” *Photogrammetric Engineering and Remote Sensing* 81: 345–54.
- Inamine, Hidetoshi, Stephen P. Ellner, James P. Springer, and Anurag A. Agrawal. 2016. “Linking the Continental Migratory Cycle of the Monarch Butterfly to Understand Its Population Decline.” *Oikos* 125 (8): 1081–91. <https://doi.org/10.1111/oik.03196>.
- Jin, Suming, Limin Yang, Patrick Danielson, Collin Homer, Joyce Fry, and George Xian. 2013. “A Comprehensive Change Detection Method for Updating the National Land Cover Database to circa 2011.” *Remote Sensing of Environment* 132 (May): 159–75. <https://doi.org/10.1016/j.rse.2013.01.012>.
- Koh, Insu, Eric V. Lonsdorf, Neal M. Williams, Claire Brittain, Rufus Isaacs, Jason Gibbs, and Taylor H. Ricketts. 2016. “Modeling the Status, Trends, and Impacts of Wild Bee Abundance in the United States.” *Proceedings of the National Academy of Sciences* 113 (1): 140–45. <https://doi.org/10.1073/pnas.1517685113>.
- Laingen, Chris. 2015. “Measuring Cropland Change: A Cautionary Tale.” *Papers in Applied Geography* 1 (1): 65–72. <https://doi.org/10.1080/23754931.2015.1009305>.

- Lemoine, Nathan P. 2015. "Climate Change May Alter Breeding Ground Distributions of Eastern Migratory Monarchs (*Danaus Plexippus*) via Range Expansion of *Asclepias* Host Plants." Edited by Robert B. Srygley. PLOS ONE 10 (2): e0118614. <https://doi.org/10.1371/journal.pone.0118614>.
- Maeckle, Monika. 2016. "US Fish and Wildlife Service Gets Three More Years to Evaluate Monarch Butterfly ESA Status." Texasbutterflyranch. July 6, 2016. <http://texasbutterflyranch.com/2016/07/06/us-fish-and-wildlife-service-gets-three-more-years-to-evaluate-monarch-butterfly-esa-status/>.
- Rick Mueller, and Mark Harris. 2013. "Reported Uses of CropScape and the National Cropland Data Layer Program." Conference Report. Sixth International Conference on Agricultural Statistics. Fairfax, VA: USDA National Agricultural Statistics Service, Research and Development Division. https://www.nass.usda.gov/Research_and_Science/Cropland/docs/MuellerICASVI_CDL.pdf.
- "National Value Added Look Up (Valu) Table Database for the Gridded Soil Survey Geographic (GSSURGO) Database for the United States of America and the Territories, Commonwealths, and Island Nations Served by the USDA-NRCS." 2017. United States Department of Agriculture, Natural Resources Conservation Service. <https://gdg.sc.egov.usda.gov/>.
- Navarrete, José-Luis, M. Isabel Ramírez, and Diego R. Pérez-Salicrup. 2011. "Logging within Protected Areas: Spatial Evaluation of the Monarch Butterfly Biosphere Reserve, Mexico." *Forest Ecology and Management* 262 (4): 646–54. <https://doi.org/10.1016/j.foreco.2011.04.033>.
- Oberhauser, Karen, Ruscena Wiederholt, Jay E. Diffendorfer, Darius Semmens, Leslie Ries, Wayne E. Thogmartin, Laura Lopez-Hoffman, and Brice Semmens. 2017. "A Trans-National Monarch Butterfly Population Model and Implications for Regional Conservation Priorities: Regional Monarch Conservation Priorities." *Ecological Entomology* 42 (1): 51–60. <https://doi.org/10.1111/een.12351>.
- Pleasants, John M., Ernest H. Williams, Lincoln P. Brower, Karen S. Oberhauser, and Orley R. Taylor. 2016. "Conclusion of No Decline in Summer Monarch Population Not Supported." *Annals of the Entomological Society of America* 109 (2): 169–71. <https://doi.org/10.1093/aesa/sav115>.
- President Barack Obama. 2014. "Presidential Memorandum -- Creating a Federal Strategy to Promote the Health of Honey Bees and Other Pollinators." Whitehouse.Gov. June 20, 2014. <https://www.whitehouse.gov/the-press-office/2014/06/20/presidential-memorandum-creating-federal-strategy-promote-health-honey-b>.
- "Protected Areas Database of the United States (PAD-US)." 2016. Combined Feature Class. Gap Analysis Program (GAP). U.S. Geological Survey. <https://gapanalysis.usgs.gov/padus/data/download/>.
- Rader, Romina, Ignasi Bartomeus, Lucas A. Garibaldi, Michael P. D. Garratt, Brad G. Howlett, Rachael Winfree, Saul A. Cunningham, et al. 2016. "Non-Bee Insects Are Important Contributors to Global Crop Pollination." *Proceedings of the National Academy of Sciences* 113 (1): 146–51. <https://doi.org/10.1073/pnas.1517092112>.

- Reitsma, Kurtis D., David E. Clay, Sharon A. Clay, Barry H. Dunn, and Cheryl Reese. 2016. "Does the U.S. Cropland Data Layer Provide an Accurate Benchmark for Land-Use Change Estimates?" *Agronomy Journal* 108 (1): 266. <https://doi.org/10.2134/agronj2015.0288>.
- Ries, Leslie, Douglas J. Taron, and Eduardo Rendón-Salinas. 2015. "The Disconnect Between Summer and Winter Monarch Trends for the Eastern Migratory Population: Possible Links to Differing Drivers." *Annals of the Entomological Society of America* 108 (5): 691–99. <https://doi.org/10.1093/aesa/sav055>.
- "Roads National Geodatabase." 2016. Geodatabase. TIGER Geodatabases. US Census Bureau Geography. <https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html>.
- Rudel, Thomas K., Oliver T. Coomes, Emilio Moran, Frederic Achard, Arild Angelsen, Jianchu Xu, and Eric Lambin. 2005. "Forest Transitions: Towards a Global Understanding of Land Use Change." *Global Environmental Change* 15 (1): 23–31. <https://doi.org/10.1016/j.gloenvcha.2004.11.001>.
- Sáenz-Romero, Cuauhtémoc, Gerald E. Rehfeldt, Pierre Duval, and Roberto A. Lindig-Cisneros. 2012. "Abies Religiosa Habitat Prediction in Climatic Change Scenarios and Implications for Monarch Butterfly Conservation in Mexico." *Forest Ecology and Management* 275 (July): 98–106. <https://doi.org/10.1016/j.foreco.2012.03.004>.
- The Center for Biological Diversity. 2016a. "Agreement: Monarch Butterfly to Get Endangered Species Act Protection Decision by 2019." August 3, 2016. https://www.biologicaldiversity.org/news/press_releases/2016/monarch-butterfly-07-05-2016.html.
- . 2016b. "Lawsuit Launched for Endangered Species Act Protection of Monarch Butterflies." August 3, 2016. https://www.biologicaldiversity.org/news/press_releases/2016/monarch-butterfly-01-05-2016.html.
- Thogmartin, Wayne E, Laura Lopez-Hoffman, Jason Rohweder, Jay Diffendorfer, Ryan Drum, Darius Semmens, Scott Black, et al. 2017. "Restoring Monarch Butterfly Habitat in the Midwestern US: ?All Hands on Deck?" *Environmental Research Letters* 12 (7): 074005. <https://doi.org/10.1088/1748-9326/aa7637>.
- "Two-Way Monarch Migration Map." 2010. Graphic interpretation. Monarch Butterfly: Fall and Spring Migration. MonarchWatch.org. <http://monarchwatch.org/blog/2010/05/13/two-way-monarch-migration-map/>.
- USDA National Agricultural Statistics Service. 2017. "Cropscape and Cropland Data Layer." August 3, 2016. https://www.nass.usda.gov/Research_and_Science/Cropland/Release/.
- USDA U.S. Forest Service. 2017. "Pollinators." October 28, 2017. <https://www.fs.fed.us/wildflowers/pollinators/>.
- U.S. Fish and Wildlife Service. 2015. "U.S. Monarch Butterfly Conservation Strategy: An Overview." PowerPoint presented at the XX Annual Trilateral Committee Meeting, San Diego, California, April 15. <https://www.trilat.org/index.php?view=download&alias=500-2015-ec->

usmonbutconsstra&category_slug=2015-annual-meeting&option=com_docman&Itemid=254.

———. 2017. “Status of the Monarch Butterfly.” August 29, 2017.
<https://www.fws.gov/savethemonarch/SSA.html>.

U.S. Fish and Wildlife Service Midwest Region. 2016. “Service Initiates Status Review of Monarch Butterfly under the Endangered Species Act.” August 3, 2016.
<https://www.fws.gov/midwest/news/764.html>.

Vidal, Omar, and Eduardo Rendón-Salinas. 2014. “Dynamics and Trends of Overwintering Colonies of the Monarch Butterfly in Mexico.” *Biological Conservation* 180 (December): 165–75. <https://doi.org/10.1016/j.biocon.2014.09.041>.

Wright, C. K., and M. C. Wimberly. 2013. “Recent Land Use Change in the Western Corn Belt Threatens Grasslands and Wetlands.” *Proceedings of the National Academy of Sciences* 110 (10): 4134–39. <https://doi.org/10.1073/pnas.1215404110>.

Appendix

1. Data Layers Specifics Required for Models

Cropland Data Layers 2008 – 2017 raster data sets. In December 2017, CDL for 2008 and 2009 was re-released at 30 meter pixel resolution. In January 2018, CDL 2017 was released. Models were updated to include these data given the updated data availability. USDA National Agricultural Statistics Service. 2017. “Cropscape and Cropland Data Layer.” August 3, 2016. https://www.nass.usda.gov/Research_and_Science/Cropland/Release/.

Cropland Data Layer National Confidence Layer 2008, 2009, 2016, and 2017 raster data sets. In December 2017, CDL for 2008 and 2009 was re-released at 30 meter pixel resolution, in addition to the re-release of CDL for these years National Confidence Layers were generated. Data were originally organized as single state layers; for the purposes of this research the National Confidence Layers for each year were mosaicked. USDA National Agricultural Statistics Service. 2017. “Cropscape and Cropland Data Layer.” August 3, 2016. https://www.nass.usda.gov/Research_and_Science/Cropland/Release/.

Cropland Data Layer Equal to Frequency (Derived) 2017 raster data set. Calculated using Cropland Data Layers 2008 – 2017 using the Equal to Frequency tool in ArcGIS Pro 2.1.2 where 2017 is the “Input Value Raster” and 2016 – 2008 are each members of the list of raster data sets that will be compared against 2017.

Electric Power Transmission Lines (2014 – 2017), specifically using data for spatial presence of transmission lines throughout the study area. <https://hifld-geoplatform.opendata.arcgis.com/datasets/electric-power-transmission-lines>.

Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States 2017 used with the National Value Added Look Up (Valu1) Table Database for the gSSURGO Database for the United States of America and the Territories, Commonwealths, and

Island Nations Served by the USDA-NRCS 2017. Specifically, accessing the 30m gSSURGO grid data set for the areas within the broad study area delineation and the following “ready-to-map” crop productivity indices for the areas within the broad study area. Fields include: nccpi2co, nccpi2cs, and nccpi2sg. <https://gdg.sc.egov.usda.gov/>.

National Land Cover Database 2011 (Homer et al. 2015) and National Land Cover Database 2006 (Fry et al. 2011). Select classes were used from the NLCD data, including 52 – Shrub/Scrub; 71 – Grassland/Herbaceous; and 81 – Pasture/Hay. <https://www.mrlc.gov/finddata.php>

Protected Areas Database of the United States (PAD-US) 2016, specifically the extent designations within the ‘\PADUS4\textunderscore1Combined Feature Class’ that serves as the nation’s official inventory of protected areas in the standard PAD-US framework, including management, easements, fee, and marine protected areas. <https://gapanalysis.usgs.gov/padus/data/download/>.

Roads National Geodatabase 2016, specifically using data for spatial presence of transportation data throughout study area. <https://www.census.gov/geo/maps-data/data/tiger-geodatabases.html>.

2. List of all Categories in CDL, original and enhanced

Table 10: Table of all CDL original and enhanced versions of data.

CDL Code	Original CDL Category	Enhanced CDL Category
1	Corn	**see corn NCCPI**
2	Cotton	Cotton
3	Rice	Rice
4	Sorghum	Sorghum
5	Soybeans	**see soybean NCCPI**
6	Sunflower	Sunflower
10	Peanuts	Peanuts
11	Tobacco	Tobacco
12	Sweet Corn	**see corn NCCPI**
13	Pop or Orn Corn	**see corn NCCPI**
14	Mint	Mint
21	Barley	**see small grains NCCPI**
22	Durum Wheat	**see small grains NCCPI**
23	Spring Wheat	**see small grains NCCPI**
24	Winter Wheat	**see small grains NCCPI**
25	Other Small Grains	**see small grains NCCPI**
26	Dbl CropWinWht/Soybeans	Dbl Crop WinWht/Soybeans
27	Rye	**see small grains NCCPI**
28	Oats	**see small grains NCCPI**
29	Millet	Millet
30	Speltz	Speltz
31	Canola	Canola
32	Flaxseed	Flaxseed
33	Safflower	Safflower
34	Rape Seed	Rape Seed
35	Mustard	Mustard
36	Alfalfa	Alfalfa
37	Other Hay/Non Alfalfa	Other Hay/Non Alfalfa
38	Camelina	Camelina
39	Buckwheat	Buckwheat
41	Sugarbeets	Sugarbeets
42	Dry Beans	Dry Beans
43	Potatoes	Potatoes
44	Other Crops	Other Crops
45	Sugarcane	Sugarcane
46	Sweet Potatoes	Sweet Potatoes
47	Misc Veggies & Fruits	Misc Veggies & Fruits
48	Watermelons	Watermelons
49	Onions	Onions
50	Cucumbers	Cucumbers
51	Chick Peas	Chick Peas

Table 10 (Cont.)

Value	Original CDL Category	Enhanced CDL Category
52	Lentils	Lentils
53	Peas	Peas
54	Tomatoes	Tomatoes
55	Caneberries	Caneberries
56	Hops	Hops
57	Herbs	Herbs
58	Clover/Wildflowers	Clover/Wildflowers
59	Sod/Grass Seed	Sod/Grass Seed
60	Switchgrass	Switchgrass
61	Fallow/Idle Cropland	Fallow/Idle Cropland
62	Pasture/Grass	Pasture/Grass
63	Forest	Forest
64	Shrubland	Shrubland
65	Barren	Barren
66	Cherries	Cherries
67	Peaches	Peaches
68	Apples	Apples
69	Grapes	Grapes
70	Christmas Trees	Christmas Trees
71	Other Tree Crops	Other Tree Crops
72	Citrus	Citrus
74	Pecans	Pecans
75	Almonds	Almonds
76	Walnuts	Walnuts
77	Pears	Pears
81	Clouds/No Data	Clouds/No Data
82	Developed	Developed
83	Water	Water
87	Wetlands	Wetlands
88	Nonag/Undefined	Nonag/Undefined
92	Aquaculture	Aquaculture
111	Open Water	Open Water
112	Perennial Ice/Snow	Perennial Ice/Snow
120		Linear Developed/Open Space
121	Developed/Open Space	Core Developed/Open Space
122	Developed/Low Intensity	Developed/Low Intensity
123	Developed/Med Intensity	Developed/Med Intensity
124	Developed/High Intensity	Developed/High Intensity
125		TIGER Primary and Secondary Roads
126		Local Neighborhood Road, Rural Road, City Street, Alley
127		Vehicular Trail (4WD), Walkway/Pedestrian Trail, Private Road for service vehicles
128		Bike Path or Trail, Bridle Path

Table 10 (Cont.)

Value	Original CDL Category	Enhanced CDL Category
131	Barren	Barren
141	Deciduous Forest	Deciduous Forest
142	Evergreen Forest	Evergreen Forest
143	Mixed Forest	Mixed Forest
150		Protected Shrubland (CDL)
151		Protected Shrubland (NLCD-derived)
152	Shrubland	Shrubland
153		Shrubland (NLCD)
154		Corn (CDL) Low Productivity (NCPPI)
155		Corn (CDL) Marginal Productivity (NCPPI)
156		Corn (CDL) Moderate Productivity (NCPPI)
157		Corn (CDL) Moderate-to-High Productivity (NCPPI)
158		Corn (CDL) High Productivity (NCPPI)
159		Cotton (CDL) Low Productivity (NCPPI)
160		Cotton (CDL) Marginal Productivity (NCPPI)
161		Cotton (CDL) Moderate to High Productivity (NCPPI)
162		Small Grains (CDL) Low Productivity (NCPPI)
163		Small Grains (CDL) Marginal Productivity (NCPPI)
164		Small Grains (CDL) Moderate Productivity (NCPPI)
165		Small Grains (CDL) Moderate to High Productivity (NCPPI)
166		Small Grains (CDL) High Productivity (NCPPI)
167		Soybeans (CDL) Low Productivity (NCPPI)
168		Soybeans (CDL) Marginal Productivity (NCPPI)
169		Soybeans (CDL) Moderate Productivity (NCPPI)
170		Soybeans (CDL) Moderate-to-High Productivity (NCPPI)
171		Soybeans (CDL) High Productivity (NCPPI)
175		Protected Grassland/Pasture (CDL)
176	Grassland/Pasture	Grassland/Pasture (CDL)
177		Protected Pasture/Hay (NLCD-derived)
178		Pasture/Hay (NLCD-derived)
179		Protected Grassland (NLCD-derived)
180		Grassland (NLCD-derived)
190	Woody Wetlands	Woody Wetlands
195	Herbaceous Wetlands	Herbaceous Wetlands
204	Pistachios	Pistachios
205	Triticale	Triticale
206	Carrots	Carrots
207	Asparagus	Asparagus

Table 10 (Cont.)

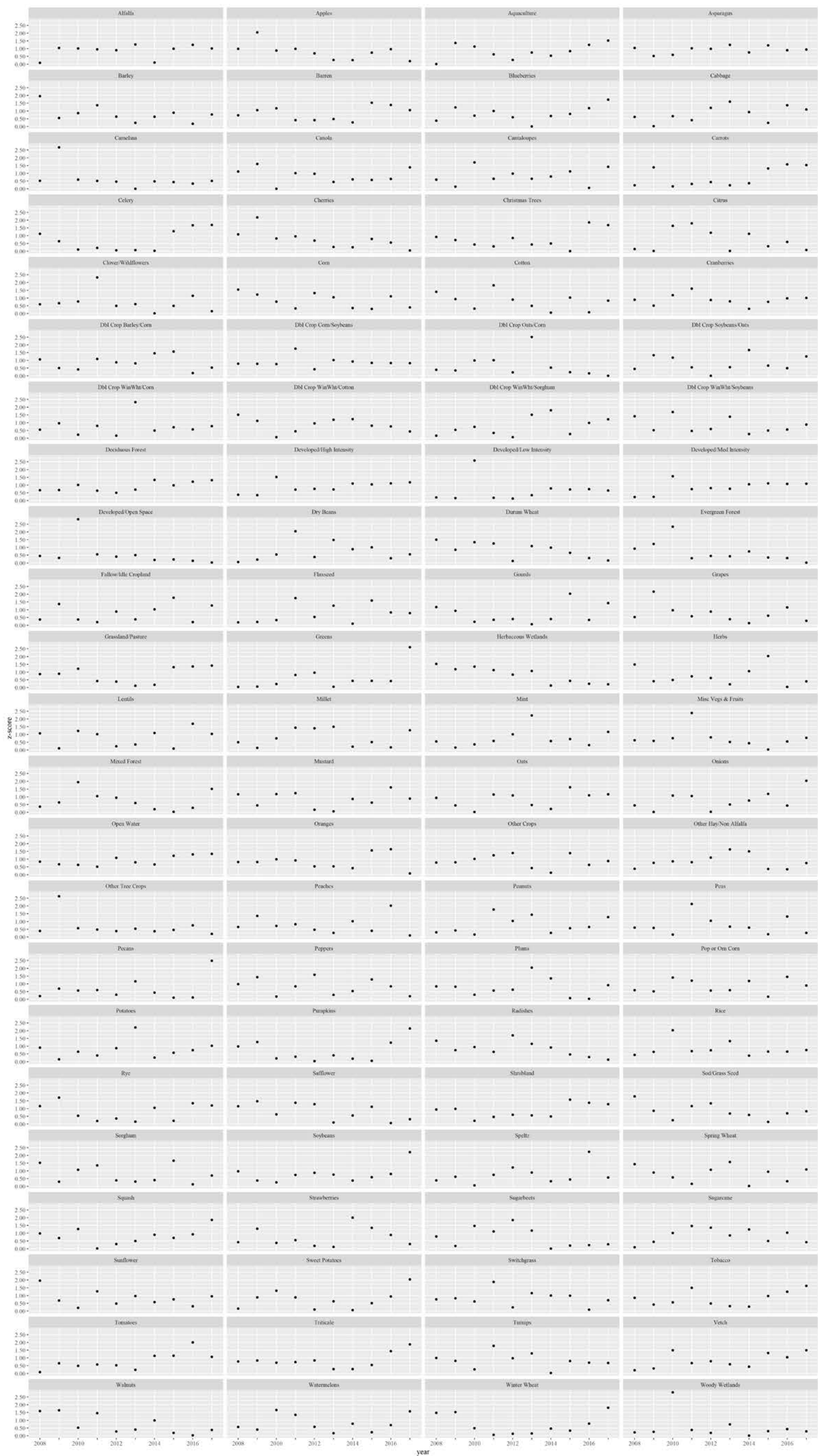
Value	Original CDL Category	Enhanced CDL Category
208	Garlic	Garlic
209	Cantaloupes	Cantaloupes
210	Prunes	Prunes
211	Olives	Olives
212	Oranges	Oranges
213	Honeydew Melons	Honeydew Melons
214	Broccoli	Broccoli
215		
216	Peppers	Peppers
217	Pomegranates	Pomegranates
218	Nectarines	Nectarines
219	Greens	Greens
220	Plums	Plums
221	Strawberries	Strawberries
222	Squash	Squash
223	Apricots	Apricots
224	Vetch	Vetch
225	Dbl Crop WinWht/Corn	Dbl Crop WinWht/Corn
226	Dbl Crop Oats/Corn	Dbl Crop Oats/Corn
227	Lettuce	Lettuce
229	Pumpkins	Pumpkins
230	Dbl Crop Lettuce/Durum Wht	Dbl Crop Lettuce/Durum Wht
231	Dbl Crop Lettuce/Cantaloupe	Dbl Crop Lettuce/Cantaloupe
232	Dbl Crop Lettuce/Cotton	Dbl Crop Lettuce/Cotton
233	Dbl Crop Lettuce/Barley	Dbl Crop Lettuce/Barley
234	Dbl Crop Durum Wht/Sorghum	Dbl Crop Durum Wht/Sorghum
235	Dbl Crop Barley/Sorghum	Dbl Crop Barley/Sorghum
236	Dbl Crop WinWht/Sorghum	Dbl Crop WinWht/Sorghum
237	Dbl Crop Barley/Corn	Dbl Crop Barley/Corn
238	Dbl Crop WinWht/Cotton	Dbl Crop WinWht/Cotton
239	Dbl Crop Soybeans/Cotton	Dbl Crop Soybeans/Cotton
240	Dbl Crop Soybeans/Oats	Dbl Crop Soybeans/Oats
241	Dbl Crop Corn/Soybeans	Dbl Crop Corn/Soybeans
242	Blueberries	Blueberries
243	Cabbage	Cabbage
244	Cauliflower	Cauliflower
245	Celery	Celery
246	Radishes	Radishes
247	Turnips	Turnips
248	Eggplants	Eggplants
249	Gourds	Gourds

Table 10 (Cont.)

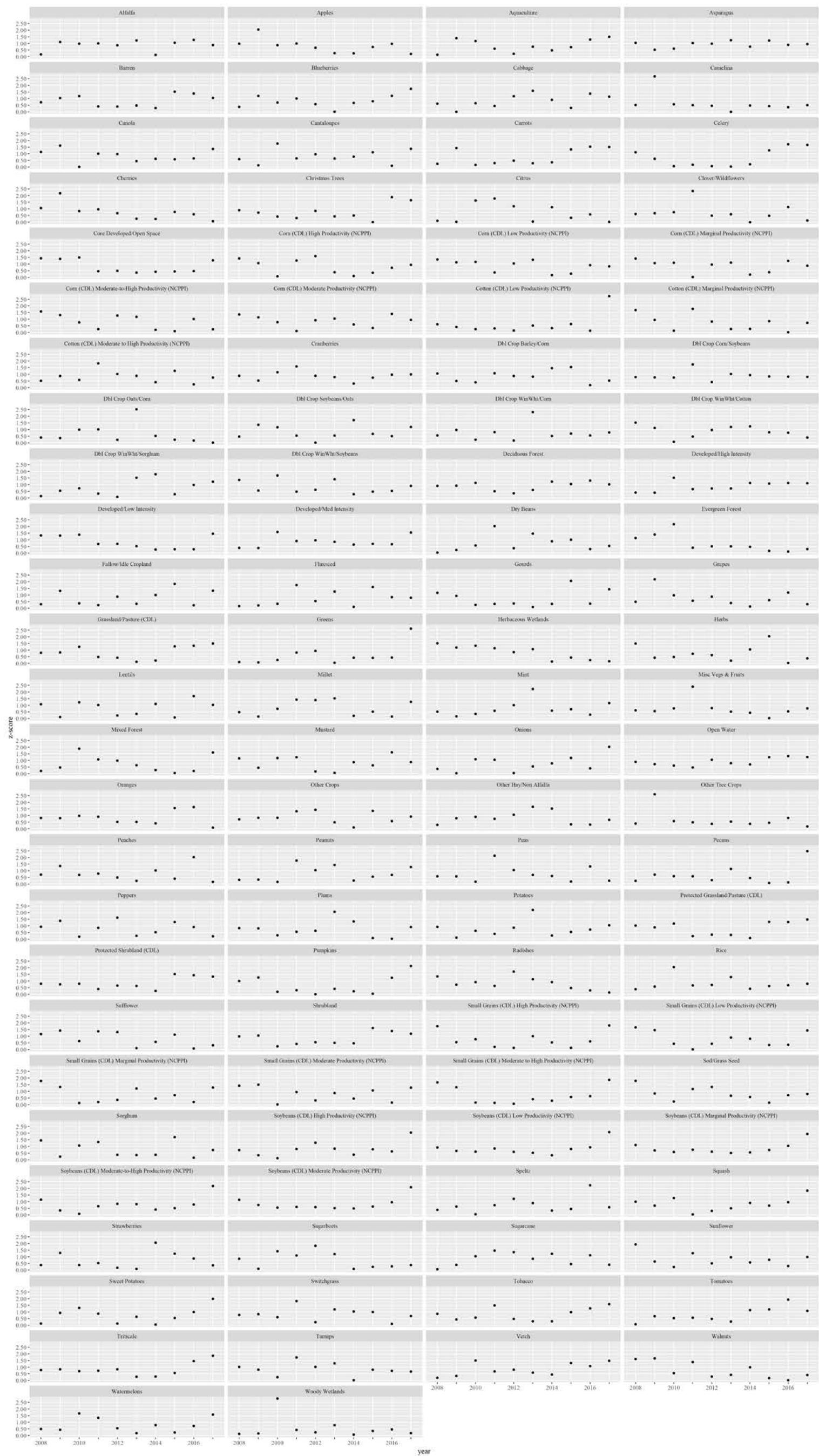
Value	Original CDL Category	Enhanced CDL Category
250	Cranberries	Cranberries
254	Dbl Crop Barley/Soybeans	Dbl Crop Barley/Soybeans

3. Plots by Category

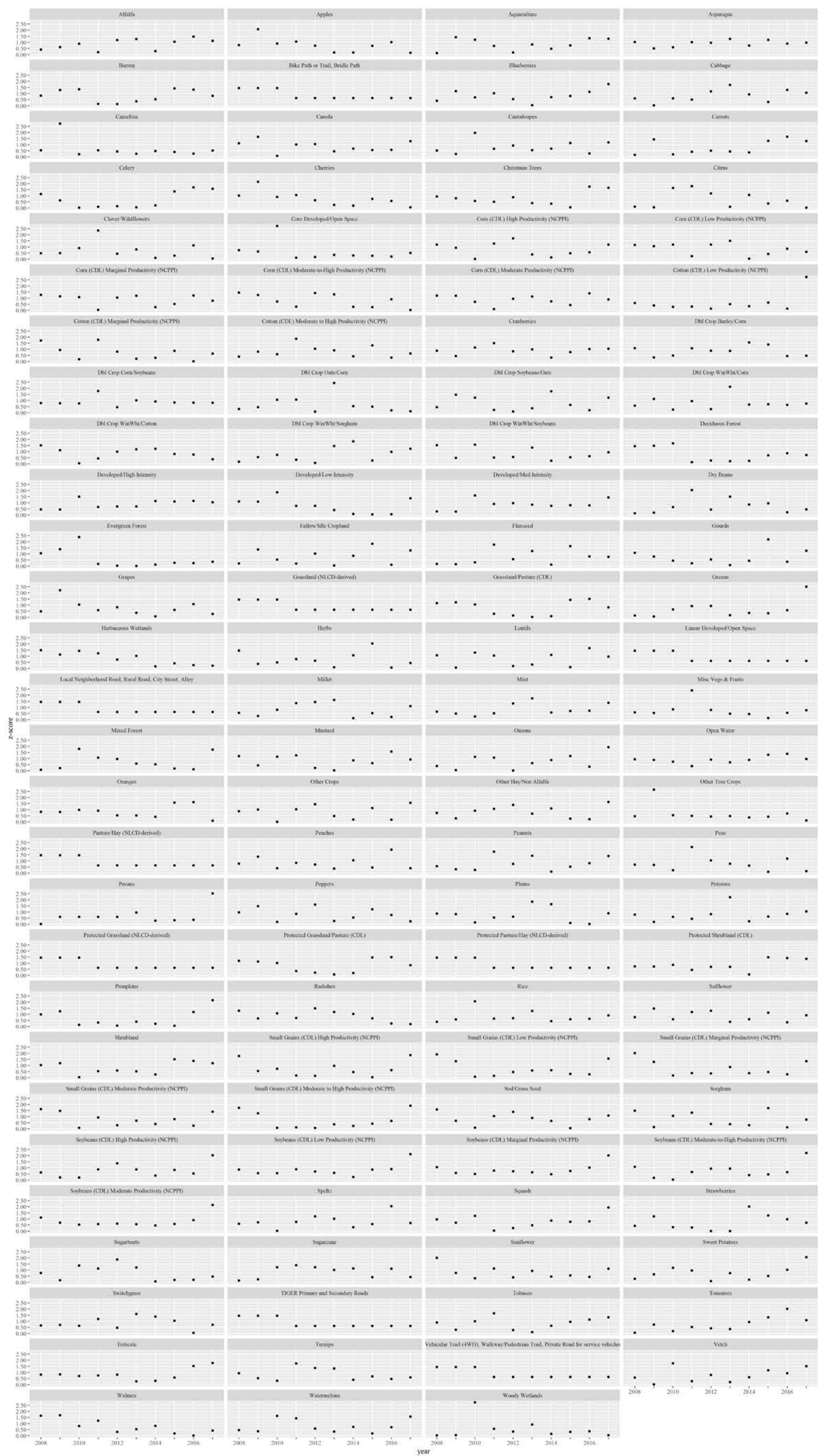
3.1 Model 1 – Z-Score Plot by Year, 2008 -2017



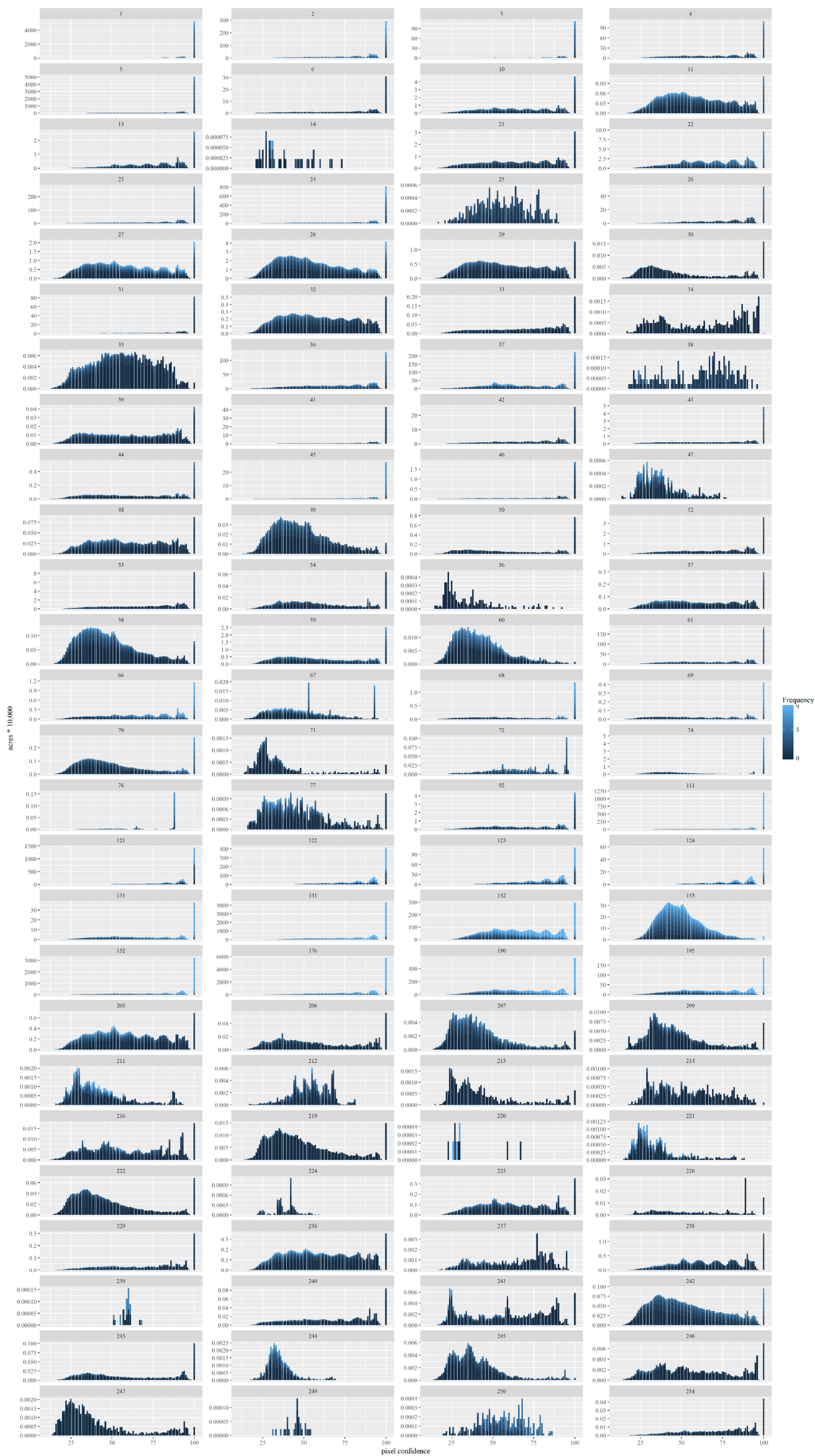
3.2 Model 2 – Z-Score Plot by Year, 2008 -2017



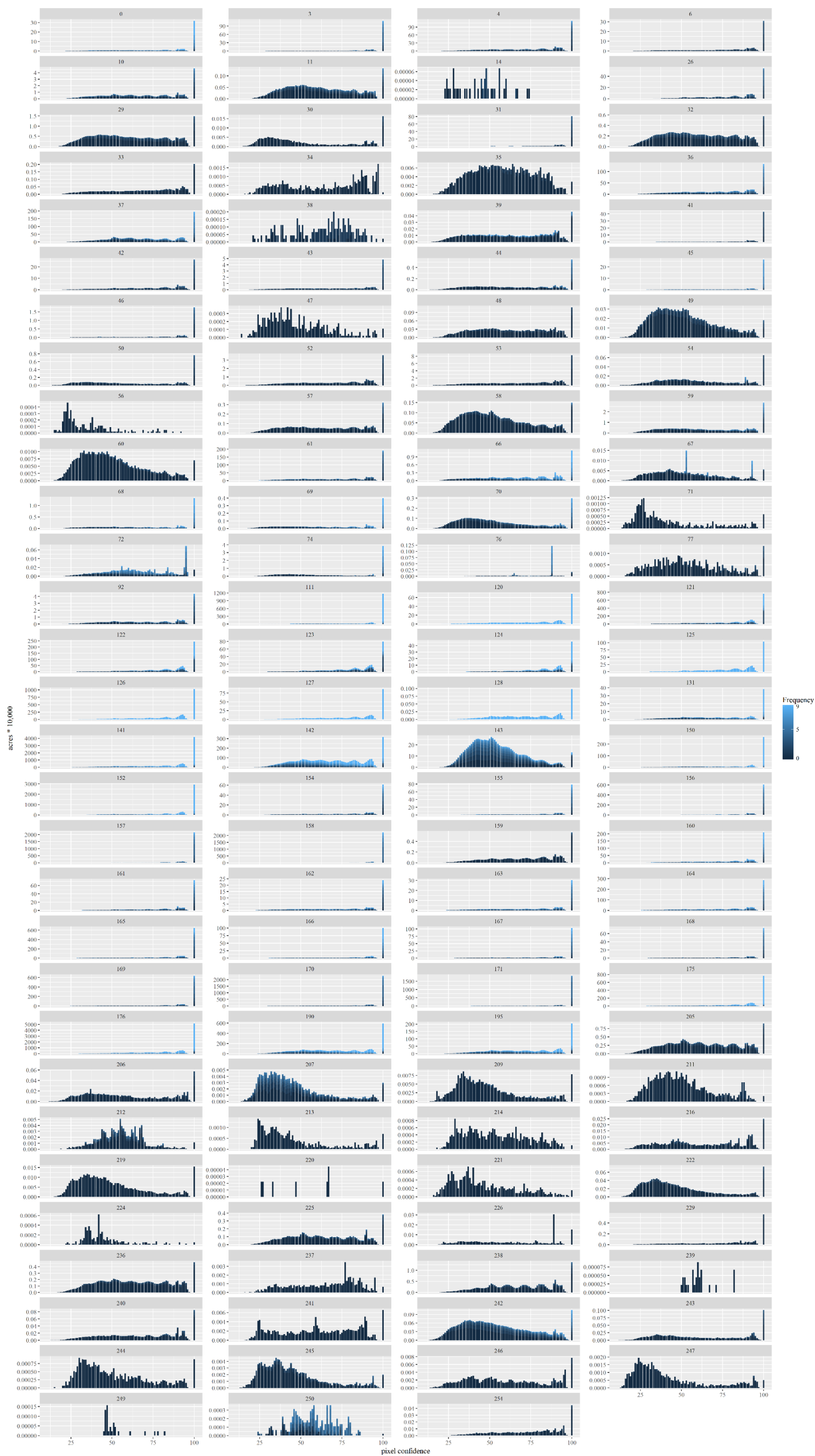
3.3 Model 3 – Z-Score Plot by Year, 2008 -2017



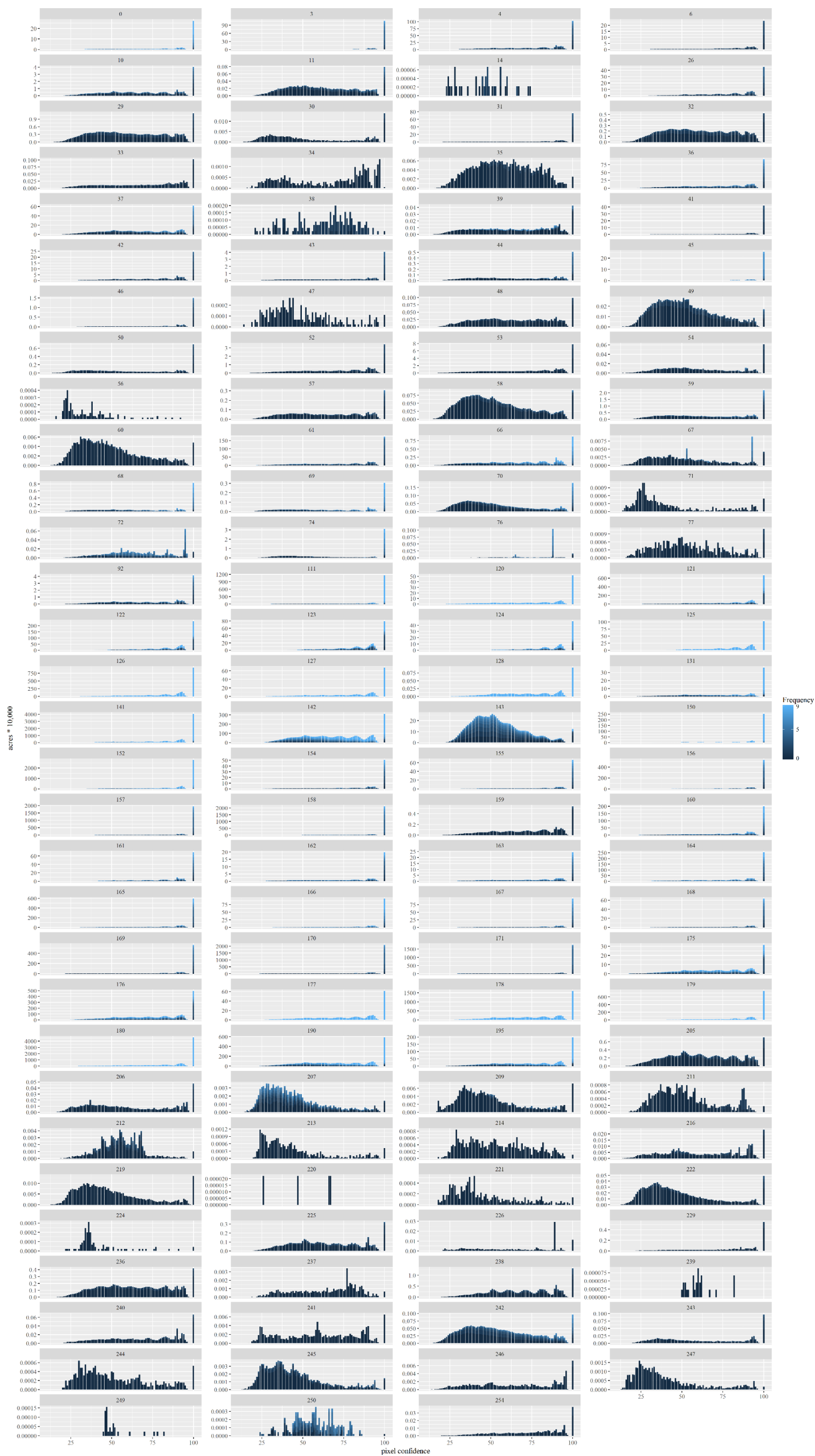
3.4 Model 1 – 2017 Frequency and Confidence Plots by Category



3.5 Model 2 – 2017 Frequency and Confidence Plots by Category



3.6 Model 3 – 2017 Frequency and Confidence Plots by Category



4. R script for Exploratory Data Analysis

```
## Clear Environments
rm(list=ls())

## Load Required Libraries
library(fBasics)
#library(ggplot) ##ggplot should not be loaded until after all arc.open() and arc.select() commands
library(tidyr)
library(tidyverse)
library(plyr)
library(reshape2)
library(readxl)
library(scales)

## unload packages when needed
#detach("package:fBasics", unload=TRUE)
#detach("package:ggplot", unload=TRUE)
#detach("package:tidyr", unload=TRUE)
#detach("package:tidyverse", unload=TRUE)
#detach("package:plyr", unload=TRUE)
#detach("package:reshape2", unload=TRUE)
#detach("package:scales", unload=TRUE)
```

To read tables directly from the file geodatabases, library(arcgisbinding) must be loaded. Afterwards, the arc.check_product() command must be entered for the ArcGIS license to be verified.

```
## Getting going with R Bridge for ArcGIS
library(arcgisbinding)

*** Please call arc.check_product() to define a desktop license.

arc.check_product()

product: ArcGIS Pro ( 12.1.0.10257 )
license: Advanced
```

Use arc.open() to select tables from file geodatabases.

```
#arc.open the tables for noModelMods version
input <- file.path(getwd(), 'model-tests.gdb', 'T2017_30m_cdls_BroadR')
T2017_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2016_30m_cdls_BroadR')
T2016_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2015_30m_cdls_BroadR')
T2015_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2014_30m_cdls_BroadR')
T2014_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2013_30m_cdls_BroadR')
T2013_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2012_30m_cdls_BroadR')
T2012_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2011_30m_cdls_BroadR')
```

```

T2011_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2010_30m_cdls_BroadR')
T2010_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2009_30m_cdls_BroadR')
T2009_30m_cdls_BroadR.SA <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2008_30m_cdls_BroadR')
T2008_30m_cdls_BroadR.SA <- arc.open(input)

#arc.open the table for the without NLCD version
input <- file.path(getwd(), 'model-tests.gdb', 'T2008_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2008_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2009_30m_cdls_Broad_RC_SummStats_I35_NoNLCD01')
T2009_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2010_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2010_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2011_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2011_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2012_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2012_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2013_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2013_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2014_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2014_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2015_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2015_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2016_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2016_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2017_30m_cdls_Broad_SummStats_I35_NoNLCD01')
T2017_30m_cdls_Broad_SummStats_I35_withoutNLCD <- arc.open(input)

#arc.open the tables for the with NLCD version
input <- file.path(getwd(), 'model-tests.gdb', 'T2008_30m_cdls_Broad_SummStats_I35_0201')
T2008_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2009_30m_cdls_Broad_RC_SummStats_I35_0201')
T2009_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2010_30m_cdls_Broad_SummStats_I35_0201')
T2010_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2011_30m_cdls_Broad_SummStats_I35_0201')
T2011_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2012_30m_cdls_Broad_SummStats_I35_0201')
T2012_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2013_30m_cdls_Broad_SummStats_I35_0201')
T2013_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2014_30m_cdls_Broad_SummStats_I35_0201')
T2014_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2015_30m_cdls_Broad_SummStats_I35_0201')
T2015_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2016_30m_cdls_Broad_SummStats_I35_0201')
T2016_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)
input <- file.path(getwd(), 'model-tests.gdb', 'T2017_30m_cdls_Broad_SummStats_I35_0201')
T2017_30m_cdls_Broad_SummStats_withNLCD <- arc.open(input)

```

Use `arc.select()` to specifying the fields wanted from each table.

#tables with no Model Modifications

```
T2017_30m_cdlS.SA <- arc.select(T2017_30m_cdlS_BroadR.SA, c('Value', 'Count', 'Class_Name'))
T2016_30m_cdlS.SA <- arc.select(T2016_30m_cdlS_BroadR.SA, c('Value', 'Count', 'Class_Name'))
T2015_30m_cdlS.SA <- arc.select(T2015_30m_cdlS_BroadR.SA, c('Value', 'Count', 'Class_Name'))
T2014_30m_cdlS.SA <- arc.select(T2014_30m_cdlS_BroadR.SA, c('Value', 'Count', 'Class_Names'))
T2013_30m_cdlS.SA <- arc.select(T2013_30m_cdlS_BroadR.SA, c('Value', 'Count', 'Class_Names'))
T2012_30m_cdlS.SA <- arc.select(T2012_30m_cdlS_BroadR.SA, c('Value', 'Count'))
T2011_30m_cdlS.SA <- arc.select(T2011_30m_cdlS_BroadR.SA, c('Value', 'Count'))
T2010_30m_cdlS.SA <- arc.select(T2010_30m_cdlS_BroadR.SA, c('Value', 'Count'))
T2009_30m_cdlS.SA <- arc.select(T2009_30m_cdlS_BroadR.SA, c('Value', 'Count'))
T2008_30m_cdlS.SA <- arc.select(T2008_30m_cdlS_BroadR.SA, c('Value', 'Count', 'Class_Name'))
```

#tables without NLCD

```
T2017_withoutNLCD_model.SA <- arc.select(T2017_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2016_withoutNLCD_model.SA <- arc.select(T2016_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2015_withoutNLCD_model.SA <- arc.select(T2015_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2014_withoutNLCD_model.SA <- arc.select(T2014_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2013_withoutNLCD_model.SA <- arc.select(T2013_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2012_withoutNLCD_model.SA <- arc.select(T2012_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2011_withoutNLCD_model.SA <- arc.select(T2011_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2010_withoutNLCD_model.SA <- arc.select(T2010_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2009_withoutNLCD_model.SA <- arc.select(T2009_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2008_withoutNLCD_model.SA <- arc.select(T2008_30m_cdlS_Broad_SummStats_I35_withoutNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
```

#tables with NLCD

```
T2017_withNLCD_model.SA <- arc.select(T2017_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2016_withNLCD_model.SA <- arc.select(T2016_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2015_withNLCD_model.SA <- arc.select(T2015_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2014_withNLCD_model.SA <- arc.select(T2014_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2013_withNLCD_model.SA <- arc.select(T2013_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2012_withNLCD_model.SA <- arc.select(T2012_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2011_withNLCD_model.SA <- arc.select(T2011_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2010_withNLCD_model.SA <- arc.select(T2010_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
T2009_withNLCD_model.SA <- arc.select(T2009_30m_cdlS_Broad_SummStats_withNLCD, c('New_Value', 'New_Category', 'SUM_Count'))
```

```
T2008_withNLCD_model.SA <- arc.select(T2008_30m_cdls_Broad_SummStats_withNLCD, c('New_Value',
New_Category','SUM_Count'))
```

For each model, join the tables so that all years are in the same table First, import the base tables that will support the joins

```
oriValues <- read_excel("J:\\Box Sync\\CAST_SUS_Monarch\\data\\LookUpTable-Development.xlsx", range
= "OriValues!A1:B256" , col_names = TRUE)
```

```
newValues <- read_excel("J:\\Box Sync\\CAST_SUS_Monarch\\data\\LookUpTable-Development.xlsx", range
= "NewValues!A1:B256" , col_names = TRUE)
```

Model 1 Joins

#tables with no Model Modifications

```
j.1 <- (full_join(oriValues, T2008_30m_cdls.SA, by = "Value"))
j.2 <- (full_join(j.1, T2009_30m_cdls.SA, by = "Value"))
j.3 <- (full_join(j.2, T2010_30m_cdls.SA, by = "Value"))
j.4 <- (full_join(j.3, T2011_30m_cdls.SA, by = "Value"))
j.5 <- (full_join(j.4, T2012_30m_cdls.SA, by = "Value"))
j.6 <- (full_join(j.5, T2013_30m_cdls.SA, by = "Value"))
j.7 <- (full_join(j.6, T2014_30m_cdls.SA, by = "Value"))
j.8 <- (full_join(j.7, T2015_30m_cdls.SA, by = "Value"))
j.9 <- (full_join(j.8, T2016_30m_cdls.SA, by = "Value"))
CompareAddFinalnoModelMods.SA <- (full_join(j.9, T2017_30m_cdls.SA, by = "Value"))
```

Model 1 field names have become redundant and difficult to differentiate, so they are renamed.

#rename the fields, model 1

```
names(CompareAddFinalnoModelMods.SA) <- c("Value", "Category", "Count.2008", "Category.2008", "Count.
2009", "Count.2010", "Count.2011", "Count.2012", "Count.2013", "Category.2013", "Count.2014", "Category.201
4", "Count.2015", "Category.2015", "Count.2016", "Category.2016", "Count.2017", "Category.2017")
```

Model 2 Joins

#tables without NLCD

```
j.1 <- (full_join(newValues, T2008_withoutNLCD_model.SA, by = "New_Value"))
j.2 <- (full_join(j.1, T2009_withoutNLCD_model.SA, by = "New_Value"))
j.3 <- (full_join(j.2, T2010_withoutNLCD_model.SA, by = "New_Value"))
j.4 <- (full_join(j.3, T2011_withoutNLCD_model.SA, by = "New_Value"))
j.5 <- (full_join(j.4, T2012_withoutNLCD_model.SA, by = "New_Value"))
j.6 <- (full_join(j.5, T2013_withoutNLCD_model.SA, by = "New_Value"))
j.7 <- (full_join(j.6, T2014_withoutNLCD_model.SA, by = "New_Value"))
j.8 <- (full_join(j.7, T2015_withoutNLCD_model.SA, by = "New_Value"))
j.9 <- (full_join(j.8, T2016_withoutNLCD_model.SA, by = "New_Value"))
CompareAddFinalwithoutNLCD.SA <- (full_join(j.9, T2017_withoutNLCD_model.SA, by = "New_Value"))
```

Model 2 field names have become redundant and difficult to differentiate, so they are renamed.

#rename the fields, model 2

```
names(CompareAddFinalwithoutNLCD.SA) <- c("New_Value", "New_Category", "Category.2008", "Count.20
08", "Category.2009", "Count.2009", "Category.2010", "Count.2010", "Category.2011", "Count.2011", "Category.
2012", "Count.2012", "Category.2013", "Count.2013", "Category.2014", "Count.2014", "Category.2015", "Count.2
015", "Category.2016", "Count.2016", "Category.2017", "Count.2017")
```

Model 3 Joins

#tables with NLCD

```
j.1 <- (full_join(newValues, T2008_withNLCD_model.SA, by = "New_Value"))
j.2 <- (full_join(j.1, T2009_withNLCD_model.SA, by = "New_Value"))
j.3 <- (full_join(j.2, T2010_withNLCD_model.SA, by = "New_Value"))
j.4 <- (full_join(j.3, T2011_withNLCD_model.SA, by = "New_Value"))
j.5 <- (full_join(j.4, T2012_withNLCD_model.SA, by = "New_Value"))
j.6 <- (full_join(j.5, T2013_withNLCD_model.SA, by = "New_Value"))
j.7 <- (full_join(j.6, T2014_withNLCD_model.SA, by = "New_Value"))
j.8 <- (full_join(j.7, T2015_withNLCD_model.SA, by = "New_Value"))
j.9 <- (full_join(j.8, T2016_withNLCD_model.SA, by = "New_Value"))
CompareAddFinalwithNLCD.SA <- (full_join(j.9, T2017_withNLCD_model.SA, by = "New_Value"))
```

Model 3 field names have become redundant and difficult to differentiate, so they are renamed.

#rename the fields, model 3

```
names(CompareAddFinalwithNLCD.SA) <- c("New_Value", "New_Category", "Category.2008", "Count.2008",
"Category.2009", "Count.2009", "Category.2010", "Count.2010", "Category.2011", "Count.2011", "Category.2012", "Count.2012", "Category.2013", "Count.2013", "Category.2014", "Count.2014", "Category.2015", "Count.2015", "Category.2016", "Count.2016", "Category.2017", "Count.2017")
```

Remove NAs from the datasets to avoid complications in processes down the line.

#eliminate NAs from dataset

```
Compare2008thru2017noModelMods.SA.OMIT <- na.omit(CompareAddFinalnoModelMods.SA)
Compare2008thru2017withNLCD.SA.OMIT <- na.omit(CompareAddFinalwithNLCD.SA)
Compare2008thru2017withoutNLCD.SA.OMIT <- na.omit(CompareAddFinalwithoutNLCD.SA)
```

Calculate acres in a new column for each year

#model 1

```
Compare2008thru2017noModelMods.SA.OMIT$Acres.2008 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2008*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2009 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2009*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2010 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2010*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2011 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2011*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2012 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2012*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2013 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2013*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2014 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2014*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2015 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2015*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2016 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2016*0.222394))
Compare2008thru2017noModelMods.SA.OMIT$Acres.2017 <- ((Compare2008thru2017noModelMods.SA.OMIT$Count.2017*0.222394))
```

#model 2

```
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2008 <- ((Compare2008thru2017withoutNLCD.SA.OMIT$Count.2008*0.222394))
```

```

MIT$Count.2008*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2009 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2009*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2010 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2010*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2011 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2011*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2012 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2012*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2013 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2013*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2014 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2014*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2015 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2015*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2016 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2016*0.222394))
Compare2008thru2017withoutNLCD.SA.OMIT$Acres.2017 <- ((Compare2008thru2017withoutNLCD.SA.O
MIT$Count.2017*0.222394))

```

#model 3

```

Compare2008thru2017withNLCD.SA.OMIT$Acres.2008 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2008*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2009 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2009*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2010 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2010*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2011 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2011*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2012 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2012*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2013 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2013*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2014 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2014*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2015 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2015*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2016 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2016*0.222394))
Compare2008thru2017withNLCD.SA.OMIT$Acres.2017 <- ((Compare2008thru2017withNLCD.SA.OMIT$C
ount.2017*0.222394))

```

Calculate standard deviation, mean, and coefficient of variation (cv)

#model 1

```

Compare2008thru2017noModelMods.SA.OMIT$Acres_SD <- rowStats(subset(Compare2008thru2017noMod
elMods.SA.OMIT, select = c(Acres.2008, Acres.2009 ,Acres.2010, Acres.2011, Acres.2012, Acres.2013, Acre
s.2014, Acres.2015, Acres.2016, Acres.2017)), sd, na.rm = TRUE)

```

```

Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean <- rowStats(subset(Compare2008thru2017noM
odelMods.SA.OMIT, select = c(Acres.2008, Acres.2009, Acres.2010, Acres.2011, Acres.2012, Acres.2013, A
cres.2014, Acres.2015, Acres.2016, Acres.2017)), mean, na.rm = TRUE)

```

```

Compare2008thru2017noModelMods.SA.OMIT$Acres_CV <- ((Compare2008thru2017noModelMods.SA.O

```

```
MIT$Acres_SD/Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean))
```

```
#model 2
```

```
Compare2008thru2017withoutNLCD.SA.OMIT$Acres_SD <- rowStats(subset(Compare2008thru2017withoutNLCD.SA.OMIT, select = c(Acres.2008, Acres.2009 ,Acres.2010, Acres.2011, Acres.2012, Acres.2013, Acres.2014, Acres.2015, Acres.2016, Acres.2017)), sd, na.rm = TRUE)
```

```
Compare2008thru2017withoutNLCD.SA.OMIT$Acres_Mean <- rowStats(subset(Compare2008thru2017withoutNLCD.SA.OMIT, select = c(Acres.2008, Acres.2009, Acres.2010, Acres.2011, Acres.2012, Acres.2013, Acres.2014, Acres.2015, Acres.2016, Acres.2017)), mean, na.rm = TRUE)
```

```
Compare2008thru2017withoutNLCD.SA.OMIT$CV <- ((Compare2008thru2017withoutNLCD.SA.OMIT$Acres_SD/Compare2008thru2017withoutNLCD.SA.OMIT$Acres_Mean))
```

```
#model 3
```

```
Compare2008thru2017withNLCD.SA.OMIT$Acres_SD <- rowStats(subset(Compare2008thru2017withNLCD.SA.OMIT, select = c(Acres.2008, Acres.2009 ,Acres.2010, Acres.2011, Acres.2012, Acres.2013, Acres.2014, Acres.2015, Acres.2016, Acres.2017)), sd, na.rm = TRUE)
```

```
Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean <- rowStats(subset(Compare2008thru2017withNLCD.SA.OMIT, select = c(Acres.2008, Acres.2009 ,Acres.2010, Acres.2011, Acres.2012, Acres.2013, Acres.2014, Acres.2015, Acres.2016, Acres.2017)), mean, na.rm = TRUE)
```

```
Compare2008thru2017withNLCD.SA.OMIT$CV <- ((Compare2008thru2017withNLCD.SA.OMIT$Acres_SD/Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean))
```

Calculate z-score for each year

```
#model 1
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2008 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2008 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2009 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2009 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2010 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2010 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2011 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2011 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2012 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2012 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2013 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2013 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2014 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2014 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2015 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2015 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```

```
Compare2008thru2017noModelMods.SA.OMIT$Zscore.2016 <- abs((Compare2008thru2017noModelMods.SA.OMIT$Acres.2016 - Compare2008thru2017noModelMods.SA.OMIT$Acres_Mean) / Compare2008thru2017noModelMods.SA.OMIT$Acres_SD)
```



```

D.SA.OMIT$Acres_SD)
Compare2008thru2017withNLCD.SA.OMIT$Zscore.2013 <- abs((Compare2008thru2017withNLCD.SA.OMIT$Acres.2013 - Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean) / Compare2008thru2017withNLCD.SA.OMIT$Acres_SD)
Compare2008thru2017withNLCD.SA.OMIT$Zscore.2014 <- abs((Compare2008thru2017withNLCD.SA.OMIT$Acres.2014 - Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean) / Compare2008thru2017withNLCD.SA.OMIT$Acres_SD)
Compare2008thru2017withNLCD.SA.OMIT$Zscore.2015 <- abs((Compare2008thru2017withNLCD.SA.OMIT$Acres.2015 - Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean) / Compare2008thru2017withNLCD.SA.OMIT$Acres_SD)
Compare2008thru2017withNLCD.SA.OMIT$Zscore.2016 <- abs((Compare2008thru2017withNLCD.SA.OMIT$Acres.2016 - Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean) / Compare2008thru2017withNLCD.SA.OMIT$Acres_SD)
Compare2008thru2017withNLCD.SA.OMIT$Zscore.2017 <- abs((Compare2008thru2017withNLCD.SA.OMIT$Acres.2017 - Compare2008thru2017withNLCD.SA.OMIT$Acres_Mean) / Compare2008thru2017withNLCD.SA.OMIT$Acres_SD)

```

Plot mean of acres against SD of acres using CV as sizing of point; highlighting greatest standard deviation

#First, load the ggplot library

library(ggplot2) ##ggplot should not be loaded until after all arc.open() and arc.select() commands

#model 1

```

ggplot(Compare2008thru2017noModelMods.SA.OMIT, aes(Acres_Mean/100000, Acres_SD/100000)) + geom_point(aes(size = Compare2008thru2017noModelMods.SA.OMIT$Acres_CV, colour = Compare2008thru2017noModelMods.SA.OMIT$Acres_SD > 2500000)) + scale_colour_manual(values = c("black", "red")) + labs(title = "CDL 2008 - 2017", x = "Mean (Acres) in Millions", y = "Standard Deviation (Acres) in Millions", caption = "") + scale_size_continuous(name = "Coefficient of Variation (StDev/Mean)") + guides(colour=FALSE) + scale_y_continuous(labels = unit_format()) + scale_x_continuous(labels = unit_format())

```

```

model1.select <- subset(Compare2008thru2017noModelMods.SA.OMIT, Acres_SD > 2500000)
model1.select

```

A tibble: 4 x 41

	Value Category	Count.2008	Category.2008	Count.2009	Count.2010
	<dbl> <chr>	<dbl>	<chr>	<dbl>	<dbl>
1	1.00 Corn	333416160	Corn	339183032	347543788
2	5.00 Soybeans	297295629	Soybeans	313591248	316805484
3	152 Shrubland	350468484	Shrubland	349325375	370428107
4	176 Grassland/Pasture	1050501987	Grassland/Pasture	1052155338	1077026486

... with 35 more variables: Count.2011 <dbl>, Count.2012 <dbl>,
Count.2013 <dbl>, Category.2013 <chr>, Count.2014 <dbl>,
Category.2014 <chr>, Count.2015 <dbl>, Category.2015 <chr>,
Count.2016 <dbl>, Category.2016 <chr>, Count.2017 <dbl>,
Category.2017 <chr>, Acres.2008 <dbl>, Acres.2009 <dbl>,
Acres.2010 <dbl>, Acres.2011 <dbl>, Acres.2012 <dbl>,
Acres.2013 <dbl>, Acres.2014 <dbl>, Acres.2015 <dbl>,
Acres.2016 <dbl>, Acres.2017 <dbl>, Acres_SD <dbl>, Acres_Mean <dbl>,
Acres_CV <dbl>, Zscore.2008 <dbl>, Zscore.2009 <dbl>,
Zscore.2010 <dbl>, Zscore.2011 <dbl>, Zscore.2012 <dbl>,

```
# Zscore.2013 <dbl>, Zscore.2014 <dbl>, Zscore.2015 <dbl>,
# Zscore.2016 <dbl>, Zscore.2017 <dbl>

#model 2
ggplot(Compare2008thru2017withoutNLCD.SA.OMIT, aes(Acres_Mean/1000000, Acres_SD/1000000)) + geom_point(aes(size = Compare2008thru2017withoutNLCD.SA.OMIT$CV, colour = Compare2008thru2017withoutNLCD.SA.OMIT$Acres_SD > 2500000)) + scale_colour_manual(values = c("black", "red")) + labs(title = "CDL 2008 - 2017 without NLCD", x = "Mean (Acres) in Millions", y = "Standard Deviation (Acres) in Millions", caption = "") + scale_size_continuous(name = "Coefficient of Variation (StDev/Mean)") + guides(colour=FALSE) + scale_y_continuous(labels = unit_format()) + scale_x_continuous(labels = unit_format())
```

```
model2.select <- subset(Compare2008thru2017withoutNLCD.SA.OMIT, Acres_SD > 2500000)
model2.select
```

```
# A tibble: 3 x 45
```

```
  New_Value New_Category Category.2008 Count.2008 Category.2009 Count.2009
    <dbl> <chr>      <chr>      <dbl> <chr>      <dbl>
1    121 Core Develop~ Core Develop~ 116096425 Core Develop~ 115539834
2    170 Soybeans (C~ Soybeans (CD~ 137981297 Soybeans (CD~ 147888539
3    176 Grassland/P~ Grassland/Pa~ 1218216465 Grassland/Pa~ 1219131791
# ... with 39 more variables: Category.2010 <chr>, Count.2010 <dbl>,
# Category.2011 <chr>, Count.2011 <dbl>, Category.2012 <chr>,
# Count.2012 <dbl>, Category.2013 <chr>, Count.2013 <dbl>,
# Category.2014 <chr>, Count.2014 <dbl>, Category.2015 <chr>,
# Count.2015 <dbl>, Category.2016 <chr>, Count.2016 <dbl>,
# Category.2017 <chr>, Count.2017 <dbl>, Acres.2008 <dbl>,
# Acres.2009 <dbl>, Acres.2010 <dbl>, Acres.2011 <dbl>,
# Acres.2012 <dbl>, Acres.2013 <dbl>, Acres.2014 <dbl>,
# Acres.2015 <dbl>, Acres.2016 <dbl>, Acres.2017 <dbl>, Acres_SD <dbl>,
# Acres_Mean <dbl>, CV <dbl>, Zscore.2008 <dbl>, Zscore.2009 <dbl>,
# Zscore.2010 <dbl>, Zscore.2011 <dbl>, Zscore.2012 <dbl>,
# Zscore.2013 <dbl>, Zscore.2014 <dbl>, Zscore.2015 <dbl>,
# Zscore.2016 <dbl>, Zscore.2017 <dbl>
```

```
#model 3
```

```
ggplot(Compare2008thru2017withNLCD.SA.OMIT, aes(Acres_Mean/1000000, Acres_SD/1000000)) + geom_point(aes(size = Compare2008thru2017withNLCD.SA.OMIT$CV, colour = Compare2008thru2017withNLCD.SA.OMIT$Acres_SD > 2500000)) + scale_colour_manual(values = c("black", "red")) + labs(title = "CDL 2008 - 2017 with NLCD", x = "Mean (Acres) in Millions", y = "Standard Deviation (Acres) in Millions", caption = "") + scale_size_continuous(name = "Coefficient of Variation (StDev/Mean)") + guides(colour=FALSE) + scale_y_continuous(labels = unit_format()) + scale_x_continuous(labels = unit_format())
```

```
model3.select <- subset(Compare2008thru2017withNLCD.SA.OMIT, Acres_SD > 2500000)
model3.select
```

```
# A tibble: 3 x 45
```

```
  New_Value New_Category Category.2008 Count.2008 Category.2009 Count.2009
    <dbl> <chr>      <chr>      <dbl> <chr>      <dbl>
1    141 Deciduous F~ Deciduous Fo~ 564204207 Deciduous Fo~ 563656789
2    152 Shrubland Shrubland 277335666 Shrubland 275285166
3    176 Grassland/P~ Grassland/Pa~ 233565371 Grassland/Pa~ 236029882
```

```
# ... with 39 more variables: Category.2010 <chr>, Count.2010 <dbl>,
# Category.2011 <chr>, Count.2011 <dbl>, Category.2012 <chr>,
# Count.2012 <dbl>, Category.2013 <chr>, Count.2013 <dbl>,
# Category.2014 <chr>, Count.2014 <dbl>, Category.2015 <chr>,
# Count.2015 <dbl>, Category.2016 <chr>, Count.2016 <dbl>,
# Category.2017 <chr>, Count.2017 <dbl>, Acres.2008 <dbl>,
# Acres.2009 <dbl>, Acres.2010 <dbl>, Acres.2011 <dbl>,
# Acres.2012 <dbl>, Acres.2013 <dbl>, Acres.2014 <dbl>,
# Acres.2015 <dbl>, Acres.2016 <dbl>, Acres.2017 <dbl>, Acres_SD <dbl>,
# Acres_Mean <dbl>, CV <dbl>, Zscore.2008 <dbl>, Zscore.2009 <dbl>,
# Zscore.2010 <dbl>, Zscore.2011 <dbl>, Zscore.2012 <dbl>,
# Zscore.2013 <dbl>, Zscore.2014 <dbl>, Zscore.2015 <dbl>,
# Zscore.2016 <dbl>, Zscore.2017 <dbl>
```

Melt the data for year-to-year plotting

```
Compare2008thru2017noModelMods.SA.OMIT.MELT <- melt(Compare2008thru2017noModelMods.SA.OMIT, id=c("Value", "Category"))
Compare2008thru2017withNLCD.SA.OMIT.MELT <- melt(Compare2008thru2017withNLCD.SA.OMIT, id=c("New_Value", "New_Category"))
Compare2008thru2017withoutNLCD.SA.OMIT.MELT <- melt(Compare2008thru2017withoutNLCD.SA.OMIT, id=c("New_Value", "New_Category"))
```

#convert value currently stored as a factor to numeric

```
Compare2008thru2017noModelMods.SA.OMIT.MELT$value=as.numeric(Compare2008thru2017noModelMods.SA.OMIT.MELT$value)
```

Warning: NAs introduced by coercion

```
Compare2008thru2017withNLCD.SA.OMIT.MELT$value=as.numeric(Compare2008thru2017withNLCD.SA.OMIT.MELT$value)
```

Warning: NAs introduced by coercion

```
Compare2008thru2017withoutNLCD.SA.OMIT.MELT$value=as.numeric(Compare2008thru2017withoutNLCD.SA.OMIT.MELT$value)
```

Warning: NAs introduced by coercion

#clear NAs that are created as a result (these are records where category was stored as a value during melt)

#eliminate NAs from dataset

```
Compare2008thru2017noModelMods.SA.OMIT.MELT <- na.omit(Compare2008thru2017noModelMods.SA.OMIT.MELT)
```

```
Compare2008thru2017withNLCD.SA.OMIT.MELT <- na.omit(Compare2008thru2017withNLCD.SA.OMIT.MELT)
```

```
Compare2008thru2017withoutNLCD.SA.OMIT.MELT <- na.omit(Compare2008thru2017withoutNLCD.SA.OMIT.MELT)
```

Plot zscores by year and category

#decim pts

```
scaleFUN <- function(x) sprintf("%.2f", x)
```

#model 1

```
ggplot(data = subset(Compare2008thru2017noModelMods.SA.OMIT.MELT, variable %in% c("Zscore.2008",
```

```
"Zscore.2009", "Zscore.2010", "Zscore.2011", "Zscore.2012", "Zscore.2013", "Zscore.2014", "Zscore.2015", "
Zscore.2016", "Zscore.2017")), aes(x = variable, y = value)) + geom_point(data = subset(Compare2008thru201
7noModelMods.SA.OMIT.MELT, variable %in% c("Zscore.2008", "Zscore.2009", "Zscore.2010", "Zscore.201
1", "Zscore.2012", "Zscore.2013", "Zscore.2014", "Zscore.2015", "Zscore.2016", "Zscore.2017"))) + geom_sm
ooth(method = "lm", se = FALSE, lwd = .5, col = "black") + facet_wrap(~ Category, shrink = TRUE, ncol = 5)
+ scale_x_discrete(labels = c(2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017)) + scale_y_conti
nuous(labels=scaleFUN, breaks = c(0,0.5,1,1.5,2,2.5,3,3.5,4))
```

```
ggsave("model-1.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi = 300)
```

#model 2

```
ggplot(data = subset(Compare2008thru2017withoutNLCD.SA.OMIT.MELT, variable %in% c("Zscore.2008",
"Zscore.2009", "Zscore.2010", "Zscore.2011", "Zscore.2012", "Zscore.2013", "Zscore.2014", "Zscore.2015", "
Zscore.2016", "Zscore.2017")), aes(x = variable, y = value)) + geom_point(data = subset(Compare2008thru201
7withoutNLCD.SA.OMIT.MELT, variable %in% c("Zscore.2008", "Zscore.2009", "Zscore.2010", "Zscore.201
1", "Zscore.2012", "Zscore.2013", "Zscore.2014", "Zscore.2015", "Zscore.2016", "Zscore.2017"))) + geom_sm
ooth(method = "lm", se = FALSE, lwd = .5, col = "black") + facet_wrap(~ New_Category, shrink = TRUE, nc
ol = 5) + scale_x_discrete(labels = c(2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017)) + scale_y
_continuous(labels=scaleFUN, breaks = c(0,0.5,1,1.5,2,2.5,3,3.5,4))
```

```
ggsave("model-2.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi = 300)
```

#model 3

```
ggplot(data = subset(Compare2008thru2017withNLCD.SA.OMIT.MELT, variable %in% c("Zscore.2008", "Zs
core.2009", "Zscore.2010", "Zscore.2011", "Zscore.2012", "Zscore.2013", "Zscore.2014", "Zscore.2015", "Zs
core.2016", "Zscore.2017")), aes(x = variable, y = value)) + geom_point(data = subset(Compare2008thru2017wi
thNLCD.SA.OMIT.MELT, variable %in% c("Zscore.2008", "Zscore.2009", "Zscore.2010", "Zscore.2011", "Zs
core.2012", "Zscore.2013", "Zscore.2014", "Zscore.2015", "Zscore.2016", "Zscore.2017"))) + geom_smooth(m
ethod = "lm", se = FALSE, lwd = .5, col = "black") + facet_wrap(~ New_Category, shrink = TRUE, ncol = 5)
+ scale_x_discrete(labels = c(2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017)) + scale_y_conti
nuous(labels=scaleFUN, breaks = c(0,0.5,1,1.5,2,2.5,3,3.5,4))
```

```
ggsave("model-3.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi = 300)
```

Build a change matrix of all possible combinations between 2017 and 2008 to use as a look-up table

#for model 1

#expand the grid for all possible values

```
orivalues.allp <- expand.grid(as.numeric(oriValues$Value), as.numeric(oriValues$Value))
orivalues.allp1 <- orivalues.allp[complete.cases(orivalues.allp), ]
```

#compute the value difference code between 2008 and 2017

```
orivalues.allp1$var3 <- (orivalues.allp1$Var1 - orivalues.allp1$Var2)^2
```

#rename for clarity

```
names(orivalues.allp1) <- c("Value", "Value2", "ValueDiff")
```

#full join to 2017 to re-capture category names for descriptive text

```
orivalues.allp2 <- (full_join(orivalues.allp1, oriValues, by = c("Value")))
```

#rename for clarity

```
names(orivalues.allp2) <- c("Value2017", "Value", "ValueDiff", "Category2017")
```

#full join to 2008 to re-capture category names for descriptive text

```

orivalues.allp3 <- (full_join(orivalues.allp2, oriValues, by = c("Value")))

#rename for clarity
names(orivalues.allp3) <- c("Value2017", "Value2008", "ValueDiff", "Category2017", "Category2008")

#create descriptive text that describe change
orivalues.allp3$descriptor <- paste(orivalues.allp3$Category2008, " to ", orivalues.allp3$Category2017)
orivalues.allp4 <- transform(orivalues.allp3, ChangeCode=paste0(17,orivalues.allp3$Value2017, 8, orivalues.a
llp3$Value2008, abs(orivalues.allp3$ValueDiff)))

#check for duplicates
n_occur <- data.frame(table(orivalues.allp4$ChangeCode))
n_occur[n_occur$Freq > 1,]

      Var1 Freq
20281 1717282222500  2
21811  171782225  2
22820 171818132400  2
27809 1719881334225  2
37140 172318252441  2
57819 1774844900  2
59615 1780855625  2
61156 1786866400  2
61930 1788886400  2
62951 1792877225  2
64735 1798888100  2

dups <- orivalues.allp4[orivalues.allp4$ChangeCode %in% n_occur$Var1[n_occur$Freq > 1],]

#using these change codes has greatly reduced duplication; however 22 instances of potential duplication still
exist. These change types seem unlikely and will be monitored for in the model seperately.

#add a column where the Change Code is stored as character values so arcgisbinding does not change the val
ues on export to fgdb
orivalues.allp4$ChangeCodeNum <- as.character(orivalues.allp4$ChangeCode)

#for models 2 and 3
#models 2 and 3 require a different lookup table as there are more possible combinations due to the introducti
on of additional data and categories.
#expand the grid for all possible values
newvalues.allp <- expand.grid(as.numeric(newValues$New_Value), as.numeric(newValues$New_Value))
newvalues.allp1 <- newvalues.allp[complete.cases(newvalues.allp), ]

#compute the value difference code between 2008 and 2017
newvalues.allp1$var3 <- (newvalues.allp1$Var1 - newvalues.allp1$Var2)^2

#rename for clarity
names(newvalues.allp1) <- c("New_Value", "New_Value2", "New_ValueDiff")

#full join to 2017 to re-capture category names for descriptive text
newvalues.allp2 <- (full_join(newvalues.allp1, newValues, by = c("New_Value")))

#rename for clarity
names(newvalues.allp2) <- c("New_Value2017", "New_Value", "New_ValueDiff", "New_Category2017")

```

```

#full join to 2008 to re-capture category names for descriptive text
newvalues.allp3 <- (full_join(newvalues.allp2, newValues, by = c("New_Value")))

#rename for clarity
names(newvalues.allp3) <- c("New_Value2017", "New_Value2008", "New_ValueDiff", "New_Category2017",
"New_Category2008")

#create descriptive text that describe change
newvalues.allp3$descriptor <- paste(newvalues.allp3$New_Category2008, " to ", newvalues.allp3$New_Category2017)
newvalues.allp4 <- transform(newvalues.allp3, ChangeCode=paste0(17,newvalues.allp3$New_Value2017, 8,
newvalues.allp3$New_Value2008, abs(newvalues.allp3$New_ValueDiff)))

#check for duplicates
n_occur <- data.frame(table(newvalues.allp4$ChangeCode))
n_occur[n_occur$Freq > 1,]

      Var1 Freq
20281 1717282222500  2
21811  171782225  2
22820 171818132400  2
27809 1719881334225  2
37140 172318252441  2
57819  1774844900  2
59615  1780855625  2
61156  1786866400  2
61930  1788886400  2
62951  1792877225  2
64735  1798888100  2

dups2 <- newvalues.allp4[newvalues.allp4$ChangeCode %in% n_occur$Var1[n_occur$Freq > 1],]

#using these change codes has greatly reduced duplication; however 22 instances of potential duplication still
exist. These change types seem unlikely and will be monitored for in the model seperately.

#add a column where the Change Code is stored as character values so arcgisbinding does not change the val
ues on export to fgdb
newvalues.allp4$ChangeCodeNum <- as.character(newvalues.allp4$ChangeCode)

```

The next section uses the arcgisbinding package, the ggplot package can sometimes cause difficulties with arc.open(), arc.select(), and arc.write() so we will detach ggplot2 to start with and load it again when we are ready to use the ggplot2 package.

```

detach("package:tidyverse", unload=TRUE)
#detach("package:ggplot2", unload=TRUE)

```

First, export the lookup tables created above to the geodatabase where they are needed

```

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'orivalues_lut')
arc.write(input, orivalues.allp4)

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'newvalues_lut')
arc.write(input, newvalues.allp4)

```

Import the combined files for confidence, frequency and category

```
#arc open the tables for the confidence - frequency models for 2017; frequency and model output are dynamic and dependent on model parameters. Confidence is static.
input <- file.path(getwd(), 'cdl-comparisons.gdb', 'cdl_30m_r_2017_albers_noModelMods_Combine_Broad')
T2017_noModelMods_Conf <- arc.open(input)

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'cdl_30m_r_2017_albers_noNLCD_Combine_Broad_tbl')
T2017_withoutNLCD_Conf <- arc.open(input)

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'cdl_30m_r_2017_albers_withNLCD_Combine_Broad_tbl')
T2017_withNLCD_Conf <- arc.open(input)

T2017_noModelMods_Conf.SA <- arc.select(T2017_noModelMods_Conf)
T2017_withoutNLCD_Conf.SA <- arc.select(T2017_withoutNLCD_Conf)
T2017_withNLCD_Conf.SA <- arc.select(T2017_withNLCD_Conf)

#arc.open the tables for the confidence - frequency models for 2008: frequency from 2017 is used; confidence is from 2008 CDL and is static; however the frequency and model output are dynamic and dependent on model parameters.

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'cdl_30m_r_2008_albers_noModelMods_Combine_Broad')
T2008_noModelMods_Conf <- arc.open(input)

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'cdl_30m_r_2008_albers_noNLCD_Combine_Broad_tbl')
T2008_withoutNLCD_Conf <- arc.open(input)

input <- file.path(getwd(), 'cdl-comparisons.gdb', 'cdl_30m_r_2008_albers_withNLCD_Combine_Broad_tbl')
T2008_withNLCD_Conf <- arc.open(input)

T2008_noModelMods_Conf.SA <- arc.select(T2008_noModelMods_Conf)
T2008_withoutNLCD_Conf.SA <- arc.select(T2008_withoutNLCD_Conf)
T2008_withNLCD_Conf.SA <- arc.select(T2008_withNLCD_Conf)

#rename for clarity, MaskValue 0 = areas of change; 1 = areas of little or no change
names(T2017_withNLCD_Conf.SA) <- c("ObjectID", "Value", "Count", "Category", "Frequency", "Confidence", "MaskValue")
names(T2017_withoutNLCD_Conf.SA) <- c("ObjectID", "Value", "Count", "Category", "Frequency", "Confidence", "MaskValue")
names(T2017_noModelMods_Conf.SA) <- c("ObjectID", "Value", "Count", "Category", "Frequency", "Confidence", "MaskValue")

names(T2008_withNLCD_Conf.SA) <- c("ObjectID", "Value", "Count", "Category", "Frequency", "Confidence", "MaskValue")
names(T2008_withoutNLCD_Conf.SA) <- c("ObjectID", "Value", "Count", "Category", "Frequency", "Confidence", "MaskValue")
names(T2008_noModelMods_Conf.SA) <- c("ObjectID", "Value", "Count", "Category", "Frequency", "Confidence", "MaskValue")

#add acres calculation
T2017_withNLCD_Conf.SA$Acres <- ((T2017_withNLCD_Conf.SA$Count*0.222394))
T2017_withoutNLCD_Conf.SA$Acres <- ((T2017_withoutNLCD_Conf.SA$Count*0.222394))
T2017_noModelMods_Conf.SA$Acres <- ((T2017_noModelMods_Conf.SA$Count*0.222394))
```

```

T2008_withNLCD_Conf.SA$Acres <- ((T2008_withNLCD_Conf.SA$Count*0.222394))
T2008_withoutNLCD_Conf.SA$Acres <- ((T2008_withoutNLCD_Conf.SA$Count*0.222394))
T2008_noModelMods_Conf.SA$Acres <- ((T2008_noModelMods_Conf.SA$Count*0.222394))

#subset data before melting
T2017_withNLCD_Conf.SA.subset <- data.frame(T2017_withNLCD_Conf.SA$Value, T2017_withNLCD_Conf.SA$Category, T2017_withNLCD_Conf.SA$Confidence, T2017_withNLCD_Conf.SA$Frequency, T2017_withNLCD_Conf.SA$Acres)
T2017_withoutNLCD_Conf.SA.subset <- data.frame(T2017_withoutNLCD_Conf.SA$Value, T2017_withoutNLCD_Conf.SA$Category, T2017_withoutNLCD_Conf.SA$Confidence, T2017_withoutNLCD_Conf.SA$Frequency, T2017_withoutNLCD_Conf.SA$Acres)
T2017_noModelMods_Conf.SA.subset <- data.frame(T2017_noModelMods_Conf.SA$Value, T2017_noModelMods_Conf.SA$Category, T2017_noModelMods_Conf.SA$Confidence, T2017_noModelMods_Conf.SA$Frequency, T2017_noModelMods_Conf.SA$Acres)

T2008_withNLCD_Conf.SA.subset <- data.frame(T2008_withNLCD_Conf.SA$Value, T2008_withNLCD_Conf.SA$Category, T2008_withNLCD_Conf.SA$Confidence, T2008_withNLCD_Conf.SA$Frequency, T2008_withNLCD_Conf.SA$Acres)
T2008_withoutNLCD_Conf.SA.subset <- data.frame(T2008_withoutNLCD_Conf.SA$Value, T2008_withoutNLCD_Conf.SA$Category, T2008_withoutNLCD_Conf.SA$Confidence, T2008_withoutNLCD_Conf.SA$Frequency, T2008_withoutNLCD_Conf.SA$Acres)
T2008_noModelMods_Conf.SA.subset <- data.frame(T2008_noModelMods_Conf.SA$Value, T2008_noModelMods_Conf.SA$Category, T2008_noModelMods_Conf.SA$Confidence, T2008_noModelMods_Conf.SA$Frequency, T2008_noModelMods_Conf.SA$Acres)

#rename for clarity
names(T2017_withNLCD_Conf.SA.subset) <- c("Value", "Category", "Confidence", "Frequency", "Acres")
names(T2017_withoutNLCD_Conf.SA.subset) <- c("Value", "Category", "Confidence", "Frequency", "Acres")
names(T2017_noModelMods_Conf.SA.subset) <- c("Value", "Category", "Confidence", "Frequency", "Acres")

names(T2008_withNLCD_Conf.SA.subset) <- c("Value", "Category", "Confidence", "Frequency", "Acres")
names(T2008_withoutNLCD_Conf.SA.subset) <- c("Value", "Category", "Confidence", "Frequency", "Acres")
names(T2008_noModelMods_Conf.SA.subset) <- c("Value", "Category", "Confidence", "Frequency", "Acres")

#melt and sort each table
T2017_withNLCD_Conf.SA.MELT <- melt(T2017_withNLCD_Conf.SA.subset, id.var="Value", na.rm=TRUE)
T2017_withoutNLCD_Conf.SA.MELT <- melt(T2017_withoutNLCD_Conf.SA.subset, id.var="Value", na.rm=TRUE)
T2017_noModelMods_Conf.SA.MELT <- melt(T2017_noModelMods_Conf.SA.subset, id.var="Value", na.rm=TRUE)

T2008_withNLCD_Conf.SA.MELT <- melt(T2008_withNLCD_Conf.SA.subset, id.var="Value", na.rm=TRUE)
T2008_withoutNLCD_Conf.SA.MELT <- melt(T2008_withoutNLCD_Conf.SA.subset, id.var="Value", na.rm=TRUE)
T2008_noModelMods_Conf.SA.MELT <- melt(T2008_noModelMods_Conf.SA.subset, id.var="Value", na.rm=TRUE)

#new value column as numeric
T2017_noModelMods_Conf.SA.MELT$value.num <- as.numeric(T2017_noModelMods_Conf.SA.MELT$value)
T2017_withNLCD_Conf.SA.MELT$value.num <- as.numeric(T2017_withNLCD_Conf.SA.MELT$value)
T2017_withoutNLCD_Conf.SA.MELT$value.num <- as.numeric(T2017_withoutNLCD_Conf.SA.MELT$value)

```



```

ue)

T2008_noModelMods_Conf.SA.MELT$value.num <- as.numeric(T2008_noModelMods_Conf.SA.MELT$value
ue)
T2008_withNLCD_Conf.SA.MELT$value.num <- as.numeric(T2008_withNLCD_Conf.SA.MELT$value)
T2008_withoutNLCD_Conf.SA.MELT$value.num <- as.numeric(T2008_withoutNLCD_Conf.SA.MELT$value
ue)

#sort data before plotting

library(data.table)

setorder(T2017_noModelMods_Conf.SA.subset, Category, Confidence, Frequency)
setorder(T2017_withNLCD_Conf.SA.subset, Category, Confidence, Frequency)
setorder(T2017_withoutNLCD_Conf.SA.subset, Category, Confidence, Frequency)

setorder(T2008_noModelMods_Conf.SA.subset, Category, Confidence, Frequency)
setorder(T2008_withNLCD_Conf.SA.subset, Category, Confidence, Frequency)
setorder(T2008_withoutNLCD_Conf.SA.subset, Category, Confidence, Frequency)

detach("package:data.table", unload=TRUE)

#first, reload tidyverse/ggplot2
library(tidyverse)

#ggplot data -- stacked columns showing acres by confidence and frequency for each category
ggplot(T2017_noModelMods_Conf.SA.subset, aes(x = Confidence, y = Acres)) + geom_col(aes(fill = Frequen
cy)) + facet_wrap(~ Category, shrink = TRUE, scales = "free_y", ncol=5)

ggsave("model-1-2017-confidence-frequency.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi
= 300)

ggplot(T2017_withoutNLCD_Conf.SA.subset, aes(x = Confidence, y = Acres)) + geom_col(aes(fill = Frequen
cy)) + facet_wrap(~ Category, shrink = TRUE, scales = "free_y", ncol=5)

ggsave("model-2-2017-confidence-frequency.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi
= 300)

ggplot(T2017_withNLCD_Conf.SA.subset, aes(x = Confidence, y = Acres)) + geom_col(aes(fill = Frequency)
) + facet_wrap(~ Category, shrink = TRUE, scales = "free_y", ncol=5)

ggsave("model-3-2017-confidence-frequency.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi
= 300)

ggplot(T2008_noModelMods_Conf.SA.subset, aes(x = Confidence, y = Acres)) + geom_col(aes(fill = Frequen
cy)) + facet_wrap(~ Category, shrink = TRUE, scales = "free_y", ncol=5)

ggsave("model-1-2008-confidence-frequency.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi
= 300)

ggplot(T2008_withNLCD_Conf.SA.subset, aes(x = Confidence, y = Acres)) + geom_col(aes(fill = Frequency)
) + facet_wrap(~ Category, shrink = TRUE, scales = "free_y", ncol=5)

ggsave("model-2-2008-confidence-frequency.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi
= 300)

```

```
ggplot(T2008_withoutNLCD_Conf.SA.subset, aes(x = Confidence, y = Acres)) + geom_col(aes(fill = Frequency)) + facet_wrap(~ Category, shrink = TRUE, scales = "free_y", ncol=5)

ggsave("model-3-2008-confidence-frequency.png", plot = last_plot(), width = 24, height = 32, units = "in", dpi = 300)
```