Journal of Data Science 24 (2), 436–454
April 2026
DOI: 10.6339/25-JDS1215
Computing in Data Science

Quantifying the Sensitivity of Land Use Land Cover Metrics Through Simulation Techniques

Haley Burger[1],* and Brennan Bean[1]
[1]Department of Mathematics and Statistics, Utah State University, United States

Abstract

Land use land cover (LULC) change in the agriculture, is a critical area of concern as it directly impacts food security, environmental health, and economic stability. One of the leading LULC data products is the U.S. Department of Agriculture's (USDA) Cropland Data Layer (CDL). Produced annually by the USDA National Agricultural Statistics Service (NASS) using satellite imagery, the CDL provides crop-specific data with an estimated classification accuracy of 85% to 95% for major crop types across the U.S. However, several limitations inherent to the CDL, such as crop underestimation bias, pixel misclassification, and difficulty distinguishing certain vegetation types, have raised questions about the accuracy of LULC change estimates derived from this dataset. In this paper, we introduce the `R` package `cdlsim`, designed to quantify the sensitivity of CDL-derived metrics through simulations of CDL data at the patch level using NASS published accuracy statistics. We present a case study utilizing landscape metrics calculated with the popular `landscapemetrics` `R` package to demonstrate the utility of `cdlsim` in quantifying the sensitivity of metrics to random perturbations in the data. The case study examines a mixed agricultural and grassland landscape in South Dakota, illustrating how our package enables researchers to achieve a more nuanced representation of land-use change.

Keywords *Cropland Data Layer (CDL); landscape metrics; sensitivity analysis; simulation*

1 Introduction

1.1 Background

Land use land cover (LULC) change in watershed areas poses one of the greatest threats to water quality (Siqueira et al., 2023). One of the leading LULC data products is the U.S. Department of Agriculture (USDA) Cropland Data Layer (CDL). This crop specific data layer is produced annually by the USDA National Agricultural Statistics Service (NASS) from satellite imagery and referenced with ground data. The strength of CDL data lies in its crop-specific categories, with the classification accuracy for major crop types in any state estimated to range between 85% and 95% (NASS, 2023). Given its comprehensive coverage and high accuracy, the CDL is frequently used to calculate landscape-level metrics. Landscape metrics or patterns provide a simple way to analyze landscape structure and can be calculated with free open access software such as `FRAGSTATS` (McGarigal and Marks, 1995) and the equivalent `R` implementation in the `landscapemetrics` package (Hesselbarth et al., 2019). `FRAGSTATS` is a software program originally developed for landscape ecologists, and it is driven by the theory that environmental

*Corresponding author. Email: A02425259@usu.edu.

Received August 15, 2025; Accepted December 19, 2025

spatial patterns strongly influence ecological processes (McGarigal and Marks, 1995). For example, structural changes in a landscape, such as when land is converted from forest to cropland, can impede or encourage ecological flow depending on the situation. Considering the impact landscape structure has on ecosystem health, the methods defined in `FRAGSTATS` fill an important role for ecologists, allowing them to quantify landscape patterns and further the understanding of landscape pattern-process relationships. Using the `R` package `landscapemetrics`, it is easy to quantify patterns like patch density, dominant class type, total area, and total edge length.

1.2 Literature Review

The accuracy of landscape metrics is inherently tied to the quality of the underlying data. Literature consistently highlights the need for post-processing when using CDL data to evaluate LULC change due to inherent data limitations such as region-dependent producer and user accuracies (Reitsma et al., 2016), pixel misclassification (Laingen, 2015), temporal crop area underestimation bias (Copenhaver et al., 2021; Pritsolas and Pearson, 2021), and difficulty distinguishing native versus cultivated grass categories (Lark et al., 2017). Moreover, the absence of standardized best practices for post-processing has led to confusion among scholars and may have propagated problematic methodologies. This variability in approaches has sometimes resulted in conflicting outcomes, even for studies conducted in the same region. To illustrate, the validity of Wright and Wimberly (2013a) research on grassland conversion in the western Corn Belt (e.g., Iowa, Minnesota, Nebraska, and South Dakota) from 2006 to 2011 was called into question due to their decision to include the alfalfa, fallow/idle, and other hay/non-alfalfa classes as grassland instead of crops (Kline et al., 2013; Wright and Wimberly, 2013b). Such discrepancies, could undermine the credibility of CDL-based studies in informing policy and management decisions if left unaddressed.

A few post-processing methods are important to consider when using CDL data to calculate LULC change, as they help address inherent limitations of the CDL in capturing pixel-level changes and accurately calculating landscape metrics. Two of these limitations are outlined in Lark et al. (2017), and include the following:

- The first limitation arises from the CDL's 30-meter resolution, which is too coarse to capture land use changes in a small area. Additionally, edge effects can inaccurately show change along field boundaries over time due to variation within pixels. For example, the boundary between rural roads and adjacent fields may be inconsistently mapped from year to year since these roads are typically less than 10 meters wide, leading to classifier uncertainty in mixed-class pixels. Considering this, the CDL data is more suited to detect changes in larger plot areas, such as whole fields or patches.
- The second limitation in assessing change in landscape metrics originates from evolving CDL classification methodologies, which have improved data accuracy in recent years. The CDL has historically under-predicted the total crop area relative to the Farm Service Agency (FSA) Common Land Unit (CLU) Program reference data. Improvements in the CDL's negative crop mapping bias have most notably increased homogeneity of mapped fields, reducing the occurrence of within field speckle (i.e., misclassified pixels scattered within a mono crop field). As a result, the fields mapped in early years of the CDL as a checkerboard of two crops have since been correctly mapped as single-crop fields in later years. This improvement in accuracy may give the false impression of reduced landscape diversity and increased patch size. Moreover, uncorrected yearly mapping biases can inflate crop acreage estimates

based on pixel areas, falsely indicating crop expansion due to reduced negative bias rather than actual conversion of non-crop land to crops. Therefore, researchers should consider calibrating CDL data with external sources, such as Census of Agriculture acreage estimates or the National Resources Inventory (NRI), to account for these temporal improvements in accuracy.

Another important post-processing consideration is reclassification of the original CDL classes, as this can dramatically influence results. For example, Laingen (2015) analyzed the use of different remote sensing products including the CDL and the National Land Cover Dataset (NLCD) to characterize cropland expansion in South Dakota. The discrepancies in cropland expansion primarily stem from the decision to classify grass-like categories as either crops or non-crops. The most problematic classes are alfalfa, fallow/idle, and other hay/non-alfalfa. The choice of whether to define these categories as crops is typically left up to researchers, which can greatly alter study outcomes. As shown in the analysis by Laingen (2015), cropland expansion in South Dakota between 2006 and 2012 was examined using two CDL classification systems – one with the other hay/non-alfalfa class as cropland and one with this class as non-crop. This revealed a nearly 4 million-acre increase in cropland in the CDL data when the class was included as a crop. Therefore, using the CDL to draw conclusions on LULC change warrants careful examination and requires researchers to provide appropriate justification for their reclassification choices.

Given the inherent errors and noise in the CDL, several techniques have been employed in the literature to improve data quality, including reclassifying data into aggregated categories, using independent sources to verify observed trends, and setting a minimum or maximum unit of change (Lark et al., 2021, 2017). More recent methods have leveraged machine learning techniques, computer vision, and a pixel's spatial neighborhood to identify and correct misclassifications (Lin et al., 2022; Zhang et al., 2023, 2021). The advantage of these methods over more traditional techniques is that they can more efficiently be scaled to the entire U.S and offer a consistent approach to data correction, which can potentially improve analysis across different studies.

In one such approach, Lin et al. (2022) used a decision tree algorithm to identify pixels prone to misclassification and reassign their values based on spatial-temporal crop trends. The improvements in this study addressed some of the concerns raised by Lark et al. (2017) and could help reduce inaccuracies in boundary mapping. The authors reclassified vulnerable pixels based on the historically dominant classification and the spatial relationship of a pixel to its neighbors. Pixels were grouped into major classes based on how similar their classification was to their eight surrounding neighbors. For instance, a pixel with all neighbors of the same class as itself was considered likely to be accurately classified and was not further refined. The most challenging type of spatial neighborhood to correct involved pixels where fewer than four of the eight surrounding cells shared the same class. These pixels, termed "mixed pixels", were not improved by the methods used by Lin et al. (2022) because historical classification and spatial relationships do not provide useful information for such pixels.

Computer vision, using Meta's Segment Anything Model (SAM), is one of the newest methods for reassigning pixel values based on satellite imagery of crop regions. The advantage of SAM in this application is its ability to process images without the need for additional training. In Zhang et al. (2023), the group used SAM to delineate crop field boundaries and reclassify pixels within a field. This strategy reduces noisy pixels within fields that are probably a result of cloudy or mixed spectral signals. The segmented cropland units are superimposed on CDL data and refined to be the mode within a given field's boundaries. Initial results are promising, with the major limitation being the ability of this method to accurately segment fields. However, this could be improved in future studies with pre-training images. As field segmentation becomes more accurate, it could be incorporated into reclassification workflows with minimal effort.

The previously mentioned methods provide ways to improve LULC classification in future data products. However, these methods do not necessarily address the misclassifications inherent in existing products, nor do these methods remove the need to understand the influence of misclassification on LULC metric calculations. From the literature, it is evident that researchers choose to apply a variety of post-processing techniques to the CDL to improve data quality, yet the same metrics are often compared across studies without considering the effect of variability in the underlying data. Further, the results of such studies may have real world implications informing policy and land management decisions, making it critical to understand this sensitivity. Accordingly, this paper provides a programmatic way to quantify the sensitivity of various LULC metrics to potential misclassifications in the data. We do this by leveraging accuracy statistics published by NASS to simulate perturbations in the landscape at the patch level. Unlike prior methods that aim to correct misclassified pixels or directly adjust for error in the data, our simulation characterizes the variability in landscape metrics calculated on CDL data, highlighting the strengths and limitations of the CDL data for specific applications. Importantly, our aim is not to compare these metrics relative to some "ground truth," but rather to quantify the imprecision of the various LULC metrics relative to the reported misclassification rates for the CDL and other remotely sensed datasets.

To conduct our simulations, we utilize the `R` package `cdlsim`, which was developed as part of this research and represents a primary contribution of this work. The remainder of this paper outlines the methodology of the `cdlsim` package and explores its practical applications (Burger, 2025).

2 Data

The Cropland Data Layer is an annual land cover product that integrates satellite imagery and a decision tree classifier to provide detailed land cover information for the conterminous United States (CONUS). CDL classifications are based on imagery from multiple satellite sources, including Landsat and Sentinel satellites. Since 2006, data derived from the Farm Service Agency (FSA) Common Land Unit (CLU) Program has been used to train and validate the agricultural classifications. To train non-agricultural categories, the CDL program relies on reference data derived from the United States Geological Survey (USGS) National Land Cover Dataset. This is because the CLU contains little information on non-crop categories. The CDL currently has 256 possible land cover classifications, with the main focus of the program being large area summer crops. It also contains information to identify fields with multiple crop types planted in the same year. Currently, categories only exist for double-planted fields such as winter wheat followed by soybean, and the program is not equipped to capture more complex crop rotations.

The CDL is publicly available to download at a standard 30-meter resolution from the USDA's `CroplandCROS` web application or with the `R` package `cdlTools` (Lisic et al., 2024; USDA, 2024). The CDL products used in this study are downloaded using the `cdlTools` function `getCDL` to get data for South Dakota for 2008 to 2023. To perform our analysis we aggregate the data to a 90-meter resolution to smooth over small inconsistencies such as roads and streams that could otherwise introduce noise into simulation results. During aggregation, we assign each

3x3 group of 30-meter grid cells to the most frequently occurring class within its corresponding 90-meter square. The CDL also includes information on non-agricultural classes from the USGS NLCD, this data can be obtained from the USDA Natural Resources Conservation Service Geospatial Data Gateway (United States Department of Agriculture, Natural Resources Conservation Service, 2025). The CDL data comes in a standard raster image format, but we convert the data to `SpatRaster` objects (Hijmans, 2025). The `terra` package allows for efficient manipulation of large data sets without loading the objects into memory. This is particularly advantageous for statewide or regional analyses.

3 The R Package

We create the R package `cdlsim` to provide a practical method to asses the sensitivity of common landscape metrics in response to pixel classification uncertainty in the CDL data. This uncertainty is in part derived from the fact that at a 30-meter resolution, the data is too coarse to capture variation within individual pixels (Lark et al., 2021). For instance, pixels containing multiple unique categories produce distinct spectral signatures in the raw imagery, which may lead to misclassification by the CDL program's algorithm (Boryan et al., 2011). These misclassification errors are then propagated when metrics are calculated across a landscape and over time. Our package provides a way to estimate the sensitivity of landscape metrics by simulating a patch's class based on the user and producer accuracy provided in the CDL's confusion matrices. Patches here are defined as in the `landscapemetrics` package as groups of pixels that are all of the same class. The probability that a cell will transition to a new class is determined by confusion matrices provided with the CDL for each state. Diagonal values represent correctly classified pixels, column sums represent the total number of reference pixels per category, and row sums represent the total number of pixels classified as each category. We generate transition matrices from these confusion matrices by normalizing each cell by the row sums, which are then used for the remainder of our analysis. For study areas spanning multiple states, we have chosen to generate combined transition matrices by calculating a weighted average of the transition probabilities for each pixel category. This approach allows us to quantify how metrics respond to classification uncertainty, providing insight into the sensitivity of commonly used metrics. We emphasize that this assessment is not the same as a confidence interval, which tries to bound a true and unknowable population parameter. Rather this assessment evaluates the relative precision of a selected metric to perturbations in the data. Evaluating the appropriateness of these standard metrics lies beyond the scope of this paper.

The two major purposes of this package are to download and format the transition matrices and to simulate the CDL data provided as the input. To accomplish the former, we have developed three main functions: `download_cdl_mat_files`, `get_mat_data`, and `get_trans_mat`. These functions download, extract, and format the CDL confusion matrices for all US states, ensuring compatibility with downstream simulation processes. The transition matrices are organized such that the cells in a row describe the probability that a unique cell value has to transition to any of the possible values. The row names represent the input value of a cell and the column names represent the possible output values (See Table 1).

The second major purpose and the bulk of the `cdlsim` package is the simulation capability available in the `simulate_raster_patch` function. To make the simulations fast and efficient we exploit the `terra` package `app` function. This function is particularly useful because it allows

Table 1: An example transition matrix. Input classes (original patch values) are shown by row; output classes (simulated patch values) by column.

		Output Class					
		0	1	2	3	4	Total
Input Class	0	1.0	0	0	0	0	1.0
	1	0	0.90	0.01	0.07	0.02	1.0
	2	0	0.01	0.91	0.02	0.06	1.0
	3	0	0.03	0.03	0.92	0.02	1.0
	4	0	0.01	0.05	0.01	0.92	1.0

users to apply a custom function directly to the cells of each layer in a `SpatRaster` (i.e., a gridded map of classifications suitable for analysis in R's `terra` package), optimizing computational time and memory usage. To take full advantage of this functionality, we structure the input to our simulations as a two-layer `SpatRaster` stack, meaning a single object that contains two layers each representing a map of values for the same spatial area. In this object the first layer represents the original landscape values, while the second layer contains unique patch ID values. We create the second layer by identifying patches using several functions from the `landscapemetrics` package.

Our structured approach ensures that each cell within a patch transitions as a unit, maintaining spatial integrity, while still allowing each patch to transition independently based on the probabilities defined in the transition matrix. This approach also enables the direct generation of simulated patch values from the original landscape for each specified iteration, ensuring that the simulated values in each iteration are treated independently. To achieve this for each unique patch, a class value is assigned for each simulation iteration by sampling with replacement from the possible categories, with likelihood determined by the probabilities defined in the associated transition matrix. The result is a `SpatRaster` object with a layer for each iteration of the simulation. In addition to the previously mentioned arguments, such as the input `SpatRaster`, the transition matrix, and the number of iterations, certain values can optionally be set to remain unchanged during the simulation. The list below details the major elements of the simulation methodology.

3.1 Simulation Process

1. **Identifying and Structuring Patches**
   (a) Use `get_patches` from `landscapemetrics` to assign unique patch IDs to the landscape.
   (b) Combine outputs into a two-layer `SpatRaster`:
   - Layer 1: Original classifications.
   - Layer 2: Corresponding unique patch IDs.
2. **Preparing for Simulation**
   (a) Apply transitions efficiently using `terra::app`.
   (b) Implement transition logic in C++ via `Rcpp` for improved computational speed.
3. **Transitioning Patches**
   (a) Treat each patch as an independent unit during transitions.
   (b) Sample new class values according to transition matrix probabilities.
   (c) Store sampled values in a list of matrices, where:

- Rows represent patch IDs, and columns represent simulated class outcomes.

4. **Generating Simulated Landscapes**
   (a) Assign simulated values to patches while preserving spatial structure.
   (b) Output a `SpatRaster` with a layer for each simulation iteration.
   (c) By default, the background class (0) does not transition.

The limits of our simulation function, `simulate_raster_patch`, depend on both the number of classes and size of the area being simulated. In order to accommodate larger study areas, the `cdlsim` package also includes a memory-efficient option in the simulation function to handle larger raster files by forcing intermediate rasters to be written to disk avoiding limitations in RAM. In both types of simulations, increasing the number of classes reduces the total area that can be simulated given limited RAM, but further testing would be needed to determine exactly how this relationship scales. For additional technical details, including function descriptions and argument specifications, we refer readers to the `cdlsim` GitHub repository (Burger, 2025).

4 South Dakota Case Study

4.1 Study Area and Background

To demonstrate the utility of our simulation procedure, we focus on a case study in three southern South Dakota counties. This region is characterized by a diverse mix of cropland and grassland categories providing an opportunity to evaluate land cover transition over time. We were particularly interested in areas like this due to recent studies emphasizing the need for a nuanced understanding of the grassland to cropland transitions under low classification accuracy scenarios (Wright and Wimberly, 2013a,b; Kline et al., 2013). These papers highlight how different grassland classification schemes and large uncertainties in the data can confound the results of land cover change-analysis. According to the USDA, the classification accuracy of grassland categories is often low because the satellite imagery used lacks sufficient resolution to distinguish between different grassy land cover types (NASS, 2023). Additionally, these categories have not been consistently defined across states or over time. To address this issue, in 2014 the USDA retroactively updated CDL maps from 1997 to 2013 to ensure consistent category definitions across all states. Given these challenges, the `cdlsim` package can help estimate the sensitivity of landscape metrics in regions with a high prevalence of grassland categories. For the purpose of this analysis the categories that fall into grassy land cover types in the CDL code include the NLCD derived class Grassland/Pasture, CDL derived non-crops such as Fallow/Idle Cropland, Pasture/Grass, and CDL crop categories Sod/Grass Seed, Switchgrass, Alfalfa, and Other Hay/Non Alfalfa.

To evaluate `cdlsim`'s performance in a region representative of both agricultural dominance and mixed grassland cover, we selected Brule, Lyman, and Buffalo counties in South Dakota for analysis from the years 2008 to 2023. This period was chosen because 2008 marked the first year the CDL released maps for all of CONUS, while 2023 is the most recent year available at the time of writing. This tri-county area highlights the scalability of our package to a diverse landscape comprising both cropland and grassland to allow us to examine how metric sensitivity varies in scenarios with mixed classification accuracy, within a single year and across time (Figure 1). Further, this region spans three Level III ecoregions as defined by the Environmental Protection Agency (EPA), illustrating the transition from native prairies in the Great Plains to cropland in the Glaciated Plains (U.S. Environmental Protection Agency, 2025).

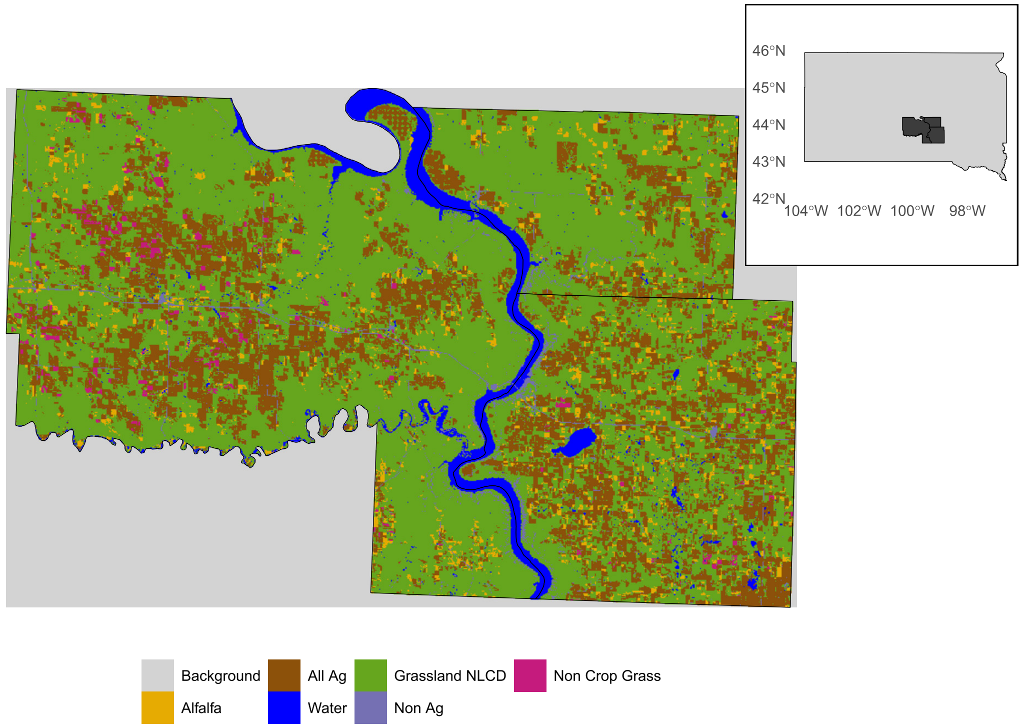

Figure 1: Brule, Lyman, and Buffalo counties in South Dakota, reclassified into the seven main land cover categories. The inset map in the top right shows the location of these counties (in gray) within the state of South Dakota.

To explore metric sensitivity in this area, we adopted a custom classification scheme that emphasizes the characteristic mix of grassland and agriculture found in South Dakota and much of the Great Plains. This scheme consists of seven categories, with agriculture and grassland types making up the majority of classes. We treat water as a separate category to designate it as a non-transitioning class, ensuring perennial water bodies do not transition to illogical land cover types (Table 2). It is important to note that, according to NASS, the accuracy of non-agricultural classes in the CDL is dependent on the NLCD data used as ground truth during model training. As a result, NASS recommends using the NLCD for studies focusing on non-crop areas. However, even when the primary focus is not on non-agricultural land, nearly

Table 2: Land use classification scheme for the South Dakota study area.

Class	Original CDL Values	New Value
Background	0	0
All Ag	1-59, 66-77, 204-254	1
Grass NLCD	176	2
Non Crop Grass	61-62	3
Alfalfa	36-37	4
Water	83, 111, 190, 195	5
Non Ag	63-65, 78-82, 84-110, 112-175, 177-189, 191-194	6

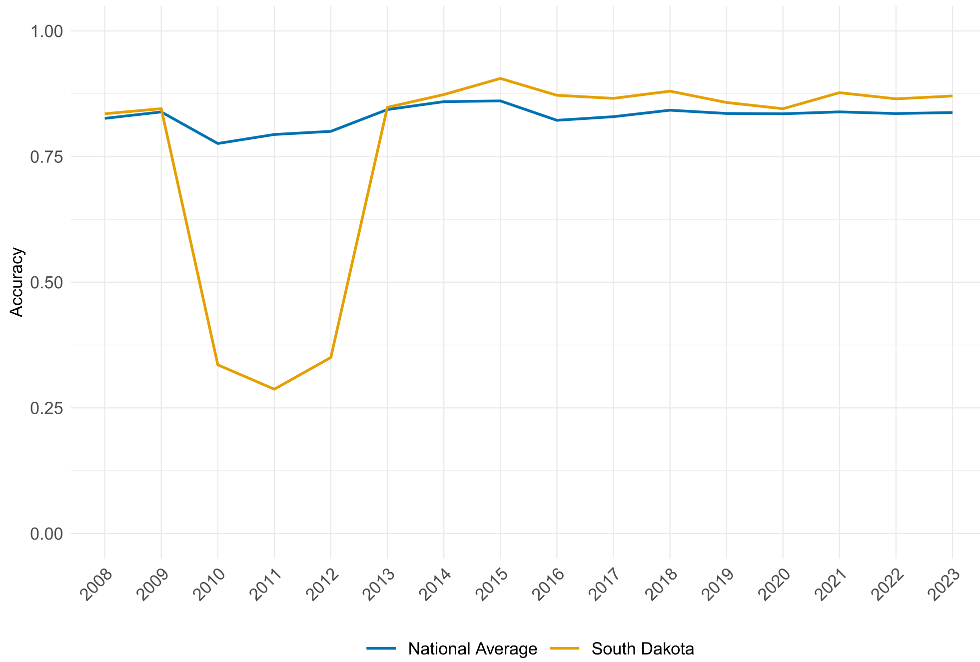

Figure 2: National overall classification accuracy averaged across all CDL states for each study year versus yearly overall accuracy in South Dakota.

all studies utilizing the CDL categorize data into crop and non-crop classes to some extent. Therefore, understanding the variability in non-agricultural classifications in the CDL is crucial for accurately analyzing crop-focused research questions.

One notable feature in the CDL data for South Dakota is a significant drop in overall classification accuracy from 2010 to 2012 (Figure 2). While the CDL generally maintains an average overall classification accuracy of 80%, during this period the overall classification accuracy in South Dakota dropped to 28–35% (NASS, 2023). State-level accuracies were obtained directly from the NASS provided confusion matrices, reported as the proportion of correctly classified pixels across all categories. These reported values were used as the yearly overall accuracy for South Dakota. National average accuracy was calculated as the mean of the state-level accuracies for each year. An unweighted mean was used because the number of pixels sampled in each state is not proportional to geographic area. Additionally, the number of pixels sampled during these low-accuracy years was significantly higher than in other years. For example, from 2010 to 2012 the number of pixels sampled ranged from 19 million to 28 million, while for all other years there was only 1 million to 10 million pixels sampled. This increase in sample size was largely driven by a few categories, particularly the NLCD Grassland/Pasture category (i.e., Grass NLCD), which had poor classification accuracy, driving the overall accuracy down.

To evaluate the sensitivity of metrics in our tri-county region, we applied common diversity and class level metrics from the `landscapemetrics` package. Additionally, we compare simulated metrics to those of the original landscape over time. This analysis demonstrates the utility of the `cdlsim` package in characterizing the variability and sensitivity of landscape metrics in multi-year change studies. Understanding how sensitive these metrics are to random perturbations over time can help researchers distinguish meaningful land cover changes from other trends inherent to the data. Given the fluctuations in accuracy and sample sizes in South Dakota, we performed simulations using two different transition matrix approaches:

- **Aggregated Simulation:** The confusion matrices for all years were summed element-wise and then normalized by row totals, smoothing over fluctuations in classification accuracy, including the drop seen in the early 2010's.
- **Yearly Simulation:** Individual confusion matrices for each year were used to represent transition probabilities, allowing some years to exhibit higher rates of change while others align more closely with the aggregated matrix.

During the creation of our transition matrices, we identified several classes—171, 181, and 182—that appeared in the raw confusion matrices but were not documented in the CDL data. Upon investigation, we determined that these classes originated from the NLCD data and were used to generate the CDL Class 176 (Grass/Pasture). To prevent double-counting pixels, we removed these classes from the transition matrices. For the analysis, we constructed 95% sensitivity intervals using the 2.5th and 97.5th percentiles of the simulated metrics' distribution, using 100 simulated layers per year to reduce computational time. Additionally, we calculate the 50th percentile for reference.

Sensitivity intervals characterize the variability of metrics after random perturbations and are distinct from confidence intervals, as they are not centered on a true parameter value but rather aim to describe the range of metric values under simulated uncertainty. Indeed, it is a known problem that bootstrap sensitive measures applied to "richness" indicators (i.e., indicators that rely on the abundance of a diversity of species, or in this case, land use types) are known to be biased when there is a large number of rare groups (Smith and van Belle, 1984). Alternatives to the bootstrap for richness indicators that mitigate this bias have been explored (Chiarucci et al., 2003), and future work may consider extending the patch-based simulation framework presented in this paper to other estimation procedures. In this paper, the key interest in the sensitivity metric is not the center of the interval, which is expected to be biased low in the presence of a large number of rare classes, but rather in range of simulated values and/or distances between key percentiles such as the quartile range. These measures of interval length provide insight into the relative imprecision of a LULC metric for a particular landscape and LULC classification scheme.

4.2 Metric Sensitivity Results

First we compare the results of Shannon's Diversity Index (SHDI) and Simpson's Diversity Index (SIDI) between the two simulation approaches. Shannon's Diversity Index (SHDI) is a popular metric in landscape ecology and biodiversity studies, as it accounts for both the richness and abundance of classes in a landscape. This metric is represented mathematically as

$$\mathrm{SHDI} = -\sum_{i=1}^{m} (P_i \cdot \ln P_i), \tag{1}$$

where P_i is the proportion of class i and m is the number of classes. The SHDI value is lowest at 0 when $P_i = 1$, meaning the landscape consists of only one class. The index is maximized when $P_i = \frac{1}{m}$ for all i, which occurs when the number of classes increases, and the proportions remain evenly distributed (Hesselbarth et al., 2019). We also calculate the Simpson's Diversity Index (SIDI) for our simple landscapes using the `landscapemetric R` package. SIDI is represented mathematically as

$$\text{SIDI} = 1 - \sum_{i=1}^{m} P_i^2, \tag{2}$$

where P_i is the proportion of class i and m is the number of classes. Unlike Shannon's Diversity Index, SIDI values are less sensitive to rare classes. However, both metrics are maximized when all classes are proportioned evenly, and minimized when the landscape consists entirely of a single class. A key distinction is that SIDI values are bounded between 0 and 1, making comparison across landscapes easier. Additionally, SIDI reflects the probability that two cells randomly selected with replacement belong to different classes, providing a robust measure of diversity.

In the aggregated simulation, the simulated landscape's median more closely follows the value of the original landscape for both diversity metrics. While the yearly simulation also tracks the original landscape's value, it tends to deviate more in the earlier years compared to the aggregated approach. Additionally for most years, both simulation types produce 95% sensitivity intervals where the upper bound is considerably farther from the median than the lower bound (Figure 3). This suggests that the simulations in this study area more readily generate landscapes with higher diversity values rather than reducing diversity below the original levels. This behavior is likely influenced by the distribution of patches in the study area. Although only seven land cover classes are present, the patches are generally small and distinct, which may allow the simulated landscapes to have a more even distribution of classes. As a result, the pixel counts become more uniform, leading to increased SHDI and SIDI values. One notable exception to this trend occurs in the yearly sensitivity intervals for 2008–2009 and 2013–2014, where the lower bounds extend considerably below the original landscape's value. This may be due to the behavior of the Grass NLCD Class, which contains the largest patches in the landscape. During these periods, the transition probabilities for the Grass NLCD Class are more evenly distributed among all classes, whereas in other years, it is concentrated in only a few classes. As a result, the simulated landscapes exhibit greater variability, with more classes potentially occupying the largest patches, leading to both low and high diversity landscapes.

We begin our analysis of class-level metrics with PLAND, which represents the percentage of the landscape occupied by each class. The Water Class, which does not transition, serves as a control in the simulations. As expected, no sensitivity intervals appear around the median value for either simulation type, confirming that this class remains unchanged. Moreover, water bodies in this region of South Dakota are stable over time; the percent land cover for this class is consistently maintained at around 4% year over year. The Non Ag Class, which has the lowest PLAND value at 1%, also exhibits relatively narrow sensitivity intervals, with simulation medians closely following the original landscape's value. Similarly, Alfalfa, which represents a small fraction of the landscape with a median close to 1%, shows greater sensitivity in the early years under the yearly simulation approach. During this period, the upper bound of its sensitivity intervals reaches nearly 60%, whereas the aggregated simulation exhibits more stable intervals across all years. The All Ag Class, covering approximately 20% of the landscape, shows similar sensitivity intervals across both simulation types. In both cases, the upper bounds of the intervals consistently extend further from the median than the lower bounds, suggesting a bias toward increased land cover in the simulations. The Grass NLCD and Non Crop Grass Classes, covering around 50% and 1% of the landscape, respectively, exhibit the most variability in their PLAND sensitivity intervals. Interestingly, their values tend to mirror each other, particularly after 2010—when Grass NLCD increases, Non Crop Grass decreases, and vice versa (Figure 4). Additionally, for these classes the sensitivity intervals for the yearly simulation type align more closely with the original landscape's median, suggesting that year-to-year variability is not well captured by the aggregated transition matrix.

Patch density (PD), is an aggregation metric describing the number of patches in each class per 100 hectares. This metric proved to be fairly robust to random perturbations, with very short sensitivity intervals across all classes and years (Figure 5). This metric had several original landscape values from 2010 to 2012 that fell outside the sensitivity intervals of both simulation types. This is not unexpected, as sensitivity intervals are designed to quantify variability, or imprecision, rather than to provide a distributional estimate of a true population parameter. Additionally, the aggregated simulation medians tended to deviate further from the original landscape's true value compared to the yearly simulations, a trend most evident in the Non Crop Grass and Alfalfa Classes. The lack of sensitivity suggests that although patches may be transitioning between classes during simulation, the landscape likely consists of classes with relatively uniform patch sizes, preventing any single class from dominating the simulation. This could explain why the density of patches within each class remains relatively stable across the landscape even after simulation.

Lastly, we examined mean patch area by class. Across all classes, sensitivity intervals gen-

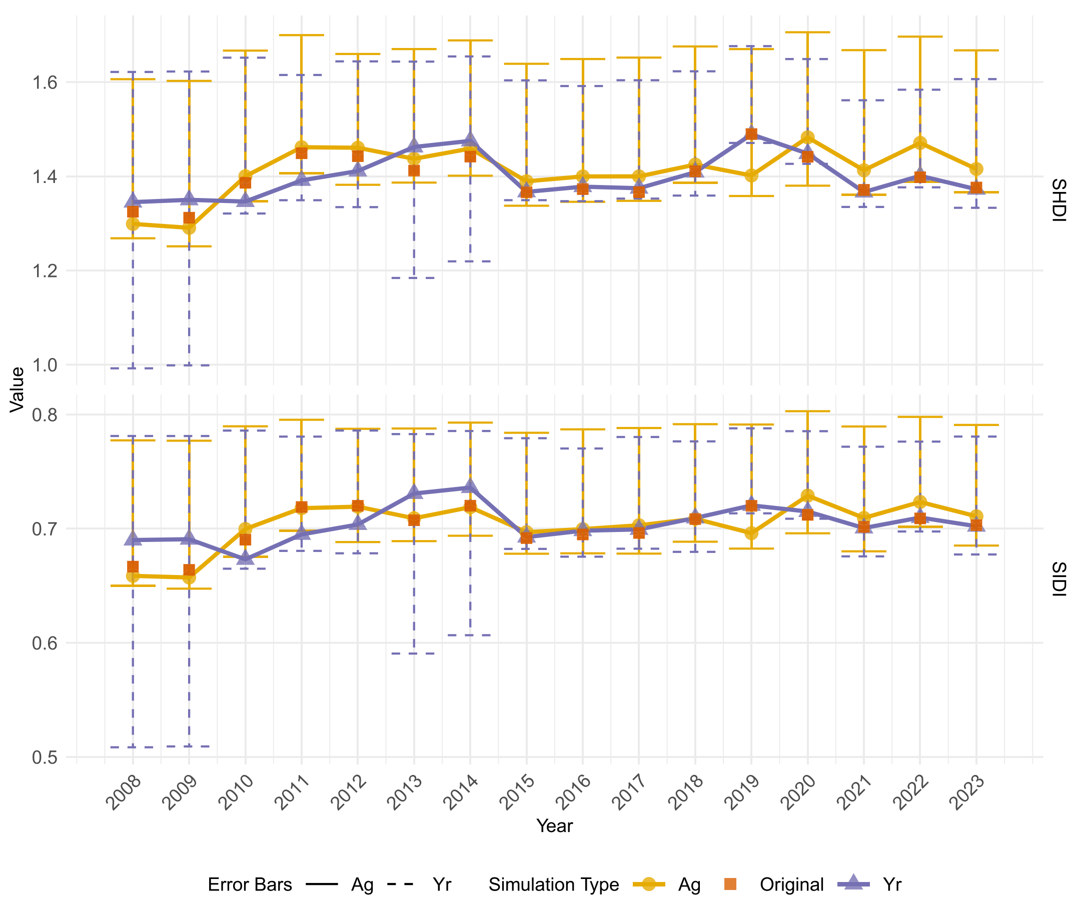

Figure 3: Shannon's Diversity Index (SHDI) and Simpson's Diversity Index (SIDI) metrics for the original landscape and both simulation types: Aggregated (Ag) and Yearly (Yr). The median along with 95% sensitivity intervals are included for each year from 2008 to 2023.

erally had upper bounds that extended further from the median than the lower bounds, which remained close to the median values. This indicates that simulations often created landscapes with more large patches than small patches. Among the four classes analyzed, All Ag and Alfalfa exhibited the most stability, with little change in mean patch area from 2008 to 2023 in both the simulations and the original landscape (Figure 6). In contrast, the Grass NLCD and Non Crop Grass Classes displayed a mirrored relationship in their simulated values, while the original landscape values remained stable over time. After 2012 this trend is particularly noticeable, where we see the yearly simulation medians aligned closely with the original landscape values and the aggregated simulation medians remain above or below in each of these two classes. These

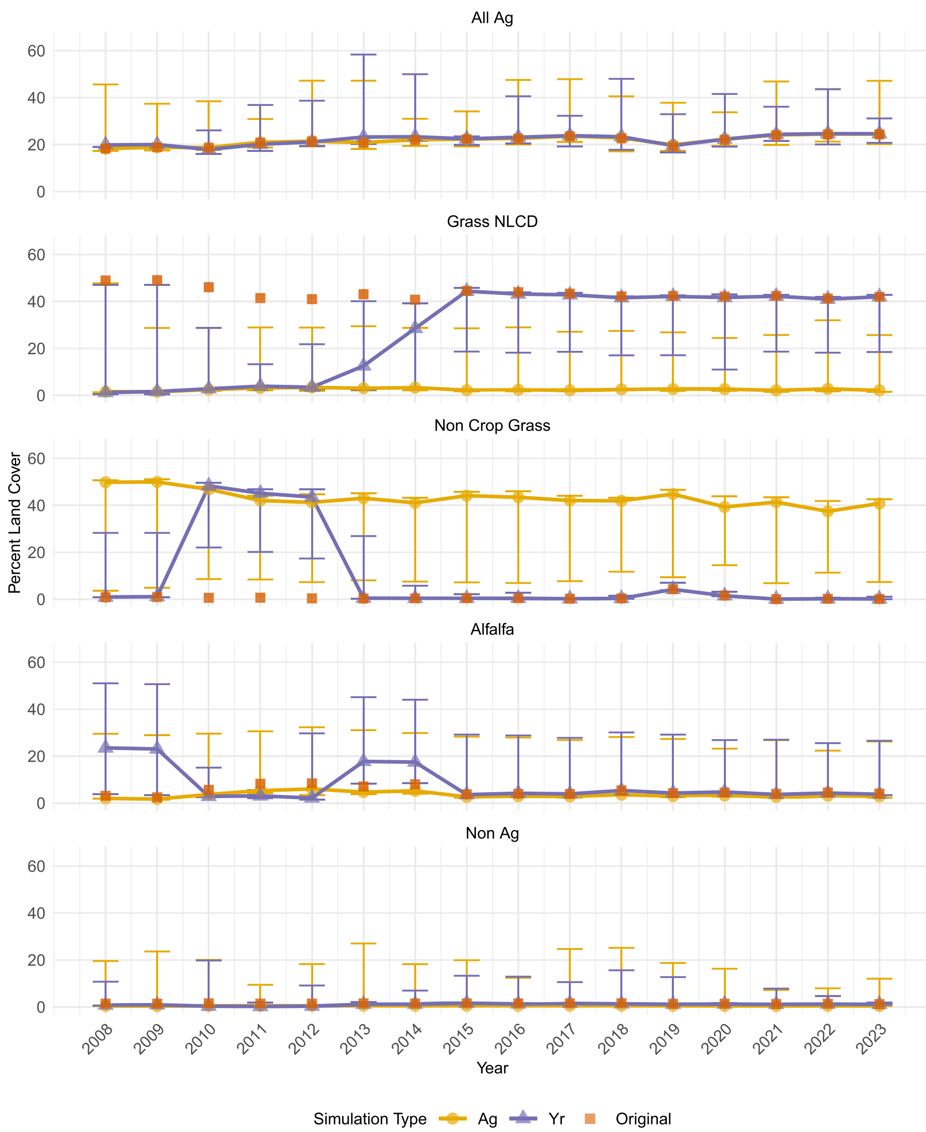

Figure 4: Percent Land Cover (PLAND) for the original landscape and both simulation types: Aggregated (Ag) and Yearly (Yr). The median and 95% sensitivity intervals are included for all study years.

results together suggest that the aggregated simulations produce sensitivity intervals that tend to deviate further from the original landscape values in a class-dependent manner.

Our case study in these three South Dakota counties demonstrates the utility of sensitivity intervals in evaluating how landscape metrics respond to data variability and random perturbations introduced through simulation. Through the use of both aggregated and yearly transition matrices, we assess the extent to which landscape metrics remain stable or fluctuate under different assumptions about classification accuracy. Across our two diversity metrics, we observe that the upper bounds of sensitivity intervals generally extended further away from the median value relative to the lower bounds, suggesting that in these counties simulated landscapes were more likely to exhibit increased diversity rather than reduced diversity. This may be due to configuration of patches in this area specifically, allowing for more diversity during simulation. For class-level metrics, PLAND results show that the Grass NLCD and Non Crop Grass categories exhibit mirrored trends, reflecting each class's change over time. In contrast, patch density remains relatively stable, indicating robustness to perturbations. For mean patch area the results vary by simulation type and class. Aggregated simulations introduce more deviation from the original landscape than yearly simulations, particularly in the earlier years of data. This analysis highlights the importance of considering metric sensitivity when interpreting long-term

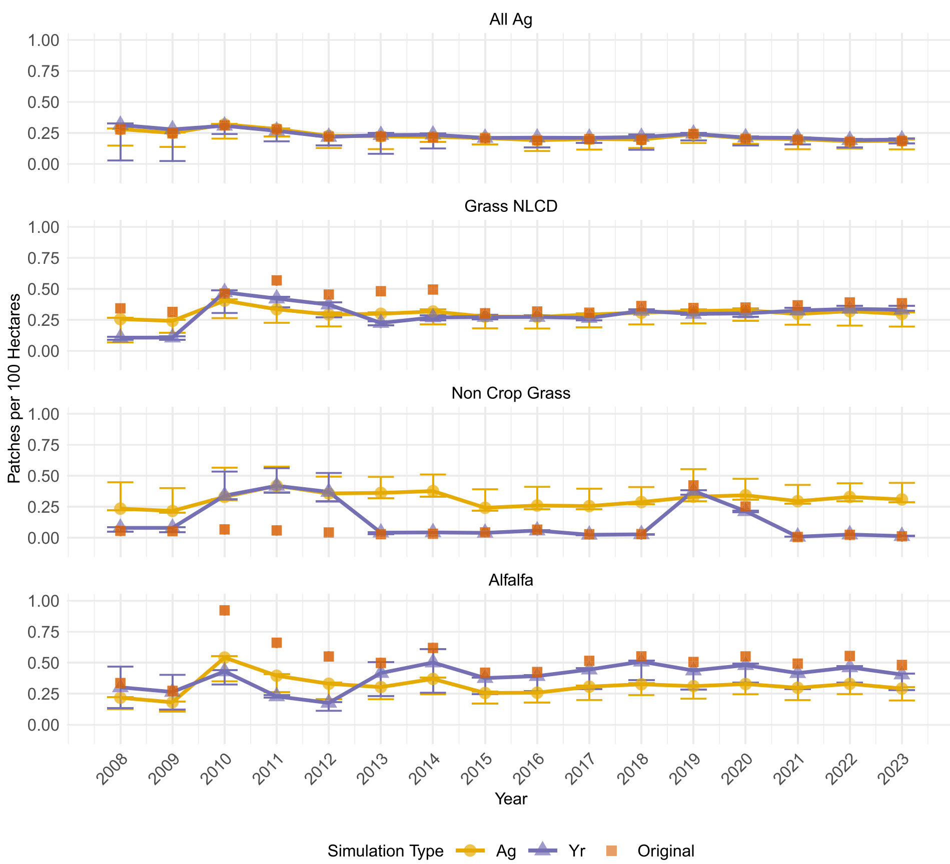

Figure 5: The patch density by class for the original landscape and both simulation types: Aggregated (Ag) and Yearly (Yr). The median and 95% sensitivity intervals are included for all study years.

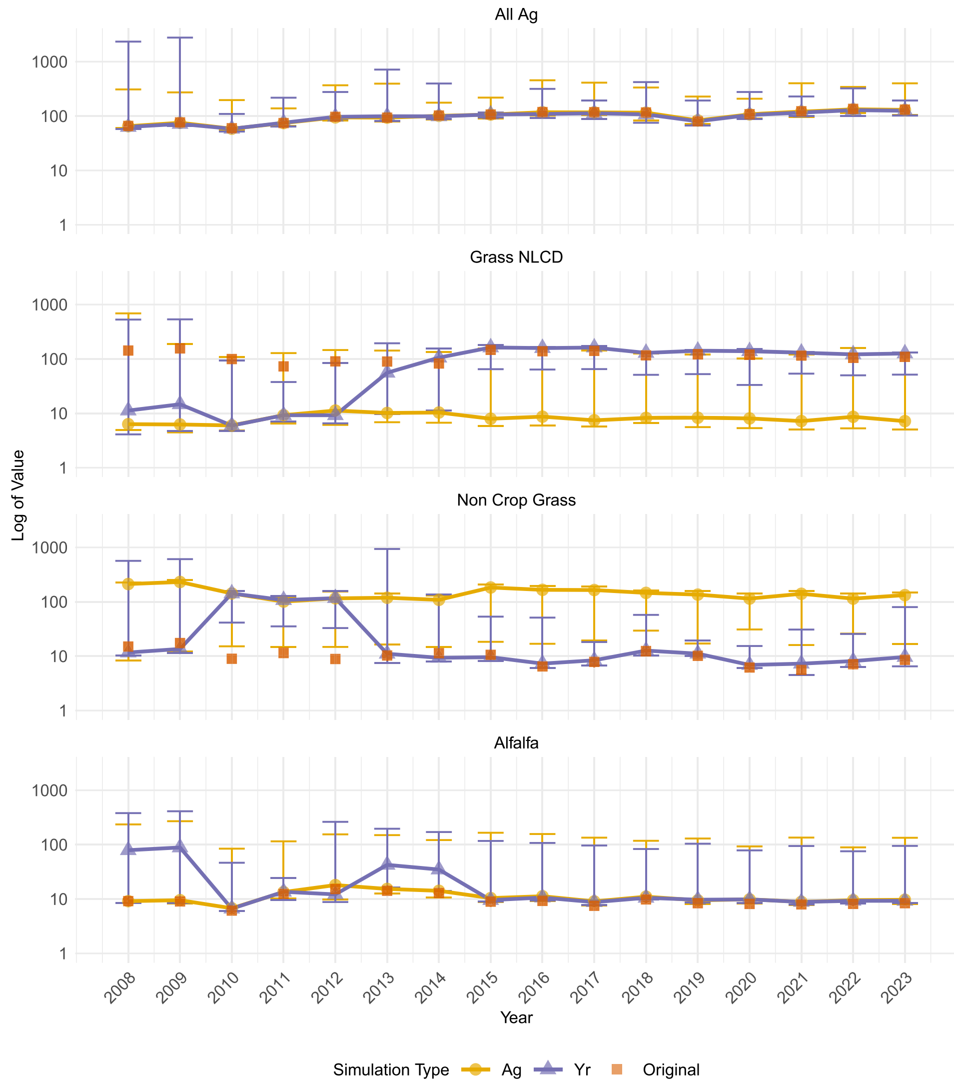

Figure 6: The mean patch area (hectares) by class for the original landscape and both simulation types: Aggregated (Ag) and Yearly (Yr). The median and 95% sensitivity intervals are included for all study years.

land cover trends, particularly in regions where land cover categories or data years exhibit high variability.

5 Conclusions

5.1 Summary of Findings

Given the increase in demand for LULC data analysis, new methods need to be developed to quantify the sensitivity of landscape metrics to random perturbations in the input data. Providing an estimate of the sensitivity of these calculations will give researchers greater context when making decisions using LULC products. Our results show how a simulation of CDL data at the patch level may better quantify the sensitivity of landscape metrics in particular. Additionally, we produced a fast and scalable simulation, our current version using both `terra` functions and a C++ helper function, allowing application to large geographical areas. Another key strength of the `cdlsim` package is its ability to use real confusion matrices from each year of CDL data or allow users to supply their own transition matrices; this increases the versatility and application to other geospatial datasets like the National Land Cover Dataset.

To assess simulation performance in a real world application, we conducted a case study in Brule, Lyman, and Buffalo counties in South Dakota (see Section 4). This case study incorporated data from 2008 to 2023 and showcased the scalability or our simulation to the county level. Additionally, we saw unique accuracy trends in the CDL data for South Dakota with a substantial deviation from the national average in 2010 to 2012 highlighting the need for a nuanced approach to analysis in this region. Sensitivity intervals generated from the simulations provided key insights into how landscape metrics respond to inherent data variability and random perturbations introduced through simulation. Class-level PLAND results revealed that the Grass NLCD and Non Crop Grass categories exhibited a mirrored pattern, emphasizing the difference between simulations with the aggregated and yearly transition matrix types. The patch density metric was generally more robust to random perturbations with shorter sensitivity intervals observed over time in both simulations. The sensitivity intervals for mean patch area were dependent on both the type of simulation type and the year of CDL data. We again saw a mirrored asymmetric pattern between the Grass NLCD and Non Crop Grass classes where the aggregated simulation intervals had upper bounds far below the median value, the opposite of interval patterns for the yearly simulations. These results emphasize the importance of considering sensitivity when interpreting long-term land cover trends, particularly in regions with classification inconsistencies. Further research could refine these approaches by exploring alternative transition matrix strategies or incorporating additional sources of uncertainty into the sensitivity analysis framework.

5.2 Discussion and Future Work

This research highlights the importance of quantifying the sensitivity of landscape metrics to random perturbations in LULC data, providing policy makers and private industry professionals with a framework to assess metric reliability. Our simulation approach, which integrates CDL data at the patch level, has demonstrated its potential to capture metric sensitivity while maintaining scalability across large geographic areas. By leveraging both `terra` functions and C++ optimization, we ensured that the `cdlsim` package remains computationally efficient and adaptable to diverse land cover and land use datasets. The flexibility to use real confusion matrices from CDL data or custom transition matrices further expands the applications of our approach.

One limitation of our simulation process is the potential to under represent or omit some class types, particularly low frequency ones, in simulated landscapes leading to a decrease in overall diversity. Though crop type distributions in our landscapes are not strictly Poisson-distributed, resampling from a Poisson-like distribution can similarly result in a loss of species but not the creation of new ones. This issue has been well researched in the field of species distribution modeling (SDM), where resampling approaches can produce sensitivity intervals that are skewed below the median (Araújo and Guisan, 2006; Dubos et al., 2022). A similar bias may also occur in LULC simulations such as ours, particularly in diversity indices or metrics that depend on low frequency classes disproportionately. To address this, bias correction approaches from SDM such as undersampling the majority class using spatial filtering could be incorporated into the simulation framework to improve the reliability of metric sensitivity estimates (Robinson et al., 2018).

Another limitation of the current simulation framework is that all patches, regardless of size, are given an equal chance of transitioning. As discussed in section 1.1, the CDL displays known edge effects stemming from its coarse 30-meter resolution, generally increasing probability of misclassification in smaller patches. This patches have a greater edge-to-area ratio and therefore a larger proportion of potentially mixed class pixels due to the transition of classes near patch boundaries. However, this relationship may not hold true across CDL classes, particularly when comparing agricultural versus non-agricultural categories. For example, classes like barren land or shrubland can form large patches, depending on the landscape, but still exhibit relatively low classification accuracy. To account for this variation across classes, the current method applies transition probabilities uniformly within in class. Exploring the impact of weighting transitions probabilities by patch size on simulated landscape configuration and metric sensitivity estimates would be a valuable extension of the current methodology, but falls outside the scope of the current effort.

Future research should also aim to refine the simulation framework by investigating additional data preprocessing techniques, exploring alternative methods for the construction of transition matrices, and expanding the geographical scale. In particular, it would be valuable to test a preprocessing method for the CDL data using another spatial product developed by NASS called the Crop Sequence Boundaries (CBS) (Hunt et al., 2024). This vector-based dataset provides automated field polygons, defined as areas that display the same crop rotation patterns. Simulation at the field level, as opposed to the patch level, would allow for finer detail in simulation results, as larger areas of the same class could be broken down into distinct fields and simulated independently. Additionally, incorporating other advanced preprocessing techniques—such as those discussed in Section 1.1—could improve the accuracy of the CDL data, allowing for a direct comparison of metric sensitivity before and after data refinement. Furthermore, the crop rotation history of each field, provided by the CBS polygons, could be used to inform the probabilities in the transition matrices, creating a more robust and realistic simulation method. Expanding the spatial extent of the simulation to cover the CONUS would enable large-scale studies to incorporate metric sensitivity analysis at a national level. Extending this approach to other regions or LULC datasets with varying classification accuracies would further validate `cdlsim`'s applicability and utility. Addressing these areas would help overcome current limitations in the simulation methodology, ultimately enhancing the precision of landscape metric sensitivity assessments and supporting more informed land management and conservation decisions.

Supplementary Material

The development version of the `cdlsim` package is available at the following link: https://github.com/burgerhaley97/cdlsim. The data used can be found at the USDA Cropland Data Layer (CDL) website: https://www.nass.usda.gov/Research_and_Science/Cropland/SARS1a.php.

References

Araújo MB, Guisan A (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10): 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x

Boryan C, Yang Z, Mueller R, Craig M (2011). Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, 26(5): 341–358. https://doi.org/10.1080/10106049.2011.562309

Burger H (2025). cdlsim: Simulation of classification uncertainty in usda cropland data layer (cdl) products. https://github.com/burgerhaley97/cdlsim

Chiarucci A, Enright NJ, Perry GLW, Miller BP, Lamont BB (2003). Performance of non-parametric species richness estimators in a high diversity plant community. *Diversity and Distributions*, 9(4): 283–295. https://doi.org/10.1046/j.1472-4642.2003.00027.x

Copenhaver K, Hamada Y, Mueller S, Dunn JB (2021). Examining the characteristics of the Cropland Data Layer in the context of estimating land cover change. *ISPRS International Journal of Geo-Information*, 10(5): 281. https://doi.org/10.3390/ijgi10050281

Dubos N, Préau C, Lenormand M, Papuga G, Monsarrat S, Denelle P, et al. (2022). Assessing the effect of sample bias correction in species distribution models. *Ecological Indicators*, 145: 109487. https://doi.org/10.1016/j.ecolind.2022.109487

Hesselbarth MH, Sciaini M, With KA, Wiegand K, Nowosad J (2019). landscapemetrics: An open-source R tool to calculate landscape metrics. R package version 2.2.1.

Hijmans RJ (2025). terra: Spatial Data Analysis. *R package version 1.8-42*.

Hunt K, Abernethy J, Beeson P, Bowman M, Wallander S, Williams R (2024). Crop sequence boundaries using usda national agricultural statistics service historic cropland data layers. *Statistical Journal of the IAOS*, 40: 1–10. https://doi.org/10.3233/SJI-230078

Kline K, Singh N, Dale V (2013). Cultivated hay and fallow/idle cropland confound analysis of grassland conversion in the western Corn Belt. *PNAS*, 110(31): E2863.

Laingen C (2015). Measuring cropland change: A cautionary tale. *Papers in Applied Geography*, 1(1): 65–72. https://doi.org/10.1080/23754931.2015.1009305

Lark TJ, Mueller RM, Johnson DM, Gibbs HK (2017). Measuring land-use and land-cover change using the U.S. Department of Agriculture's Cropland Data Layer: Cautions and recommendations. *International Journal of Applied Earth Observation and Geoinformation*, 62: 224–235. https://doi.org/10.1016/j.jag.2017.06.007

Lark TJ, Schelly I, Gibbs H (2021). Accuracy, bias, and improvements in mapping crops and cropland across the United States using the USDA Cropland Data Layer. *Remote Sensing*, 13(5): 968. https://doi.org/10.3390/rs13050968

Lin L, Di L, Zhang C (2022). Validation and refinement of Cropland Data Layer using a spatial-temporal decision tree algorithm. *Sci Data*, 9(63).

Lisic J, Chen L, Stachelek J (2024). Tools to download and work with USDA Cropscape Data. R package version 1.13.

McGarigal K, Marks BJ (1995). FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. United States Department of Agriculture Pacific Northwest Research Station.

NASS (2023). Frequently anticipated questions. Accessed: 2024-07-01.

Pritsolas J, Pearson R (2021). A cautionary tale: A recent paper's use of research based on the USDA Cropland Data Layer to assess the environmental impacts of claimed cropland expansion.

Reitsma KD, Clay DE, Clay SA, Dunn BH, Reese C (2016). Does the U.S. Cropland Data Layer provide an accurate benchmark for land-use change estimates? *Agronomy*, 108(1): 266–272. https://doi.org/10.2134/agronj2015.0288

Robinson OJ, Ruiz-Gutierrez V, Fink D (2018). Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions*, 24(4): 460–472. https://doi.org/10.1111/ddi.12698

Siqueira T, Pessoa L, Vieira L (2023). Evaluating land use impacts on water quality: Perspectives for watershed management. *Sustainable Water Resources Management*, 9(192). https://doi.org/10.1007/s40899-023-00968-2

Smith EP, van Belle G (1984). Nonparametric estimation of species richness. *Biometrics*, 40(1): 119–129. https://doi.org/10.2307/2530750

United States Department of Agriculture, Natural Resources Conservation Service (2025). Geospatial data gateway. Accessed: 2024-05-19.

US Environmental Protection Agency (2025). Ecoregions of North America. Accessed: 2025-02-01.

USDA (2024). Cropland Data Layer: USDA NASS, Marketing and Information Services Office. Washington, D.C.

Wright CK, Wimberly MC (2013a). Recent land use change in the western Corn Belt threatens grasslands and wetlands. *PNAS*, 110(10): 4134–4139. https://doi.org/10.1073/pnas.1215404110

Wright CK, Wimberly MC (2013b). Reply to Kline et al.: Cropland Data Layer provides a valid assessment of recent grassland conversion in the western Corn Belt. *PNAS*, 110(31): E2853.

Zhang C, Marfatia P, Farhan H, Di L, Lin L, Zhao H, et al. (2023). Enhancing USDA NASS Cropland Data Layer with Segment Anything Model. In: *2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*. (Agro-Geoinformatics 2023 Organizing Committee, eds.), 1–5.

Zhang C, Yang Z, Di L, Lin L, Hao P, Guo L (2021). Applying machine learning to Cropland Data Layer for agro-geoinformation discovery. In: *IEEE International Geoscience and Remote Sensing Symposium IGARSS*. (D Miralles, C Persello, K Beenen, eds.), Brussels, Belgium. 2021, 9554628.