

Determining Latitudinal Shifts of Grass Plant Functional Types within the Great Plains

Paul Lin

Abstract

Grass-based biomes are one of the most important ecosystems and provide numerous crucial functions and services, yet remain as one of the world's most vulnerable ecosystems today. In particular, the distribution of Plant Functional Types (PFT) between C₃ grasses and C₄ grasses are expected to shift dramatically due to grass' sensitivity to changes in environmental factors. This study analyzed the potential shifts in the spatiotemporal distribution of grass PFTs and its correlations with temperature and precipitation changes within the North American Great Plains. A set of twelve annual phenological metrics were extracted from the Normalized Difference Vegetation Index (NDVI) profiles along five training clusters and three testing latitudinal transects between 2010 and 2020. Measurements of the phenological metric values from the training sites were used as features to construct six base classifiers for predicting site PFT classification (C₃-majority or C₄-majority) from phenological metric values. The three ensemble method classifiers (Random Forest, Adaboost, and Gradient Boost) demonstrated the highest classification accuracy amongst the base classifiers. Amongst the three latitudinal transects analyzed, only the East Transect demonstrated a significant shift in PFT distribution. In particular, correlations between year and the mean latitude of testing sites classified as C₃-majority yielded a slope value of 0.1556 and R² value of 0.6410, indicating a northward migration of 0.1556 degrees latitude per year by the centroid of C₃-majority sites. However, this distribution shift appeared to be independent of temperature and precipitation changes, as nearly all correlations between changes in PFT distribution and changes in temperature and precipitation values were found to be insignificant. These preliminary findings demonstrated that the degree of northward migration of grass PFTs along latitudinal transects varied across geographical regions. Further analysis may be conducted to reveal the strengths of environmental factors for driving PFT distribution shifts and to provide projections of PFT migrations under a changing climate.

Introduction

Grasses and grass-based biomes play crucial roles in ecological functions and human development. Grass-based biomes (savannas, grasslands, and shrublands) span an area of over 4000 million ha and yield an annual total net primary productivity of 13.7 Pg C, approximately one-third of the Earth's total land surface and total terrestrial productivity (Bengtsson et al., 2019; Del Grosso et al., 2008). Grass-based ecosystems serve important ecological functions as carbon storage reservoirs, nitrogen fixation sources, wildlife habitat, and water flow and quality regulators (Bengtsson et al., 2019; Carlier, Rotar, Vlahova, & Vidican, 2009). Human have long capitalized on the multitude of services and resources provided by grass-based ecosystems. Human civilizations use grasses as biofuel feedstock, livestock grazing sites, and agricultural products, amongst other uses (Blair, Nippert, Briggs, & Monson, 2014).

Despite the importance of grass-based biomes, these ecosystems remain one of the most vulnerable ecosystems today. Accelerated changes in climate and land use patterns have drastically reduced global grasslands cover in recent years. (Sterling & Ducharne, 2008) found natural grasslands areas to have shrunk by 92%, the most significant reduction of any natural biomes. Within grasslands, changes in regional climate patterns have also induced significant shifts in the PFT distribution of C₃ and C₄ grasses. Such internal shifts within the grasslands also have major global and regional implications. First, domesticated C₃ and C₄ grasses are critical global agricultural productivity, with wheat (C₃), rice (C₃), and maize (C₄) accounting for 20%, 19%, and 5% of the world's global caloric intake respectively (Shiferaw et al., 2013). Second, C₃ grasses generally live longer and carry more biomass than C₄ grasses, and so serve as better carbon reservoirs than C₄ grasses (Still, Berry, Collatz, & DeFries, 2003). As such, even minor expansions or contractions of C₃ and C₄ grass territorial boundaries could have substantial impacts on global food security and the global carbon cycle.

Background: PFT Distribution Response to Environmental Variables

During geological periods with high temperatures or low atmospheric CO₂ concentration conditions, elevated photorespiration rates reduce the carboxylation efficiency of C₃ plants (Blair et al., 2014). In more than 60 independent plant lineages, physiological adaptations of the C₄ photosynthetic pathway have evolved to increase local mesophyll CO₂ concentrations to counteract high photorespiration rates (Sage, Christin, & Edwards, 2011). The primary physiological difference between C₃ plants and C₄ plants is the introduction of the phosphor-enolpyruvate carboxylase (PEP-C) enzyme within C₄ plants (Blair et al., 2014). In C₄ plants, CO₂ is converted into oxaloacetate by the PEP-C enzyme then continuously pumped from the plant's mesophyll cells into its bundle-sheath cells (Gowik & Westhoff, 2011). This process allows C₄ plants to maintain high internal CO₂ concentration and diffusion rates, thereby reducing photorespiration rates (Gowik & Westhoff, 2011). Moreover, high internal CO₂ concentration also allow C₄ plants to reorient stomatal opening times around periods of cooler temperatures and reduce water loss from transpiration (Blair et al., 2014). However, such pumping mechanisms are also energetically expensive and hinder the photosynthetic efficiency of C₄ plants under low photorespiration rates. Consequentially, modern grass PFTs are partitioned along temperature and precipitation gradient with C₃ grasses occupying cool and moist, high-latitude landscapes where photorespiration rates are low and C₄ grasses occupying warm and dry, low-latitude landscapes where photorespiration rates are high (Goodin & Henebry, 1997; Pau, Edwards, & Still, 2013).

Current patterns in climate change appear to stimulate paradoxical effects on PFT distribution. All other environmental conditions being held constant, elevated temperatures are predicted to favor C₄ expansion due to C₄ plants' ability to minimize photorespiration and water loss (Collatz, Berry, & Clark, 1998; Sage & Kubien, 2007; Zhang et al., 2014). Conversely, elevated atmospheric CO₂ concentration are predicted to favor C₃ expansions due to C₃ plants' higher photosynthetic efficiency under optimal growing conditions. Precipitation and water availability are also expected to influence PFT distribution. C₄ plants are predicted to become more competitive under low water availability conditions while C₃ plants are predicted to become more competitive under high water availability conditions (Still et al., 2003; Witwicki, Munson, & Thoma, 2016).

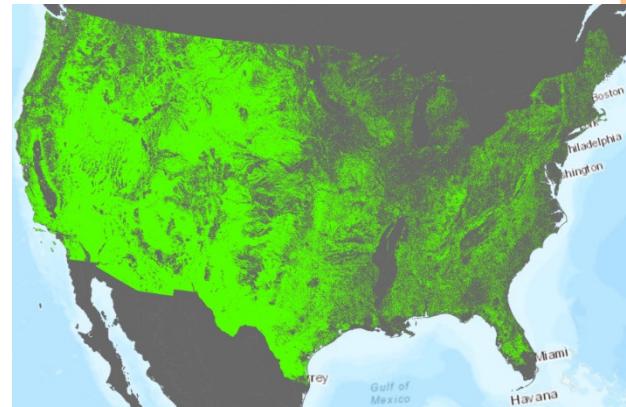
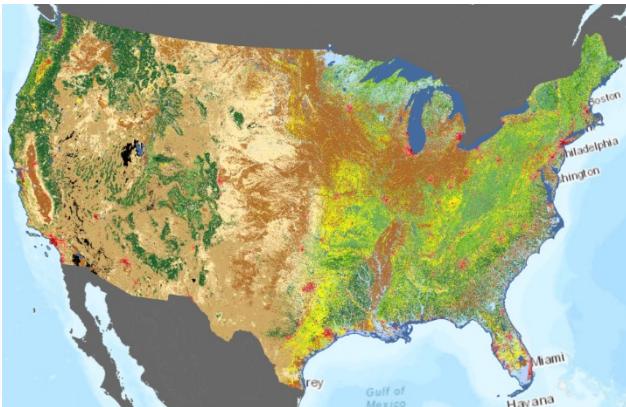
Background: PFT Distribution Analysis Methods

Past methods for analyzing PFT distribution have typically utilized field-collected measurements as classification metrics to classify sites as C₃-majority or C₄-majority (Collatz et al., 1998; Griffith et al., 2017). A standard PFT classification method capitalized on the difference in carbon isotope composition between C₃ and C₄ plants (Still et al., 2003). Under this method, site PFT class could be determined by the site's ¹³C and ¹²C concentration values (Still et al., 2003). However, field-collected measurements are often constrained by high costs, infrequent and limited spatial coverage, and coarse temporal resolution. As such, PFT classification methods classification metrics derived from remotely-sensed measurements have become increasingly popular due to the low cost, regular and high spatial coverage, and fine temporal resolution produced by remotely-sensed measurements.

Another popular method for site PFT classification is based on the asynchronous seasonal profiles between C₃ and C₄ plants. The asynchronous seasonal profiles between C₃ and C₄ plants have been well-documented and reveal that C₃ plants exhibit greater photosynthetic activity in the moist and cool spring and late fall while C₄ plant exhibit greater photosynthetic activity in the dry and hot summer (Goodin & Henebry, 1997; Tieszen, Reed, Bliss, Wylie, & DeJong, 1997). Consequentially, C₃-majority and C₄-majority sites yield different ranges of values for temporal-based phenological metrics such as date of onset of greenness (season start), date of end of greenness (season end), date of maximum greenness, and duration of greenness (season length) that can be used for site PFT classification (Reed et al., 1994; Tieszen et al., 1997). For example, (Foody & Dash, 2007) found that the time of maximum greenness strongly correlated with site PFT composition and could be used as metric for site PFT classification. Likewise, studies conducted by (Liu, X., Bo, Zhang, & He, 2015) and (Wang et al., 2013) utilized data from the Moderate Resolution Imaging and Spectroradiometer (MODIS) to generate NDVI time-series profiles, extracted phenological metrics from the NDVI time-series, and built decision boundary models based on the distinct values of seasonal phenological metrics between C₃ and C₄ plants. In both instances, phenology-based models have proved to be successful classifiers, with the Support Vector Machine (SVM) model (Liu et al., 2015) yielding a classification accuracy of 85.75% and the decision tree model (Wang et al., 2013) yielding a classification accuracy of 80%.

Dataset: NLCD

- Open Water
- Perennial Snow/Ice
- Developed, Open Space
- Developed, Low Intensity
- Developed, Medium Intensity
- Developed, High Intensity
- Barren Land
- Deciduous Forest
- Evergreen Forest
- Mixed Forest
- Shrub/Scrub
- Herbaceous
- Hay/Pasture
- Cultivated Crops
- Woody Wetlands
- Emergent Herbaceous Wetlands



Figures 1a – 1b. NLCD 2016 classification of the contiguous United States (1a). Herbaceous land cover mask of the contiguous United States (1b).

National Land Cover Database (NLCD): The National Land Cover Database (NLCD) is a 30-m resolution nationwide land cover dataset based on the Landsat satellite. Amongst the land cover classes delineated by NLCD 2016 (Figure 1a), four land classes are considered herbaceous: Shrub/Scrub, Grassland/Herbaceous, Pasture/Hay, and Barren Land. These four land cover classes were extracted to generate an herbaceous land cover mask (Figure 1b) to ensure that training and testing sites only occupied herbaceous lands.

Dataset: NEON, MODIS, Copernicus

- National Ecological Network (NEON)

NEON is a nationwide research program dedicated to “discovering the understanding the impacts of climate change, land-use change, and invasive species on ecology”. NEON product DP1.10058.001 (Plant Species Presence and Percent Cover) documents the existing plant species and their respective percent cover at multiple sample plots located within NEON Terrestrial Observations System (TOS) sites. NEON product DP1.10058.001 data were used in this study to prescribe PFT classification to training plots as C₃-majority or C₄-majority plots.

- Moderation Resolution Imaging Spectroradiometer (MODIS)

MODIS surface reflectance data was collected from the Terra Surface Reflectance 8-Day Global 250m (MOD09Q1.006) dataset hosted by the Land Process Distributed Active Archive Center (Vermote, 2015). Between 2010 to 2020, MODIS surface reflectance data was collected at each training and testing plot on an 8-day interval. For training plots spanning multiple pixels, the average reflectance value of the pixels within the training plot’s boundaries was used instead. NDVI can be calculated from MODIS surface reflectance data as $(\text{Band}_1 - \text{Band}_2) / (\text{Band}_1 + \text{Band}_2)$, with Band₁ representing the near-infrared band and Band₂ representing the red band. Using the 8-day surface reflectance value, an 8-day NDVI value was calculated at each training and testing plot between 2010 to 2020.

- Copernicus

Temperature and precipitation data was collected from Copernicus’ ERA5 daily aggregate dataset (Copernicus Climate, 2019). Between 2010 to 2020, total precipitation and mean air temperature at 2m height were collected as precipitation and temperature data at each testing plot on a daily interval.

Study Area: Great Plains Ecoregion

The North American Great Plains (Great Plains) occupies an area of approximately 3.5 million km² and is the world's second largest contiguous grasslands ecoregion (Commission for Environmental Cooperation, (CEC), 1997; Dixon, Faber-Langendoen, Josse, Morrison, & Loucks, 2014). Stretching approximately 600 km horizontally and 1,500 km vertically from northern Mexico to Southern Canada, the Great Plains traverses over the longest latitudinal gradient of any North American ecoregion (Commission for Environmental Cooperation, (CEC), 1997). Due to extensive variations in climatic and topographic properties, the Great Plains hosts a variety of grass species from multiple PFTs, largely delineated along latitudinal temperature and longitudinal precipitation gradients. From a latitudinal perspective, temperature gradients decrease from north to South and profoundly affect PFT distributions along a latitudinal gradient. When traversing from north to south, C₃ distribution decrease while C₄ distribution increase, with the cool northern regions exhibiting the highest C₃ & lowest C₄ concentrations and the warm southern regions exhibiting the highest C₄ distribution & lowest C₃ concentrations (Tieszen et al., 1997). From a longitudinal perspective, precipitation gradients decrease from east to west and result in tallgrass prairies occupying the wet eastern regions, mixed-grass prairies occupying the central regions, and shortgrass prairies occupying the dry western regions (Commission for Environmental Cooperation, (CEC), 1997; Wang et al., 2013).

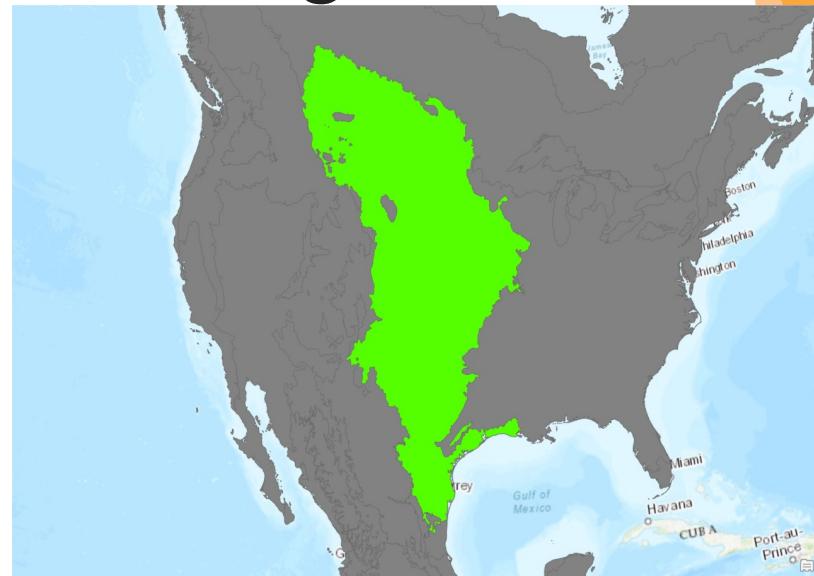
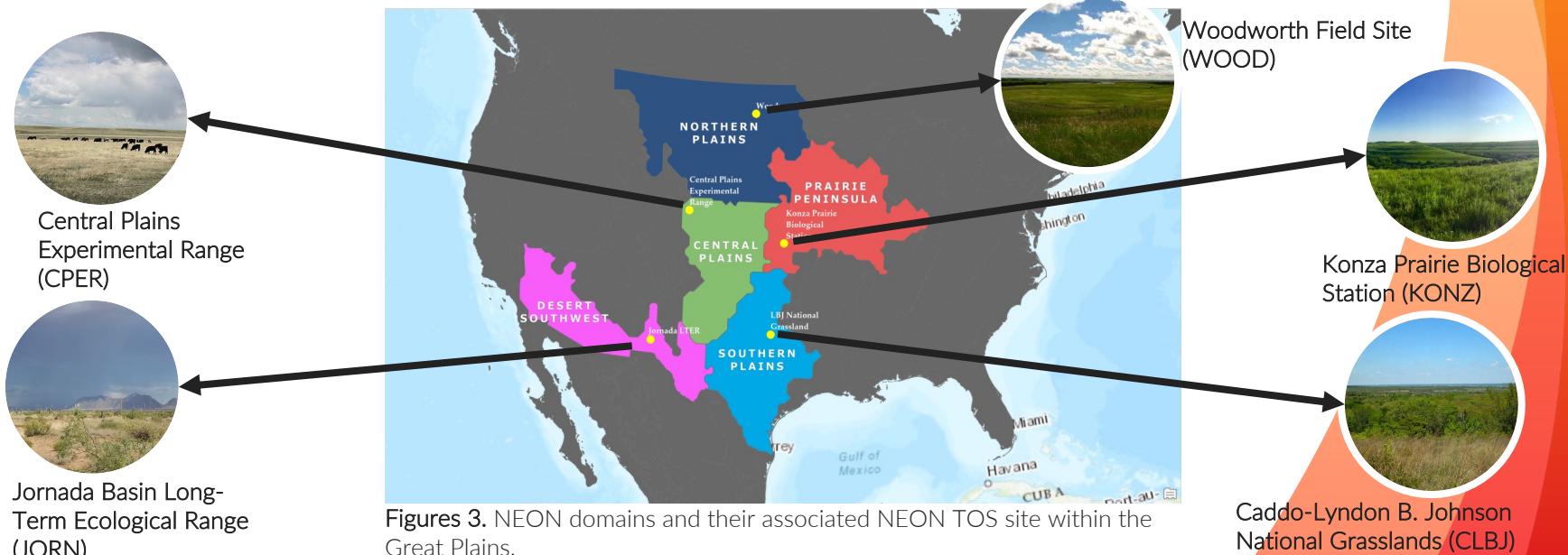


Figure 2. The Great Plains ecoregion (as determined by the CEC) is highlighted in green.

Training Sites

Five NEON domains are associated with the Great Plains ecoregion: Prairie Peninsula (Domain 6), Northern Plains (Domain 9), Central Plains (Domain 10), Southern Plains (Domain 11), and Desert Southwest (Domain 14). For each of the five domains, a NEON TOS site was chosen to represent the associated domain. Environmental and biological observations were collected from sampling plots within these TOS sites to be used as training data. Altogether, these five TOS sites (**Figure 3**) span across a wide climatic and topographic range and broadly represent the habitats of various C₃ and C₄ grass species within the Great Plains. A summary of each TOS site's land cover distribution and seasonal phenological metric values are described in **Table 1** and **Table 2**.



Training Site Characteristics

	GH	S	DF	EHW	Other	Total
CLBJ	37	0	56	0	2	95
C PER	94	0	0	0	2	96
JORN	1	91	0	0	2	94
KONZ	75	5	12	0	2	94
WOOD	78	0	0	13	2	93
Total	285	96	68	13	10	379

Table 1. NLCD Distribution of the number of plots at each TOS site.

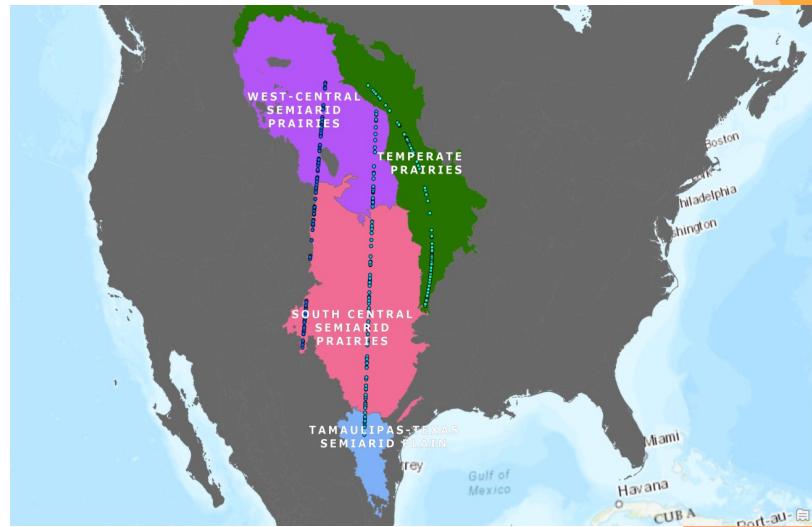
Column Keys: GH (Grassland / Herbaceous), S (Shrub / Scrub), DF (Deciduous Forest), EHW (Emergent Herbaceous Wetlands). Only the Grassland / Herbaceous and Shrub / Scrub plots are considered herbaceous and used within the study.

	Day of Onset of Greenness	Day of Peak Greenness	Day of End of Greenness	Season Length
CLBJ (Peak 1)	Day 60 (03/01)	Day 134 (05/15)	Day 295 (10/22)	235 days
CLBJ (Peak 2)	Day 215 (08/03)	Day 230 (08/18)	Day 320 (11/20)	115 days
C PER	Day 90 (04/01)	Day 165 (06/15)	Day 350 (12/7)	260 days
JORN	Day 80 (03/21)	Day 185 (07/04)	Day 340 (12/06)	260 days
KONZ	Day 90 (04/01)	Day 160 (06/10)	Day 300 (10/28)	210 days
WOOD	Day 120 (05/01)	Day 180 (06/30)	Day 290 (10/18)	170 days

Table 2. Seasonal phenological metric values at each TOS site. The CLBJ site experiences two phenological peaks, one in the early spring and one in the late summer.

Testing Sites

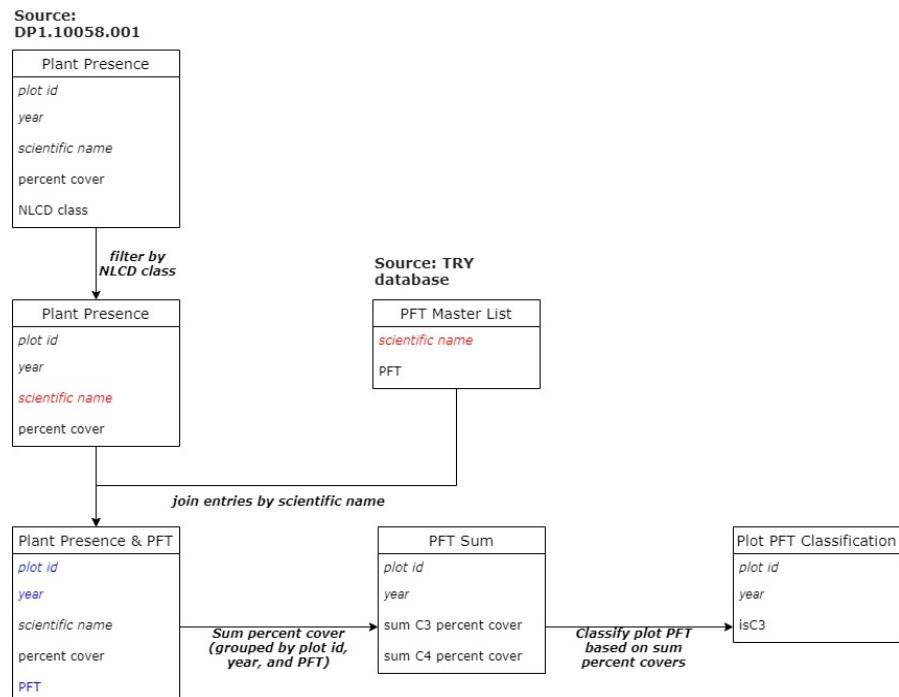
Three latitudinal transects (East, Central, West) were drawn over the Great Plains ecoregion to serve as baselines for generating herbaceous testing sites. The East Transect occupies the Temperate Prairies ecoregion, a region characterized by high precipitation and tallgrass presence. The Central Transect and West Transect, drawn along the 100°W longitude and 104°W longitude respectively, occupy the Semiarid Prairies ecoregions, a region characterized by low precipitation and shortgrass presence. 400 preliminary sites were randomly generated along each transect and an NLCD herbaceous land cover mask (**Figure 1b**) was applied to filter out sites occupying non-herbaceous lands. From the pool of remaining preliminary sites, 100 herbaceous sites were randomly chosen along each transect as testing sites for a total of 300 testing sites (**Figure 4**).



Figures 4. Testing sites along the East (navy blue), Central (light blue), and West (teal) vertical transects. Each transect consists of 100 testing sites.

Methods: Training Site (Plot PFT Classification)

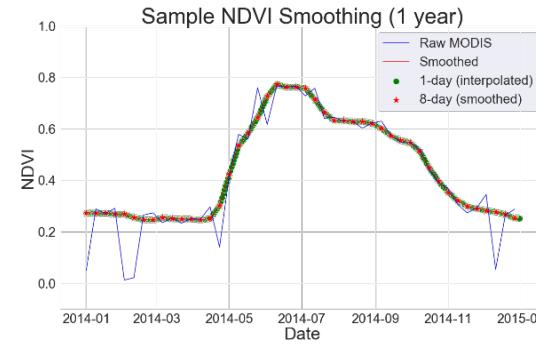
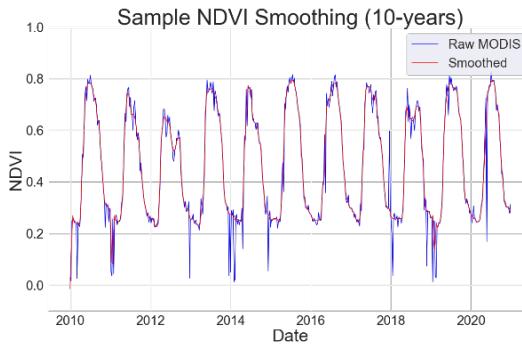
Herbaceous plots located within the five representative NEON TOS sites were selected as training plots used to build the learned model. Each DP1.10058.001 dataset was comprised of entries composed of plot id, year, NLCD class, species' scientific name, and species percent cover. Only entries representing herbaceous land class were considered. A PFT master list matching species scientific name to their corresponding PFT class was compiled from the TRY Plant Trait Database and other resources (Kattge et al., 2011). The PFT of individual entries within the DP1.10058.001 dataset was then identified by joining the DP1.10058.001 dataset with the PFT master list based on the species' scientific name key. Finally, the sum of C₃ percent cover and sum of C₄ percent cover was calculated by aggregating and summing the percent cover values of entries sharing the same plot id, year, and PFT. For a given plot, if the sum of C₃ percent cover was greater than the sum of C₄ percent cover, the plot was classified as C₃-majority. If the sum of C₄ percent cover was greater than the sum of C₃ percent cover, the plot was classified as C₄-majority. A flow diagram summarizing the process of converting DP1.10058.001 data into plot PFT classifications is specified in Figure 5.



Figures 5. Flow diagram of the conversion from raw DP1.10058.001 data (*plot id*, *year*, *scientific name*, NLCD class, and percent cover) into plot PFT classification (*plot id*, *year*, *isC3*). Table indexes are indicated by italics, join attributes by blue text, and grouping attributes by red text.

Methods: Training Site (NDVI Time-Series)

The MOD09Q1.006 product data was used to construct a continuous NDVI time-series between 2010 and 2020 at each training plot. The maximum-value-composite (MVC) technique (Huete et al., 2002; Wang et al., 2013) applied to the MOD09Q1.006 product was found to be insufficiently comprehensive for removing the effects of cloud and aerosol contaminations on NDVI values (Wang et al., 2013). As such, two additional filters were applied to smooth the NDVI time-series: 1) a five-point median filter (Wang, Fritschi, Stacey, & Yang, 2011) to reduce major cloud contamination effects; 2) a Savitzky-Golay filter (Savitzky & Golay, 1964) to reduce minor atmospheric effects. The resulting time-series exhibited relative smoothness, so no additional smoothing techniques were applied. Finally, a linear interpolation technique was applied on the smoothed time-series to increase the time-series' temporal resolution from one value per eight days to one value per day. A sample raw, smoothed, and interpolated NDVI time-series over a 10-year and 1-year time-frame is shown in Figures 6a-6b.

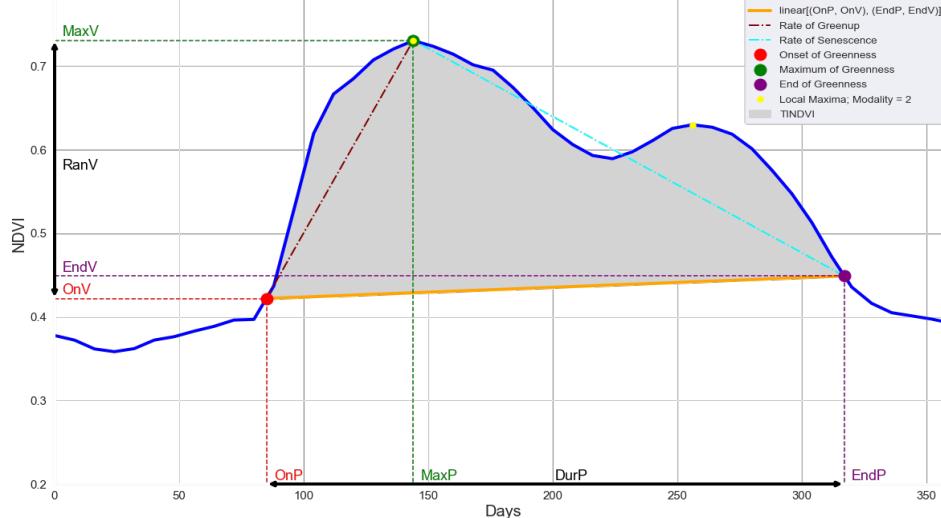


Figures 6a-6b. Sample raw and smoothed NDVI time-series over 10-year (6a) and 1-year (6b) time-frame. Raw NDVI time-series are indicated by blue lines and smoothed NDVI time-series by red lines. Within the 1-year time-frame diagram, the 8-day data (smoothed) are indicated by red stars and 1-day data points (interpolated) by green dots.

Methods: Training Site (Phenological Metrics)

A set of twelve phenological metrics were identified by (Reed et al., 1994) to be extractable from NDVI time-series and significant for measuring ecosystem performances. These metrics are divided into three categories: 1) *temporal*: event timing values (Time of Onset of Greenness – OnP, Time of End of Greenness – EndP, Duration of Greenness – DurP, and Time of Maximum NDVI – MaxP); 2) *NDVI-based*: event NDVI values (Value of Onset of Greenness – OnV, Value of End of Greenness – EndV, Value of Maximum NDVI – MaxV, and Range of NDVI – RanV); 3) *derived*: time-series derivation values (Time-Integrated NDVI – TINDVI, Rate of Greenup – RtUp, Rate of Senescence – RtDn, Modality – Mod). The twelve metrics, their phenological meanings, and their derivations are fully described in **Table 3**. Annual measurements of the set of 12 phenological metrics were extracted from each training plot's annual NDVI time-series.

An extraction for the set of phenological metrics values from a sample NDVI time-series profile is shown in **Figure 7**.



Figures 7. Sample extraction of phenological metrics values from a seasonal NDVI time-series profile. Inspired by (Reed et al., 1994).

Methods: Training Site (Phenological Metrics)

		Abbreviation	Phenological Meaning	Derivation
Temporal NDVI Metrics	Time of Onset of Greenness	OnP	Beginning of measurable photosynthesis (start of season)	First day meeting both requirements: a) $NDVI \geq early_min + threshold \times amplitude$ b) Monotonically increasing for next 15 days
	Time of End of Greenness	EndP	Cessation of measurable photosynthesis (end of season)	First day after 07/01 meeting both requirements: a) $NDVI \leq late_min + threshold \times amplitude$ b) Monotonically decreasing for past 15 days
	Duration of Greenness	DurP	Duration of photosynthetic activity (season length)	$EndP - OnP$
	Time of Maximum NDVI	MaxP	Time of maximum measurable photosynthesis (peak of season)	Day of maximum NDVI value
NDVI-Value Metrics	Value of Onset of Greenness	OnV	Level of photosynthetic activity at beginning of growing season	NDVI value on time of onset of greenness
	Value of End of Greenness	EndV	Level of photosynthetic activity at end of growing season	NDVI value of time of end of greenness
	Value of Maximum NDVI	MaxV	Maximum measurable level of photosynthetic activity	Maximum NDVI value
	Range of NDVI	RanV	Range of measurable photosynthetic activity	$MaxV - \min(OnV, EndV)$
Derived Metrics	Time-Integrated NDVI	TINDVI	Net primary production	$\int_{OnP}^{EndP} Smooth\ NDVI - \int_{OnP}^{EndP} linear[(OnP, OnV), (EndP, EndV)]$
	Rate of Greenup	RtUp	Acceleration of photosynthesis	$\frac{MaxV}{OnV}$
	Rate of Senescence	RtDn	Deceleration of photosynthesis	$\frac{MaxV}{EndV}$
	Modality	Mod	Periodicity of photosynthetic activity	Number of local maxima in 15-day windows

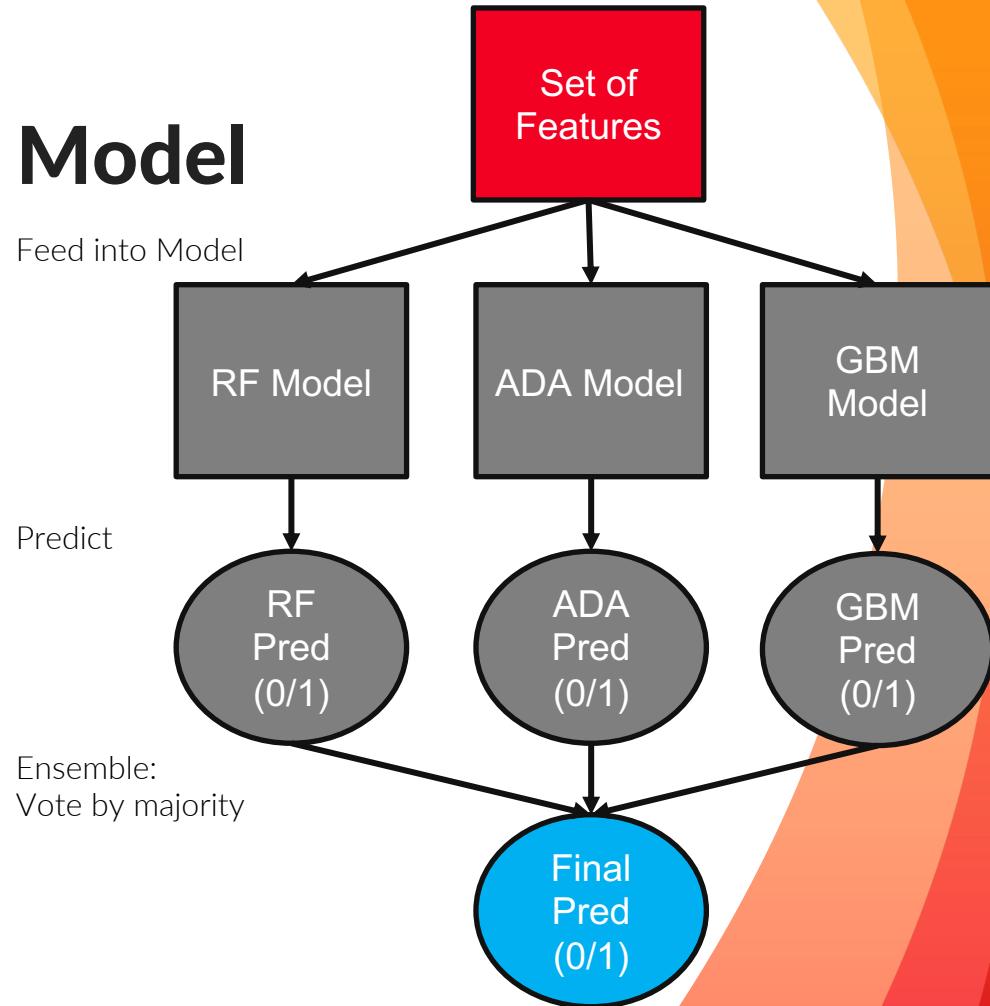
Table 3. The phenological meanings and derivation methods of the phenological metrics extracted from NDVI time-series-extracted. Phenological metric names, abbreviations, and phenological meaning reproduced from (Reed et al., 1994).

Methods: Learned Model

- **Model Input:** Training data for building the learned model was created by joining the plot PFT classification data and the plot phenological features data based on the *plot id* and year keys. The resulting entries were comprised of three components: 1) an identifier (*plot id* and *year*); 2) a PFT label (*isC3*); 3) a set of features (the 12 phenological metrics).
- **Model Selection:** A series of steps and comparisons were undertaken to determine the learned model's optimal structure. First, training data was split into 67% **input data** and 33% **validation data**. Since the input dataset contained more C₃-majority plots than C₄-majority plots, three sets of sampling techniques were performed over the **input data**: 1) no sampling techniques; 2) oversampling; 3) under-sampling. Oversampling and undersampling techniques were applied to the **input data** to generate **oversampling input data** and **undersampling input data**. Six model types were identified as potential base classifiers for the learned model: 1) Logistic Regression (LR); 2) Support Vector Machines (SVM); 3) K-Nearest Neighbors (k-NN); 4) Random Forest (RF); 5) Adaboost (ADA); 6) Gradient Boost (GBM). Each of the six base classifiers were then built three times (once from **input data**, **oversampling input data**, and **undersampling input data** each) then evaluated with the **validation data**. Models built on the **oversampling input data** and the RF, ADA, and GBM base classifiers yielded the highest classification accuracy and so were chosen as the learned model's base classifiers.
- **Model Building:** An oversampling technique was performed over the **training data** to generate **oversampling training data**. The three best-performing base classifiers (RF, ADA, GBM) were built from the **oversampling training data**. Next, hyperparameter tuning was performed on each of the base classifiers based on a 5-fold cross-validation. Each of the tuned model demonstrated an improvement in classification accuracy compared to their non-tuned counterparts. Finally, the learned model was built as an ensemble, majority-voting classifier comprised of the tuned RF base classifier, tuned ADA base classifier, and tuned GBM classifier. That is, for each set of features inputted into the learned model, three base classifiers each independently generated a PFT prediction and the final prediction of the learned model was then outputted as the majority value of the three predictions (Figure 8).

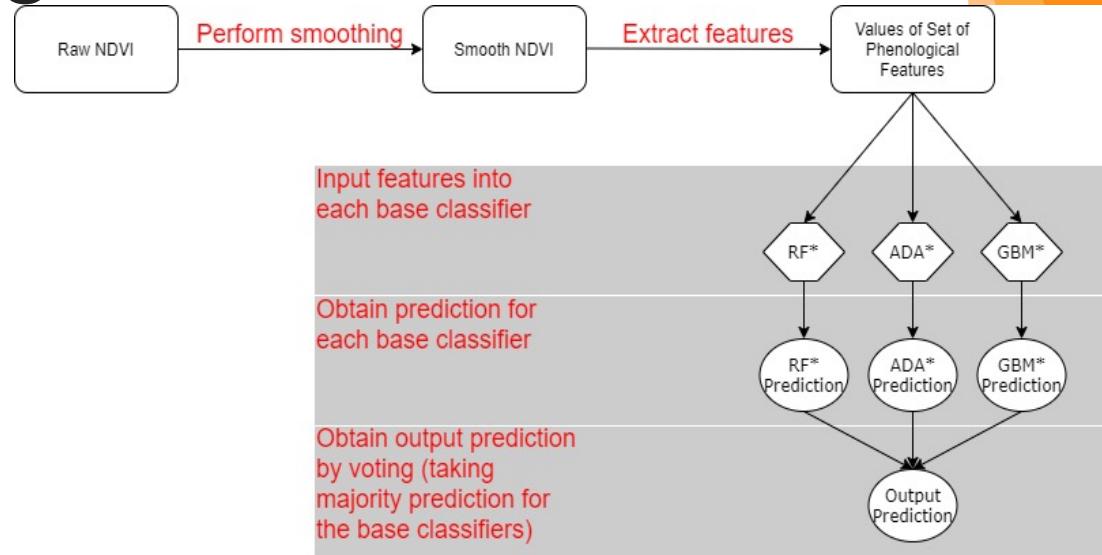
Methods: Learned Model

Figures 8. Steps undertaken to convert a set of features into a final prediction.



Methods: Testing Site

A continuous, smoothed NDVI time series between 2011 and 2020 was generated for each testing site following the steps outlined in *Training Methods (Time-Series)*. Afterwards, annual values for the set of phenological metrics mentioned in were extracted from the smoothed NDVI time-series at each testing site following the steps outlined in *Training Methods (Phenological Metrics)*. Each set of phenological features was then normalized and scaled under a normal distribution. Finally, the set of normalized features were inputted into the ensemble learned model to obtain a PFT prediction for a given site and year. A sample pipeline procedure for transforming an NDVI time-series into a PFT classification is shown in **Figure 9**. A PFT prediction was obtained for each testing site at each year between 2010 and 2020.



Figures 9. Pipeline procedure for obtaining site PFT prediction. The input is a raw NDVI time-series and the output is a PFT prediction (C_3 -majority or C_4 -majority).

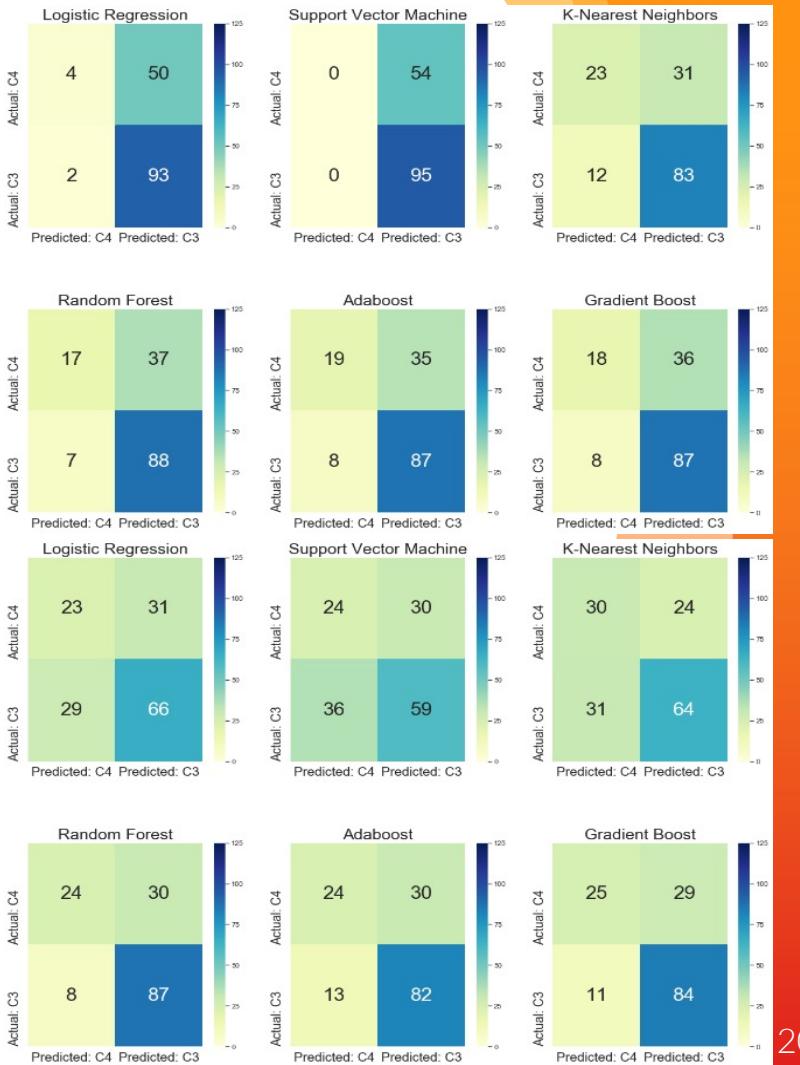
Methods: Regression Analysis

- **Precipitation:** Daily precipitation data was collected at each testing site between 2010 and 2020. The **yearly-sum precipitation** was calculated by summing all precipitation values for each testing sites at each year. The **yearly transect precipitation** was calculated by averaging all the **yearly-sum precipitation** values for the testing points along the same transect.
- **Temperature:** Daily temperature data was collected at each testing site from 2010 to 2020. Next, the **monthly-mean temperature** was calculated by aggregating averaging the daily temperature values of data entries sharing the same month and year at each testing site. Then, the **yearly-max monthly-mean temperature** was calculated by identifying the maximum **monthly-mean temperature** value for each year at each testing site. Finally, the **yearly transect temperature** was calculated by averaging all **yearly-max monthly-mean temperature** values for testing sites along the same transect.
- **Testing Site Subsets:** Testing sites were partitioned into three non-overlapping subsets: central subset (testing sites in Central Transect), east subset (testing sites in East Transect), west subset (testing sites in West Transect). Additionally, the all subset was created to include testing sites from all three transects. Each subset was then subdivided into two sub-subsets: 1) all subset testing sites classified as C₃-majority by the learned model; 2) all subset testing sites classified as C₄-majority by the learned model. In summary, testing sites were divided into four subsets (central, east, west, all) and eight sub-subsets (a C₃ and C₄ sub-subset for each of the four subsets).
- **Testing Site Regression:** Three PFT distribution attributes (count, mean latitude, and median latitude) were calculated for each sub-subset. The count attribute was calculated by counting the number of testing sites within the sub-subset. The mean and median latitude attribute were calculated by taking the average and median values of the latitude values of testing sites within the sub-subset. In essence, the latter two attributes were the centroid latitude values of the testing sites within the sub-subsets. Regression analysis tests then calculated between each sub-subset's PFT distribution attribute and time, **yearly transect temperature**, and **yearly transect precipitation**. For each sub-subset, a total of nine regressions tests were performed.

Results: Model Accuracy

Across all sampling techniques (no sampling, oversampling, undersampling), the RF, ADA, GBM models demonstrated the highest classification accuracy. Models built on the **input data** – particularly Logistic Regression and Support Vector Machine – yielded high classification accuracy by maintaining a high preference towards the C₃ class. Such a result was unsurprising, given the imbalance class distribution within the **input data**. Models built on imbalanced often learned to predict with a preference for the majority class. For example, since the **input data** was comprised of 67% C₃-majority sites and 33% C₄-majority sites, models built on the **input data** could simply always predict C₃-majority and still be correct 67% of the time. As a result, these models generated high numbers of false positive among the C₄-majority inputs and frequently classified C₄-majority sites as C₃-majority due to their oversized preference for the C₃-majority class. The oversampling technique was utilized to alleviate class imbalance and reduce the models' preferences for the majority (C₃) class. In each model instance, the number of false positives among the C₄ inputs were lower in models built on the **oversampling input data** than models built on the **input data**. For example, the number false positives among the C₄ inputs decreased from 37, 35, 36 to 30, 30, 29 amongst the RF, ADA, and GBM models, from 50 to 31 for the LR model, and from 54 to 30 for the SVM model. Moreover, the reduction of C₄ false positives was achieved while maintaining classification accuracy comparable to models built on **input data**.

Figures 10a-b. Confusion matrices of models for the **input data** (top: 10a) and the **oversampling input data** (bottom: 10b).



Results: PFT Distribution Correlations vs Year

- **East Transect:** A linear regression between time and the count of the East Transect C₃ sub-subset demonstrated a p-value of 0.001, R² value of -0.847, and slope of -3.809, suggesting the number of sites classified as C₄-majority along the East Transect increased at the expense of the number of sites classified as C₃-majority at a rate of 3.809 sites per year (**Figures 11a-11c**). Mean latitude yielded a p-value of 0.002, R² of 0.827, and slope of 0.237, while median latitude yielded a p-value of 0.003, R² of 0.799, and slope of 0.541 (**Tables 4b-4c**). These findings indicated that the centroid of testing sites classified as C₃ along the East Transect shifted northward at a rate of 0.237 – 0.541 degrees latitude per year. Altogether, the simultaneous patterns at the East Transect of 1) a decrease in count of sites classified as C₃-majority and 2) an increase in the centroid latitudinal value of sites classified as C₃-majority suggested a northward shift of the southern boundary of the C₃ territory at the East Transect. That is, over the years the most southernmost sites of the sites classified as C₃-majority along the East Transect were switching to classification to C₄-majority, thereby decreasing the C₃ count and increasing the mean and median C₃ latitude value since only the northern sites retained their C₃-majority class. Such conclusions are supported by findings in **Figure 12**, which showed the sites at the southern East Transect, which were mostly classified as C₃-majority (blue) in 2010, gradually and consistently shifting to the C₄-majority classification over the years.
- **Central & West Transect:** For both transects, regression correlations between time and the PFT attributes were all statistically insignificant and mostly yielded p-values greater than 0.50 . Within the Central Transect, there appeared to be no discernable trends in the changes in PFT distribution while the West Transect, PFT distribution appeared relatively stable.
- **All Transect:** Regression correlations between time and PFT attributes at the all subset (aggregate of all testing sites across all transects) was also mostly statistically insignificant. The one exception was the regression between time vs median latitude of the C₄ sub-subset, which yielded a p-value of 0.006, R² of 0.76, and slope of 0.127. Similar to the East Transect, this suggested that the centroid of C₄ testing sites along all transects shifted northward at a rate of 0.127 degrees latitude per year. However, such results are contradicted by the regression between time vs mean latitude, which yielded a p-value of 0.651. Nonetheless, p-values across linear regressions were consistently lower than their counterpart values at West and Central transects. For example, the p-value was 0.119 between year and C₃ testing site count, 0.519 between year and C₃ mean latitude, and 0.116 between year and C₃ median latitude. Altogether, the results indicated intense northward shift in PFT distribution within the East Transect, but a

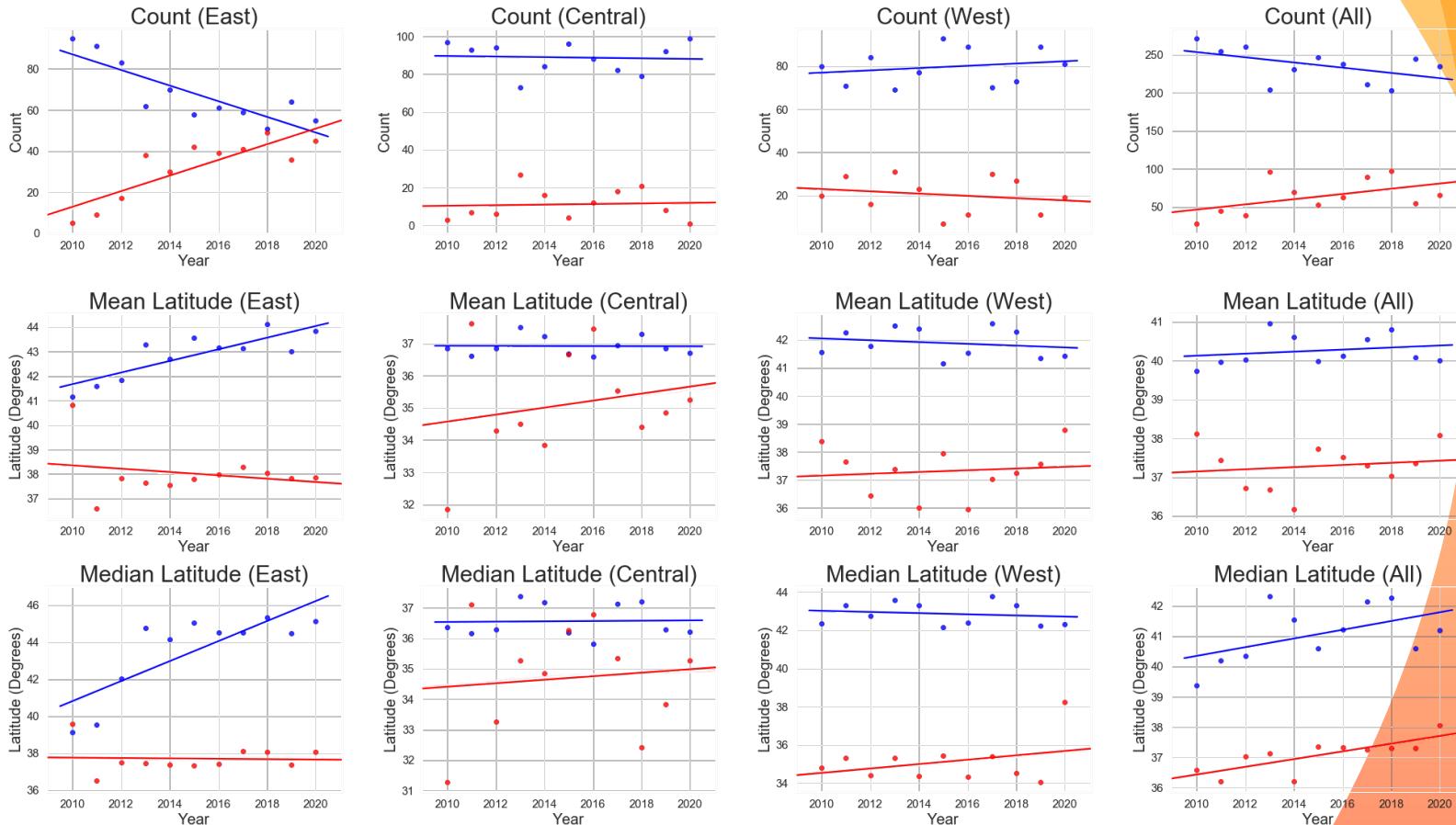
Results: PFT Distribution Correlations vs Year

	C ₃			C ₄		
	Slope	R ²	p-value	Slope	R ²	p-value
Central	-0.155	-0.062	0.857	0.155	0.062	0.857
East	-3.809	-0.847	0.001	3.809	0.847	0.001
West	0.527	0.208	0.539	-0.527	-0.208	0.539
All	-3.436	-0.497	0.119	3.436	0.497	0.119

	C ₃			C ₄		
	Slope	R ²	p-value	Slope	R ²	p-value
Central	-0.001	-0.014	0.967	0.109	0.214	0.527
East	0.237	0.827	0.002	-0.068	-0.219	0.517
West	-0.032	-0.205	0.546	0.031	0.114	0.739
All	0.026	0.218	0.519	0.028	0.154	0.651

	C ₃			C ₄		
	Slope	R ²	p-value	Slope	R ²	p-value
Central	0.005	0.031	0.927	0.057	0.104	0.760
East	0.541	0.799	0.003	-0.010	-0.043	0.900
West	-0.030	-0.168	0.622	0.115	0.330	0.322
All	0.144	0.501	0.116	0.127	0.769	0.006

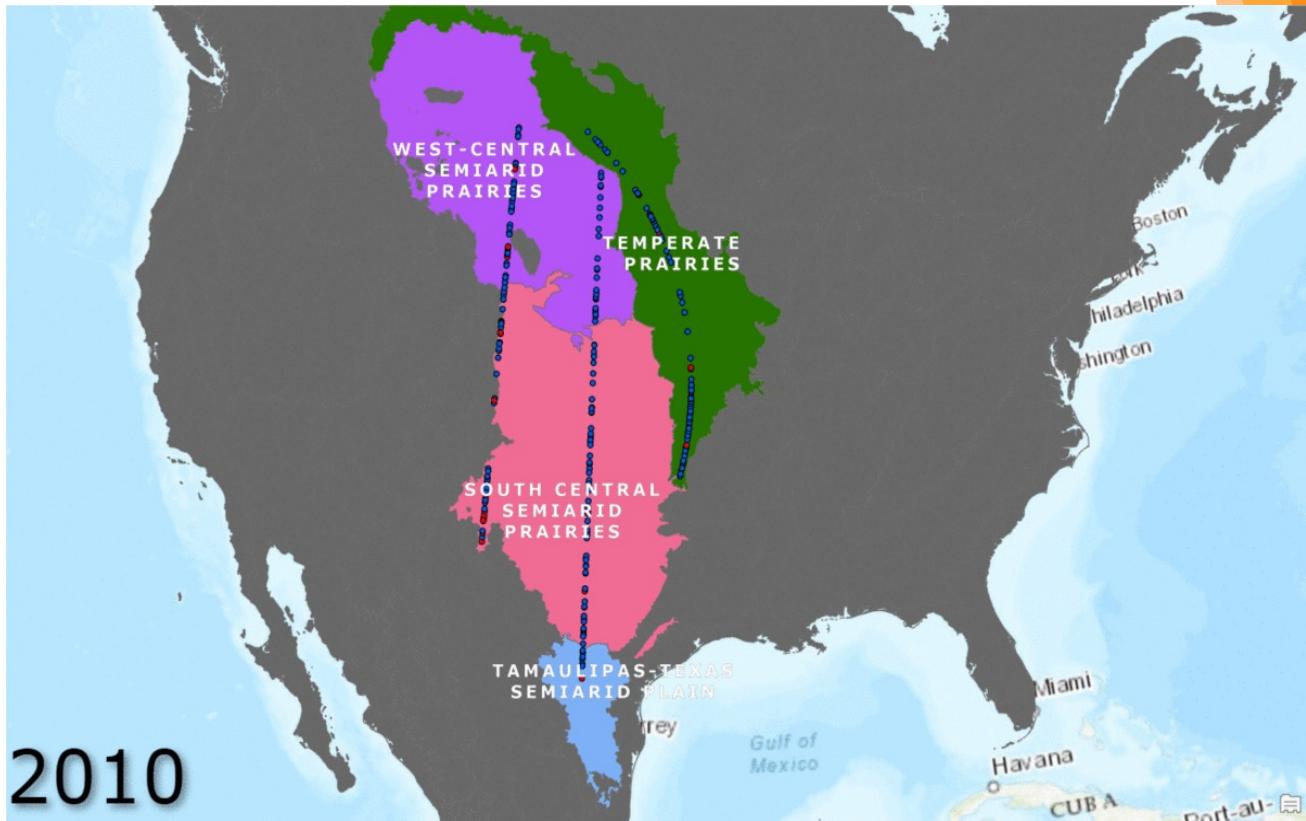
Tables 4a-4c. Regression correlation of Year vs PFT attributes. Correlations for PFT count is displayed on the top (Table 4a), PFT mean latitude in the middle (Table 4b), and PFT median latitude on the bottom (Table 4c). Significant values (p-value < 0.05) are indicated in red.



Figures 11a-11c. Scatterplots of Year vs PFT attributes. PFT count is displayed on the top row (Figure 11a), PFT mean latitude in the middle row (Figure 11b), and PFT median latitude in the bottom row (Figure 11c). Blue points and lines represent the data points and trend lines for sites classified as C₃-majority while red points and lines represent data points and trend lines for sites classified as C₄-majority.

Results: PFT Distribution Correlations vs Year

Figure 12. PFT spatiotemporal distribution at the Great Plains between 2010 and 2020. Blue dots represent sites classified as C₃-majority and red dots represent sites classified as C₄-majority. A year-by-year snapshot is shown in the Appendix.



Results: PFT Distribution Correlations vs Environmental Factors

Temperature and precipitation generally did not correlate with the PFT distribution attributes (Table 5). Across all transects and the aggregate of all transects, correlation regression between PFT attributes and temperature and precipitation variability mostly yielded p-values greater than 0.05. The one exception occurred at the West Transect between Median Latitude of C₄ and Precipitation, which yielded a p-value of 0.021, a R² value of -0.681, and a slope of -0.00684. A slope of -0.00684 indicated that for every additional 1mm of precipitation in the West Transect, the centroid of the C₄ testing sites shifts south by 0.00684 degrees latitude.

	Temperature (Count)		Precipitation (Count)		Temperature (Mean Lat.)		Precipitation (Mean Lat.)		Temperature (Median Lat.)		Precipitation (Median Lat.)	
	C ₃	C ₄	C ₃	C ₄	C ₃	C ₄	C ₃	C ₄	C ₃	C ₄	C ₃	C ₄
Central	0.05	-0.05	-0.02	0.02	-1.92	0.19	0.00	0.00	-1.00	0.05	0.00	0.00
East	0.05	-0.05	0.00	0.00	-0.74	-0.16	0.00	0.00	-0.35	-0.22	0.00	0.00
West	-0.03	0.03	-0.01	0.01	0.52	-0.30	0.00	0.00	0.54	-0.02	0.00	-0.01
All	0.02	-0.02	-0.02	0.02	-1.21	0.31	0.00	0.00	-0.42	-0.08	0.00	0.00

Tables 5. Slope values between Temperature (Celsius) and Precipitation (mm) vs PFT distribution attributes. Statistically slope values (p-value < 0.05) are indicated in red.

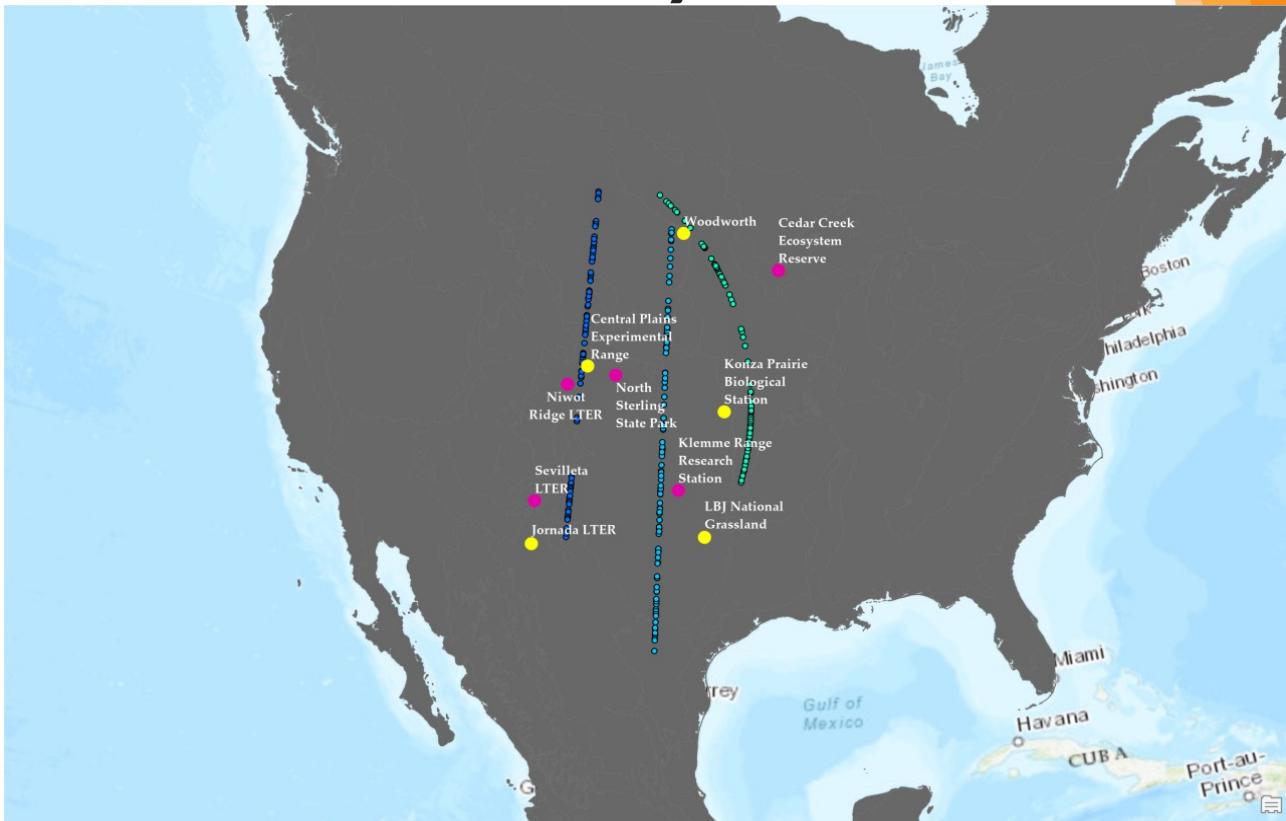
Next Steps 1: Model Diversity

The study was limited by the number of training sites used to build the learned model. In particular, the learned model frequently mistakenly classified low-latitude testing sites along the of the Central Transect. These regions are characterized by hot, dry climates and desert-like environments, conditions that are known to more inducive towards C₄ growth than C₃ growth (Tieszen et al., 1997). Nonetheless, the learned model classified testing sites within the region as C₃-majority in all years between 2010 and 2020. An overlay of the training sites over the testing transects (**Figure 13**) demonstrated that only the Central Transect lacked training sites bordering both its northern and southern regions. For example, the West Transect is bordered by the CPER site in the north-central region and JORN site in the south region and the East Transect is bordered by the WOOD site in the north region and the KONZ site in the south-central region. In contrast, the Central Transect is bordered only by the WOOD site in the north region and lacked a training site that was nearby its southern region. Without a training site at the southern region of the Central Transect, the model relied on data patterns from the southern regions at the East Transect and West Transect, but may be unable to take into account the data feature irregularities and patterns that are localized to the southern Central Transect. As such, including additional training data from different sites, particular in sparsely represented regions such as the southern Central Transect, can allow the model to better generalize to different data patterns in different regions.

Beyond limitations in training site diversity, this study was also limited by training class diversity. As previously mentioned , class imbalance between the C₃ and C₄ classes existed within the training data sample. With the training data comprised of 67% C₃ samples and 33% of C₄ samples, models built on the imbalanced training data demonstrated a clear preference for predicting the majority class: C₃. While this study used an oversampling technique to alleviate class imbalance within the training data, oversampling also induced additional issues with training data diversity. Oversampling bolstered representation of the minority class by repeatedly selecting multiple copies of the same training sample from members of the minority class. In essence, a single C₄ sample may be utilized multiple times for learning by the learned model. However, while repeatedly sampling the same data sample may reduce class imbalance, the learned model will not learn any new information. Ultimately, the inclusion of additional training data sites can increase both training site diversity and training class diversity. In particular, if additional C₄ training data are included, the learned model can depend less on the oversampling technique and reduce its preference for predicting the C₃ class. Nonetheless, collecting additional data from other training sites remains a difficult challenge, as this data are often limited in scope, quality, and availability.

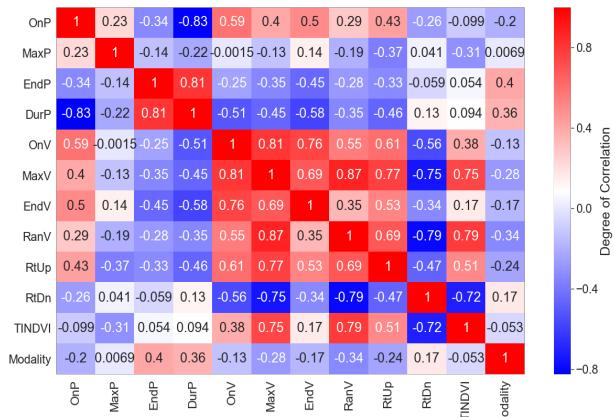
Next Steps 1: Model Diversity

Figure 13. Overlay map of training sites and vertical transects. Current training sites are indicated in yellow and suggested additional training sites by magenta.

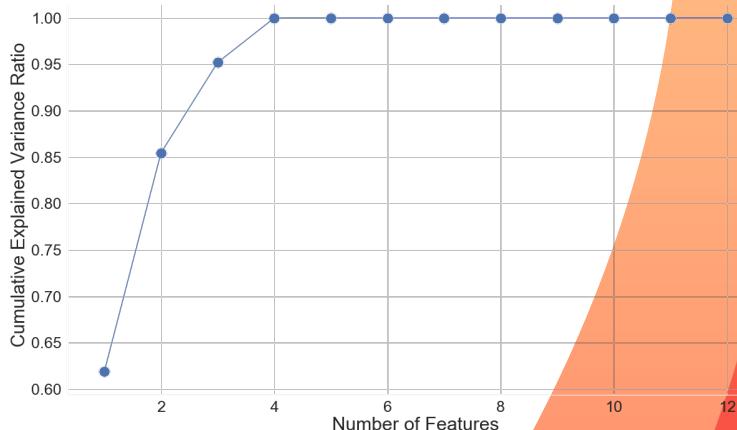


Next Steps 2: New Features

The study can reconsider the set of features utilized to build the learned model. This study utilized the set of twelve NDVI time-series extracted phenological metrics identified by Reed et al. (1994) to build the learned model. However, since many of these phenological metrics were derived from one another, this set of features suffered from a high degree of multicollinearity. A correlation matrix between the feature values of the training data is shown in **Figure 14** and demonstrated high correlations between multiple features. Multicollinearity, the phenomenon in which multiple features in a set of input features are correlated with one another, means that several features are not providing additional new information useful to the model. An analysis between the number of features versus the cumulative explained variance ratio (**Figure 15**) revealed that just four features accounted for over 99% of the explained variance in feature values between training data points. The inclusion of additional features, particularly those not derived from or correlated with existing features, could provide new information from which the learned model can base its classification decisions.



Figures 14. Correlation matrix between phenological metric features among the training data samples. Dark cells indicate high degrees of correlations between features while light cells indicate low degrees of correlations between features.

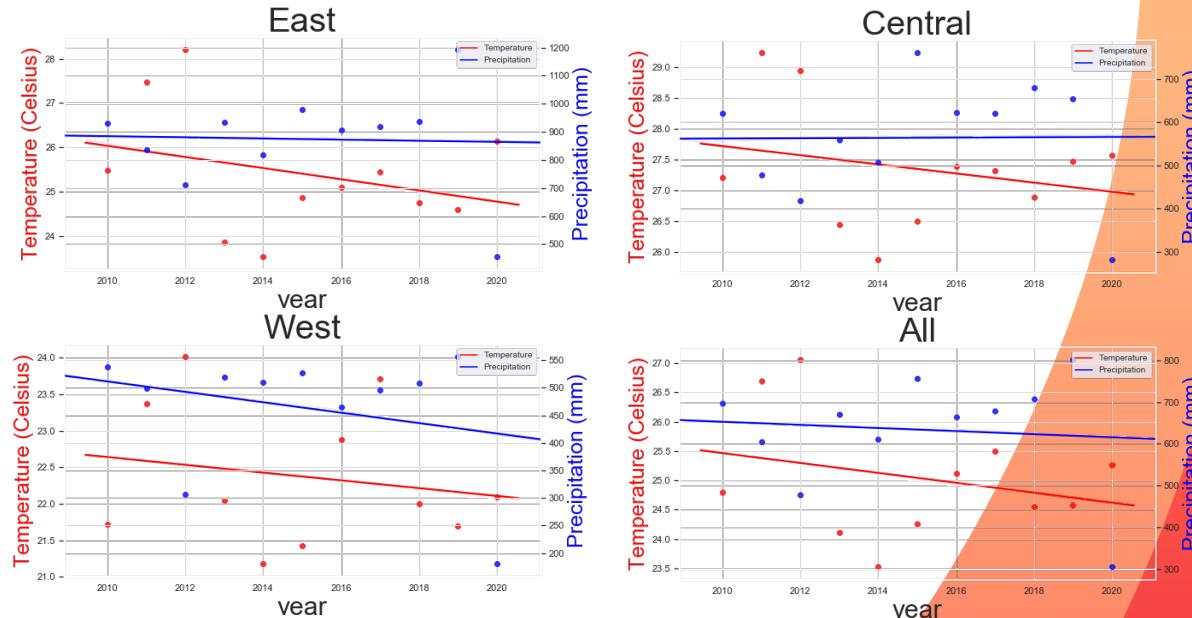


Figures 15. Line plot of the number of features vs cumulative explained variance ratio. The cumulative explained variance ratio is the ratio of explained variance in feature values among the data samples for the most

Next Steps 3: Extending Environmental Factors Analysis

Deeper analysis regarding the correlations between the PFT distribution and environmental factors can be analyzed. A possible reason that correlations between temperature and precipitation and PFT distribution appeared to be non-correlated (despite numerous past observations that indicate otherwise) could have arisen from a lack of discernable trends in environmental factors over the past decade (**Figure 156**). As such, extrapolating the study into a longer time-frame (e.g.: 20 years), may reveal stronger correlations between PFT distribution and temperature and precipitation. Additionally, further studies can also expand the number of the environmental factors considered. For example, CO₂ levels, soil factors, and fire intensity, amongst others, are all also environmental factor that impact PFT distribution.

Figure 16. Yearly transect temperature and yearly transect precipitation data at each transect between 2010 and 2020.



Next Steps 4: Model Regressor

The study can utilize a model regressor instead of a model classifier. Other studies have demonstrated the capability of learned models for transitioning from predicting site PFT classification into site PFT precent coverage (Foody & Dash, 2007; Wang et al., 2013). Model regressors (i.e.: site is $x\%$ C₃) provide more fine-grained details on PFT distribution and can better track changes in the intensity of C₃ and C₄ distribution at specific sites and along latitudinal transects. A summary of several potential additional methods (**Table 6**) that can be implemented to improve the learned model's classification accuracy.

	Action	Current State	Ideal State
Training Site Diversity	Include training sites at sparsely represented regions	5 NEON TOS sites	5 NEON TOS sites + 5 additional NEON TOS & LTER sites
Training Site Diversity	Increase the proportion of training data classified as C ₄	67% C ₃ / 33% C ₄ training data distribution	50% C ₃ / 50% C ₄ training data distribution
Temporal Range	Increase study time-frame	2010 – 2020	1990 – 2020 or 1980 – 2020
Environmental Factors	Increase the number environmental factors considered	Temperature & Precipitation	Temperature & Precipitation (+ CO ₂ levels, fire intensity, soil factors, etc.)
Model Type	Transition to regressor	Binary Classifier (C ₃ vs C ₄)	Regressor ($x\%$ C ₃ , $y\%$ C ₄)
Features	Add non-correlated features	12 features, 4 features account for 99% of cumulative explained value ratio	12+ features, 4+ features account for 99% of cumulative explained value ratio

Table 6. Sample next steps

Conclusion

An analysis on the changes in PFT distribution of the North American Great Plains over the past decade was performed to identify potential shifts in PFT distributions and its correlations with environmental factors. The results found that changes in PFT distribution in the East Transect exhibited a northward migration, while the PFT distribution in the Central Transect and West Transect appeared to remain stable. Regression analysis between changes in PFT distribution and changes in environmental variables also indicated no significant dependencies between the two. Further works seeks to improve the accuracy of the learned model classifier as well as expand the scope of the analysis into considering longer temporal ranges, additional phenological features, and correlations with new environmental variables.

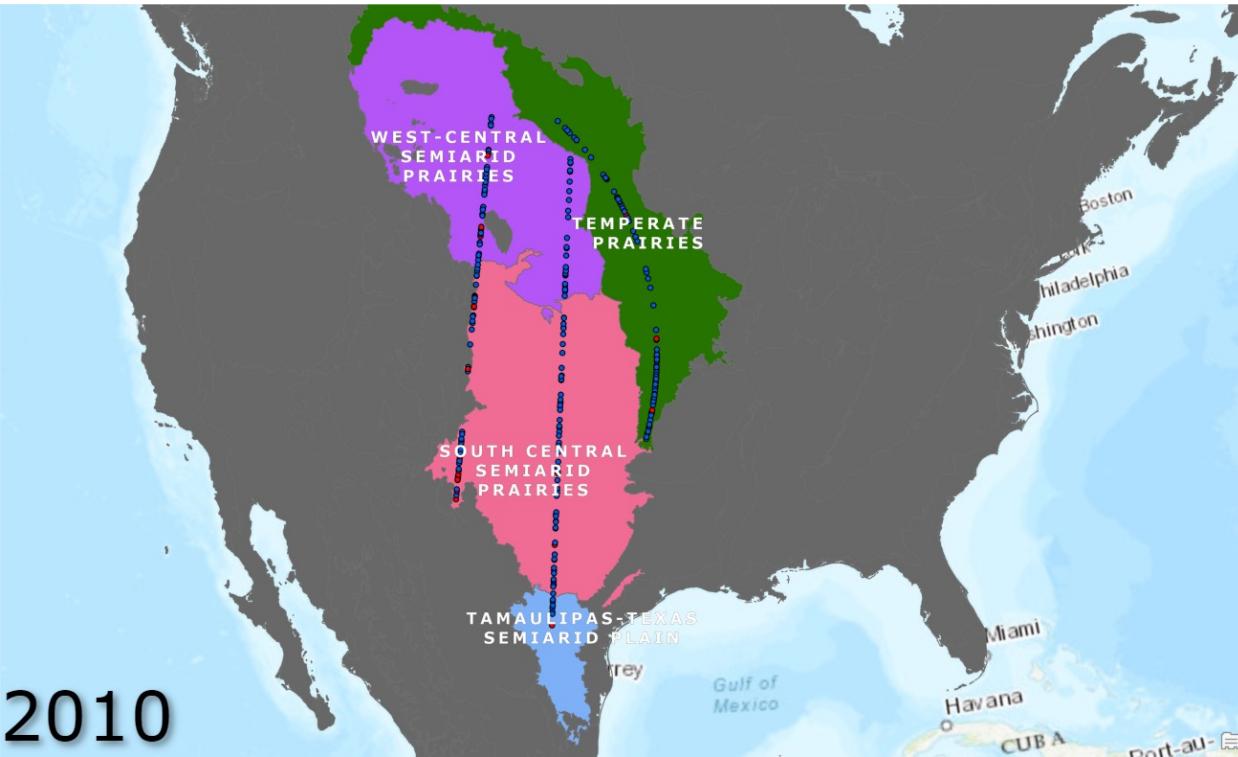
References

- Adams, R. M., Rosenzweig, C., Peart, R. M., Ritchie, J. T., McCarl, B. A., Glyer, J. D., et al. (1990). Global climate change and US agriculture. *Nature*, 345(6272), 219-224. doi:10.1038/345219a0
- Adjorlobo, C., Mutanga, O., Cho, M. A., & Ismail, R. (2013a). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 21, 535-544. doi:<https://doi.org/10.1016/j.jag.2012.07.011>
- Adjorlobo, C., Mutanga, O., Cho, M. A., & Ismail, R. (2013b). Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification. *International Journal of Applied Earth Observation and Geoinformation*, 21, 535-544. doi:<https://doi.org/10.1016/j.jag.2012.07.011>
- Bengtsson, J., Bullock, J. M., Egoh, B., Everson, C., Everson, T., O'Connor, T., et al. (2019). Grasslands—more important for ecosystem services than you might think. *Ecosphere*, 10(2), e02582. doi:<https://doi.org/10.1002/ecs2.2582>
- Blair, J., Nippert, J., Briggs, J., & Monson, R. K. (2014). *Grassland ecology*. New York, NY : Springer New York. doi:10.1007/978-1-4614-7501-9_14
- Carlier, L., Rotar, I., Vlahova, M., & Vidican, R. (2009). Importance and functions of grasslands. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 37(1), 25. Retrieved from Agricultural & Environmental Science Collection; Publicly Available Content Database database.
- Clevers, J. G. P. W., & Getelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344-351. doi:<https://doi.org/10.1016/j.jag.2012.10.008>
- Collatz, G. J., Berry, J. A., & Clark, J. S. (1998). Effects of climate and atmospheric CO₂ partial pressure on the global distribution of C4 grasses: Present, past, and future. *Oecologia*, 114(4), 441-454. doi:10.1007/s004420050468
- Commission for Environmental Cooperation, (CEC). (1997). *Ecological regions of north america: Toward a common perspective*
- Copernicus Climate, C. S. (2019a). *ERA5-land hourly data from 2001 to present ECMWF*. Retrieved from DOI.org (Datacite) database. Retrieved from <https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac>
- Copernicus Climate, C. S. (2019b). *ERA5-land hourly data from 2001 to present ECMWF*. Retrieved from DOI.org (Datacite) database. Retrieved from <https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac>
- Del Grosso, S., Parton, W., Stohlgren, T., Zheng, D., Bachelet, D., Prince, S., et al. (2008). Global potential net primary production predicted from vegetation class, precipitation, and temperature. *Ecology*, 89(8), 2117-2126. doi:<https://doi.org/10.1890/07-0850.1>
- Dixon, A. P., Faber-Langendoen, D., Josse, C., Morrison, J., & Loucks, C. J. (2014). Distribution mapping of world grassland types. *Journal of Biogeography*, 41(11), 2003-2019. doi:<https://doi.org/10.1111/jbi.12381>
- Elmendorf, S., & Barnett, D. (2021). *NEON user guide to plant presence and percent cover NEON* (National Ecological Observatory Network).
- Foody, G. M., & Dash, J. (2007). Discriminating and mapping the C3 and C4 composition of grasslands in the northern great plains, USA. *Ecological Informatics*, 2(2), 89-93. doi:<https://doi.org/10.1016/j.ecoinf.2007.03.009>
- Gamon, J. A., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: An optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, 112(4), 492-501. Retrieved from JSTOR database. Retrieved from <http://www.jstor.org.proxy.library.upenn.edu/stable/4221805>
- Goodin, G., & Henebry, G. M. (1997). A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model. *Remote Sensing of Environment*, 61(2), 270-278. doi:[https://doi.org/10.1016/S0034-4257\(97\)00043-6](https://doi.org/10.1016/S0034-4257(97)00043-6)
- Gowik, U., & Westhoff, P. (2011). The path from C₃ to C₄ photosynthesis. *Plant Physiology*, 155(1), 56-63. Retrieved from JSTOR database. Retrieved from <http://www.jstor.org.proxy.library.upenn.edu/stable/41433981>
- Griffith, D. M., Cotton, J. M., Powell, R. L., Sheldon, N. D., & Still, C. J. (2017). Multi-century stasis in C3 and C4 grass distributions across the contiguous United States since the industrial revolution. *Journal of Biogeography*, 44(11), 2564-2574. doi:<https://doi.org/10.1111/jbi.13061>
- Homer, C., Dewitz, J., Jin, S., Xian, G., Costelloe, C., Danielson, P., et al. (2020). *Conterminous United States land cover change patterns 2001–2016 from the 2016 national land cover database*. Amsterdam : Elsevier Science BV. doi:10.1016/j.isprsjprs.2020.02.019
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1), 195-213. doi:[https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Jin, S., Homer, C., Yang, L., Danielson, P., Dewitz, J., Li, C., et al. (2019). *Overall methodology design for the United States national land cover database 2016 products*. Basel, Switzerland : Molecular Diversity Preservation International. doi:10.3390/rs11242971
- Kattge, J., Diaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönnisch, G., et al. (2011). TRY - a global database of plant traits: TRY - A global database of plant traits. *Global Change Biology*, 17(9), 2905-2935. doi:10.1111/j.1365-2486.2011.02451.x
- Krauss, R., Meier, C., Patterson, M., & Smith, O. (2018a). *TOS site characterization report: Domain 06*
- Krauss, R., Meier, C., Patterson, M., & Smith, O. (2018b). *TOS site characterization report: Domain 10*
- Krauss, R., Meier, C., Patterson, M., & Smith, O. (2018c). *TOS site characterization report: Domain 11 NEON* (National Ecological Observatory Network).
- Krauss, R., Meier, C., Patterson, M., & Smith, O. (2018d). *TOS site characterization report: Domain 14*
- Krauss, R., Meier, C., Patterson, M., & Smith, O. (2018e). *TOS site characterization report: Domain 9*

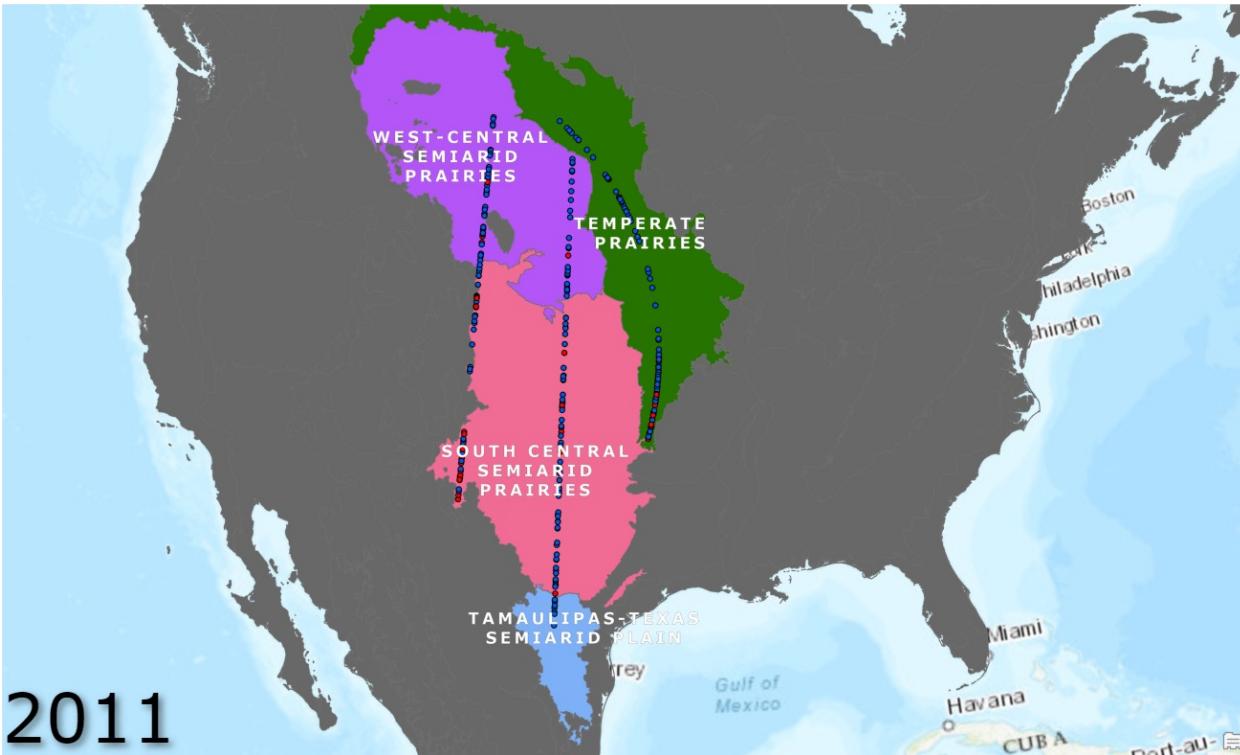
References

- Liu, L. (2010). Discriminating C3 and C4 plants from hyperspectral data. Paper presented at the pp. 4252-4255. doi:10.1109/IGARSS.2010.5648994
- Liu, X., Bo, Y., Zhang, J., & He, Y. (2015). Classification of C3 and C4 vegetation types using MODIS and ETM+ blended high spatio-temporal resolution data doi:10.3390/rs71115244
- National Ecological Observatory Network, (NEON). (2021). Plant presence and percent cover (DP1.10058.001) <https://data.neonscience.org/data-products/DP1.10058.001> doi:10.1371/journal.pone.0105139
- National Ecological Observatory Network (NEON). (2021). Plant presence and percent cover (DP1.10058.001) <https://data.neonscience.org/data-products/DP1.10058.001>
- NEON (National Ecological Observatory Network). (2019). Frequently asked questions (FAQ). Retrieved 31 March, 2021, from
- Omernik, J. M. (1987). *Ecoregions of the conterminous united states*. Washington, D.C. : The Association. doi:10.1111/j.1467-8306.1987.tb00149.x
- Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous united states: Evolution of a hierarchical spatial framework. *Environmental Management*, 54(6), 1249-66. doi:<http://dx.doi.org.proxy.library.upenn.edu/10.1007/s00267-014-0364-1>
- Paruelo, J. M., & Lauenroth, W. K. (1996). Relative abundance of plant functional types in grasslands and shrublands of north america. *Ecological Applications*, 6(4), 1212-1224. doi:<https://doi.org/10.2307/2269602>
- Pau, S., Edwards, E. J., & Still, C. J. (2013). *Improving our understanding of environmental controls on the distribution of C 3 and C 4 grasses*. Oxford, England] : Blackwell Science. doi:10.1111/gcb.12037
- Phelan, J. (2018). *What is life? : A guide to biology with physiology* (Fourth edition. ed.). New York: W.H. Freeman, Macmillan Learning.
- Reed, B. C., Brown, J. F., Vanderzee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, 5(5), 703-714. Retrieved from USGS Publications Warehouse database. Retrieved from <http://pubs.er.usgs.gov/publication/70187023>
- Risser, P. G. (1988). Diversity in and among grasslands. *Biodiversity* (pp. 176-180). Washington D.C.: National Academy of Sciences.
- Sage, R. F., Christin, P., & Edwards, E. J. (2011). *The C4 plant lineages of planet earth*. Oxford,: Clarendon Press.
- Sage, R. F., & Kubien, D. S. (2007). The temperature response of C3 and C4 photosynthesis. *Plant, Cell & Environment*, 30(9), 1086-1106. doi:<https://doi.org/10.1111/j.1365-3040.2007.01682.x>
- Savitzky, A., & Golay, M. J. E. (1964). *Smoothing and differentiation of data by simplified least squares procedures*. Washington : American Chemical Society. doi:10.1021/ac60214a047
- Shiferaw, B., Smale, M., Braun, H., Duveiller, E., Reynolds, M., & Muricho, G. (2013). Crops that feed the world 10. past successes and future challenges to the role played by wheat in global food security. *Food Security*, 5(3), 291-317. doi:10.1007/s12571-013-0263-y
- Shoko, C., Mutanga, O., & Dube, T. (2018). *Determining optimal new generation satellite derived metrics for accurate C3 and C4 grass species aboveground biomass estimation in south africa* doi:10.3390/rs10040564
- Sterling, S., & Ducharme, A. (2008). Comprehensive data set of global land cover change for land surface model applications. *Global Biogeochemical Cycles*, 22(3) doi:<https://doi.org/10.1029/2007GB002959>
- Still, C. J., Berry, J. A., Collatz, G. J., & DeFries, R. S. (2003). Global distribution of C3 and C4 vegetation: Carbon cycle implications. *Global Biogeochemical Cycles*, 17(1), 6-14. doi:<https://doi.org/10.1029/2001GB001807>
- Streck, N. A. (2005). Climate change and agroecosystems: The effect of elevated atmospheric CO₂ and temperature on crop growth, development, and yield. *Ciencia Rural*, 35, 730-740.
- Teeri, J. A., & Stowe, L. G. (1976). Climatic patterns and the distribution of C₄ grasses in north america. *Oecologia*, 23(1), 1-12. Retrieved from JSTOR database. Retrieved from <http://www.jstor.org.proxy.library.upenn.edu/stable/4215231>
- Tieszen, L. L., Reed, B. C., Bliss, N. B., Wylie, B. K., & DeJong, D. D. (1997). Ndvi, c3 and c4 production, and distributions in great plains grassland land cover classes. *Ecological Applications*, 7(1), 59-78. doi:[https://doi.org/10.1890/1051-0761\(1997\)007\(0059:NCACP\)2.0.CO;2](https://doi.org/10.1890/1051-0761(1997)007(0059:NCACP)2.0.CO;2)
- Vermote, E. (2015). *MOD09Q1 MODIS/terra surface reflectance 8-day L3 global 250m SIN grid V006 NASA EOSDIS Land Processes DAAC*.
- W. W. Hargrove, & F. M. Hoffman (1999). *Using multivariate clustering to characterize ecoregion borders* doi:10.1109/5992.774837
- Wang, C., Fritschi, F. B., Stacey, G., & Yang, Z. (2011). Phenology-based assessment of perennial energy crops in north american tallgrass prairie. *Annals of the Association of American Geographers*, 101(4), 742-751. Retrieved from JSTOR database. Retrieved from <http://www.jstor.org.proxy.library.upenn.edu/stable/27980222>
- Wang, C., Hunt, E. R., Zhang, L., Guo, H., & Hunt, E. R. (2013). *Phenology-assisted classification of C3 and C4 grasses in the U.S. great plains and their climate dependency with MODIS time series*. New York, NY] : Elsevier Science Inc. doi:10.1016/j.rse.2013.07.025
- Witwicki, D. L., Munson, S. M., & Thoma, D. P. (2016). Effects of climate and water balance across grasslands of varying C3 and C4 grass cover. *Ecosphere*, 7(11), e01577. doi:<https://doi.org/10.1002/ecs2.1577>
- Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S. M., et al. (2018). *A new generation of the united states national land cover database: Requirements, research priorities, design, and implementation strategies*. Amsterdam : Elsevier Science BV. doi:10.1016/j.isprsjprs.2018.09.006
- Zhang, H., Yu, Q., Huang, Y., Zheng, W., Tian, Y., Song, Y., et al. (2014). *Germination shifts of C3 and C4 species under simulated global warming scenario*. San Francisco, CA : Public Library of Science. doi:10.1371/journal.pone.0105139

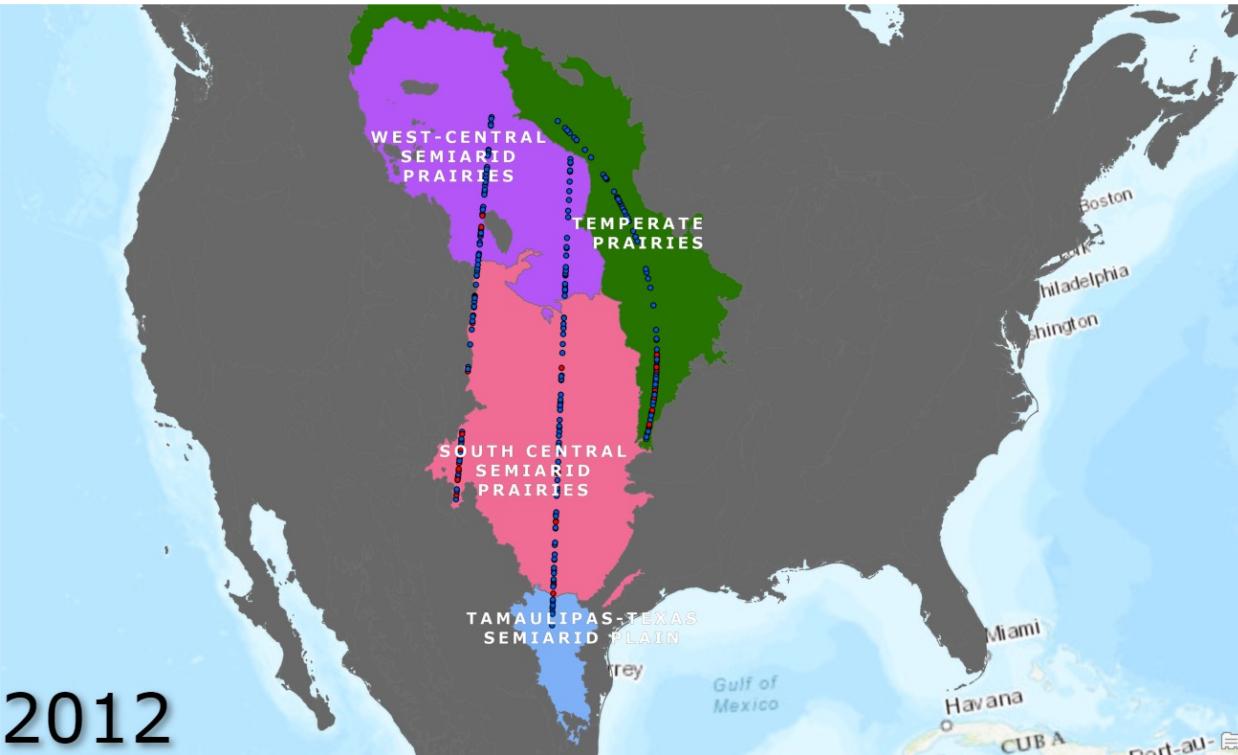
Appendix: 2010



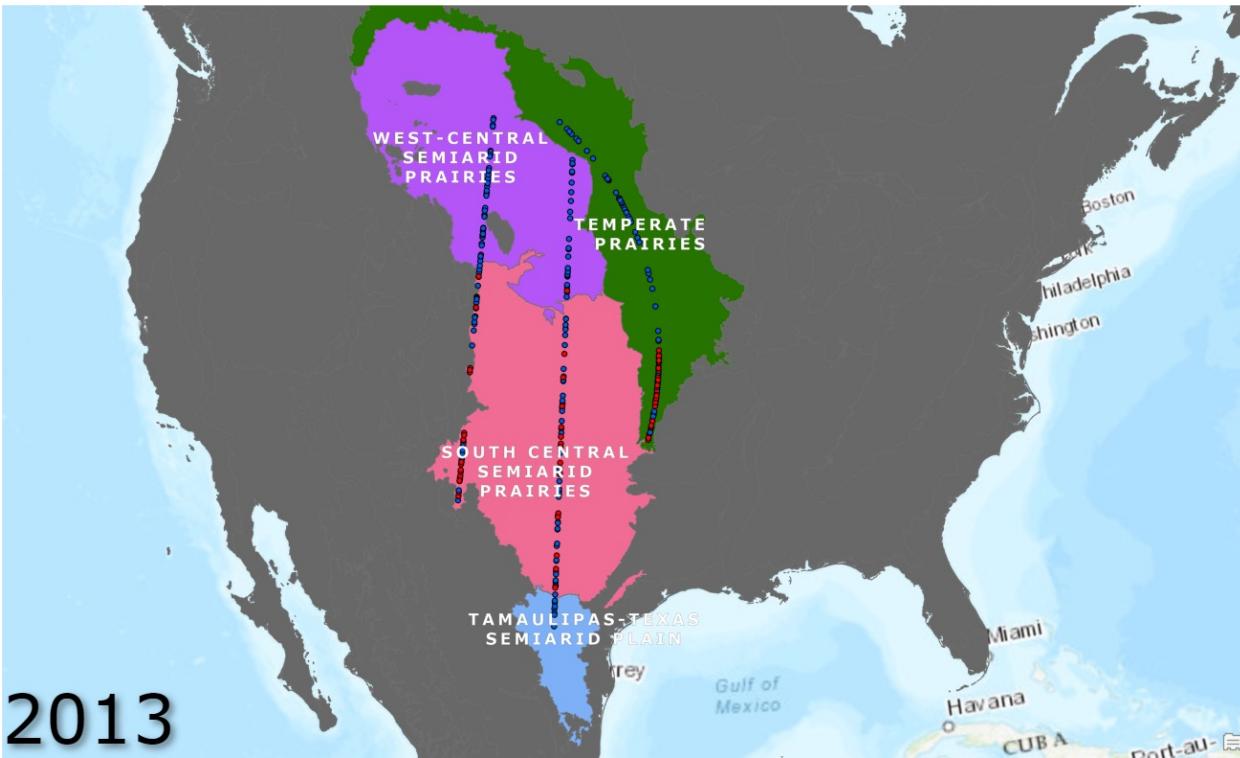
Appendix: 2011



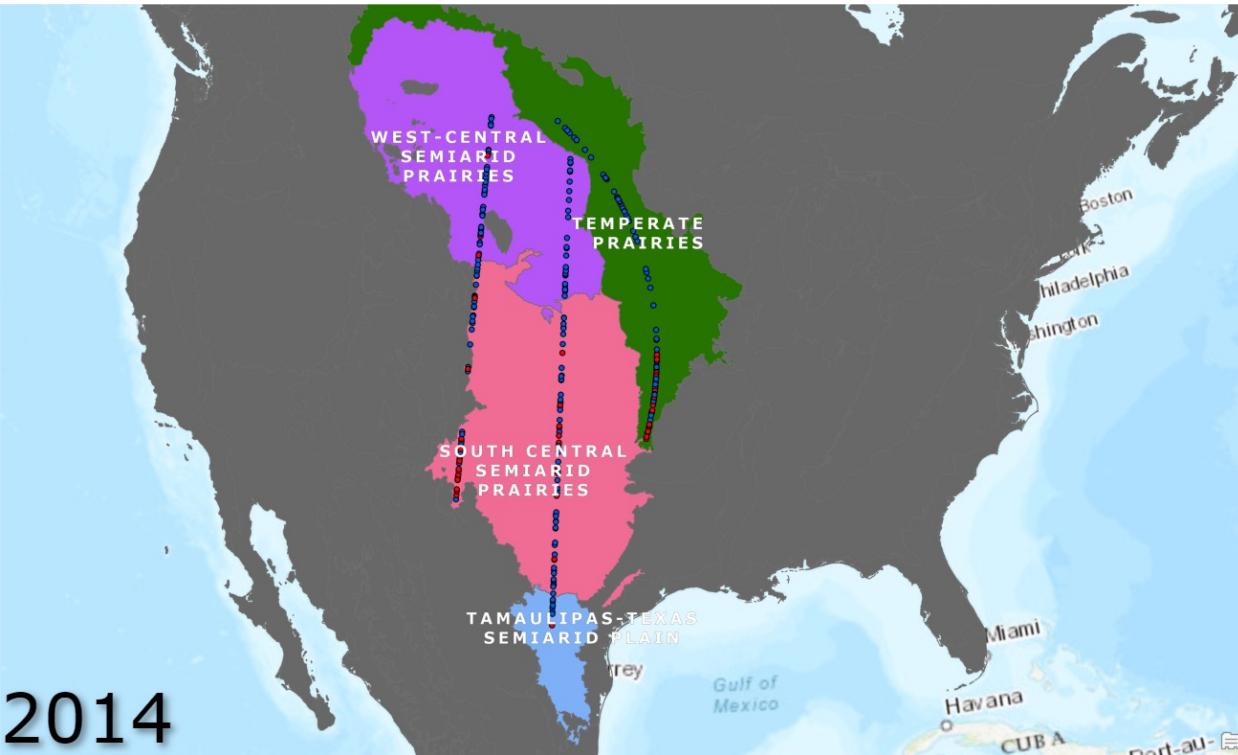
Appendix: 2012



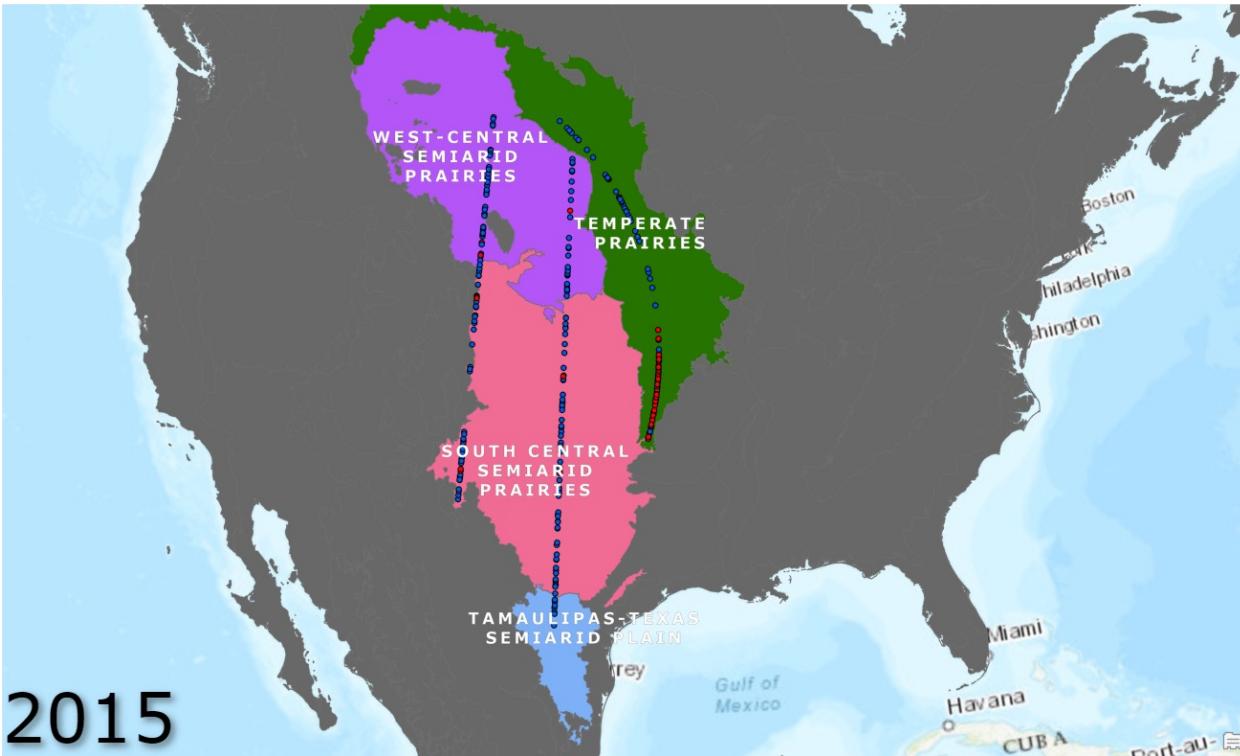
Appendix: 2013



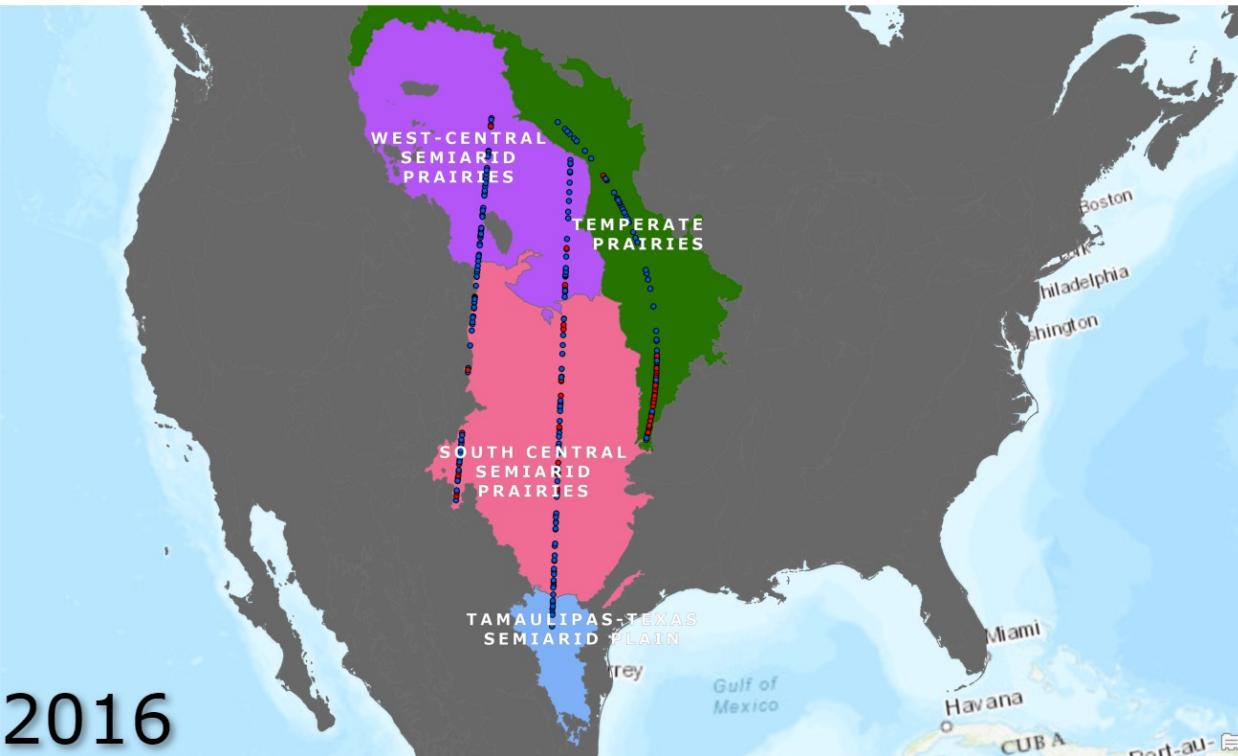
Appendix: 2014



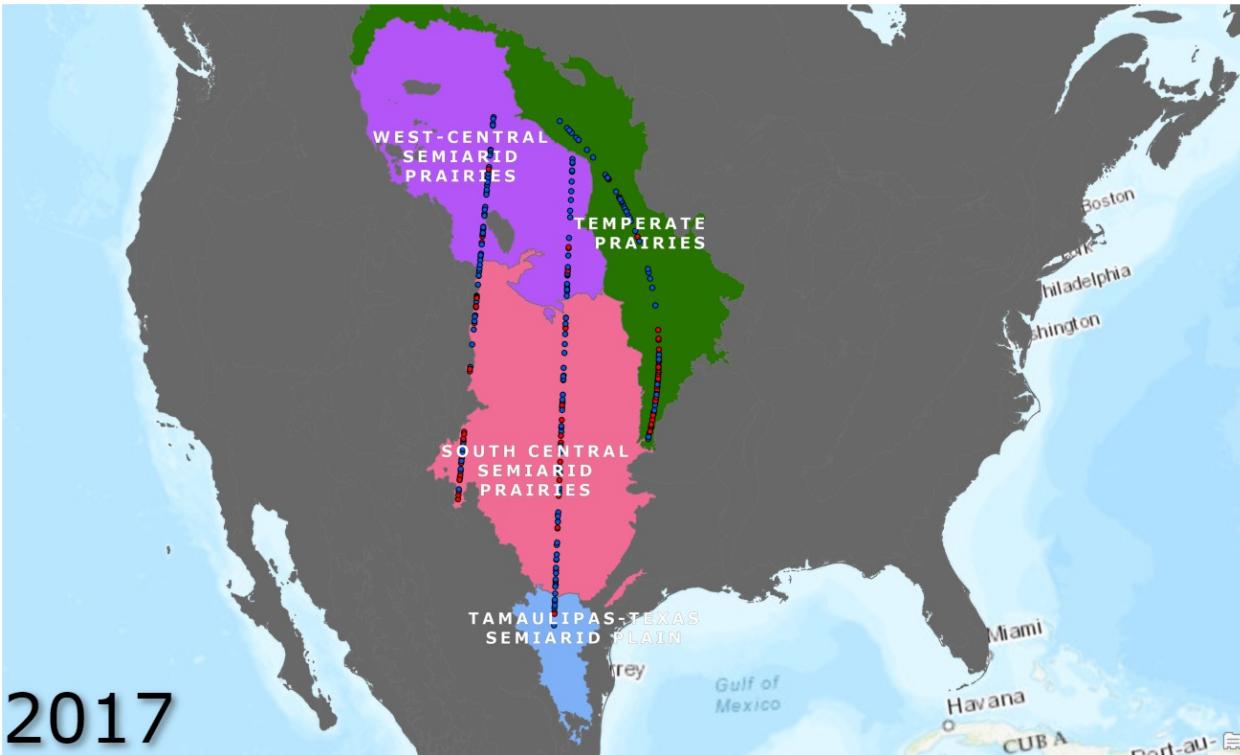
Appendix: 2015



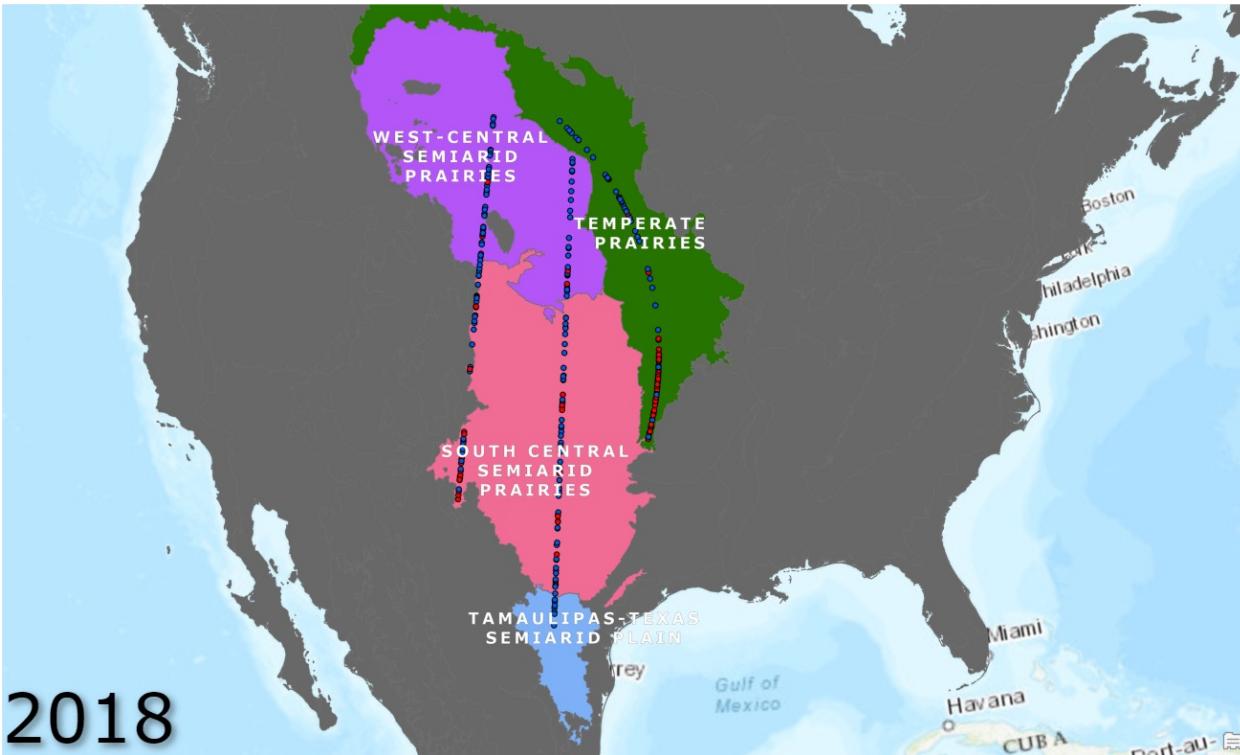
Appendix: 2016



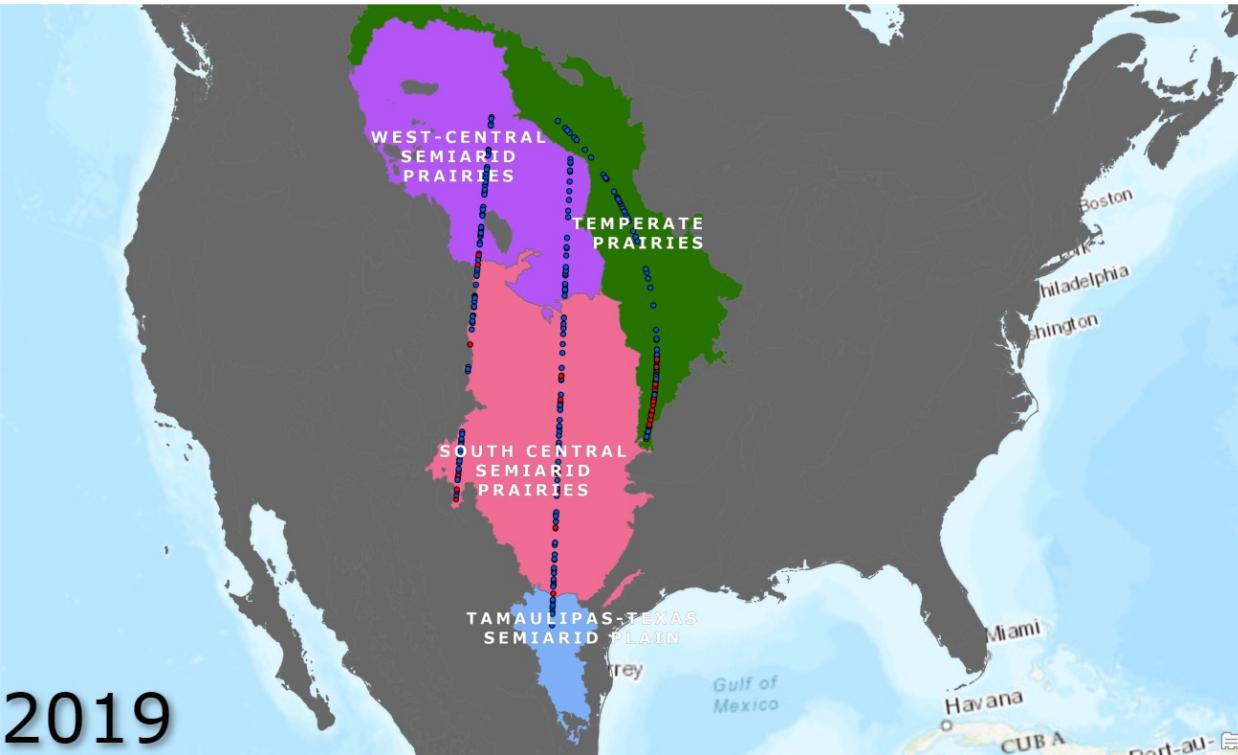
Appendix: 2017



Appendix: 2018



Appendix: 2019



Appendix: 2020

