Eastern Texas Ecological Conservation

Mapping Change in Longleaf Pine Distribution across Eastern Texas to Inform Conservation and Restoration Efforts

Fall 2025 | Idaho – Pocatello
November 21st, 2025

Authors: Seth Christiansen (Analytical Mechanics Associates), Caiden Hartrich (Analytical Mechanics Associates), Olivia Hockley-Rodes (Analytical Mechanics Associates), Jennings Leavell (Analytical Mechanics Associates)

Abstract: Once widespread in the southeastern U.S., longleaf pine (*Pinus palustris*) and its ecosystem declined to only 3% of its original extent due to logging, fire suppression, and encroachment from industrial pine plantations. The latter are typically dominated by loblolly pine (*Pinus taeda*) and frequently harvested and replanted. Distinguishing longleaf from loblolly pine is challenging; they exhibit similar structure, growth patterns and spectral signatures and share similar habitat. Current longleaf conservation action heavily relies on manual site survey. NASA DEVELOP collaborated with the Texas Longleaf Team, Texan by Nature, and Texas A&M Forest Service to map the current and historical distribution of longleaf in eastern Texas using NASA Earth observations and machine learning. We trained a Random Forest classification model on Landsat 9 Operational Land Imager (OLI)-2 reflectance data and derived spectral indices, and applied the model to historical imagery from Landsat 8 OLI and Landsat 5 Thematic Mapper (TM) dating back to 1985. Compared to reference data, the model achieved an overall agreement of 90% with a kappa index of agreement (κ) of 0.79. Consideration of current and historical models suggests a 21.5% average increase in longleaf pine per decade. While these results may depict actual conservation outcomes, the ability to apply a classifier across the study area was hindered by longleaf pine stands occurring among similar pine tree species and seral stages. The results may be improved by augmenting the training data with more sites spread across the study area. Despite training data and map accuracy uncertainties, our study demonstrated the feasibility of using NASA Earth observations with machine learning to locate potential longleaf pine stands for informing conservation and restoration efforts.

Key Terms: eastern Texas, forest ecology, Landsat 5 TM, Landsat 8 OLI, Landsat 9 OLI-2, loblolly pine, longleaf pine, remote sensing, spectral signature

Advisors: Keith Weber (Idaho State University GIS Training and Research Center), Joseph Spruce (Analytical Mechanics Associates)

Lead: Isaac Goldings (Idaho – Pocatello)

1. Introduction

Longleaf pine (*Pinus palustris*) and its ecosystem once occurred throughout the coastal plains, foothills, and uplands of the southeastern United States. Through the 1800s, longleaf pine was heavily logged for its long, straight stems that produced high-value timber and sap products, such as pitch and turpentine (Farjon, 2013; Kush et al., 2003). By 1990, only 3% of the 89 million acres of native longleaf ecosystems remained (Frost, 1993; Oswalt et al., 2012). The decline of longleaf forest in the southeastern United States was recognized as problematic by ecologists and land managers nearly a century ago (Stoddard, 1931). Longleaf pine was later ranked the third most endangered ecosystem in the United States, and the tree was listed as endangered and declining on the International Union for Conservation of Nature (IUCN) Red List in 2013 (Farjon, 2013; Noss et al., 1995).

Current longleaf ecosystems continue to be threatened by industrial loblolly (*Pinus taeda*), shortleaf (*Pinus echinata*), and slash (*Pinus elliottii*) pine plantations and associated forest management practices that exclude fire (Chen & Guo, 2024; Fagan et al., 2018). Longleaf pine ecosystems depend highly on fire; the pine's intermediate life stages are relatively intolerant to shade, relying on fire to keep nutrient and shade conditions favorable and hardwood encroachment at bay (Barnett, 1999; Boyer & Peterson, 1983). These characteristics are key to the success of the longleaf ecosystem and the numerous specialized and endangered species it protects (Jafarov et al., 2021).

The longleaf pine ecosystem is highly productive and biodiverse. Several U.S. Fish and Wildlife Service listed threatened wildlife species are endemic to longleaf forests, such as the red-cockaded woodpecker, the gopher tortoise, and the black pine snake (Chen & Guo, 2024; Means, 2007). As an ecosystem, healthy longleaf pine forests are highly productive at improving air and water quality, controlling erosion, and regulating local climate (FEMA, 2022). These ecosystem services are valued at a minimum of $5,095 per hectare per year and yield roughly $41.80 per $1 spent on habitat conservation (FEMA, 2022; Texan by Nature, 2024). The high value of longleaf pine ecosystems has led to conservation efforts across its historical geographic range, including areas in eastern Texas.

We partnered with the Texas Longleaf Team, Texan by Nature, and Texas A&M Forest Service to identify and perform current and historical analyses of longleaf pine extent in eastern Texas using NASA Earth observations (EOs) and remote sensing tools. These organizations educate and collaborate with landowners to restore longleaf pine ecosystems on public and private forestlands, as well as facilitate partnerships with conservation groups and businesses. Current localized tools for informing restoration of longleaf pine in eastern Texas were made using the broader Longleaf Pine Prioritization Model. This model was developed by the Lower Mississippi Valley Joint Venture and Texas Longleaf Implementation Team. These localized decision-making tools work to guide future conservation actions by prioritizing lands adjacent to extant longleaf sites using multiple data inputs (Texas Longleaf Implementation Team, n.d.). The Texas Longleaf Team identified the need for mapping longleaf pine stands in eastern Texas specifically to help quantify the outcome of past conservation efforts and inform future decision-making. This need for longleaf pine mapping presented an ideal opportunity to supplement existing decision-making tools with NASA EOs.

Landsat surface reflectance data have spatial, spectral, and temporal resolution characteristics that may provide a means for estimating current and historical longleaf pine extent. We hypothesized that the challenge of differentiating longleaf pine from loblolly pine was feasible with Landsat data when supplemented with sufficient training data, derived spectral indices, additional explanatory data, and a machine learning approach to image classification. Kinane et al. (2021) and Sinha et al. (2011) suggest that these species may be differentiable at Landsat's 30 m resolution via various metrics and statistics, such as calculating the Normalized Difference Moisture Index (NDMI) and Enhanced Vegetation Index (EVI), and correlation by principal component analysis. Temporal analyses have proved useful as well; seasonal composite mosaics or harmonic regression can differentiate targeted tree species based on specific phenology (Adams et al., 2020; Grigorieva et al., 2020). Other studies suggest coupling spectral data with Forest Inventory and Analysis data

or structural Light Detection and Ranging (LiDAR) data to improve tree species identification accuracy (Fagan et al., 2018; Hogland et al., 2019).

The study area consists of 20 counties in eastern Texas, each within the historical range of longleaf pine (Figure 1). The Gulf of America is the southern border of the study area while the Texas-Louisiana border is the eastern edge. Our primary goal was to use Landsat 5 Thematic Mapper (TM) Landsat 8 Operational Land Imager (OLI), and Landsat 9 OLI-2 sensor data alongside ground truth data to identify longleaf pine stands to provide more comprehensive maps and datasets to the partner organizations. Mapping the current extent of longleaf pine utilizing Landsat 9 data could further inform our partner organizations' estimates of the impacts and cost of current and future conservation efforts. The secondary goal was to generate parallel products from 2015 Landsat 8 OLI data and 1985, 1995, and 2005 Landsat 5 TM data. These additions allowed the partner organizations to better quantify how logging, fire suppression, climate change, and other factors have impacted the longleaf ecosystem.

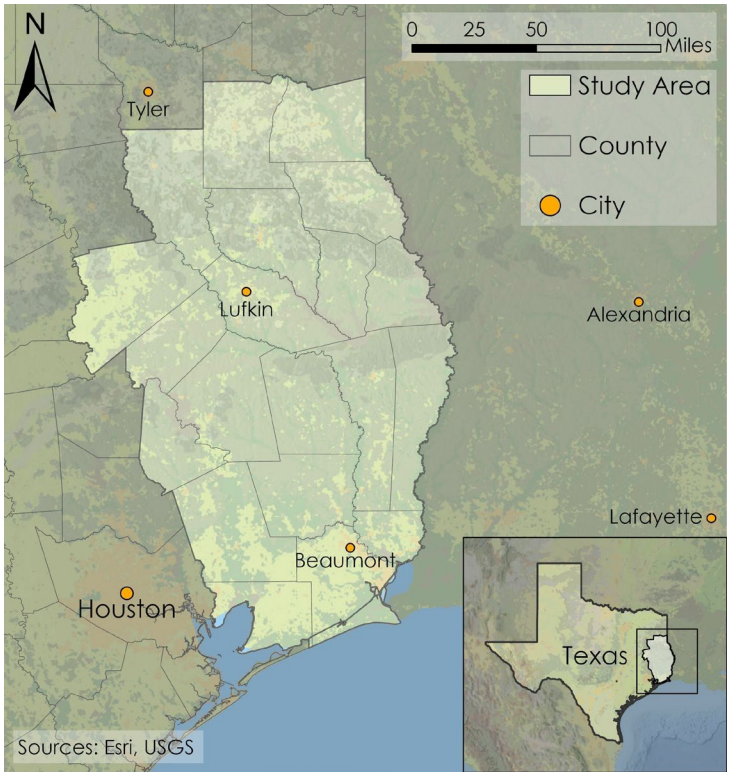


Figure 1. The Eastern Texas Eco project study area. The highlighted region consists of 20 counties within the historical range of longleaf pine and covers an area of 11.29 million acres. Cities in or near the study region are noted in orange. Texas county lines are shown in grey.

2. Methodology

2.1 Data Acquisition

2.1.1 Earth Observation Data Acquisition

We acquired satellite imagery from NASA and United States Geological Survey (USGS) data repositories to map longleaf and loblolly pine extent in eastern Texas. Landsat scenes were downloaded from the USGS EarthExplorer data portal (USGS, n.d.). Landsat 9 OLI-2 collection 2, level 2 multispectral data from scenes intersecting the study area on 21 December 2024 were manually selected for deriving the present-day longleaf distribution map. Landsat 5 TM and Landsat 8 OLI collection 2 level 2 winter imagery was obtained for historical maps through SpatioTemporal Asset Catalog query (STAC; Microsoft Open Source et al., 2022). Historical imagery was selected in decadal intervals after filtering for imagery with less than 10% cloud cover.

The study area was determined by a list of counties of interest from the partner organizations, filtered from Texas Department of Transportation (TxDOT) county boundary polygon data, and made into a GIS shapefile in ArcGIS 3.5.2 (Esri, 2025; TxDOT, 2025). Imagery from OLI bands 2-7 and TM bands 1-5 and 7 were also loaded into ArcGIS Pro for processing and classification (Table A1). We then used the USGS National Land Cover Database (NLCD) data to mask selected Landsat imagery and inform further data processing (USGS, 2024a). We downloaded NLCD data from the USGS Multi-Resolution Land Characteristics Consortium (MRLC) web viewer (MRLC, 2024), covering the study area for select years during 1985-2024 to develop both current and historical distribution maps.

2.1.2 Ground Truth Data Acquisition

We acquired ground truth data from project collaborators to support a supervised classification of longleaf and loblolly pine (Table A2). Longleaf restoration project data were provided by the Texas Longleaf Team, a project partner. Additional ground truth data were acquired from the Southeast Longleaf Ecosystem Occurrences (LEO) Geodatabase, an ancillary dataset provided by the Longleaf Alliance for this project (Longleaf Alliance, n.d.). The LEO Geodatabase contains longleaf pine survey data developed through targeted rapid field assessment across southeast United States. The data described occurrence of longleaf pine and other species, including loblolly pine, within the survey polygons. These datasets were crucial in training and validating a modelled prediction of longleaf and loblolly pine within the study area.

2.2 Data Processing

2.2.1 Processing of Earth Observations

Decadal Earth observation imagery between 1985 and 2024 was processed to prepare clean, standardized inputs for classification. This included clipping mosaiced Landsat imagery to the study area, correcting reflectance values, masking out all cloud and non-evergreen pixels, and calculating spectral indices. First, we converted integerized Landsat imagery back to its original floating-point reflectance value by applying a scale factor of 0.0000275 and an additive offset of –0.2 to each band, with a final target value range between 0 and 1 (USGS, 2024b). Next, the QA_PIXEL layer was used to remove interference from clouds, shadows, and snow. We masked each selected year of imagery to an NLCD Evergreen Forest layer from the corresponding year (USGS, 2024a).

We used this processed imagery to calculate various spectral indices to better differentiate longleaf from loblolly pine. Based on insight from Nieminen (2014) and Hogland et al. (2019), the following indices were derived: Normalized Difference Vegetation Index (NDVI; Equation 1; Kriegler et al., 1969), Normalized Difference Moisture Index (NDMI; Equation 2; Wilson & Sader, 2002), Bare Soil Index (BSI; Equation 3; Roy et al., 2002), Enhanced Vegetation Index (EVI; Equation 4; Huete et al., 1994), and Modified Soil Adjusted Vegetation Index (MSAVI2; Equation 5; Qi et al., 1994). In the equations below, NIR is the near infrared band's surface reflectance, Red is the red band's surface reflectance, SWIR is the shortwave infrared band's surface reflectance, and Blue is the blue band's surface reflectance.

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})} \tag{1}$$

$$\mathrm{NDMI} = \frac{(\mathrm{NIR} - \mathrm{SWIR})}{(\mathrm{NIR} + \mathrm{SWIR})} \tag{2}$$

$$\mathrm{BSI} = \frac{\big((\mathrm{SWIR} + \mathrm{Red}) - (\mathrm{NIR} + \mathrm{Blue})\big)}{\big((\mathrm{SWIR} + \mathrm{Red}) + (\mathrm{NIR} + \mathrm{Blue})\big)} \tag{3}$$

$$\mathrm{EVI} = 2.5 * \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + 6 * \mathrm{Red} - 7.5 * \mathrm{Blue}) + 1} \tag{4}$$

$$\text{MSAVI2} = \frac{2 * \text{NIR} + 1 - \sqrt{(2 * \text{NIR} + 1)^2 - 8 * (\text{NIR} - \text{Red})}}{2} \quad (5)$$

2.2.2. Ground Truth Data Processing

We generated point data from the ground-truth polygons to train and validate the model. Most polygons consisted of large, heterogeneous stands of forest, with trees of different species and various ages contained within one polygon. Non-vegetated areas were filtered out using the NLCD Evergreen Forest class in ArcGIS Pro. We applied a 15-meter internal buffer to the masked polygons, which ensured that minor errors in the NLCD classification were reduced. This buffer resulted in 765 refined polygons containing longleaf pine stands and 182 refined polygons containing loblolly pine stands. Within these filtered, masked, and buffered polygons, 300 random points were generated for each targeted tree species.

We loaded Landsat bands, indices, and rasterized ground truth points into TerrSet liberaGIS 20.0 (Clark CGA, 2024). Spectral signatures for the generated longleaf and loblolly points were analyzed and were not found to be highly differentiable. Ground truth points were then purified based on spectral signatures to reduce variability between training sites. We performed a nonparametric purification, assuming non-normal distribution of training point spectra, with a typicality threshold of 0.03. After purification, 185 longleaf points and 183 loblolly points remained, and their spectral signatures were more distinct than those prior to purification (Figure 2A).

2.3 Data Analysis

2.3.1 Decision Forest Classification

We used a Random Forest machine learning decision model as implemented by TerrSet to make species predictions. A Random Forest model works by combining the predictions of many individual decision trees to make a more accurate and reliable result (Belgiu & Drăguţ, 2016). Each tree is trained on a random subset of the input reference training data and considers only a random subset of features when splitting nodes, which helps reduce overfitting. For classification, the model takes a majority vote from all trees. We chose to use a Random Forest classifier because it handles complex, nonlinear relationships well, reduces overfitting through ensemble averaging, and provides high accuracy without extensive parameter tuning.

We input the assemblage of processed data into the TerrSet liberaGIS "DecisionForest" tool, which utilizes the Random Forest method to classify the masked study area. Using individual Landsat bands 2-7 and the ground truth points, we performed principal component analysis via the ArcGIS Principal Components tool. The first two components, which explained 98.93% of the variance, were initially added to the data stack of inputs to the model. Processed ground truth data were split 50-50% for training and validation. After attempting multiple iterations of the Random Forest model, we found that having more than 100 decision trees didn't significantly improve model accuracy. In assessing the performance of the model, we continually revised the input training data. EVI was removed because of its similarity to MSAVI2, which weighted the model too heavily toward values from those indices. Principal component analysis was removed from the model because its value range per component varied across different imagery dates and years.

2.3.3 Accuracy Assessment

Accuracy assessments of model runs were conducted in ArcGIS Pro and TerrSet. The ground truth points obtained from nonparametric purification were randomly split into two equal groups, with group A used for training and group B used for independent model validation. We then reversed the inputs, using group B as training data and group A for validation. It is important to note that while the points were independent, there were often multiple points sampled within large areas, which may lead to spatial autocorrelation. The model validations were compared, and the group A points validated by group B yielded an independently-assessed agreement of 90% with a κ of 0.79 (Table C5).

We iterated upon the model inputs 30 times and ultimately reached the highest agreement with training points and a moderate amount of uncertainty (a value between 0 and 1 given to each classified pixel) when run only on the NIR band, NDVI, NDMI, MSAVI2 and BSI (Table C1, C2). This model had high confidence regarding longleaf pine classification, with the majority of data classified as above 0.95 confidence or below 0.05 (Figure C1), with a κ o f 0.81 with a standard deviation of 0.34 (Table C1, C2). A run with the entire data stack showed the least amount of uncertainty (Figure C2) but had slightly less agreement with training points ($\kappa = 0.79$, $\sigma = 0.37$) (Table C3, C4). A model run on only raw bands had the lowest agreement and highest uncertainty ($\kappa = 0.77$, $\sigma = 0.31$) (Figure C3, Table C5, C6). We selected the model input with the entire data stack because of its low uncertainty and the relatively consistent explanatory power of all the input layers (Table C7). It is important to note that the model created a binary, wall-to-wall classification, so NLCD Evergreen pixels that were neither longleaf nor loblolly were still included in this binary classification.

3. Results

3.1 Analysis of Results

3.1.1 Spectral Separability Results

We evaluated differences in spectral characteristics between longleaf and loblolly pine training pixels using TerrSet. We compared the mean and standard deviation scaled reflectance values for Landsat 9 OLI bands 2–7 for each species (Figure 2A). We calculated indices using Landsat 9 OLI-2 bands (Equations 1-3, 5) and compared the mean and standard deviation index values, accounting for variability using effect size (Figure 2B). This measure allowed for a comparison across inputs with different scales and variabilities. NDMI showed the greatest separation with an effect size of 1.47, followed by BSI at 1.30 and MSAVI2 at 0.99. NDVI and band 5 were moderate, 0.63 and 0.61, respectively. These results indicated that NDMI and BSI provided the most consistent distinction between classes while visible bands were less dependable due to their high variability (Table B2).

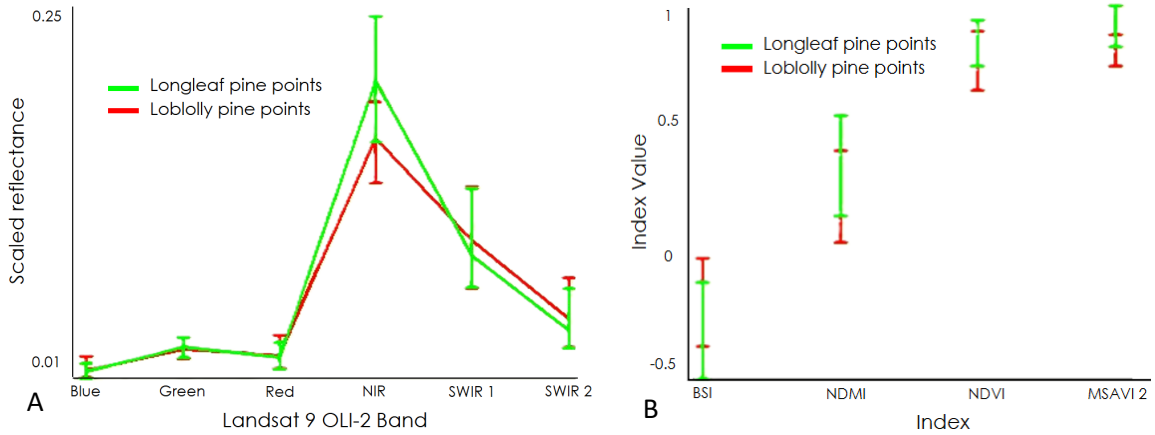


Figure 2. A spectral signature comparison of purified longleaf and loblolly training data using scaled reflectance values from Landsat 9 bands (A), and model inputs (B). Red represents the values of purified loblolly training points while green represents purified longleaf training points. The error bars represent minimum and maximum values, and the connecting line represents mean values (A).

3.1.2 Random Forest Model Results

Using 2024 winter Landsat imagery, the model classified 1.32 million acres as longleaf pine and 2.06 million acres as loblolly pine within the study area of 11.29 million acres. This equates to 12% and 19% of the study area, respectively (Figure 3A). Using historical Landsat imagery, the model predicted areas of longleaf and loblolly pine at a single point in time for each decade: 1985, 1995, 2005, and 2015 (Figure 4, Table B1). The acreage of longleaf pine increased each year, with a predicted increase in longleaf pine acreage by an average

of 21.5% every decade (Figure 5, Table B1). Linear and general logistic (i.e. Chapman-Richards) equations were fitted to the longleaf acreage data. The linear equation showed good agreement with the data: y=0.02x-42.35, $r^2 = 0.9386$. A general logistic equation, widely-used to represent growth dynamics of natural systems (Somers & Farrar, 1991), was also fitted and showed significant agreement: $y = \frac{1.35}{1+e^{-0.1(x-1990.82)}}$, R^2=0.9954 (Figure 5). With the historical data, a conservative estimate of longleaf area was developed by filtering down to pixels that were classified as longleaf at each date analyzed, which yielded 16,231 acres, 0.02% of the study area (Figure 3B).

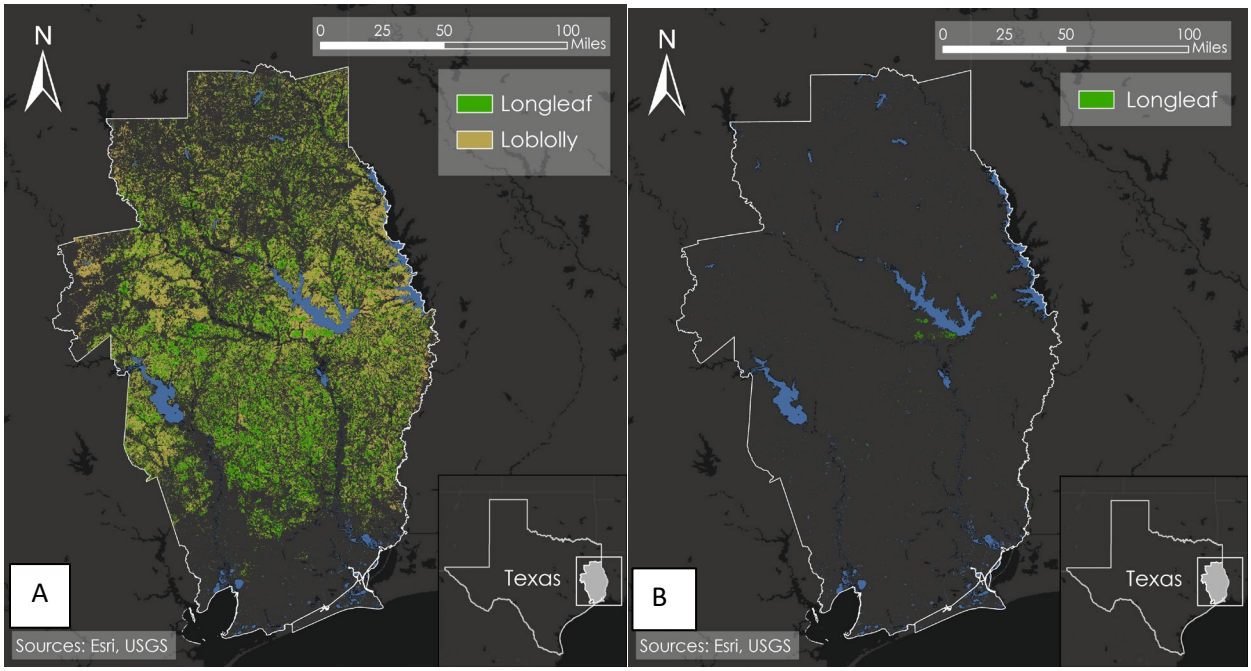
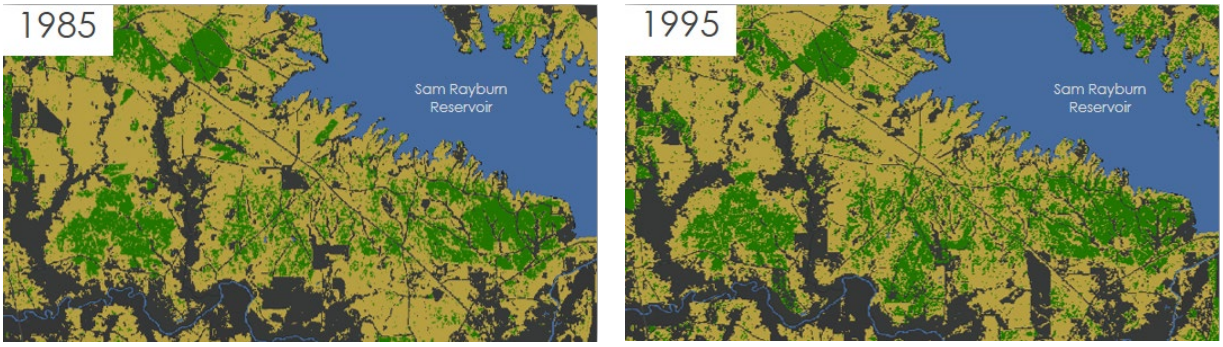


Figure 3. Random Forest classifier result maps. Longleaf pine is highlighted in green, loblolly pine in yellow, and water in blue against a dark grey study area background. The direct output of the 2024 model predicted 12% of the study area as longleaf (1,315,942 acres) and 19% (2,057,373 acres) as loblolly pine (A). Model output filtered to show only those pixels classified as longleaf pine only during all years of the study, yielding 0.02% of study area (16,231 acres) (B).



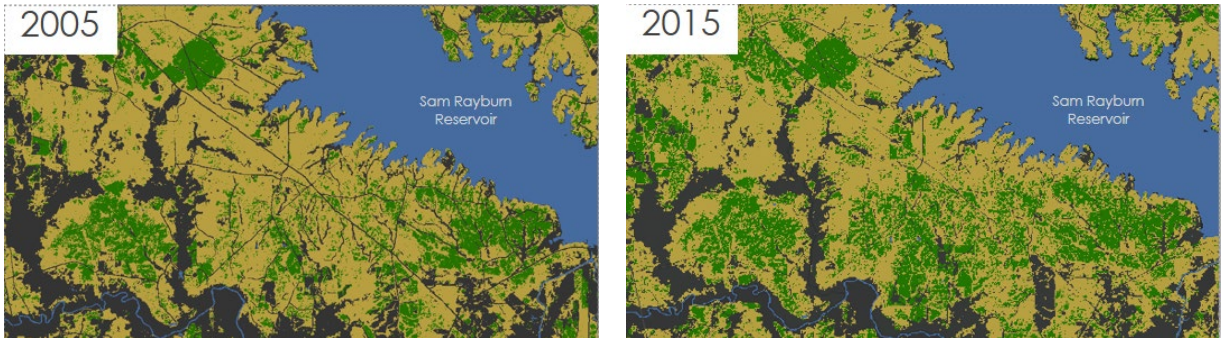


Figure 4. Historical distribution maps centered on Angelina National Forest, an area of interest where the Texas Longleaf Team and Texas A&M Forest Service have been actively restoring longleaf pine. Green represents predicted longleaf pine and gold depicts predicted loblolly pine on the above maps.

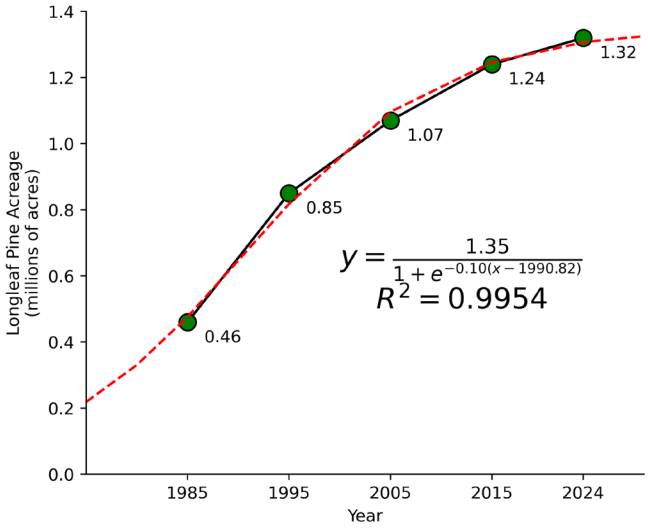


Figure 5. Longleaf pine acreage classified by the model in millions of acres over the study period. Individual classification acreages are represented by green circles with the exact values noted to the right. A best-fit general logistic equation, printed on the graph, was applied to the data with an R^2 of 0.9954, indicating significant correlation. Given that these data are model estimations, the sample size is small, and the longleaf pine ecosystems modeled are dynamic and actively managed, this best-fit line should be interpreted conservatively.

When compared to a false color composite RGB (blue=blue, green=NIR, red=SWIR1), the model appears to have predicted longleaf pine across areas of forest that appeared brighter green on the RGB, indicating more reflectance for the NIR band relative to most other cover types (Figure 6). Loblolly was often predicted in areas of darker green forest cover on the same RGB, indicating lower NIR reflectance, consistent with the spectral comparison (Figure 2A). Given the model's apparent overclassification of longleaf pine and the tendency for younger pine trees to have overall greater photosynthetic activity (higher greenness) compared to mature pines (Drake et al., 2010; Kuusk et al., 2017), stand age may be a significant confounding variable in the model. Additionally, Landsat's near infrared band is used in part to calculate each one of the selected spectral indices.

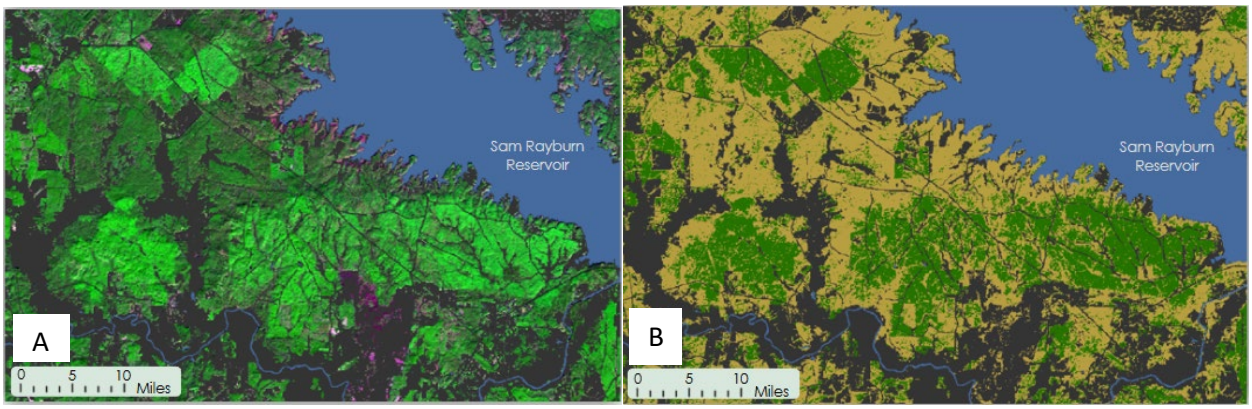


Figure 6. An area of interest centered over the Angelina National Forest depicting the possible correlation between NIR reflectance and predicted longleaf pine. A false color composite image of Landsat 9 imagery (blue=blue, green=NIR, red=SWIR1) (A). The model's prediction within the area of interest, with predicted longleaf in green and predicted loblolly in gold (B).

To determine if the model was linking the NIR band to tree age, a supervised classification was run solely on indices that used the red, green, and blue bands. These included the Blue-Green-Red Vegetation Index, the Normalized Green-Red Difference Index, the Visible Atmospherically Resistant Index, and the Green Leaf Index (Table C9). The resulting classification performed relatively poorly, with a κ of 0.50 and an independently assessed accuracy of 74.96% (Table C10), suggesting that the NIR band and the indices derived from it are very important in distinguishing longleaf from loblolly. We ran an additional model solely on inputs with high explanatory power (Table C7). However, the resulting classification also did not perform any better than the model with all inputs (Table C8, κ = 0.79, accuracy = 90%).

3.2 Errors & Uncertainties

There were multiple sources of error and uncertainty in the training data and methodology that affected the performance of the binary pine species classification model. The first source of error was from the ground-truth data polygons. Our partner organizations are longleaf pine restoration experts, so the polygons they supplied largely consisted of young stands of recently planted or restored longleaf pine stands. Loblolly polygons were derived from ground truth data from the LEO Geodatabase, which largely included areas with natural, mature trees. Therefore, most of the loblolly pine reference data were for mature forests with very few sample locations containing immature loblolly pine. In addition, the LEO Geodatabase was created to document longleaf pine, and loblolly polygons were selected from areas where loblolly was indicated as "present" in "Other Pine Species" and longleaf dominance status was indicated as "occasional_rare". Therefore, the ground truth data were biased towards young, planted longleaf stands and old, heterogenous natural loblolly stands, with a much greater number of training polygons for longleaf pine than loblolly. The limited availability of ground truth data across all common seral stages likely introduced confounding variables of age, species, and stand type into the model.

A second source of error came from the assumption that longleaf and loblolly are the only evergreen conifer species present as stands in the region. Slash pine and shortleaf pine are also present within the study area. Shortleaf pine is native to northwestern Texas and is often found mixed with longleaf and loblolly stands in the Pineywoods region (Texas A&M Forest Service, 2025). Slash pine was introduced to eastern Texas in 1926 and was once commonly planted in East Texas for timber production. Slash pine now reproduces naturally in the southern part of the study area (Texas A&M Forest Service, 2025). In addition, longleaf and loblolly pines have naturally hybridized in the region, creating a kind of tree called the Sonderegger pine. We expect that these less-common coniferous pine trees were incorporated into the binary longleaf-loblolly classification.

Using the NLCD Evergreen Forest class as a mask introduced some uncertainty into the model prediction. While the NLCD represents July landcover, the study utilized winter imagery collected approximately 6 months later. During this time, mature stands of trees may have been harvested for timber or pulpwood, creating clearcuts in areas still classified as evergreen. Also, certain young longleaf pine stands may have been too young to be classified on the NLCD as an evergreen forest. The NLCD appears to classify certain juvenile stands of pine trees as Scrub/Shrub due to low canopy cover and low tree height. Longleaf pine also likely exists within the NLCD Mixed Forest class, though longleaf within these mixed forest areas were not detected by the model because they were excluded by the NLCD evergreen forest mask.

The training and validation point purification process also introduced a source of error. While the points were randomly split into two equal groups, both sets were sampled from the same set of polygons. This introduces potential spatial autocorrelation where the validation was independent at the point level but not fully independent at the polygon level, inflating the model's accuracy. The point purification process may also have inflated the model's accuracy. While improving spectral separability, purification may have introduced bias by disproportionately selecting homogenous stand conditions that may not reflect species-wide growth patterns across the study area.

4. Conclusions

4.1 Interpretation of Results

The model, although mostly in agreement with reference data, appears to overpredict longleaf pine, likely due to uncontrolled confounding variables of stand age and stand type. Additionally, ground-truth data biased toward young, planted longleaf and older, natural loblolly stands compounded the age and stand type related bias, further impacting the model's performance. As a result, the model often categorized immature stands of both pine species as longleaf. In addition, very young, newly planted conifer forests were likely excluded from the model's prediction due to use of the NLCD Evergreen Forest mask.

The model predicted a 21.5% average increase in longleaf pine area each decade. This supposed increase may reflect the increasing number of plantation-based tree farms where pines are often cut as soon as they reach maturity (Fagan, 2018). This may also be due to classification that is responding to differences in plantation soil exposure highlighted by BSI, which would account for some classification based on the characteristics of managed forest landscapes as opposed to individual species at lower resolutions (Jia & Pang, 2023). Overall, 1.32 million acres of classified longleaf pine is far more than expected, given the total acreage of longleaf was estimated at 258,872 acres in the Southern Forests Futures Project subregion Coastal Plain (west), which includes eastern Texas and Louisiana (Oswalt et al., 2012).

It is difficult to compare this study's estimated pine acreages to existing estimates that utilized different data, study areas, and methods. The linear best-fit line for the longleaf acreage data was interpreted as unlikely to correspond with a real trend within the study area (Figure 5). The general logistic function may be interpreted as a better estimation of Texas pine plantation acreage, though likely not just longleaf. The general logistic best-fit line also shows reasonable agreement with a 1997-2027 projection for Texas pine plantation acreage on private land using USDA Forest inventory and supporting data (Zhang & Polyakov, 2010). Our estimate differs by approaching an asymptote at 1.35 million acres, while Zhang & Polyakov (2010) projected a decrease in Texas pine plantation on private lands in 2027. Disentangling stand age and type may yield better southern pine acreage estimate through more robust remote sensing model results.

Longleaf pine may be greener, moister, brighter, and more barren than loblolly pine, though it is still unclear exactly how age and stand type influence this spectral-signature-based observation. Pine's NIR reflectance may decrease with age, planted stands may have a lower BSI value than naturally regenerating ones, and the unintended bias in the ground truth data may be reflective of these possibilities. Without targeted ground truth data collection in regards to these variables, clearly determining their influence on the model is difficult.

However, NDMI and MSAVI2 had the highest explanatory power and effect sizes, with BSI and band 5 causing a slight drop in the model's dependent accuracy assessment (Tables B2 and C7). MSAVI2 controls for the reflectance of soil and NDMI may not be as clearly connected to stand age as other inputs (Kinane et al., 2021). Despite the confounding variables, a Landsat classification based on a Random Forest model shows potential for species occurrence mapping at a regional scale.

4.2 Feasibility & Partner Implementation

4.2.1 Feasibility

Remote longleaf pine identification is a cost-effective and scalable way to survey project sites. NASA EO data are free and consistently cover large areas of forest, reducing the need for time-intensive field surveys. The feasibility of remote species identification improves significantly when ground truth data accounts for a variety of confounding variables, and an intentional list of indices and ancillary data are used. The model used for this feasibility study may have overclassified longleaf pine due to a lack of training data for immature loblolly stands, though its high κ of 0.81 and 90% agreement with ground truth data suggest feasibility with continued iteration.

4.2.2 Partner Implementation

Although it is likely the Random Forest model currently overpredicted longleaf pine, its high agreement with the training points generated suggests that the model may produce highly reliable maps given further refinement of more representative training data from partner organizations. Possible model refinements include adding diverse ground truth datasets (e.g., for immature loblolly and mature longleaf pine), controlling confounding variables, incorporating additional multispectral data, and using forest structure data. Most of the ground truth data were clustered around the center of the study area, forcing the model to extrapolate on large swaths of forest. Adding new ground truth data spread across the study area would minimize extrapolation and provide a more diverse ground truth dataset to input into the model. The model could also be improved by including more specific ground truth data that spans a wider variety of tree growth patterns (e.g., seral stages). By including more loblolly ground truth data, especially for younger tree stands, the model would likely have less implicit age bias in its species classification. Additional ground truth data could include classifications to break out stands by species, age, and stand type.

Adding age and stand type classifications would help to control confounding variables during classification. By creating separate classes for young or mature stands and natural or planted stands, different spectral signatures could be determined for each class of each species. Similarly, hyperspectral imagery could be utilized to determine more precise spectral signatures for each species, with more bands available to specifically analyze reflectance characteristics for each targeted tree species. Our project results indicate there is little spectral differentiation between longleaf and loblolly pine in the visible light and SWIR bands using Landsat imagery, leaving the model to rely heavily on NIR signatures to differentiate between species. Finding greater differentiation of longleaf and loblolly pine with specific hyperspectral bands could greatly increase the effectiveness of the classifier. A smaller study area of priority longleaf restoration habitat would also allow for feasible processing of hyperspectral imagery, multi-band imagery with high spatial resolution, or LiDAR altimetry data with information on forest structure. Higher spatial resolution multispectral imagery may enable classification of individual trees as opposed to 30-meter stands, which may increase the differentiability between longleaf and loblolly pine. LiDAR-based forest structure data can be useful in addition to spectral data for classifying pine species in the absence of multiple classifications for age and stand type (Fagan et al., 2018). To aid in future implementation and refinement of our methodology by the partners, we developed a comprehensive tutorial. The tutorial walks the user through data acquisition, processing, and analysis.

The Texas Longleaf Team, Texan by Nature, and the Texas A&M Forest Service will benefit highly from the findings of this feasibility study. The Texas Longleaf Team can use the longleaf pine distribution map from the project to aid in the assessment of habitat fragmentation and prioritize future restoration efforts. The Texas A&M Forest Service can use the outcomes to supplement existing databases of targeted longleaf forest restoration, with the ability to use our methodology with newly acquired ground-truth data inputs. Texan by

Nature will benefit from the exploration of use-cases for NASA EOs to inform ecological conservation projects with future partner organizations. Though this feasibility study had limitations, the documented methodology provides a framework for partner organizations to incorporate remote sensing into their ecological conservation and restoration efforts.

5. Acknowledgements

We extend our thanks to the following individuals for their invaluable contributions to our project:

- **DEVELOP Personnel:**
  - Isaac Goldings – Node Lead (Idaho – Pocatello)
- **Advisors:**
  - Keith Weber – Idaho State University, GIS Training and Research Center GIS Director
  - Joe Spruce – Analytical Mechanics Associates, Senior Research Scientist
- **Partners:**
  - Jenny Sanders – Texas Longleaf Team Coordinator
  - Taylor Keys – Texan by Nature Director of Programs
  - Caitlin Tran – Texan by Nature Program Manager
  - Hughes Simpson – Texas A&M Forest Service Assistant Director
  - Eric Wannlund – Texas A&M Forest Service Staff Forester

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration. This material is based upon work supported by NASA through contract 80LARC23FA024.

6. Glossary

ArcGIS Pro: A geospatial analysis software developed by Esri where users can analyze satellite imagery and create maps.
BSI: Bare Soil Index. A spectral index used to identify areas of bare soil using shortwave infrared, red, near infrared, and blue bands to distinguish soil. Higher BSI values indicate the presence of bare soil.
EO: Earth observations. Satellites with sensors that collect information about the Earth's physical, chemical, and biological systems over space and time.
EVI: Enhanced Vegetation Index. A spectral index used to measure vegetation health using near infrared, red, and blue bands to distinguish vegetation. Higher EVI values indicate the presence of healthier vegetation.
MSAVI2: Modified Soil Adjusted Vegetation Index 2. A spectral index used to measure plant health and growth by minimizing the effects of bare soil using the near infrared and red bands to distinguish vegetation. Higher MSAVI2 values indicate the presence of healthier vegetation.
NDMI: Normalized Difference Moisture Index. A spectral index used to measure vegetation water content using the near infrared and short-wave infrared bands. Values of –1 to 0 indicate areas with no moisture, while values of 0 to 1 indicate areas with vegetation water content. Higher NDMI values indicate the presence of well-watered vegetation.
NDVI: Normalized Difference Vegetation Index. A spectral index used to identify areas of green vegetation using near infrared and red bands to distinguish green vegetation. Values of –1 to 0 indicate barren areas while values 0 to 1 indicate vegetated areas. Higher NDVI values indicate the presence of dense vegetation.
NLCD: National Land Cover Database. A land cover database developed by the USGS to map and monitor changes in land cover in the United States.
OLI: Operational Land Imager. We used Landsat 9 OLI-2 imagery to develop the current distribution map and Landsat 8 OLI imagery to develop the 2014 historical distribution map.
Random forest: A supervised machine learning algorithm that uses a combination of multiple decision trees to classify data.
Spectral reflectance: The ratio between the energy reflected by the surface and the energy incident on the surface. This energy typically comes from the sun's electromagnetic radiation; however, active sensors can also produce energy and measure its reflectance. It is measured as a function of wavelength.
TerrSet: An open-access integrated geospatial software system for monitoring and modeling the earth system for sustainable development developed by Clark University Center for Geospatial Analytics.
TM: Thematic Mapper. We used Landsat 5 TM imagery to develop our 1985, 1995, and 2005 historical maps.

7. References

Adams, B., Iverson, L., Matthews, S., Peters, M., Prasad, A., & Hix, D. M. (2020). Mapping forest composition with Landsat time series: An evaluation of seasonal composites and harmonic regression. *Remote Sensing, 12*(4), 610. https://doi.org/10.3390/rs12040610

Barnett, J. P. (1999). Longleaf pine ecosystem restoration: The role of fire. *Journal of Sustainable Forestry*, *9*(1-2), 89-96. https://doi.org/10.1300/J091v09n01_07

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal for Applied Earth Observation and Geoinformation, 39,* 79-86. https://doi.org/10.1016/j.jag.2015.02.012

Boyer, W. D. & Peterson, D. W. (1983). Longleaf pine. In R. M. Burns (Tech. Comp.), *Silvicultural systems for the major forest types of the United States, Agriculture Handbook 445* (pp. 153-156). U.S. Department of Agriculture, Forest Service. https://www.fs.usda.gov/rm/pubs_series/wo/wo_ah445.pdf

Chen, X. & Guo, Q. (2024). Updates on longleaf pine ecology, restoration, and management. *Forests*, *15*(8), 1399. https://doi.org/10.3390/f15081399

Clark Center for Geospatial Analytics (ClarkCGA). (2024). *TerrSet liberaGIS* (Version 20.0.0) [computer software]. Available from https://www.clarku.edu/centers/geospatial-analytics/terrset/download/

Drake, J. E., Raetz, L. M., Davis, S. C., & Delucia, E. H. (2010). Hydraulic limitation not declining nitrogen availability causes the age-related photosynthetic decline in loblolly pine (*Pinus taeda L.):* Loblolly pine photosynthesis. *Plant, Cell & Environment*, 33(10), 1756–1766. https://doi.org/10.1111/j.1365-3040.2010.02180.x

Esri. (2025). *ArcGIS Pro* (Version 3.5.2) [computer software]. Available from https://pro.arcgis.com/en/pro-app/latest/get-started/download-arcgis-pro.htm

Fagan, M. E., Morton, D. C., Cook, B. D., Masek, J., Zhao, F., Nelson, R. F., & Huang, C. (2018). Mapping pine plantations in the southeastern U.S. using structural, spectral, and temporal remote sensing data. *Remote Sensing of Environment*, *216*, 415-426. https://doi.org/10.1016/j.rse.2018.07.007

Farjon, A. (2013). Pinus palustris, longleaf pine. *The IUCN Red List of Threatened Species* (ICUN 2008: e.T39068A2886222). https://doi.org/10.2305%2FIUCN.UK.2013-1.RLTS.T39068A2886222.en

Federal Emergency Management Agency (FEMA). (2022). *FEMA ecosystem service value updates.* U.S. Department of Homeland Security, FEMA. https://www.fema.gov/sites/default/files/documents/fema_ecosystem-service-value-updates_2022.pdf

Frost, C. (1993). Four centuries of changing landscape patterns in the longleaf pine ecosystem. In Hermann, S. M. (Ed.), *Proceedings 18th tall timbers fire ecology conference* (pp. 17-43). Tall Timbers Research, Inc. https://www.talltimbers.org/wp-content/uploads/2014/03/Frost1993_op.pdf

Gitelson, A. A., Kaufman, Y. J, Stark, R., & Rundaquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment, 80*(1), 76-87. https://doi.org/10.1016/S0034-4257(01)00289-9

Grigorieva, O., Brovkina, O., & Saidov, A. (2020). An original method for tree species classification using multitemporal multispectral and hyperspectral satellite data. *Silva Fennica, 54*(2), Article 10143. https://doi.org/10.14214/sf.10143

Hogland, J., Anderson, N., Affleck, D. L. R., & St. Peter, J. (2019). Using forest inventory data with Landsat 8 imagery to map longleaf pine forest characteristics in Georgia, USA. *Remote Sensing, 11*(15), 1803. https://doi.org/10.3390/rs11151803

Huete, A., Justice, C., & Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment, 49*(3), 224-234. https://doi.org/10.1016/0034-4257(94)90018-3

Jafarov, E. E., Loudermilk, L. E., Hiers, K. J., Williams, B., Linn, R., Jones, C., Hill, S. C., & Atchley, A. L. (2021). Linking habitat suitability with a longleaf pine-hardwood model: Building a species-predictive fire-land management framework. *Ecological Modelling, 440*, 109387. https://doi.org/10.1016/j.ecolmodel.2020.109387

Jia, W., & Pang, Y. (2023). Tree species classification in an extensive forest area using airborne hyperspectral data under varying light conditions. *Journal of Forestry Research, 34*, 1359–1377. https://doi.org/10.1007/s11676-022-01593-z

Kinane, S. M., Montes, C. R., Albaugh, T. J., & Mishra, D. R. (2021). A model to estimate leaf area index in loblolly pine plantations using Landsat 5 and 7 images. *Remote Sensing, 13*, 1140. https://doi.org/10.3390/rs13061140

Kriegler, F., Malila, W., Nalepka, R., & Richardson, W. (1969). Preprocessing transformations and their effect on multispectral recognition. *Proceedings of the 6th International Symposium on Remote Sensing of Environment.* University of Michigan, 97-131. https://archive.org/details/trent_0116401042951

Kush, J. S., Meldahl, R. S., McMahon, C. K., & Boyer, W. D. (2003). Longleaf pine: A sustainable approach for increasing terrestrial carbon in the southeastern United States. *Environmental Management, 33*(1), S139-S147. https://doi.org/10.1007/s00267-003-9124-3

Kuusk, V., Niinemets, U., & Valladeres, F. (2017). A major trade-off between structural and photosynthetic investments operative across plant and needle ages in three Mediterranean pines. *Tree Physiology, 38,* 543-557. https://doi.org/10.1093/treephys/tpx139

Longleaf Alliance. (n.d.). Southeast longleaf ecosystem occurrences (LEO) geodatabase. https://longleafalliance.org/what-we-do/restoration-through-partnerships/leo/

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aereal photography for documentation of grazing impacts on wheat. *Geocarto International, 16*(1), 65-70. https://doi.org/10.1080/10106040108542184

Means, D.B. (2007). Vertebrate faunal diversity of longleaf pine ecosystems. In S. Jose, E.J. Jokela, D.L. Miller (Eds.), *The longleaf pine ecosystem*. Springer Series on Environmental Management. https://doi.org/10.1007/978-0-387-30687-2_6

Microsoft Open Source, McFarland, M., Emanuele, R., Morris, D., & Augspurger, T. (2022). *microsoft/PlanetaryComputer: October 2022* (Version 2022.10.28) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7261897

Multi-Resolution Land Characteristics Consortium (MRLC). (2024). National Land Cover Database Class Legend and Description. https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description

National Aeronautics and Space Administration (NASA). (n.d.). *Satellites*. Landsat Science. https://landsat.gsfc.nasa.gov/satellites/

Nieminen, M. F. (2014). *Spectral separability of longleaf and loblolly pines in high-resolution satellite data* [Master's thesis, Mississippi State University]. MSU Scholars Junction. https://scholarsjunction.msstate.edu/td/4025

Noss, R. F., LaRoe, E. T. III., & Scott, J. M. (1995). Endangered eco-systems of the United States: A preliminary assessment of loss and degradation. US Department of the Interior, Biological Report, 28.https://www.researchgate.net/publication/246063035

Oswalt, C. M., Cooper, J. A., Brockway, D. G., Brooks, H. W., Walker, J. L., Connor, K. F., Oswalt, S. N., & Conner, R. C. (2012). *History and current condition of longleaf pine in the southern United States.* United States Department of Agriculture, Forest Service, Southern Research Station. https://www.srs.fs.usda.gov/pubs/gtr/gtr_srs166.pdf

Qi, J., Chehbouni, A., Huete, A., Keer, Y. H., & Sorooshian, S. (1994). A modified soil vegetation adjusted index. *Remote Sensing of Environment*, *48,* 119-126. https://doi.org/10.1016/0034-4257(94)90134-1

Roy, P., Rikimaru, A. & Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology, 43*(1), 39-47. https://www.researchgate.net/publication/255718070_Tropical_forest_cover_density_mapping

Sinha, P., Kumar, L., & Reid, N. (2011). Seasonal land use/land cover mapping: accuracy comparison of various band combinations. *Ecosystem Management, University of New England.* https://www.isprs.org/proceedings/2011/isrse-34/211104015Final00280.pdf

Somers, G. L., & Farrar Jr, R. M. (1991). Biomathematical growth equations for natural longleaf pine stands. *Forest Science*, *37*(1), 227-244. https://doi.org/10.1093/forestscience/37.1.227

Stoddard, H.L. (1931). *The bobwhite quail; its habits, preservation, and increase*. Charles Scribner's Sons.

Texan by Nature. (2024). Return on Conservation™ report. *Texanbynature.org.* https://texanbynature.org/rocreport/

Texas A&M Forest Service. (2025). *Trees of Texas: List of trees. http://texastreeid.tamu.edu/content/listOfTrees/*

Texas Department of Transportation (TxDOT). (2025). Texas County Boundaries. TxDOT Open Data Portal. https://gis-txdot.opendata.arcgis.com/maps/9b2eb7d232584572ad53bad41c76b04d

Texas Longleaf Implementation Team. (n.d.). *East Texas Longleaf Prioritization Model* [Fact sheet]. Lower Mississippi Valley Joint Venture. https://static1.squarespace.com/static/5bb3865d2727be6f94acf2fc/t/6244ae23dc3c9c630a0c5633/1648668206884/TLIT_longleaf_prioritization.docx.pdf

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote*

Sensing of Environment, 8(2), 127-150. https://doi.org/10.1016/0034-4257(79)90013-0

U.S. Geological Survey (USGS). (2021). Landsat 4-7 Collection 2 (C2) Level 2 Science Product (L2SP) Guide. https://www.usgs.gov/media/files/landsat-4-7-collection-2-level-2-science-product-guide

U.S. Geological Survey (USGS). (2024a). Annual NLCD Collection 1 Science Products. U.S. Geological Survey data release. https://doi.org/10.5066/P94UXNTS.

U.S. Geological Survey (USGS). (2024b). Landsat 8-9 Collection 2 (C2) Level 2 Science Product (L2SP) Guide. https://www.usgs.gov/media/files/landsat-8-9-collection-2-level-2-science-product-guide

U.S. Geological Survey. (n.d.). EarthExplorer. U.S. Department of the Interior. Retrieved October 1, 2025, from https://earthexplorer.usgs.gov/

Wilson, E. H. & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, *80*, 385-396. https://doi.org/10.1016/S0034-4257(01)00318-2

Zhang, D., & Polyakov, M. (2010). The geographical distribution of plantation forests and land resources potentially available for pine plantations in the US South. B*iomass and Bioenergy*, *34*(12), 1643-1654. https://doi.org/10.1016/j.biombioe.2010.05.006

8. Appendices

Appendix A: *Datasets*

Table A1. List of remote sensed data sources

Data Product Title	Spatial Resolution	Temporal Resolution	Temporal Range	Imagery Dates
Landsat 9 OLI-2 Collection 2 Level-2 Surface Reflectance	30 x 30 m	8 days	27 September 2021 - Present	21 December 2024
Landsat 8 OLI Collection 2 Level-2 Surface Reflectance	30 x 30 m	8 days	11 February 2013 - Present	27 January 2015
Landsat 5 TM Collection 2 Level-2 Surface Reflectance	30 x 30 m	16 days	1 March 1984 – November 2011	14 December 2004 19 December 1994 27 January 1986
Annual NLCD	30 x 30 m	Annual	1985 - Present	1985, 1995, 2005, 2015, 2024

Table A2. List of validation and other data sources

Data Product Title	Data Product Source	Data Product Type and Brief Description
Longleaf Ecosystem Occurrences geodatabase	Longleaf Alliance in partnership with Florida Natural Areas Inventory	Polygons and attribute table for southern pine forest compiled from previous southern pine studies
Texas Longleaf Team restoration data	Texas Longleaf Team	Polygons and attribute table for multiple years of longleaf restoration projects in eastern Texas

Appendix B: *Distribution and Separability Metrics*

Table B1. Predicted longleaf and loblolly stand acreage by year

Year	1985	1995	2005	2015	2024
Longleaf	460,045	850,149	1,071,764	1,244,098	1,315,942
Longleaf percent change	n/a	46%	21%	14%	5%
Loblolly	2,772,947	2,424,294	2,119,395	2,059,399	2,057,373
Loblolly percent change	n/a	-13%	-13%	-3%	<1%

Table B2. Mean and Standard Deviation values for purified training points

Variable	Longleaf Mean	Longleaf SD	Loblolly Mean	Loblolly SD	Difference of Means	Pooled SD	Effect Size
NDMI	0.39	0.10	0.25	0.09	0.14	0.10	1.47
BSI	-0.32	0.10	-0.19	0.10	-0.13	0.10	-1.30
MSAVI2	0.84	0.08	0.77	0.06	0.07	0.07	0.99
NDVI	0.80	0.08	0.75	0.08	0.05	0.08	0.63
Band 5	0.21	0.07	0.17	0.06	0.04	0.07	0.61
Band 3	0.30	0.87	0.03	0.04	0.27	0.62	0.44
Band 2	0.13	0.84	0.01	0.08	0.12	0.60	0.20
Band 7	0.04	0.05	0.05	0.06	-0.01	0.06	-0.18
Band 6	0.09	0.09	0.10	0.06	-0.01	0.08	-0.13
Band 4	0.02	0.09	0.02	0.05	0.00	0.07	0.00

Appendix C: *Evaluations of Model Accuracy and Decision Distribution*

Table C1: Confusion Matrix of highly differentiable inputs model, points group B validates prediction A

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	76	5	81	0.06
Loblolly	12	90	102	0.12
Total	88	95	183	
Error of Omission	0.14	0.05		0.09

Overall Kappa Index of Agreement = 0.81

Table C2: Confusion Matrix of highly differentiable inputs model, points group A validates prediction B

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	80	8	88	0.09
Loblolly	17	80	97	0.18
Total	97	88	185	
Error of Omission	0.18	0.09		0.14

Overall Kappa Index of Agreement = 0.73

Table C3: Confusion Matrix of all inputs model, points group B validates prediction A

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	74	5	79	0.06
Loblolly	14	90	104	0.13
Total	88	95	183	
Error of Omission	0.16	0.05		0.10

Overall Kappa Index of Agreement = 0.79

Table C4: Confusion Matrix of all inputs model, points group A validates prediction B

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	81	9	90	0.10
Loblolly	16	79	95	0.17
Total	97	88	185	
Error of Omission	0.17	0.10		0.14

Overall Kappa Index of Agreement = 0.73

Table C5: Confusion Matrix of bands only model, points group B validates prediction A

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	75	8	83	0.10
Loblolly	13	87	100	0.13
Total	88	95	183	
Error of Omission	0.15	0.08		0.12

Overall Kappa Index of Agreement = 0.77

Table C6: Confusion Matrix of bands only model, points group A validates prediction B

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	79	8	87	0.09
Loblolly	18	80	98	0.18
Total	97	88	185	
Error of Omission	0.19	0.09		0.14

Overall Kappa Index of Agreement = 0.72

Table C7: Input Variables' Explanatory Power in the Decision Forest Model (Purified Points)

Variable	**OOB Accuracy**	**Accuracy Drop**
NDMI	83.69	0.62
MSAVI2	83.77	0.54
b4	84.25	0.06
b5	84.33	-0.02
b6	84.55	-0.24
b7	84.61	-0.3
BSI	84.76	-0.45
NDVI	84.78	-0.47
b2	84.9	-0.59
b3	85.11	-0.8
All variables	84.31	N/A

Table C8: Confusion Matrix of Variables with a positive accuracy drop (NDMI, MSAVI2, Band 4)

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	76	7	83	0.08
Loblolly	12	88	100	0.12
Total	88	95	183	
Error of Omission	0.14	0.07		0.10

Overall Kappa Index of Agreement = 0.79

Table C9: RGB-only indices used to isolate importance of NIR band

Index	Equation	Citation
Blue-Green-Red Vegetation Index	$RGBVI = \frac{(Green*Green)-(Red*Blue)}{(Green*Green)+(Red \cdot Blue)}$	Bendig et al., 2015
Normalized Green-Red Difference Index	$NGRDI = \frac{(Green-Red)}{(Green+Red)}$	Tucker, 1979
Visible Atmospherically Resistant Index	$VARI = \frac{Green-Red}{Green+Red-Blue}$	Gitelson et al., 2002
Green Leaf Index	$GLI = \frac{(2*Green-Red-Blue)}{(2*Green+Red+Blue)}$	Louhaichi et al., 2001

Table C10: Confusion Matrix of RGB indices only model, group B validates prediction A

	Longleaf	Loblolly	Total	Error of Commission
Longleaf	67	25	92	0.27
Loblolly	21	70	91	0.23
Total	88	95	183	
Error of Omission	0.24	0.26		0.25

Overall Kappa Index of Agreement = 0.50

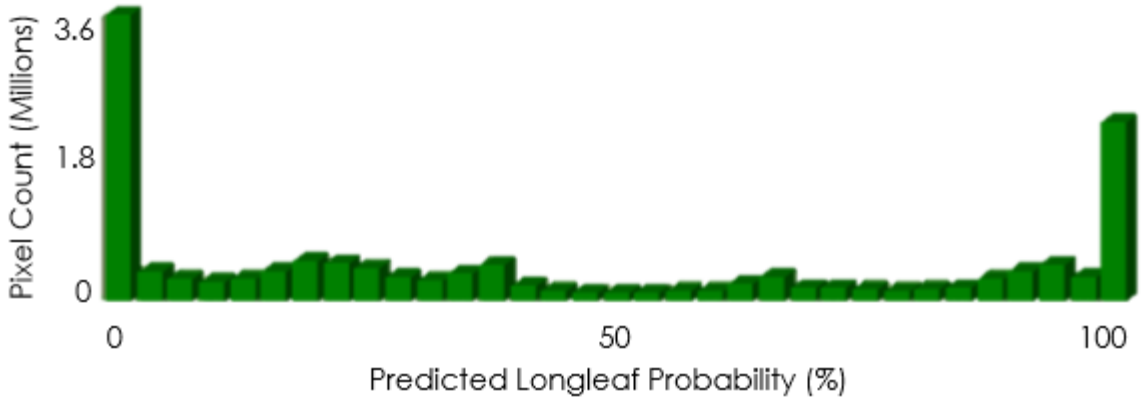


Figure C1. Summary statistics for pixel values of the 2024 longleaf probability surface, highly differentiable inputs, from ErrMat TerrSet Module. Values range from 0% confidence to 100% confidence. The mean is 0.43 and the standard deviation is 0.34.

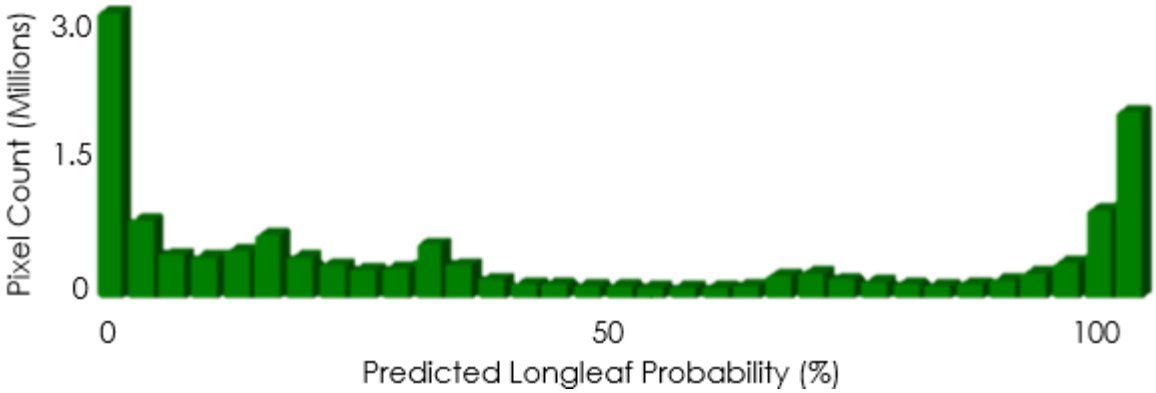


Figure C2. Summary statistics for pixel values of the 2024 longleaf probability surface, all inputs (spectral bands and indices), from ErrMat TerrSet Module. Values range from 0% confidence to 100% confidence. The mean is 0.43 and the standard deviation is 0.37.

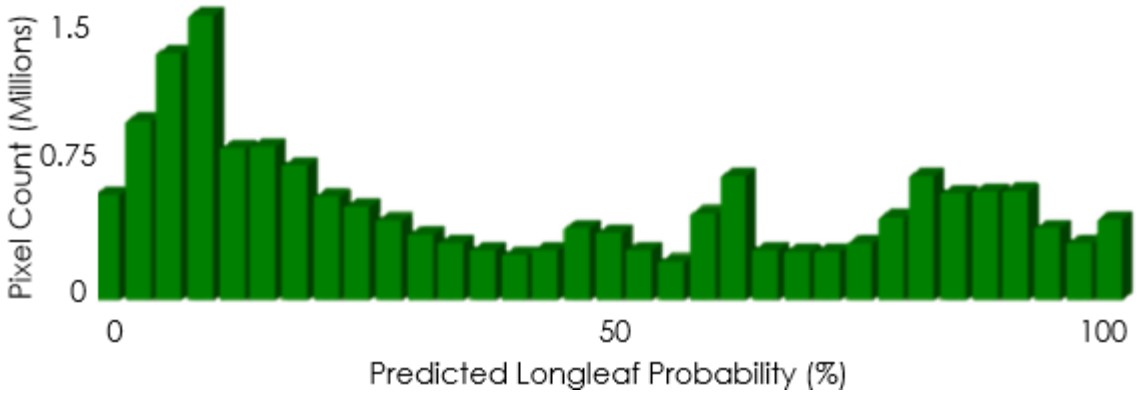


Figure C3. Summary statistics for pixel values of the 2024 longleaf probability surface, spectral bands only, from ErrMat TerrSet Module. Values range from 0% confidence to 100% confidence. The mean is 0.42 and the standard deviation is 0.31.