Forest Stand Classification in the Southeastern United States

Duke MIDS Master in Interdisciplinary Data Science
Duke Statistical Science

Nathan Bush, Nruta Choudhari, Nakiyah Dhariwala, Yixiao Wang

THE CONSERVATION FUND

2026 Duke MIDS Capstone Symposium

Motivation

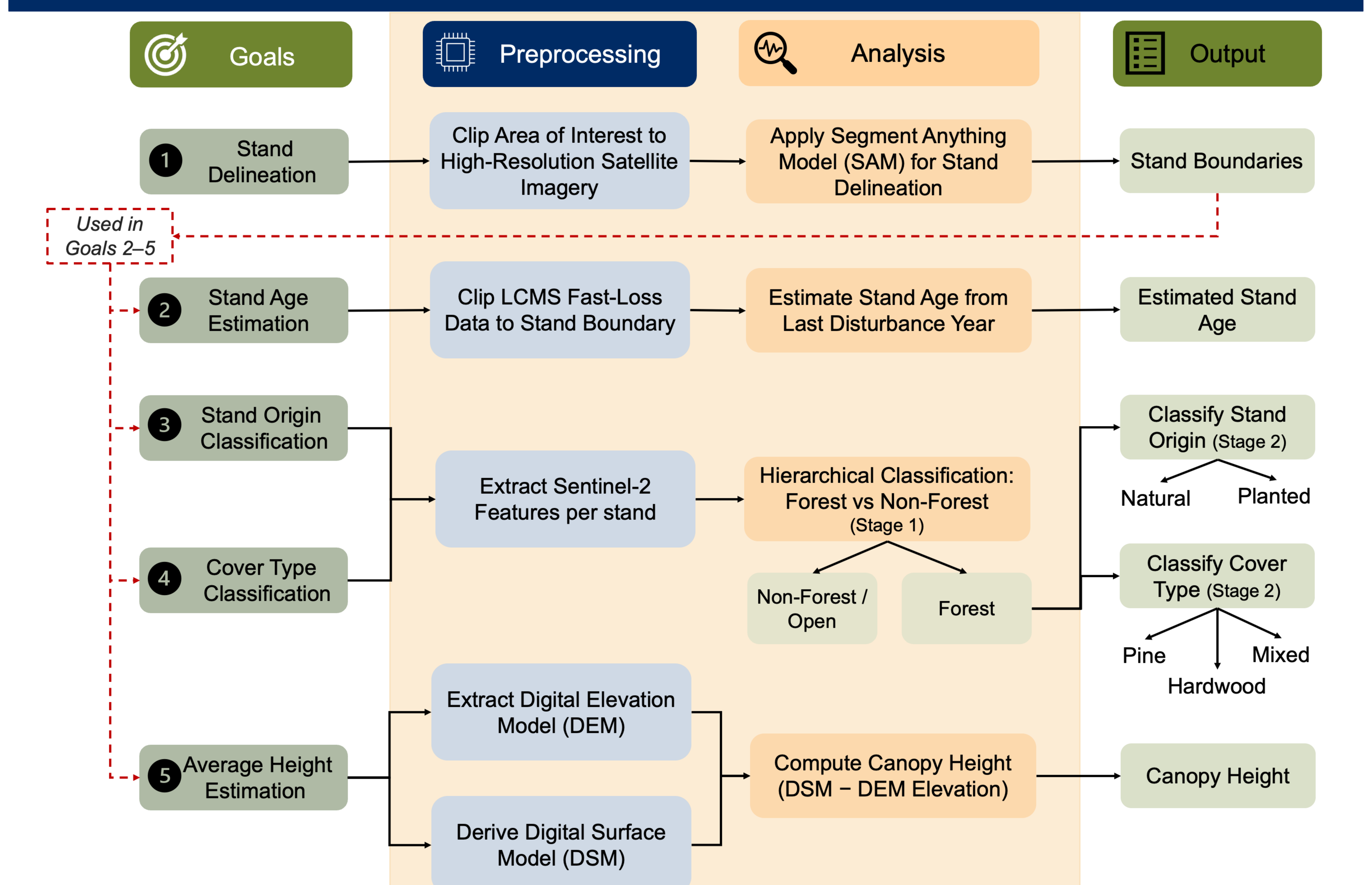
Estimating forest biomass is essential but traditionally slow and costly. Remote sensing enables scalable estimation, but it depends on understanding forest structure.

This capstone project incorporates the use of AI and remote sensing data to classify forest stands and deliver a Python package for scalable analysis, answer key questions:

- Which stands are pine plantations?
- Which stands have natural forests?
- What is the age of plantation and natural stands?

These outputs enable faster, more scalable biomass estimation for forest management.

Forest Stand Analysis Pipeline



Client-Provided Stand Boundaries

External Data & Tools

1. **ESRI World Imagery (Basemap):** High-resolution satellite and aerial imagery for spatial context and initial stand visualization
2. **Segment Anything Model (SAM):** Foundation segmentation model used for automated stand boundary extraction from imagery
3. **Segment Anything Model 2 (SAM 2):** Enhanced segmentation model with improved generalization and temporal consistency
4. **LCMS Fast Loss Dataset:** Used to detect recent vegetation loss and disturbance patterns.
5. **Sentinel-2 Multispectral imagery:** Used to derive vegetation indices and seasonal features
6. **LiDAR:** Elevation and canopy structure data used for vertical forest characterization

Goal 1: Segmenting Forest Stands

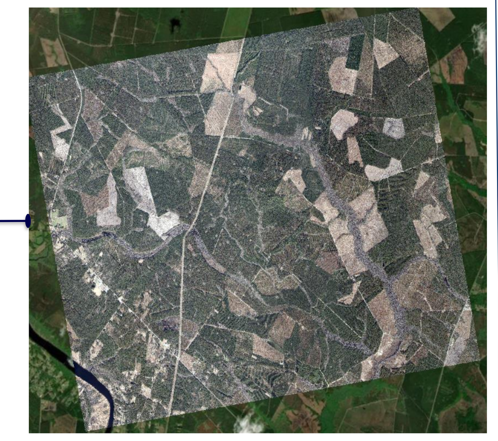
Segmentation Pipeline

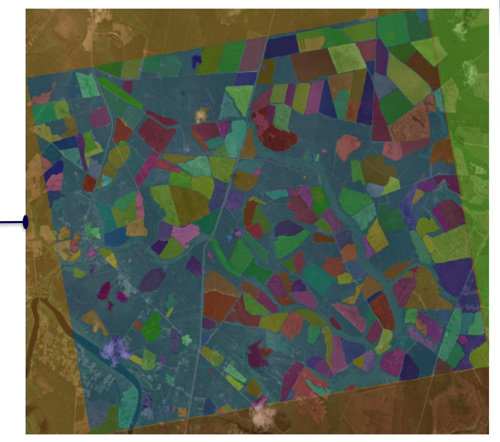
Bounding Box AOI

Download Clipped High-Res Imagery

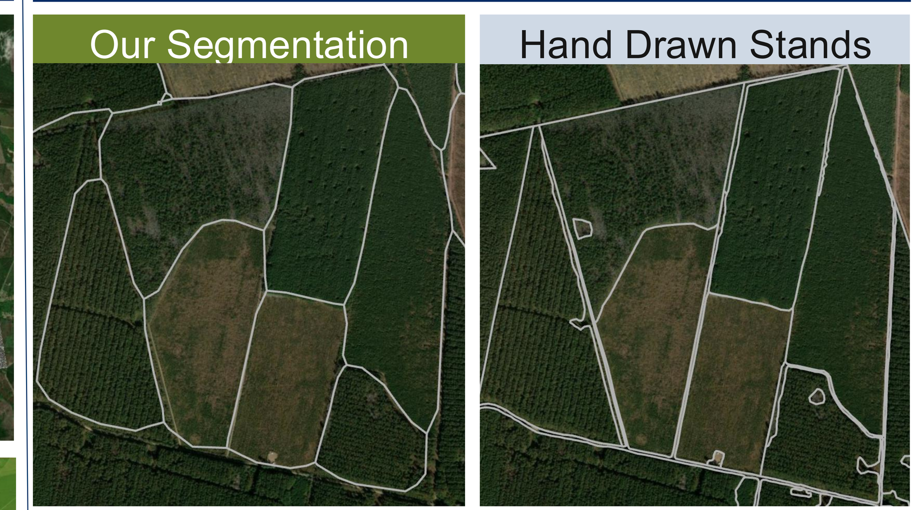
Segment Stands Using Segment Anything Model (SAM)

Output Cleaned Segmented Stand Regions





Compare/Improve Segmentation



Quantitative Evaluation with IoU

- Match predicted and hand drawn regions using Intersection over Union (IoU) with the Hungarian algorithm (optimal one-to-one matching). We then weight each pair by the hand delineated stands area to ensure large stands matter more

SAM Hyperparameter Tuning

- Optimize for IoU while maintaining manageable runtime and VRAM usage by tuning sampling density, quality thresholds, cropping depth, minimum region size, and other key parameters

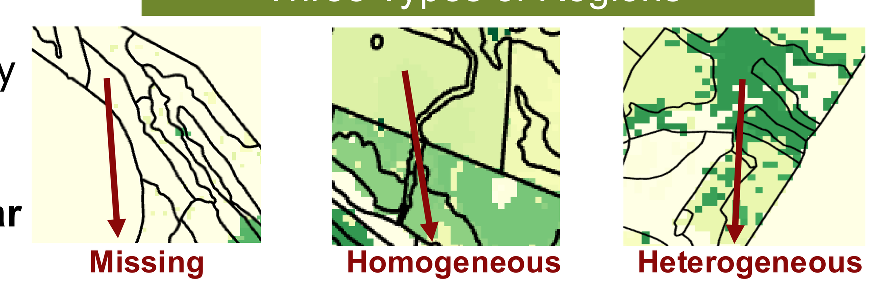
Goal 2: Predicting Stand Age

Starting Point

Most Recent Year of Fast Loss: an initial proxy for stand age derived from remote sensing data

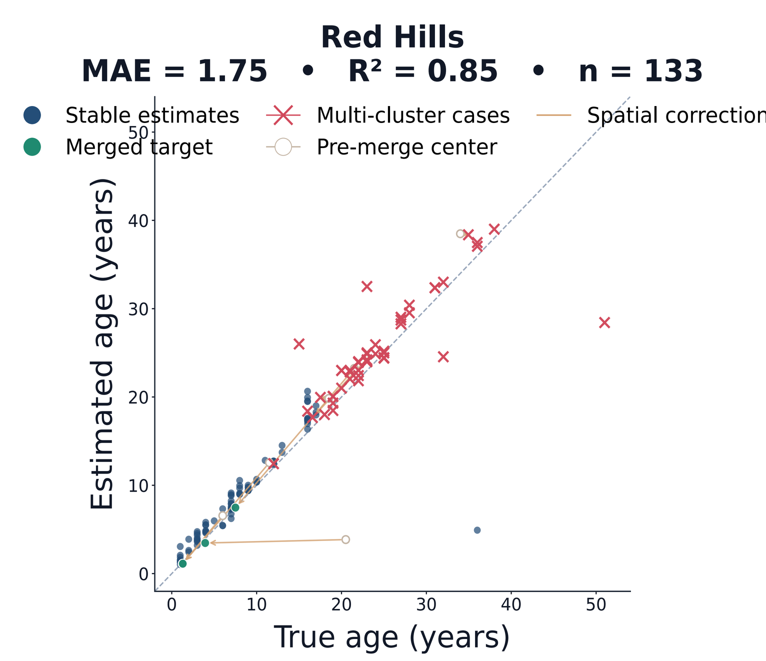
Estimated age = Current year – Fast-loss year

Three Types of Regions



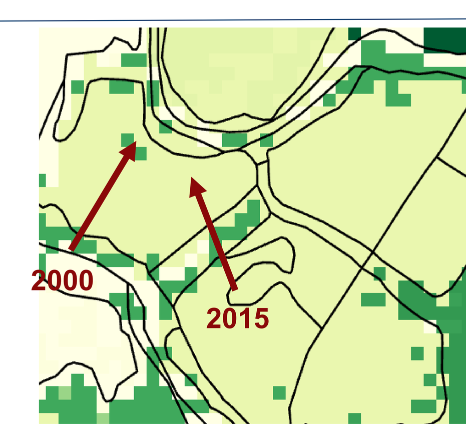
Refinement Strategies

1. **Multi-scale clustering:** identify multiple temporal clusters within each stand to capture diverse disturbance patterns
2. **Temporal correction:** correct invalid age estimates by re-estimating age using valid historical signals constrained by the current year
3. **Spatial merging:** improve spatial consistency by merging small stands into neighboring large regions



Why is Thinning Challenging?

- Thinning: partial removal of trees
- Clear-Cut: complete removal



Key Insights

- Fast-loss year is a strong but imperfect proxy for stand age
- Multi-cluster patterns reveal heterogeneous disturbance histories
- Spatial + temporal refinement significantly improves consistency and accuracy

Goal 3 and 4: Predicting Stand Origin and Cover Type

Remote sensing features derived from Sentinel-2 imagery were used to capture spectral, structural, textural and seasonal measures across the forest stands.

Feature	What It Measures (Per Stand)	Forest Interpretation
Normalized Difference Vegetation Index (NDVI) Mean	Average vegetation greenness	High values – dense, healthy canopy
Near Infrared (NIR) Mean	Reflectance in near-infrared band	High values – strong vegetation
Gray Level Co-Currence Matrix (GLCM) Entropy	Texture complexity	High values – diverse forests
NDVI Seasonal Change	NDVI change across seasons	Differences in canopy behavior
Coefficient of Variation (CV)	Relative variability of NDVI	High values – heterogenous structure

A hierarchical model first separates forest vs non-forest, then predicts origin (natural/planted) and cover type (pine/hardwood/mixed).

Key findings and challenges include:

- Classifying forest and non-forest stands was the main bottleneck
- Performance improved after isolating the forested stands
- Pine was the easiest to classify as it is more abundant in the southeastern US
- Mixed and hardwood cover type was frequently confused
- Logistic regression models performed the best overall

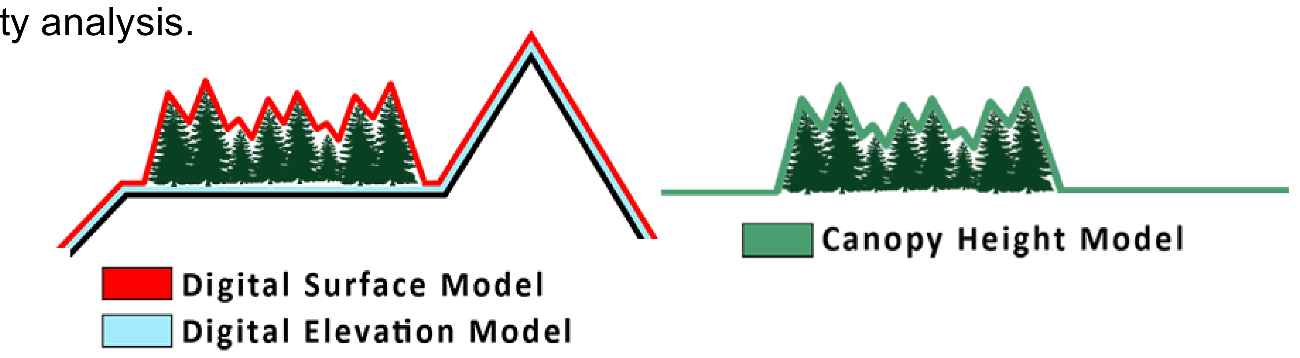
Task	Macro F1	Balanced Accuracy
Origin (Natural VS Planted)	0.76	0.82
Cover Type (Pine VS Hardwood VS Mixed)	0.58	0.71

Goal 5: Tree Height Estimation

LiDAR-derived Digital Elevation Models (DEM) and Digital Surface Models (DSM) were extracted at multiple resolutions and aligned to stand boundaries. A Canopy Height Model (CHM) was computed as:

$$CHM = DSM - DEM$$

The CHM helps capture the vegetation height above ground. It helps enable the scalable estimation of stand-level height, supporting downstream tasks such as biomass estimation and forest productivity analysis.



Acknowledgements

We would like to thank our mentor, Dr. Kyle Bradbury, and our partners, Pete McNeary and Kevin Harnish from The Conservation Fund for their continuous guidance, feedback, and support throughout the duration of this project.

References

1. ESRI World Imagery: https://tiledbasemaps.arcgis.com/arcgis/rest/services/World_Imagery/Map
2. Intermediate Remote Sensing: https://gsp.humboldt.edu/olm/Courses/GSP_326/Labs/2023/lab7.html
3. Segment Anything Model (SAM) - Meta AI: https://segment-anything.metademolab.com/
4. Segment Anything Model 2 (SAM 2) - Meta AI: https://ai.meta.com/research/sam2/
5. LCMS (Land Change Monitoring System) - U.S. Forest Service: https://www.usgs.gov/centers/eros/science/usfs-landscape-change-monitoring-system-lcms
6. Sentinel-2 Mission - European Space Agency (ESA): https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2
7. LiDAR (3DEP) - U.S. Geological Survey (USGS): https://www.usgs.gov/3d-elevation-program