Continuing Project Report

Multi-regional evaluation of new machine learning algorithms for mapping tree species distribution and abundance

CAFS.20.79

Kasey Legaard[1], Aaron Weiskittel[1], Ken Bundy[1], Erin Simons-Legaard[2]

[1] University of Maine, Center for Research on Sustainable Forests
[2] University of Maine, School of Forest Resources

Presenter: Kasey Legaard

Justification

- Remote sensing has high potential to meet spatial information needs of forest management
- But remote sensing maps obtained from empirical models trained against field plot data are often biased, leading to systematic error that degrades map value
- More work should be done to establish algorithms and workflows that reduce systematic error and produce better maps at lower cost

Objectives

High-quality, low-cost spatial data to support forest management

- 10 m species relative abundance, biomass, forest type, disturbance
- Multi-objective machine learning to reduce systematic prediction error
- Algorithms, workflows integrated into high-volume production software
- Test algorithms and workflows in multiple regions

Objectives

Maine High Resolution Land Cover Project

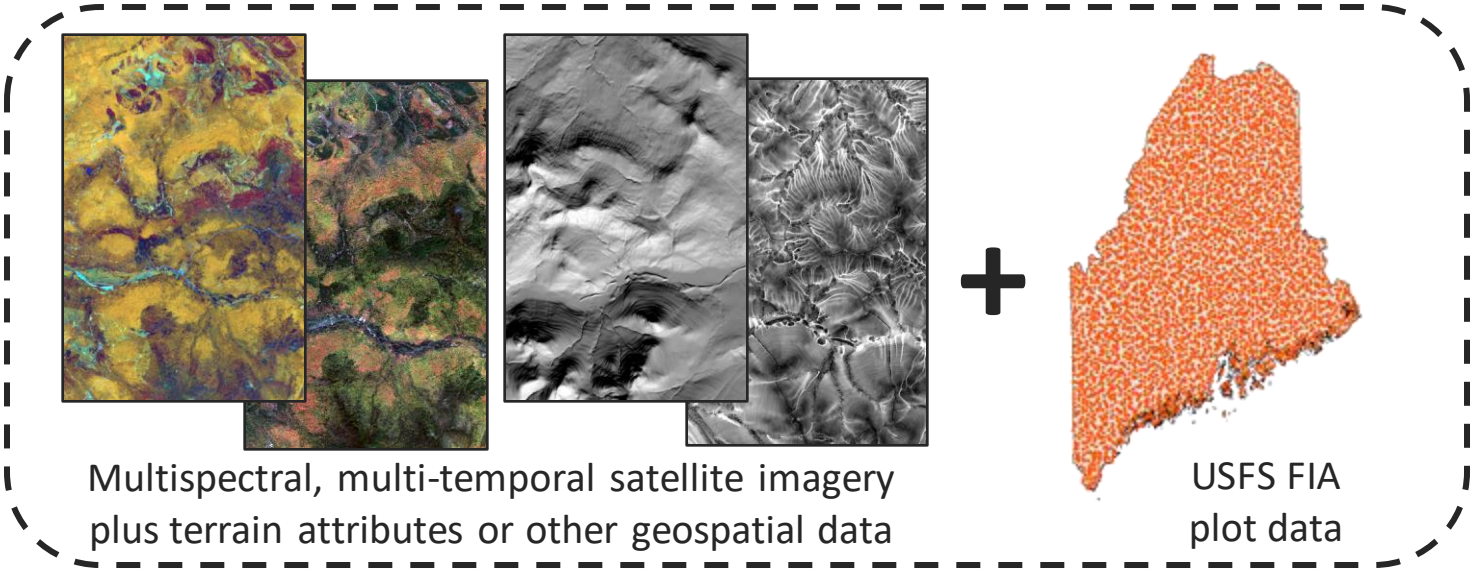
Multi-resolution land cover and forest type data for the State of Maine

- 1 m resolution land cover map – NOAA C-CAP
- 10 m resolution land cover and forest type map – NOAA C-CAP and UMaine
- 10 m resolution forest carbon map - UMaine

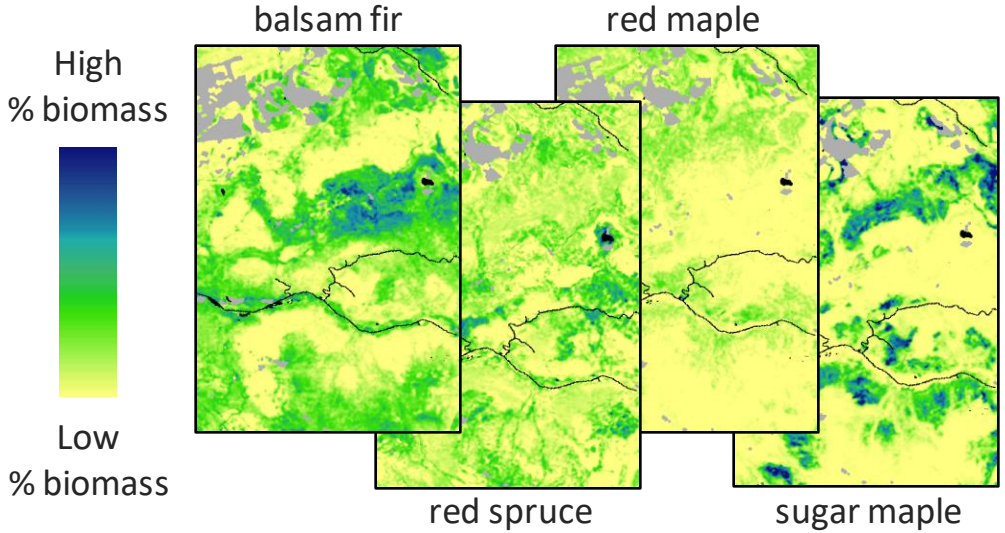
Baxter State Park
Maine Bureau of Parks and Lands
Maine Department of Environmental Protection
Maine Department of Transportation
Maine Library of Geographic Information
Maine Natural Areas Program
NOAA Office for Coastal Management
The Nature Conservancy
UMaine Advanced Computing Group
UMaine Center for Research on Sustainable Forests, Intelligent GeoSolutions group
UMaine Cooperative Forestry Research Unit
UMaine Wheatland Geospatial Lab
USFS NRS FIA Program

Species mapping:

Methods

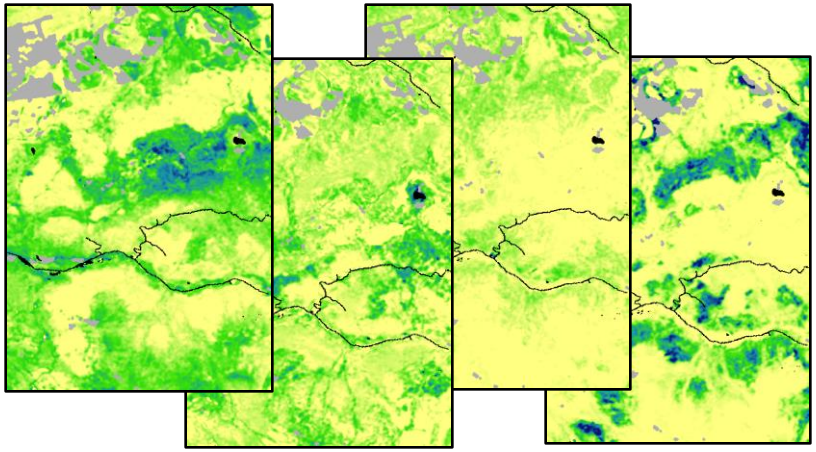


Predictions of species abundance (% of live biomass)



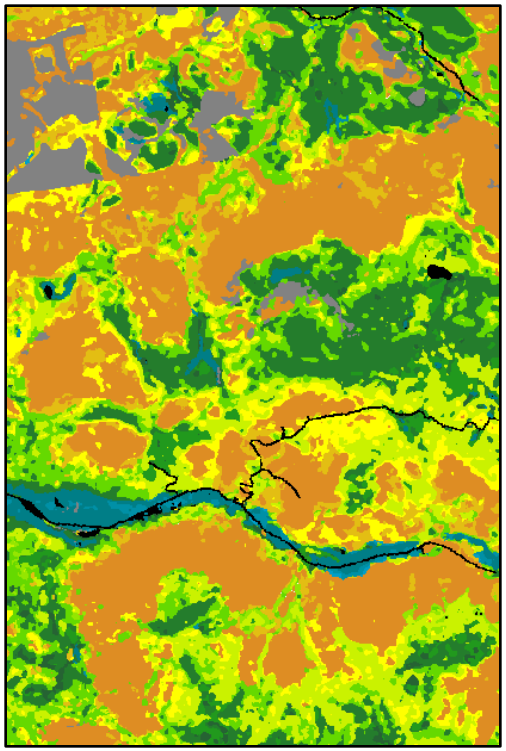
Methods

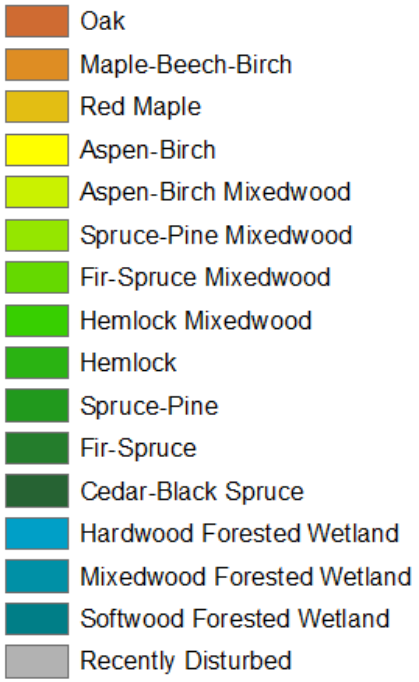
Predictions of tree species abundance (% of live biomass)



- Paper birch
- Poplar species
- Oak species
- Sugar maple
- White ash
- Yellow birch
- American beech
- Red maple
- Eastern hemlock
- Balsam fir
- Pine species
- Red spruce
- White spruce
- Black spruce
- Northern white cedar
- Softwood species

Forest type classification derived from individual species models

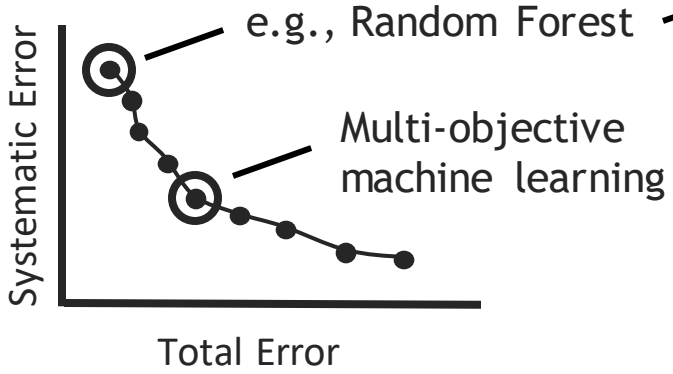


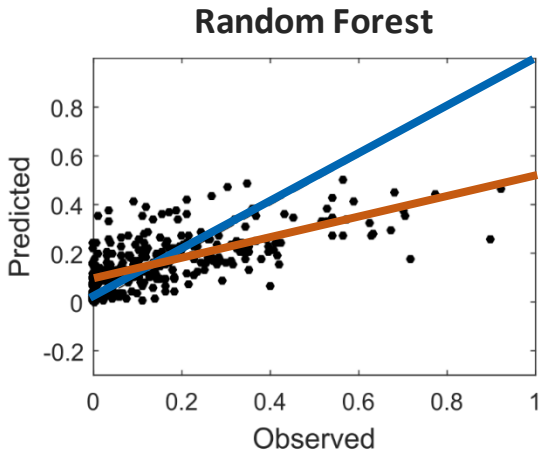


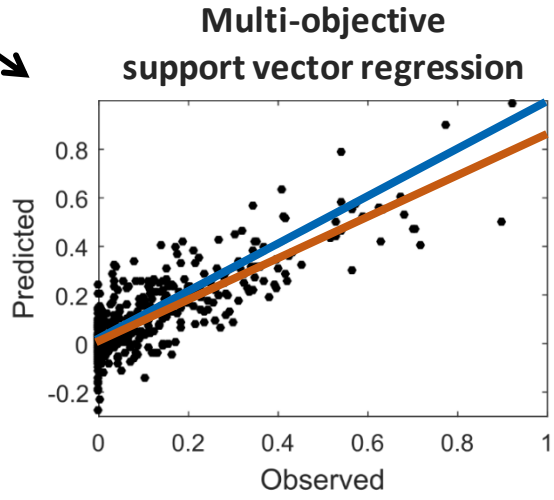
Multi-objective machine learning:

Methods

Location and measurement uncertainty introduces a tradeoff between total and systematic prediction error:







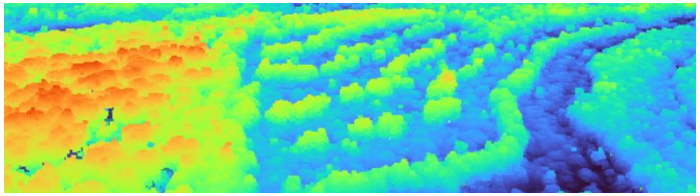
Minimization of total error causes attenuation bias in regression models and systematic error in maps.

Multi-objective machine learning minimizes both total and systematic error.

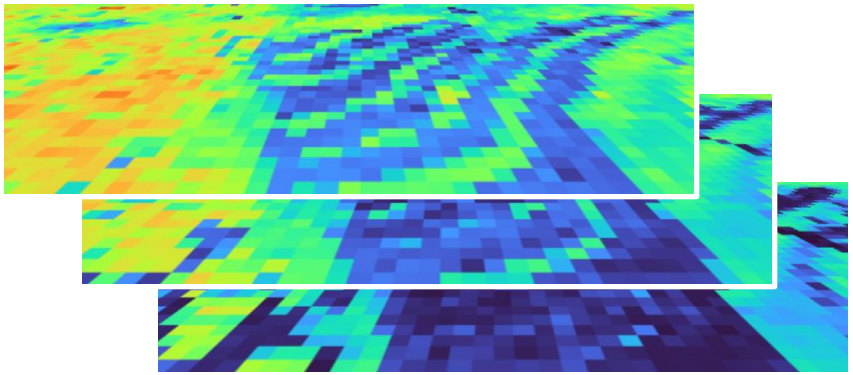
Statewide biomass mapping from 2021 NAIP

Methods

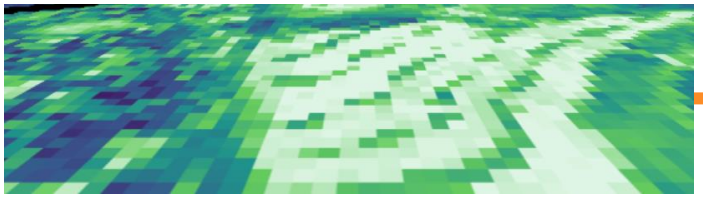
2021 NAIP DSM:

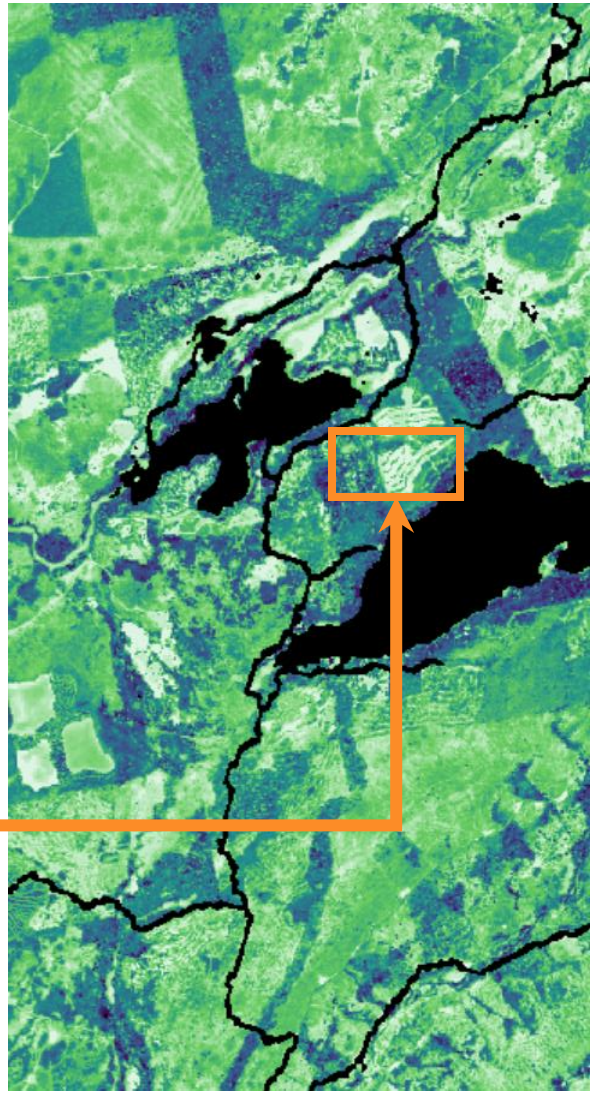


Height metrics computed over 10-meter grid:



Biomass from ML models trained at FIA plots:



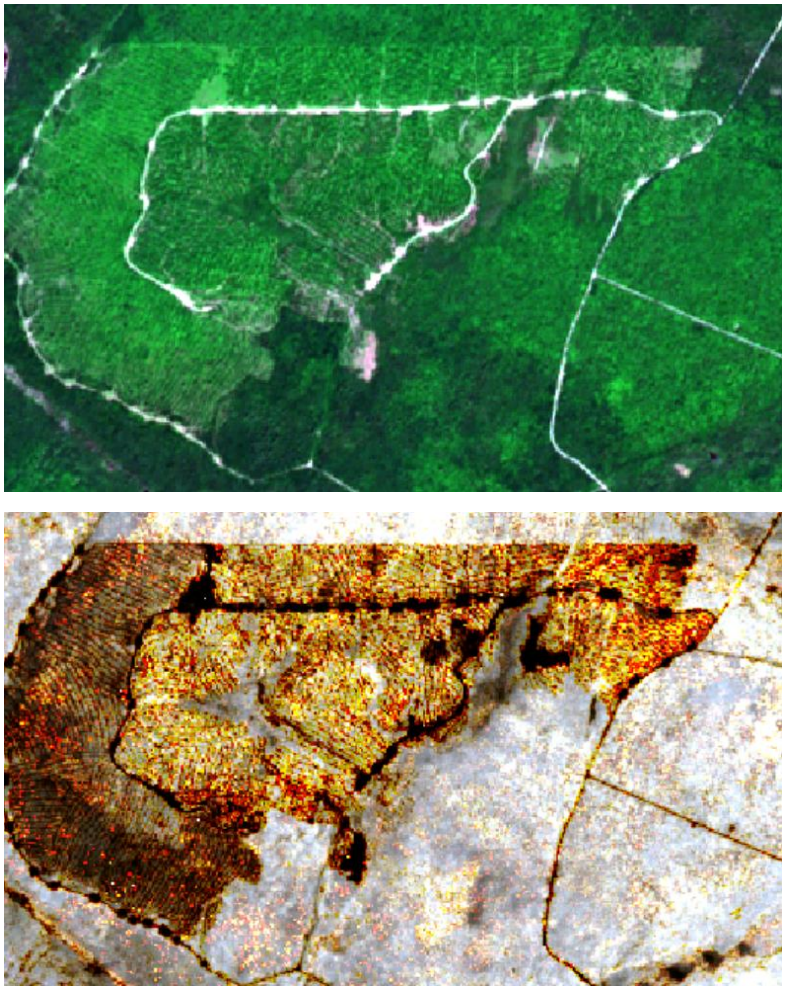


Major Findings

Large-area computation of gridded height metrics from LiDAR or photogrammetric point clouds on HPC or on the cloud

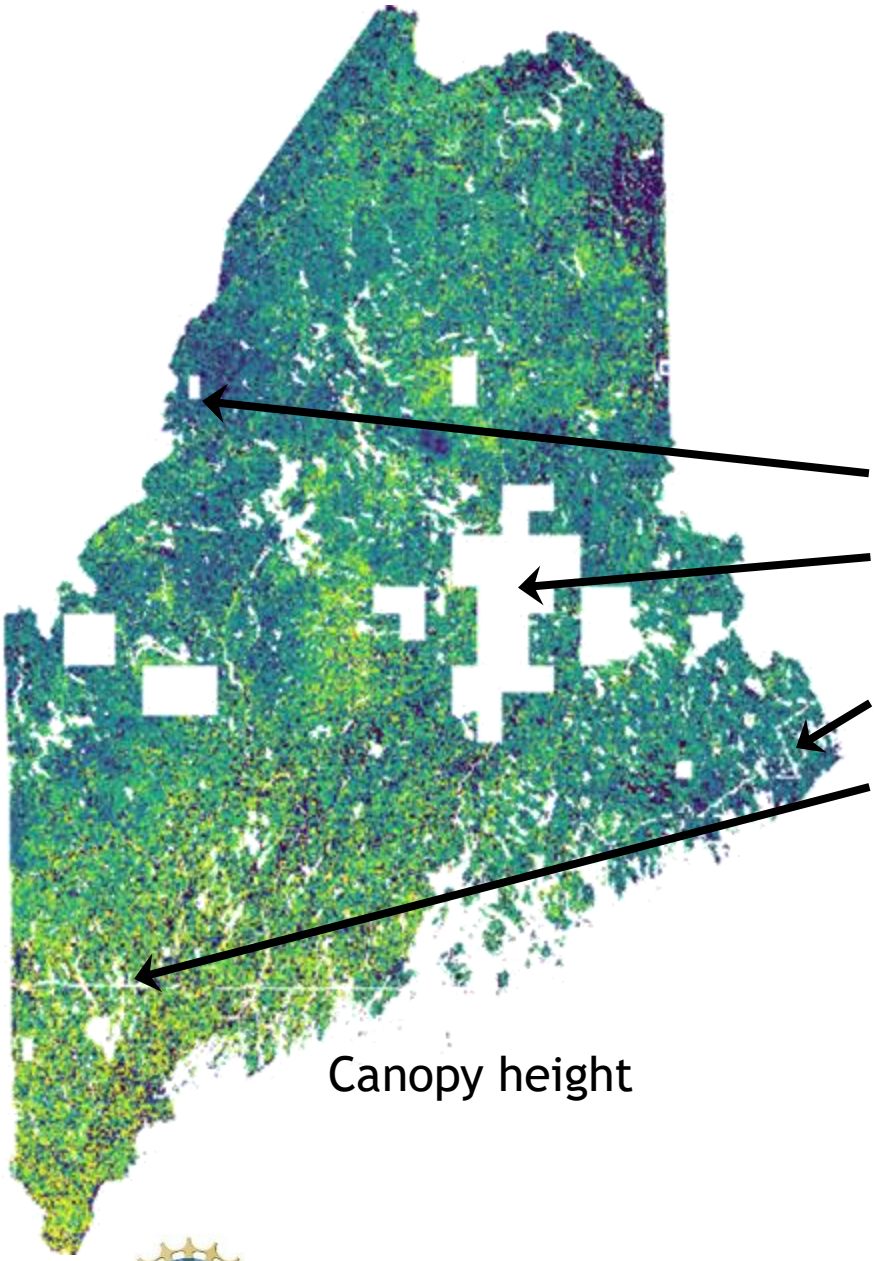
- Points assigned to predefined grid on load
- Subsequent computations are not spatial; independent jobs and unlimited scaling
- Automated job scheduling and tracking
- Computations in C++ and Rust, with job orchestration in Python
- Optimized IO and data handling
- Built-in, optional outlier fences

Fast and scalable: capable of statewide NAIP processing on the cloud in under an hour



2021 NAIP, 10-meter gridded height percentiles: RGB = 95th, 50th, 5th

Major Findings



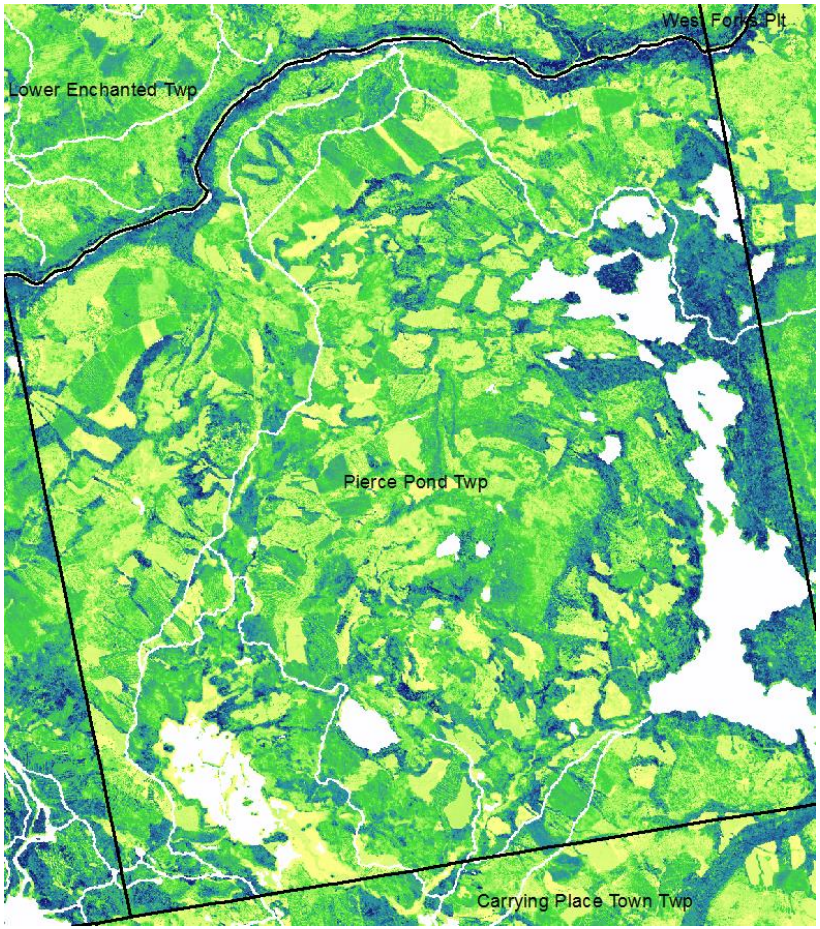
Canopy height

Remaining challenges in statewide NAIP processing pertain to data access and quality, not data processing

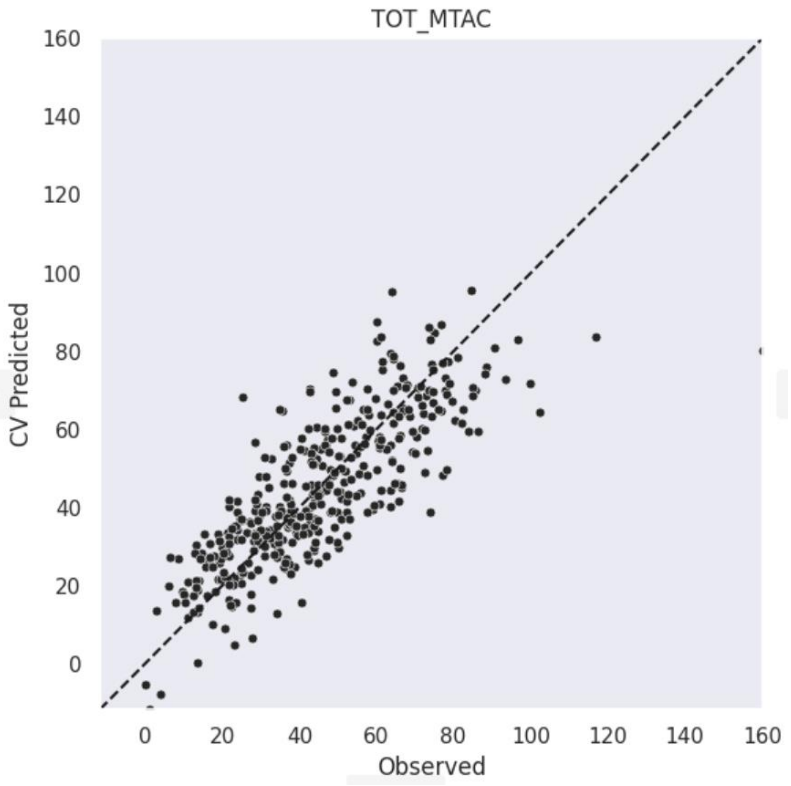
- Corrupted NAIP tiles
- Tribal data sovereignty restricts access to LiDAR-derived DEM tiles
- Gaps in LiDAR-derived elevation data
- Gaps between NAIP flight lines

NAIP DSMs distributed only to those organizations that contributed to original buy-ups

Major Findings



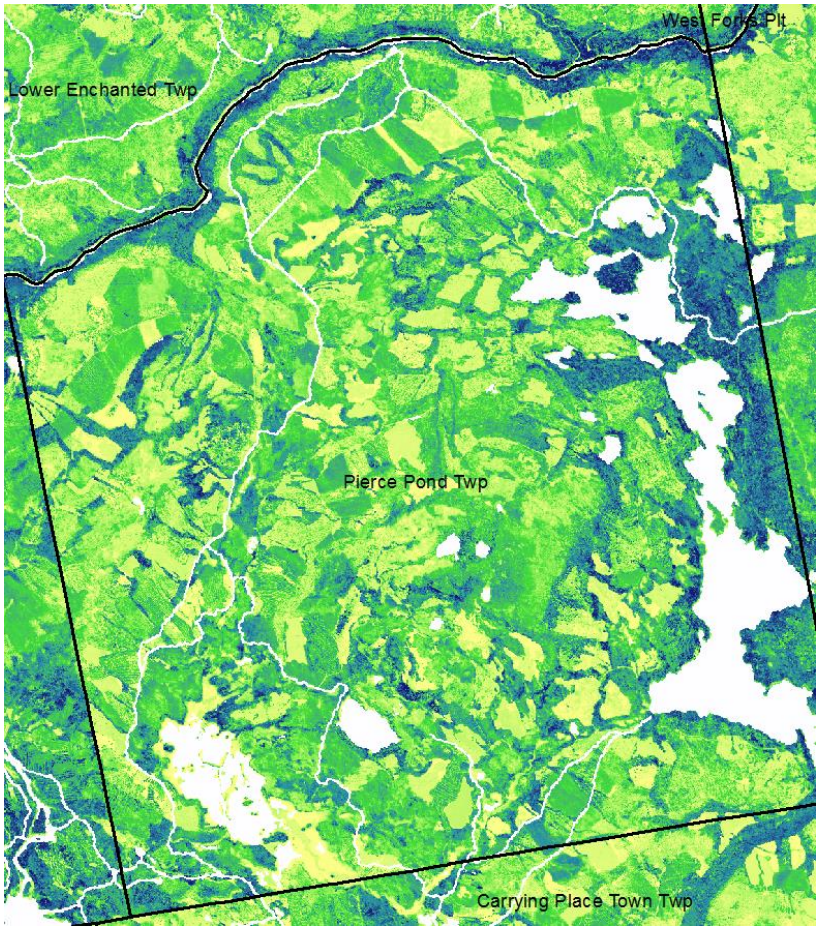
Biomass from trial of 2021 NAIP + Sentinel, using a multi-objective SVM



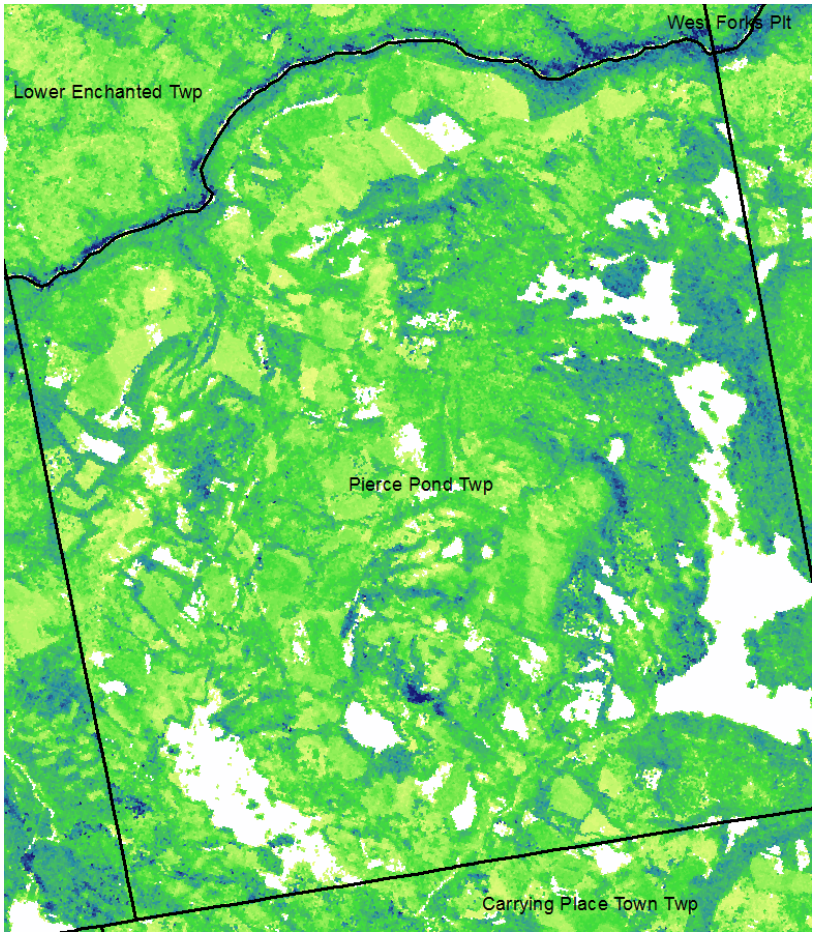
RMSE: 12.5 metric tons/ac

MAE: 9.7 metric tons/ac

Major Findings



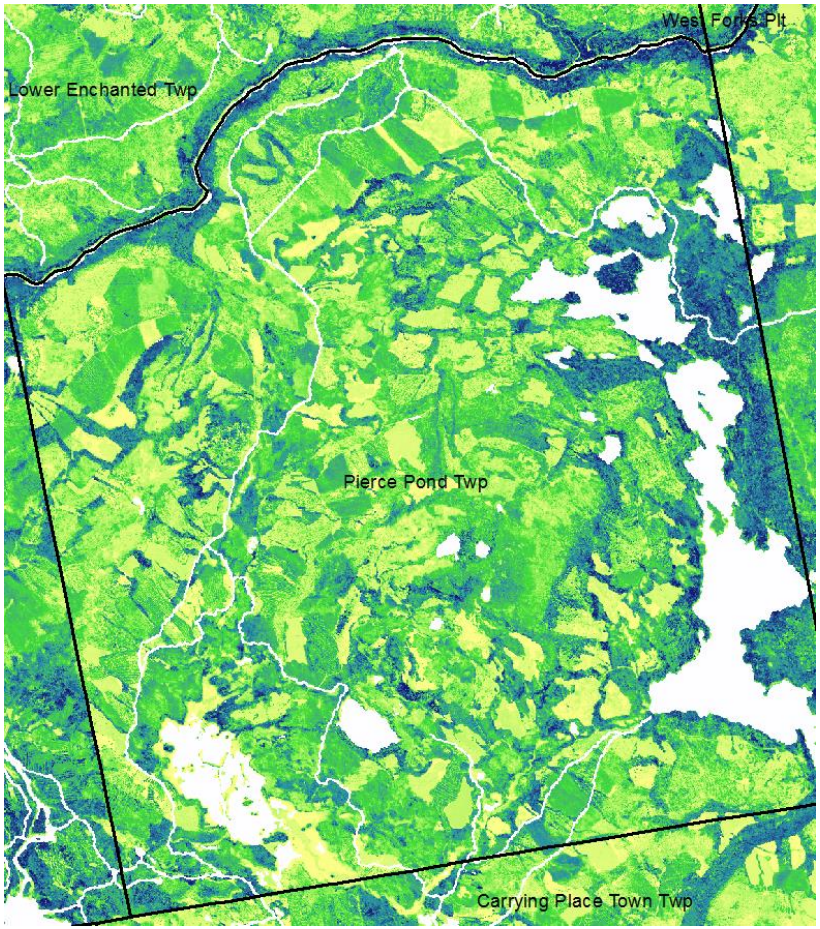
Biomass from trial of 2021 NAIP + Sentinel, using a multi-objective SVM



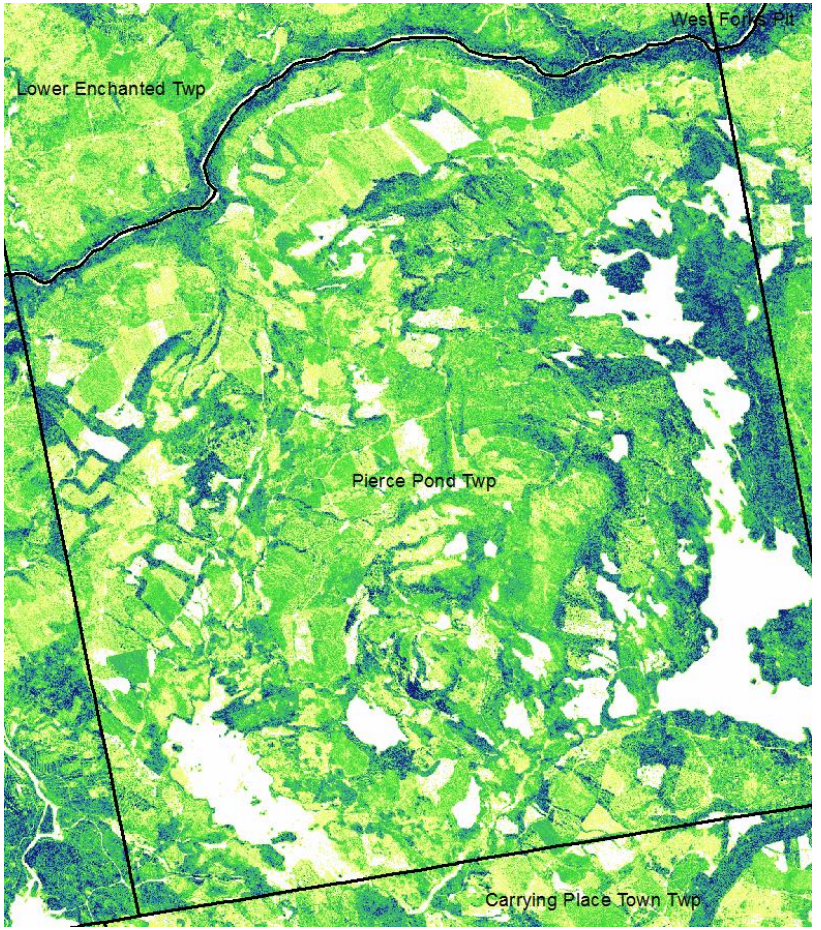
Tang et al. 2021: 3DEP LiDAR from 2015(?), using a Random Forest model

Tang, H. et al. 2021. https://doi.org/10.3334/ORNLDAAC/1854

Major Findings



Biomass from trial of 2021 NAIP + Sentinel, using a multi-objective SVM

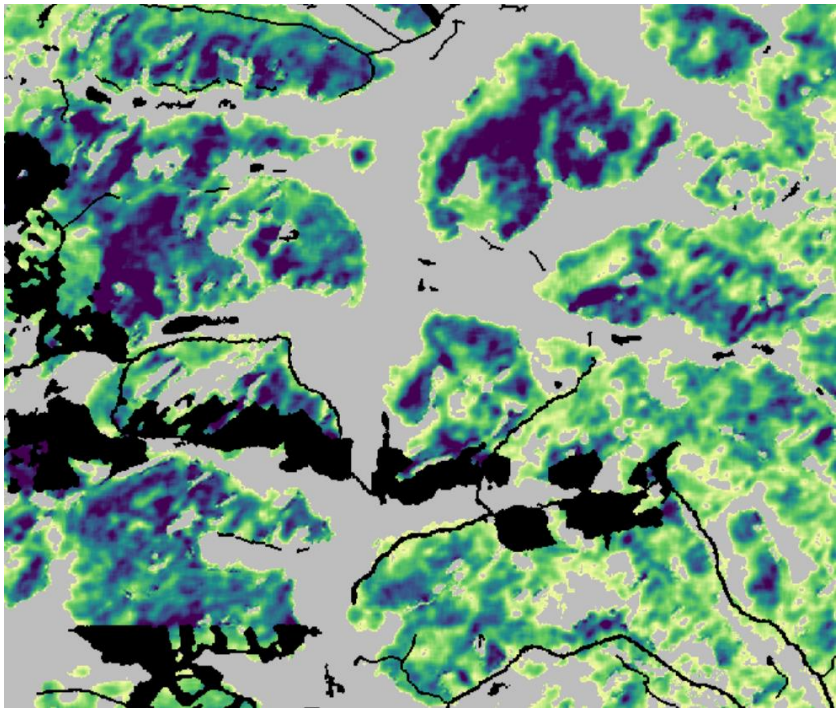


Ayrey et al. 2021: 3DEP LiDAR from 2015(?), using a 3D CNN model

Ayrey et al. 2021. Remote Sens. 2021, 13(24)

Major Findings

Sugar maple



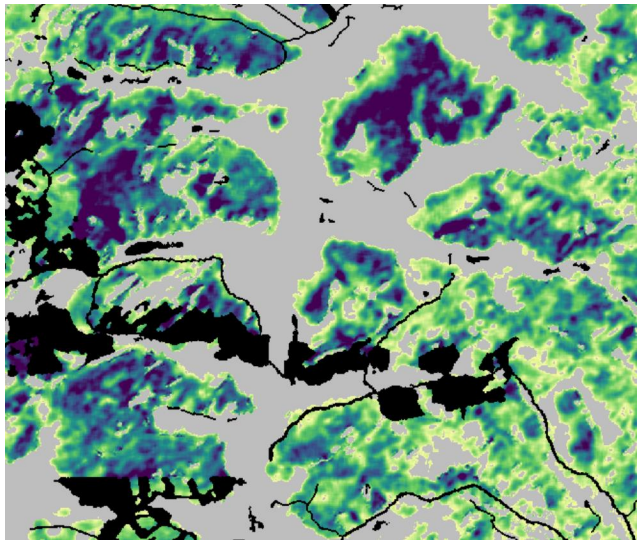
Species workflows are essentially finished

Currently evaluating our forest typing workflow and results

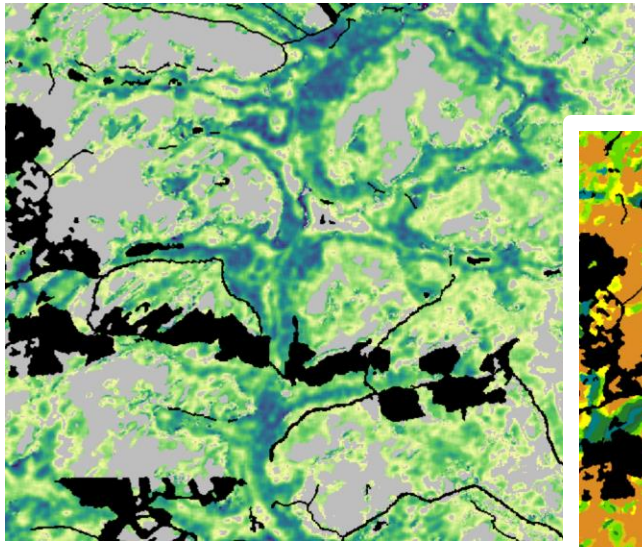
Center for Advanced Forestry Systems

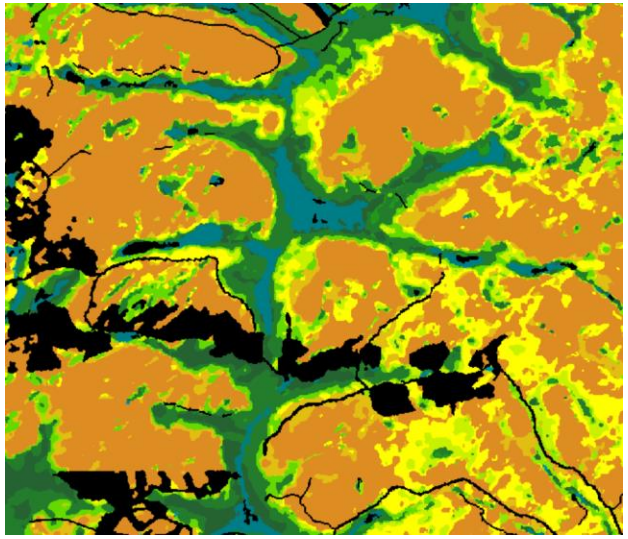
Major Findings

Sugar maple

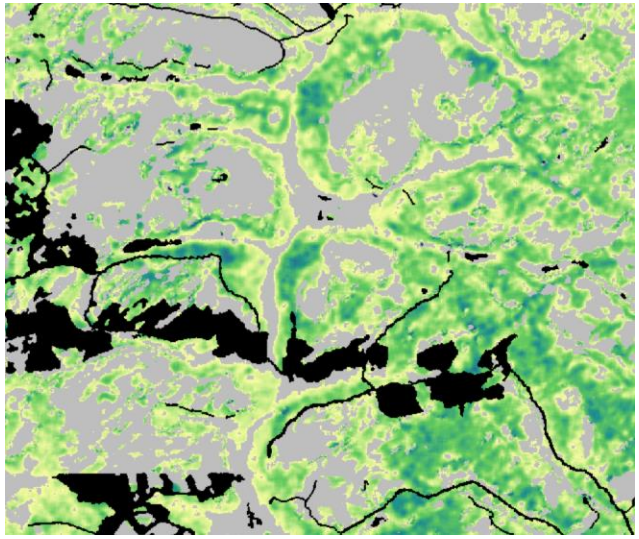


Balsam fir

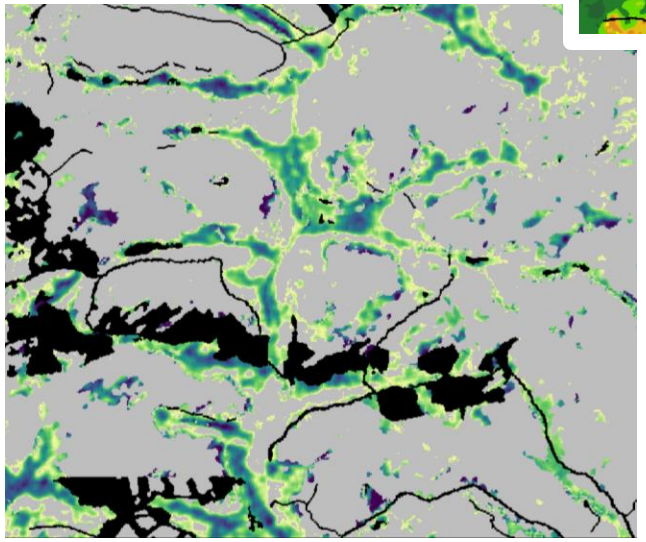




Paper birch

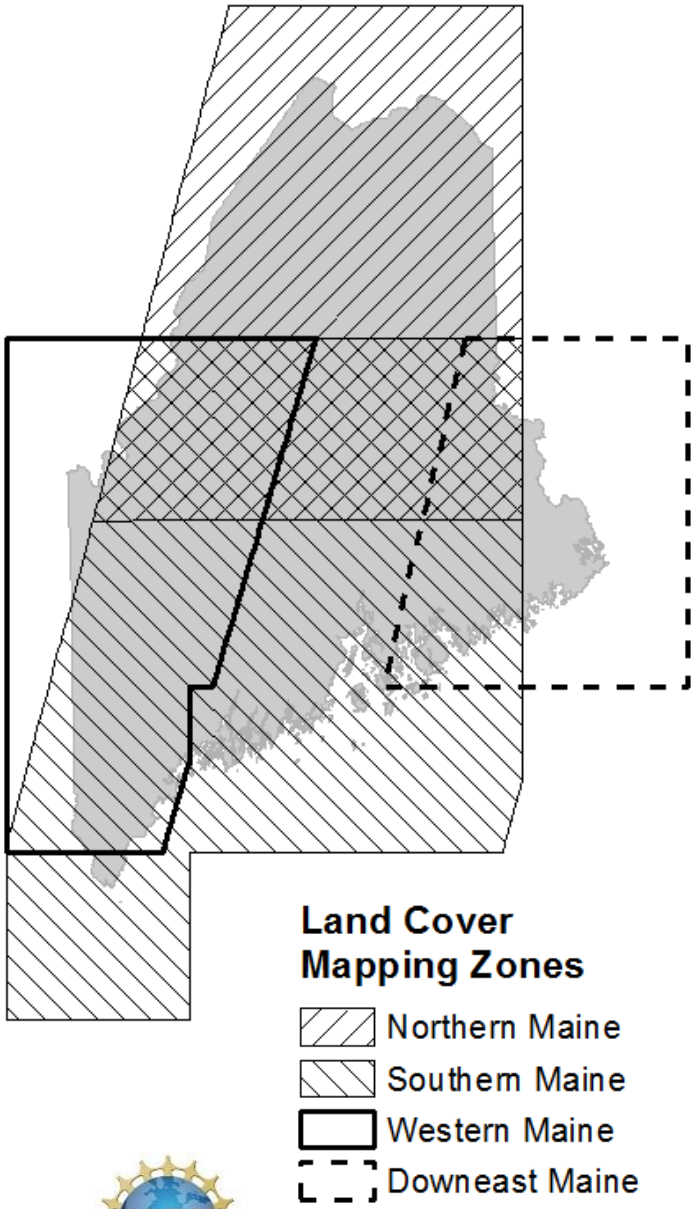


Northern white cedar



- Oak
- Maple-Beech-Birch
- Red Maple
- Aspen-Birch
- Aspen-Birch Mixedwood
- Spruce-Pine Mixedwood
- Fir-Spruce Mixedwood
- Hemlock Mixedwood
- Hemlock
- Spruce-Pine
- Fir-Spruce
- Cedar-Black Spruce
- Hardwood Forested Wetland
- Mixedwood Forested Wetland
- Softwood Forested Wetland

Major Findings



'Mapping zones' based on availability of suitable Sentinel-2 imagery

- Northern Maine: species mapping complete; working on forest type predictions and accuracy assessment
- Southern Maine: image processing complete; ready to train models
- Western Maine: image processing approaching completion (>80% done)
- Downeast Maine: image processing in progress (<20% done)

Major Findings

Student Projects:

Summer 2023:

- Hosted a summer intern from Monroe Community College (NSF ATE)
- Developed a QGIS plugin for Sentinel-2 cloud/shadow masking using Python machine learning pipelines

Summer 2023 – spring 2024:

- Hosted an NSF CAREERS Cyberteam undergraduate researcher from Cornell, funded through an NSF REU award to CAFS
- Developed and tested a prototype cloud-hosted geospatial database application to enable borderless, un-tiled raster data processing for large forest mapping projects

Spring and summer 2024:

- Hosting three interns from Monroe Community College (NSF START for ATE)
- Integrating newly developed forest maps with spruce budworm monitoring data to test for associations between forest conditions and population trends

Company Benefits

- Greater value from low-cost remote sensing and geospatial data
  - 10-meter species, forest type, disturbance, and biomass mapping workflows using free or low-cost data
  - Maine state land cover and forest type data, summer 2024
  - Maine state biomass/carbon data, summer 2024
  - New, extremely efficient methods for point cloud processing
  - Reduced time and cost for inventory and mapping
- Currently working with one large landowner in Maine to compare maps with company data
  - Happy so far, but comparing data at the stand-scale is challenging
  - Testing use of FIA vs. owner-supplied fixed area plots, with similar outcomes so far
- Interested in opportunities to test outside of Maine, starting this summer