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], Larry Whitsel[2], Erin Simons-Legaard[3]

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

Presenter: Aaron Weiskittel

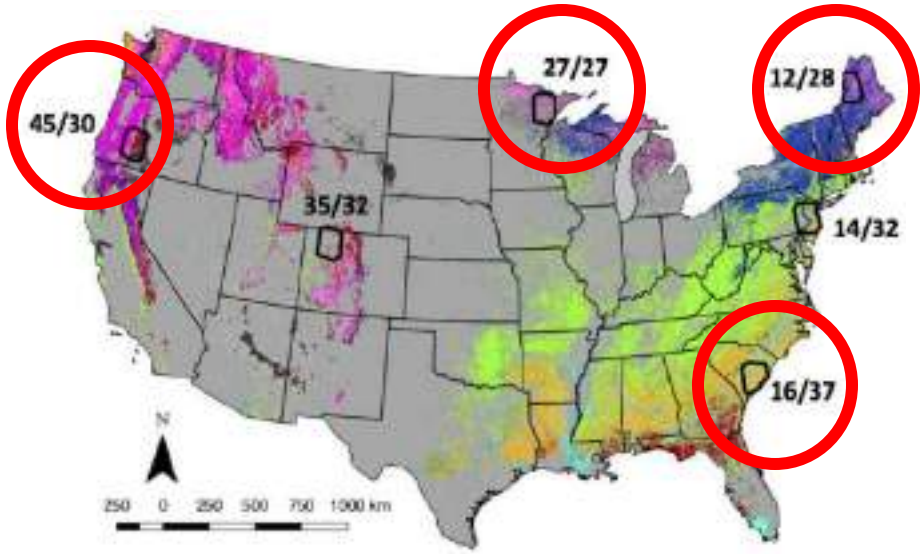
Justification

Goals:

High-quality, affordable spatial data to support data-driven management

Multi-regional validation of automated machine learning (ML) and geospatial processing methods developed at UMaine

- 10 m species, composition classes, disturbance history, and biomass
- Algorithms, workflows integrated into high-volume production software



USDA FIA and NASA Carbon Monitoring System (CMS) data for benchmarking (CMS 2013, Cohen et al.)

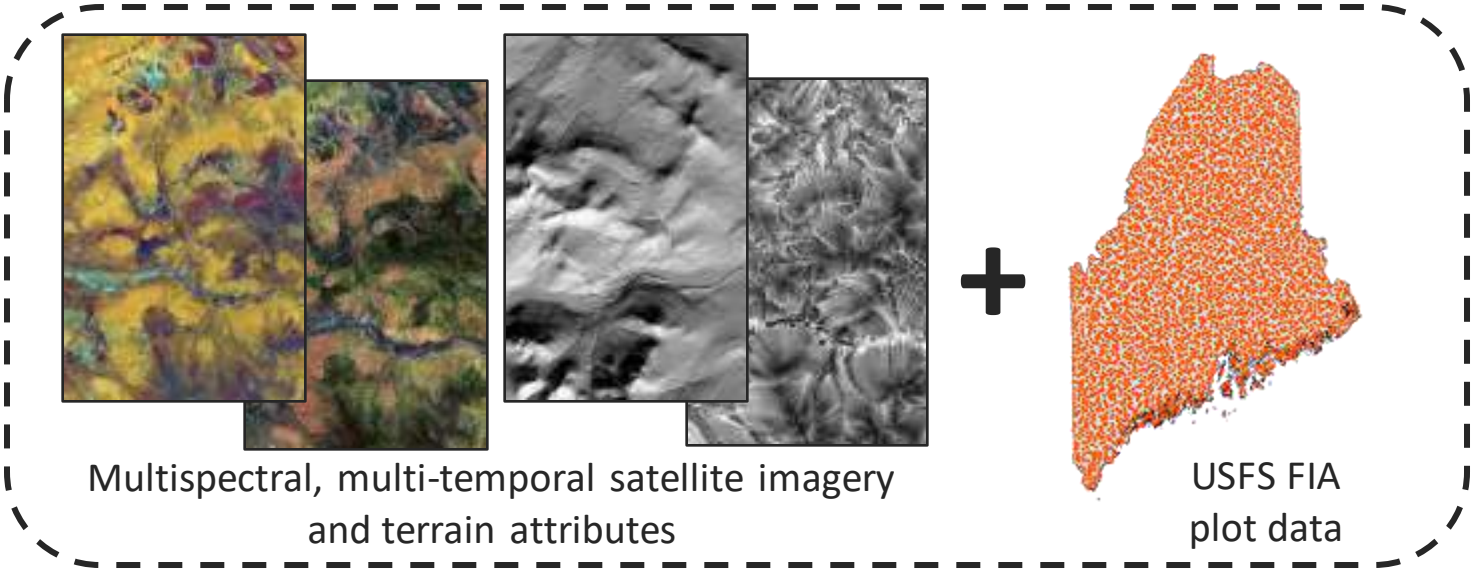
Objectives

Year 2 objectives:

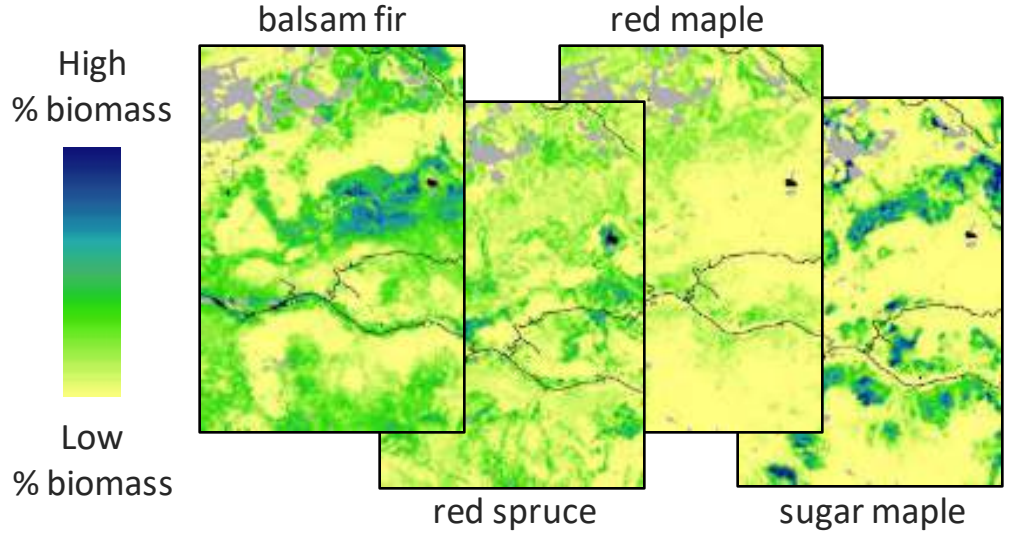
1) Scale 10 m tree species mapping (relative abundance as % AGLB) across northern Maine
   - Multi-objective ML algorithms to control systematic error
   - Larger test of software and workflows
2) Complete workflows and software for time series harvest and disturbance mapping
   - Multi-objective ML to eliminate over- and under-prediction
3) Integrate low-cost LiDAR and 3D NAIP with species and disturbance processing
   - Species biomass estimation at high resolution and low cost

Species mapping:

Methods

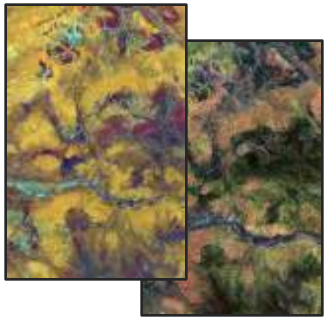


Predictions of species abundance (% of live biomass)



Methods

Forest mapping software executed on the cloud and on computing clusters



Sentinel/Landsat image download & processing

Process management database

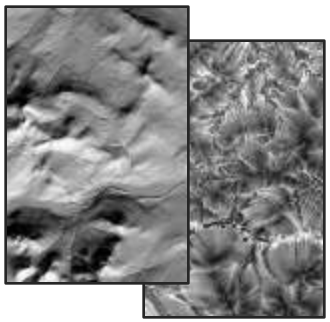
Ancillary data import & processing

Raster data storage

ML problem definition & management

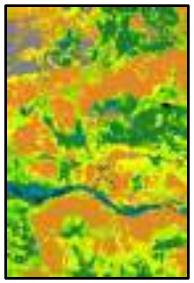
Machine learning

Map generation

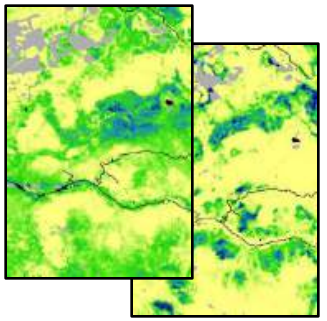


Reference data storage

USFS & other field data



Map fusion

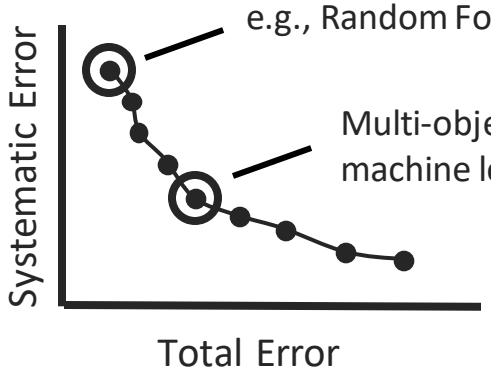


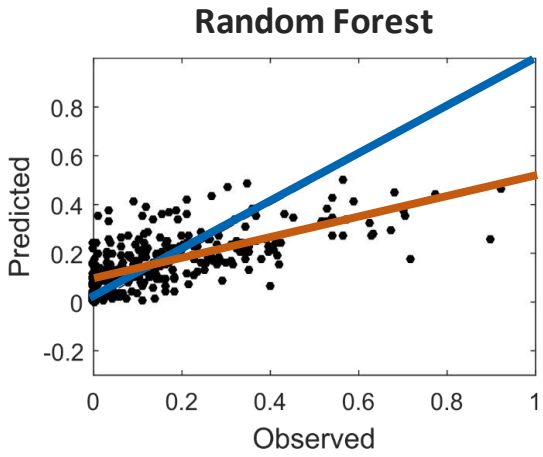
Multi-objective ML:

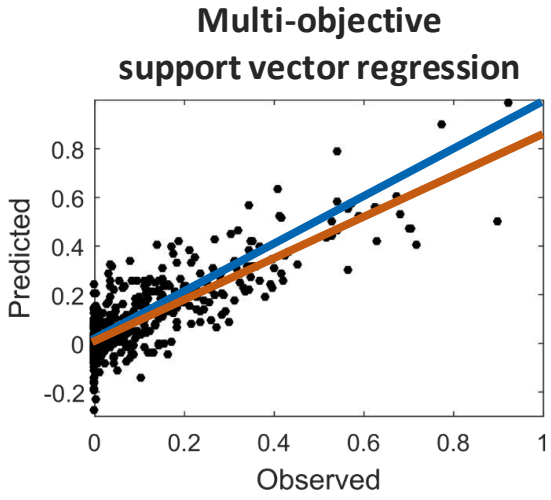
Methods

Location and measurement errors in satellite and field data introduce attenuation bias in regression models and systematic error in maps.

There is a tradeoff between total and systematic error, but conventional algorithms ignore this.





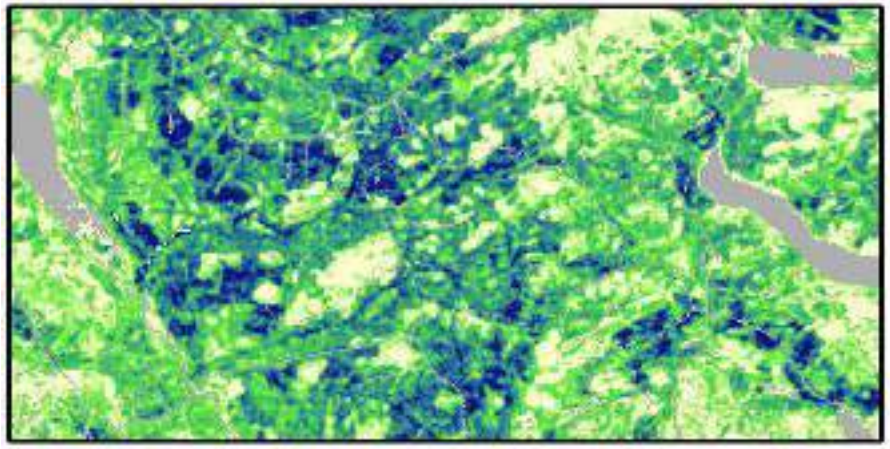


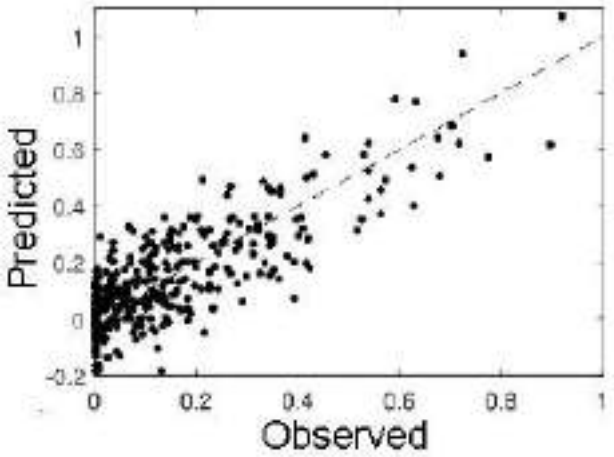
Multi-objective machine learning simultaneously minimizes total AND systematic error.

Works for both regression and classification.

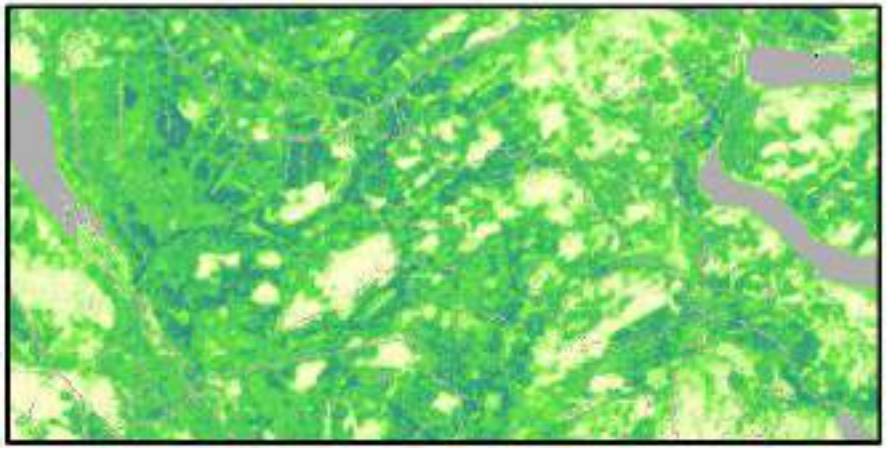
Major Findings

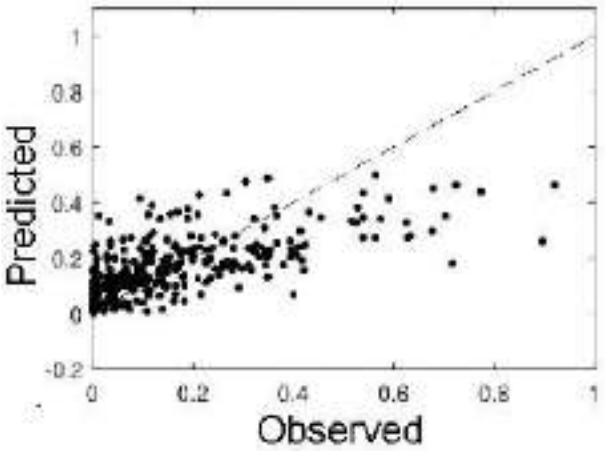
Balsam fir %AGLB, multi-objective ML:





Balsam fir %AGLB, random forest:

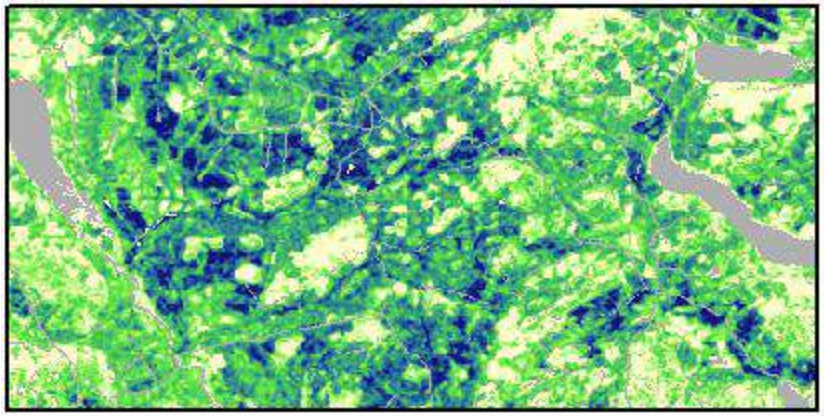




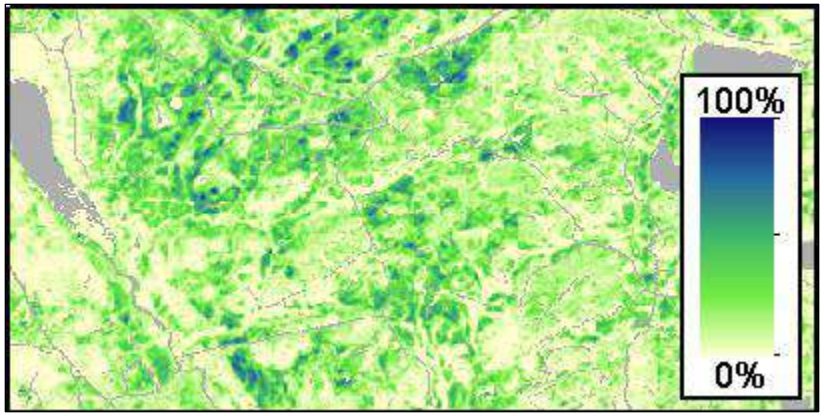
0 1 2 4
km

Major Findings

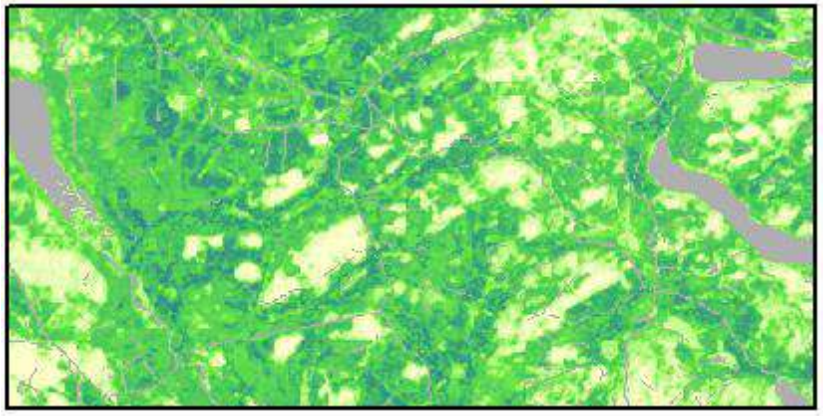
Balsam fir, multi-objective ML:



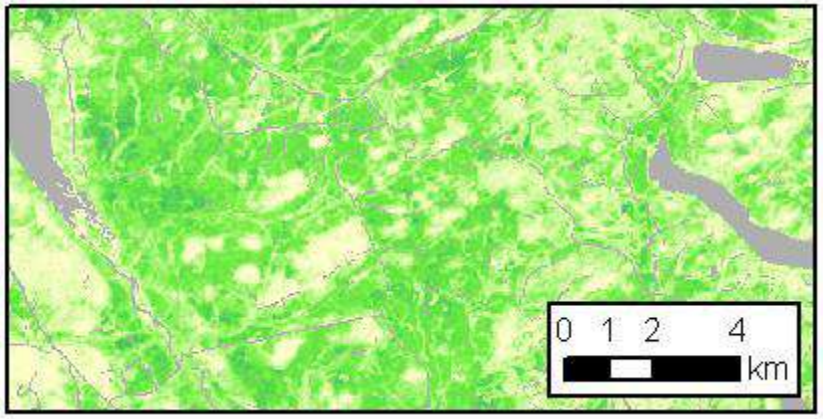
Red spruce, multi-objective ML:



Balsam fir, random forest:



Red spruce, random forest:

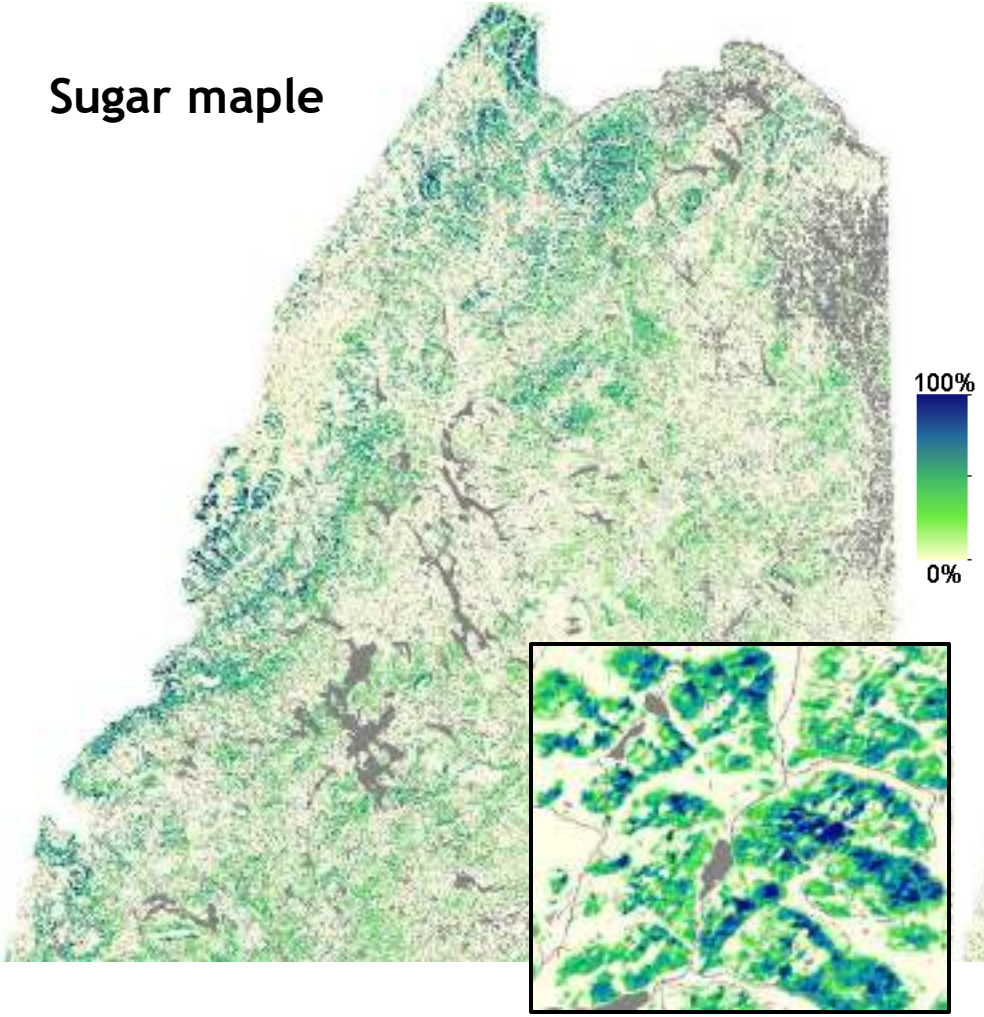


Reduced bias of multi-objective ML improves species differentiation

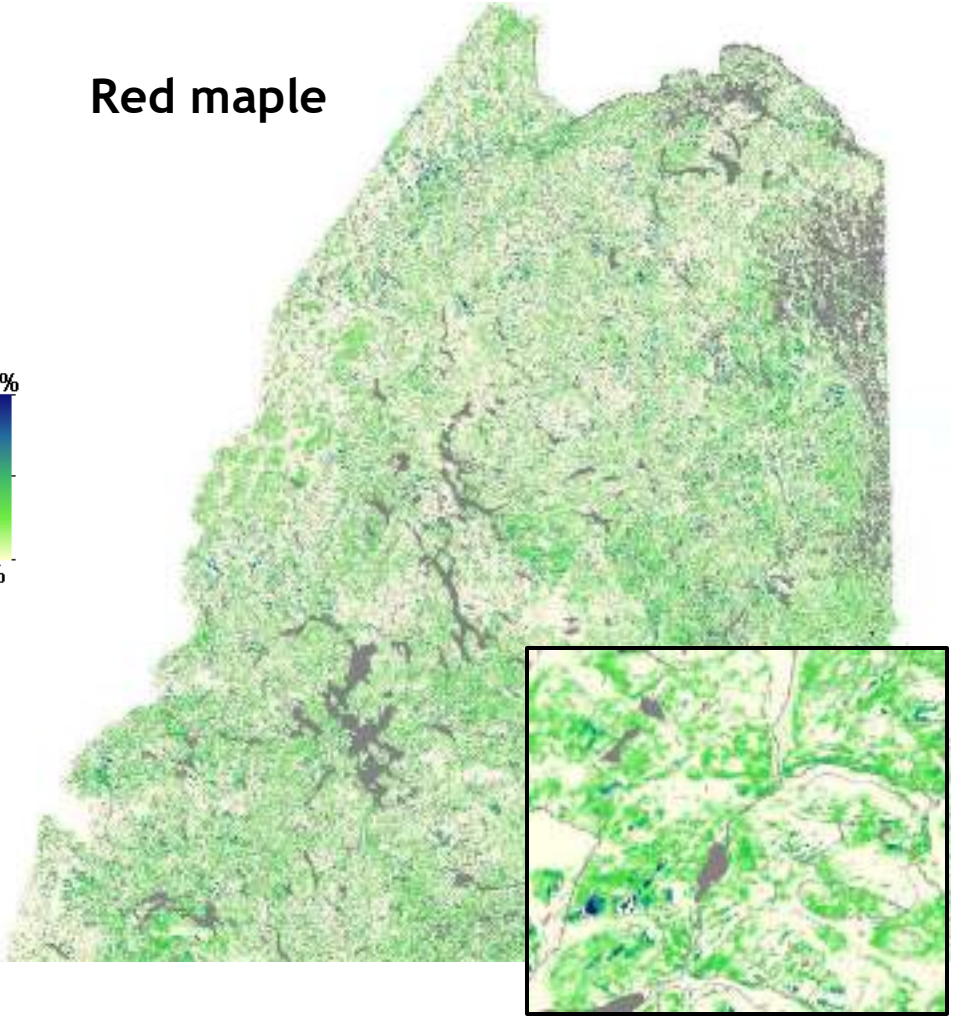
Major Findings

2021 species %AGLB mapped at 10 m resolution across 10 million acres

Sugar maple

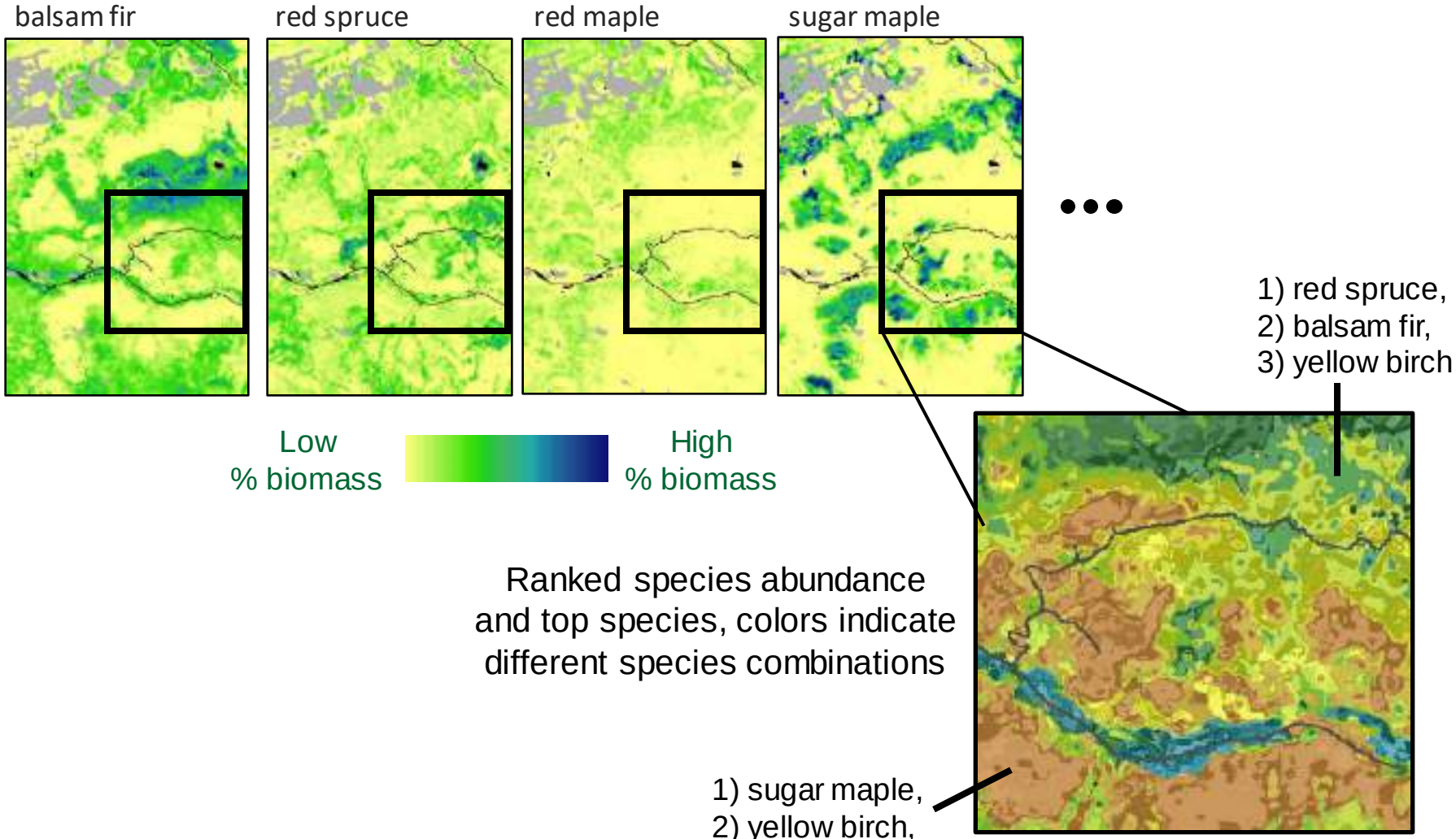


Red maple



Major Findings

Additional progress translating species into overstory composition and forest type maps

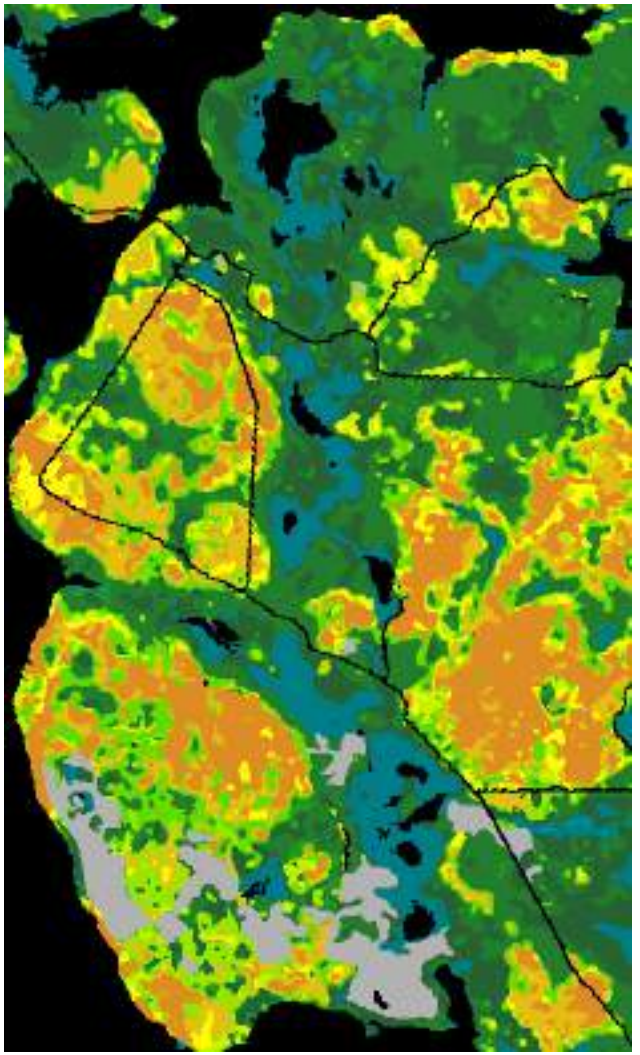


Major Findings

Cooperative agreement with the state of Maine to map forest types statewide

Cooperative agreement with the state of Maine to map forest biomass and carbon at 10 m using 3D NAIP and satellite data

New software engineer to assist with large-scale integration of LiDAR and 3D NAIP

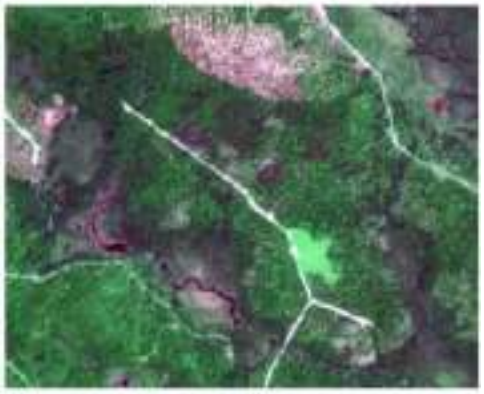


Major Findings

Complete workflows for harvest/disturbance mapping by multi-objective ML

Currently automating reference data handling for long time series, experimenting with automated post-processing and polygonization

Biennial trial, 2020 – May 2022, 10 m resolution:



Company Benefits

Continued development and proof of concept of low-cost forest mapping methods using multi-objective ML and automated geospatial processing

- High-resolution species and forest type mapping using free satellite imagery and FIA data
- High accuracy, low bias harvest and disturbance history
- Integration of low-cost LiDAR and 3D NAIP for species biomass
- Integration of open-source ML libraries for benchmarking against other methods
- Efficient production software, fully generalized for application outside of Maine
- Reduced time and cost for inventory and mapping

Summary

Year 2 Progress:

1) Proof of concept of high-resolution, large-area species mapping using multi-objective ML
   - 18 individual species, plus additional species groups, across 10 mil acres of northern Maine at 10 m resolution
   - 2-stage multi-objective ML, predicting occurrence then abundance
   - Reduced systematic error benchmarked against random forest; generally 5-10% mean error
2) Complete workflows for time series harvest and disturbance mapping using multi-objective ML
   - Low or no bias, 90+% accuracy
3) Progress on the integration of LiDAR and 3D NAIP (year 3 priority)
4) Software generalized for use in other states/regions (year 3 priority)