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

- Satellite remote sensing has high potential to meet information needs of forest management
- But integration of satellite-derived forest maps into forest inventory and management planning has been limited
- Satellite-derived maps obtained from empirical models trained against field plot data are often biased, leading to systematic map error that can severely degrade map value
- The goal of this project is to develop and test algorithms and workflows that reduce systematic error and produce better maps at lower cost

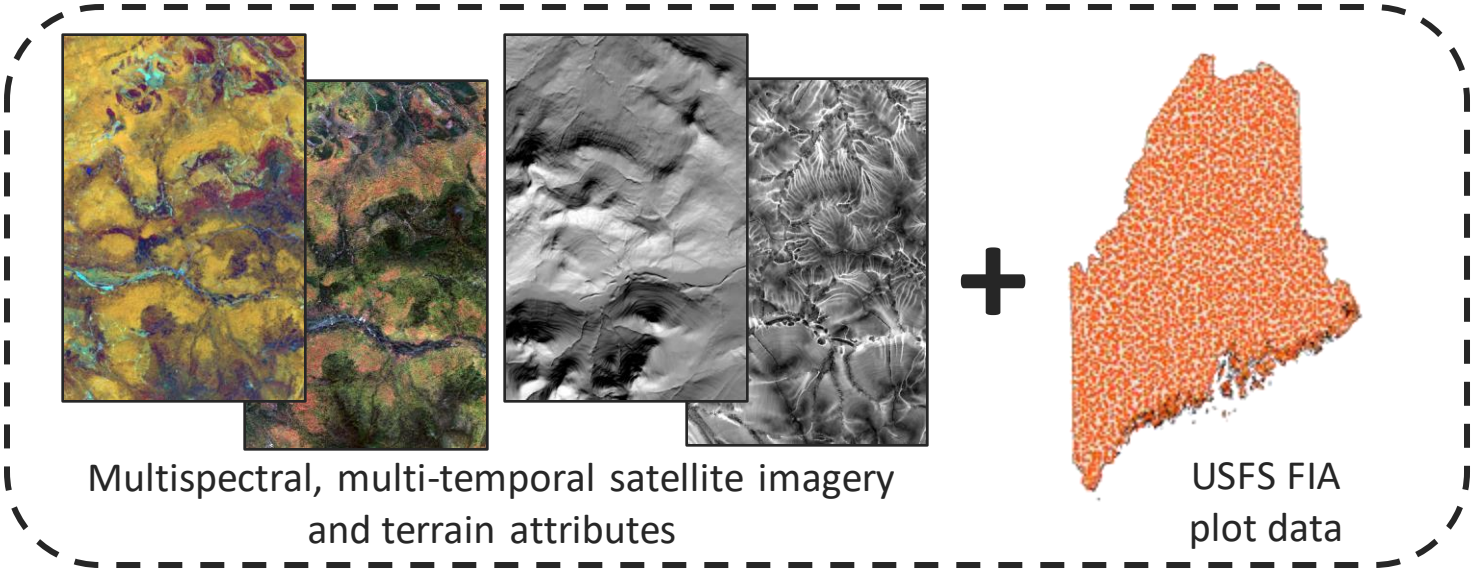
Objectives

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

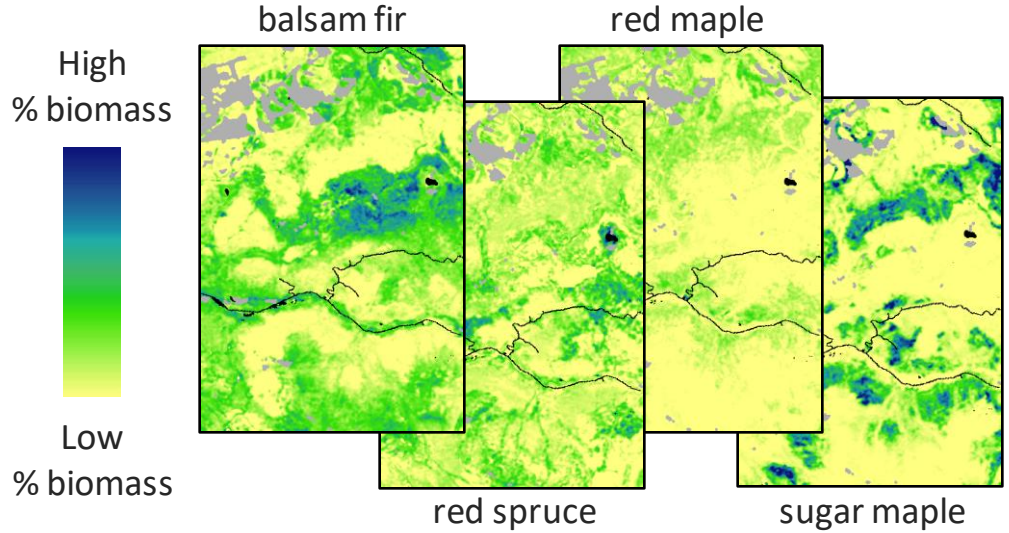
- 10 m species relative abundance, biomass, forest type, disturbance
- Multi-objective machine learning to reduce systematic prediction error
- Algorithms, workflows integrated into high-volume production software
- Test algorithms and workflows in multiple regions

Species mapping:

Methods



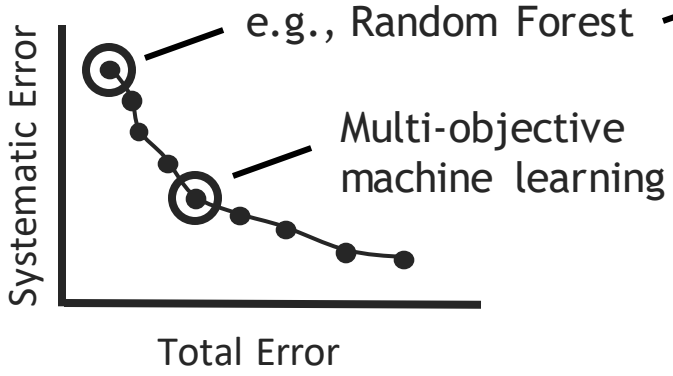
Predictions of species abundance (% of live biomass)

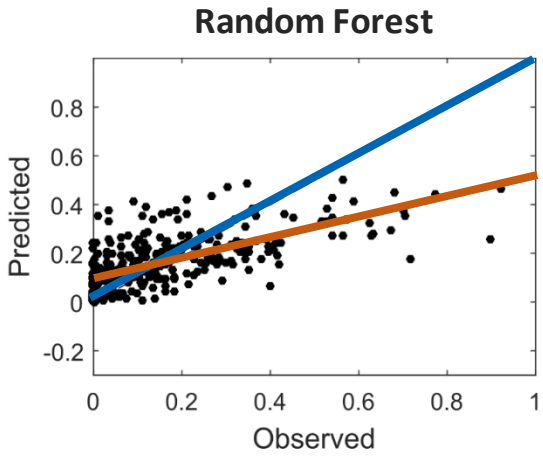


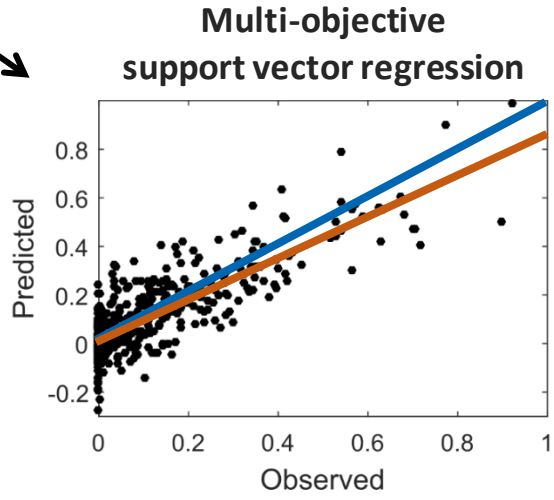
Multi-objective machine learning:

Methods

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





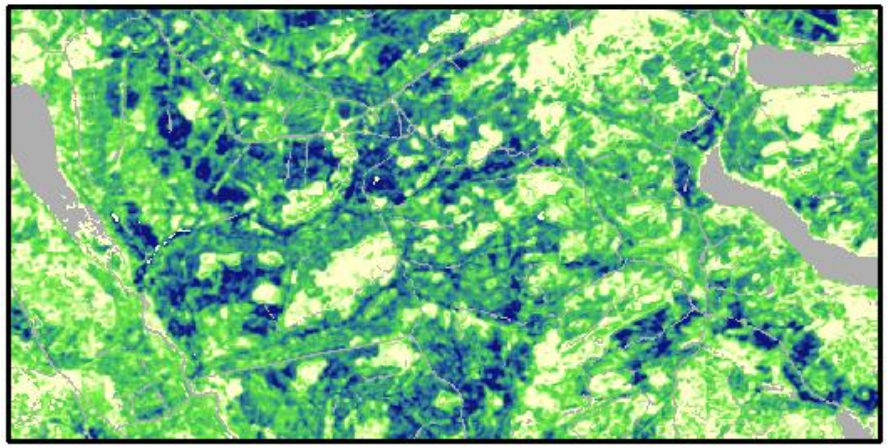


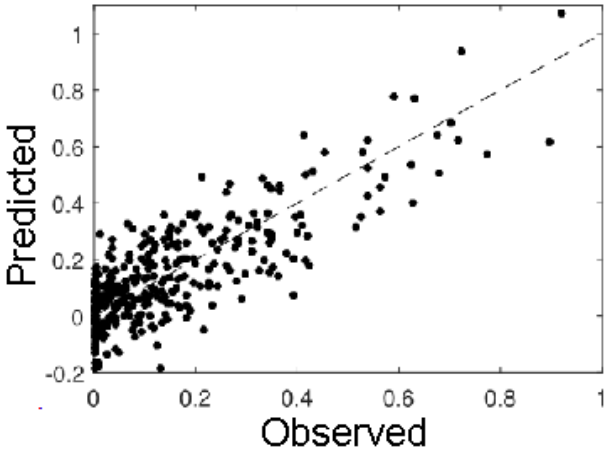
Minimization of total error causes attenuation bias in regression models and systematic error in maps.

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

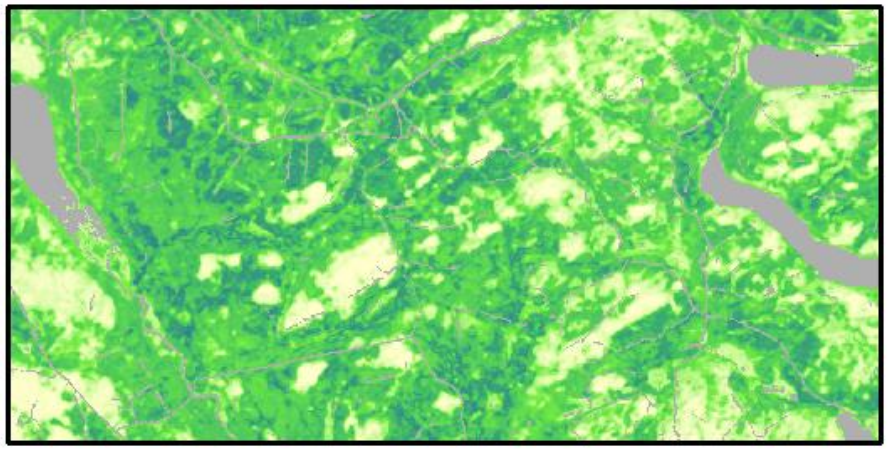
Major Findings

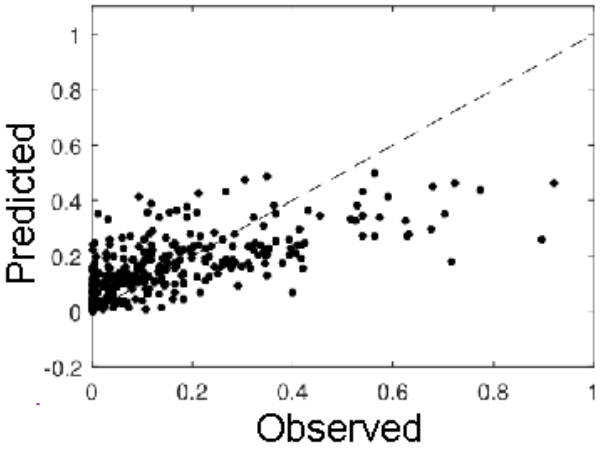
Balsam fir %AGLB, multi-objective ML:

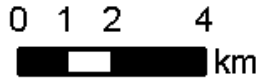




Balsam fir %AGLB, random forest:

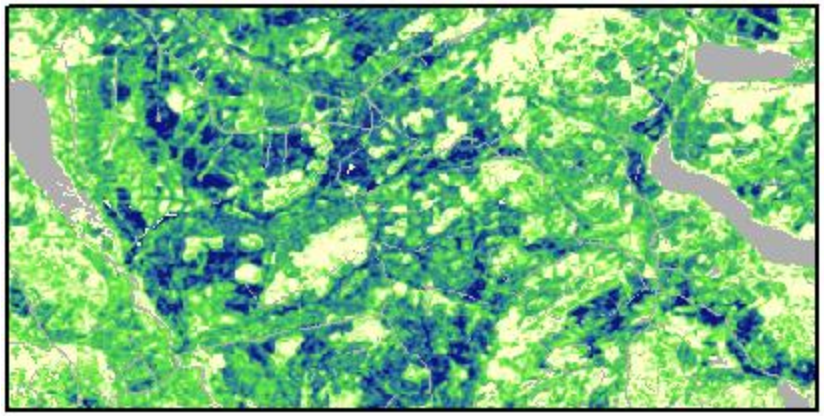




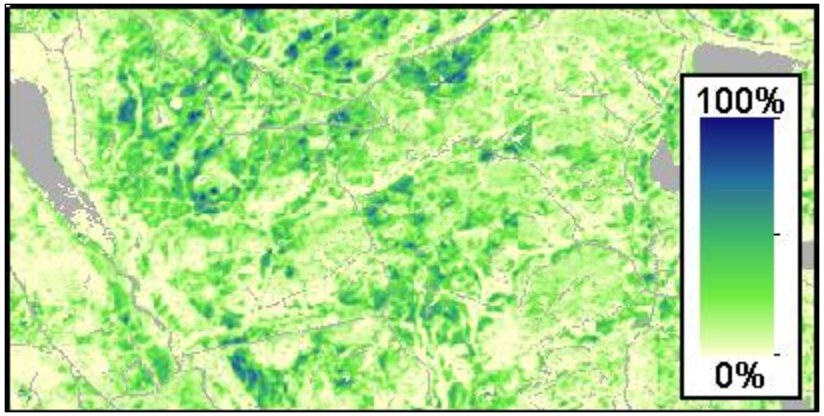


Major Findings

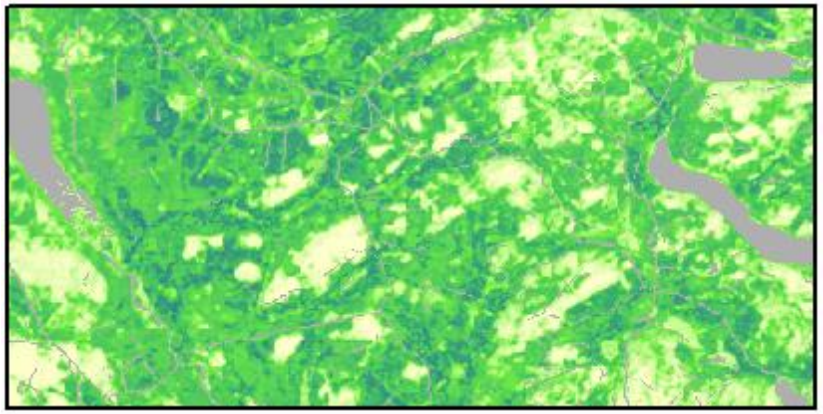
Balsam fir, multi-objective ML:



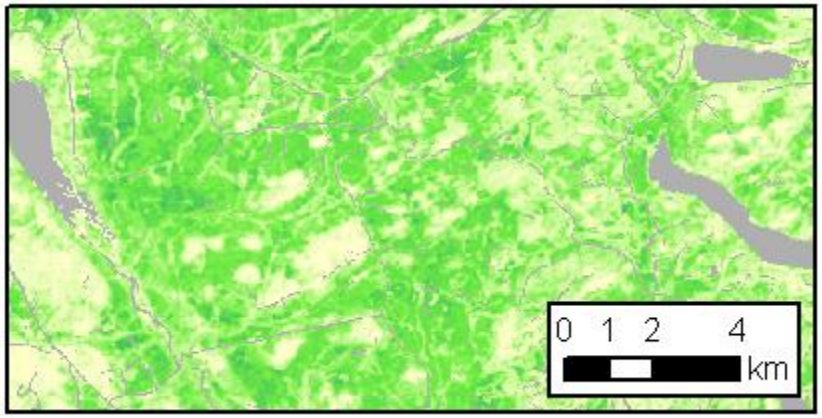
Red spruce, multi-objective ML:



Balsam fir, random forest:



Red spruce, random forest:



Reduced bias of multi-objective ML improves species differentiation

Major Findings

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

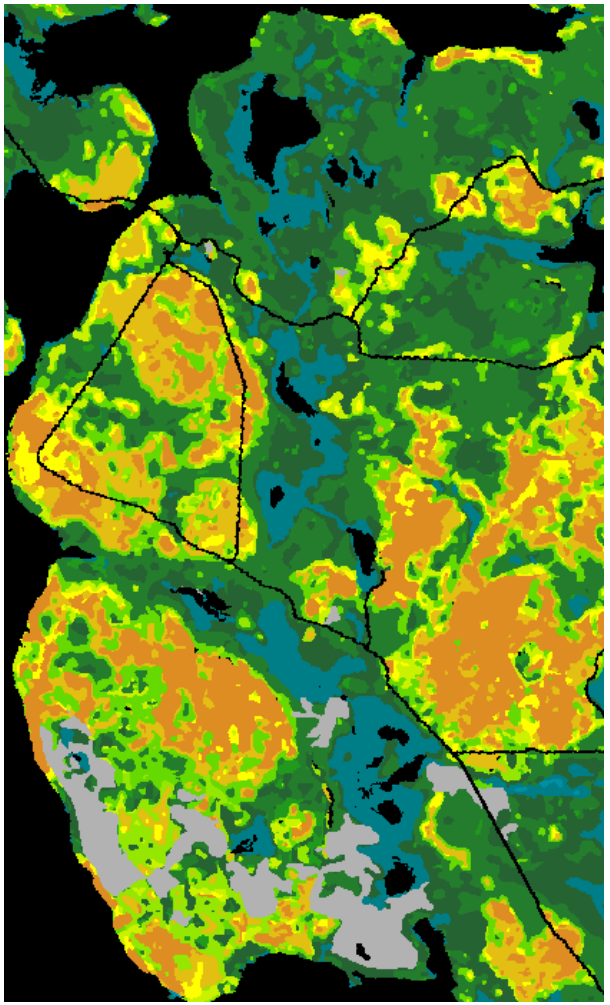
Major Findings

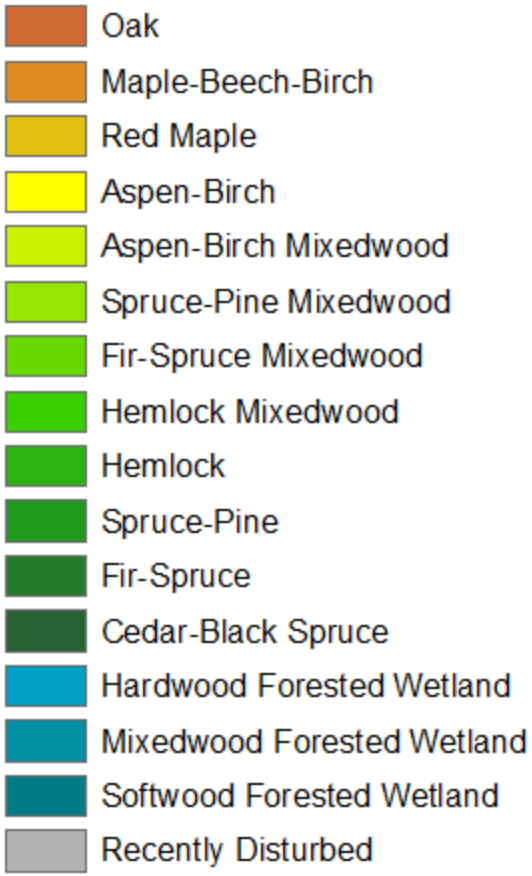
Thematic detail:
- 12 upland forest types
- 3 forested wetland types
- 1 disturbance class (2016-2021)

Spatial detail:
- 10 m pixel resolution
- ~¼ acre MMU

Updates:
- Minimum 4-6 year cycle, tied to NOAA land cover updates

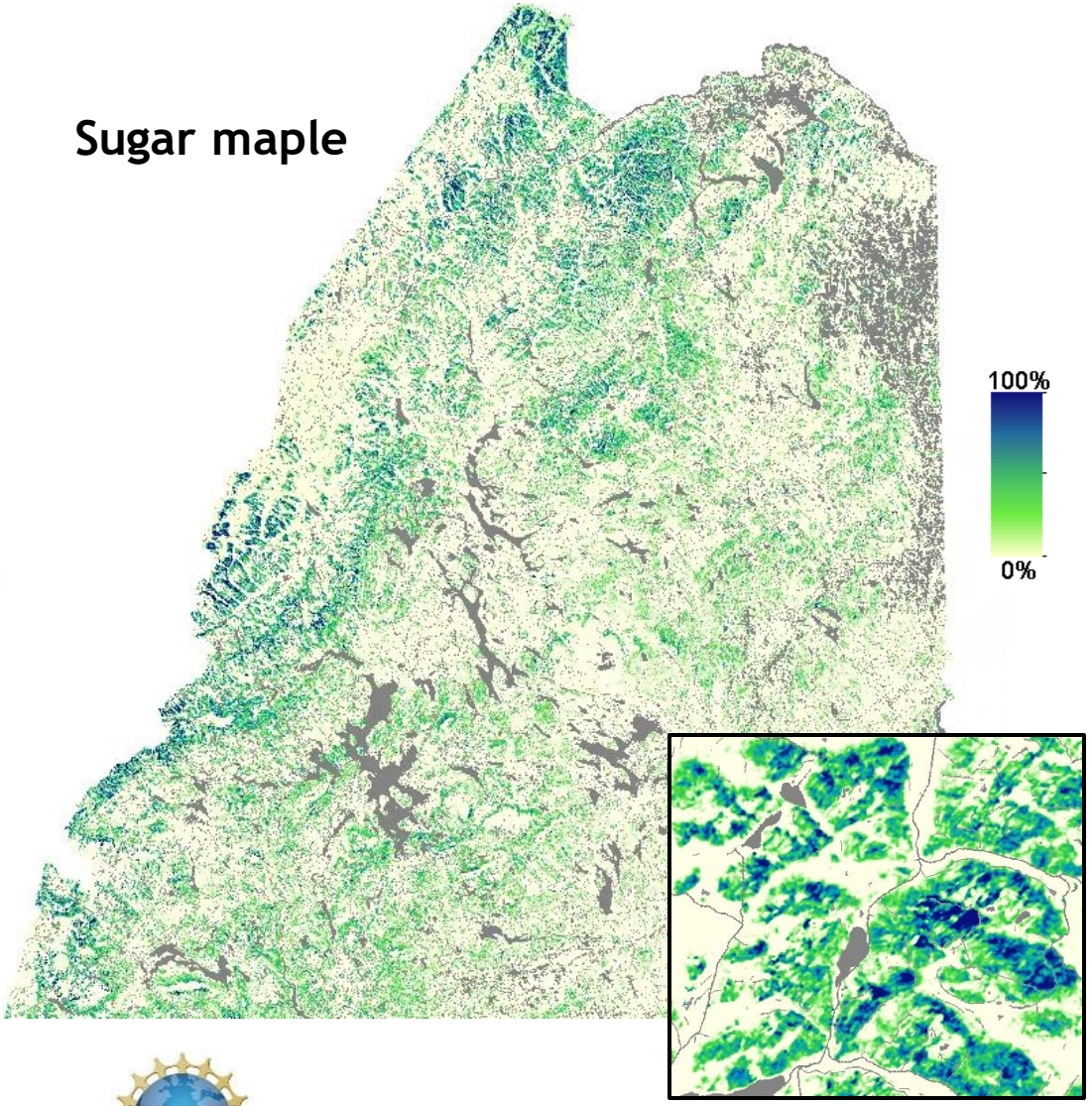
UMaine forest type map





Major Findings

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



- Paper birch
- Aspen 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

Methods

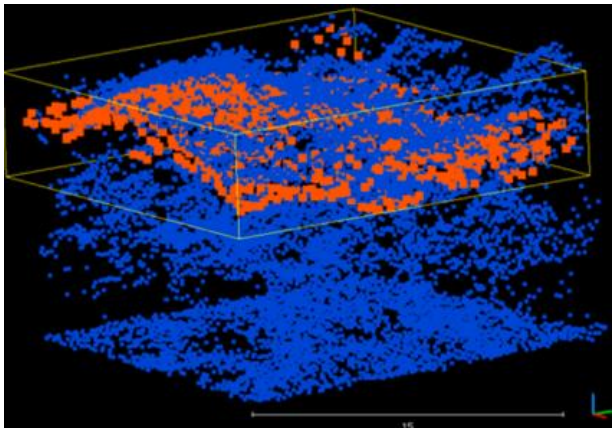
Statewide aboveground biomass and carbon density:

- Funded by the Maine DEP to support the state's Climate Action Plan
- Coordinated with the production and maintenance of the 10-meter land cover and forest type map
- Biomass predictions based on 2021 3D NAIP digital aerial photogrammetry data collected statewide
- Standard area-based approach, computing point cloud metrics over a fixed 10-meter grid

Methods

Challenges:

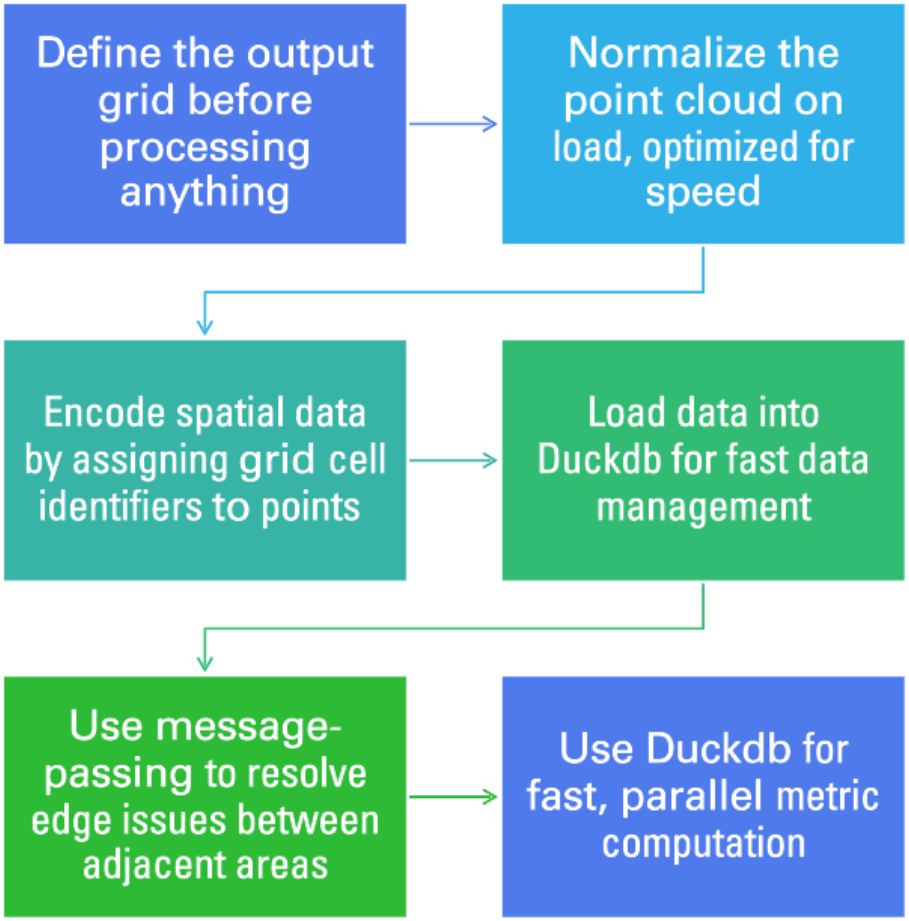
- Late delivery of NAIP data; compressed project timeline
- LARGE volume of data; MANY small tiles
- Normalization of height measurements using LiDAR-derived digital elevation data
- Difficulties scaling production statewide using existing software
  - Typically intended for desktop processing of smaller areas
  - Slow, particularly for certain steps
  - Prone to tile/grid alignment problems



Prior et al. Int J Appl Earth Obs Geoinf 110 (2022) 102813

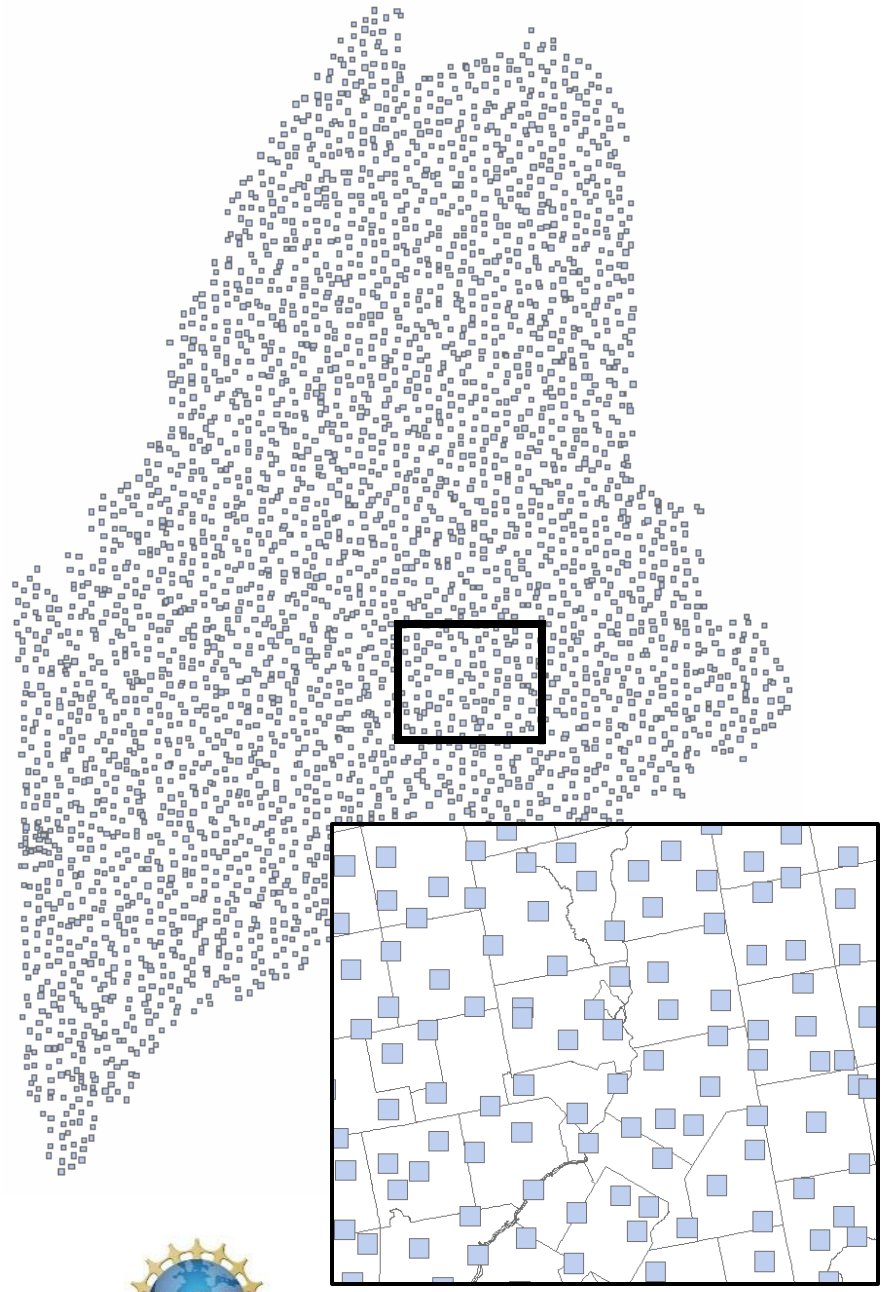
Major Findings

New data processing solution



- Extremely fast computations
- Portable (compute on the cloud, HPC cluster, or laptop)
- Easy horizontal scaling; parallel processing of tiles or regions through a task queue
- No mismatches along tile or region boundaries, achieved with no redundant computation

Major Findings



Statewide NAIP processing

- NAIP data processed around FIA plots to support biomass/C modeling
  - 2 km x 2 km blocks centered over public (fuzzed) FIA plot locations
  - Required to meet data security conditions placed on our use of true plot locations
- Complete statewide processing will progress through the summer, in parallel with modeling

Major Findings

Student involvement:

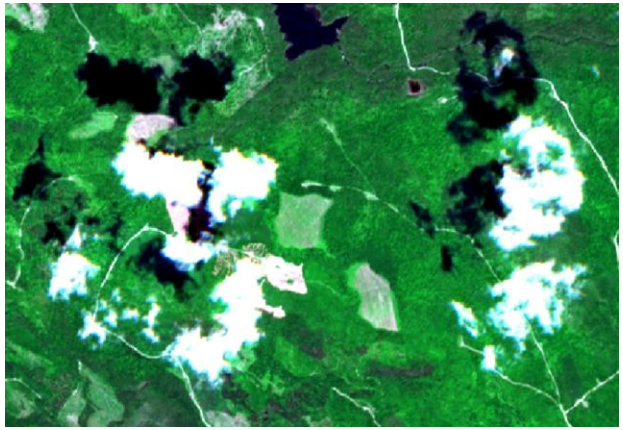
Summer 2022 – Two student interns from Monroe Community College Geospatial Information Technology (GIST) program, working at UMaine on machine learning cloud/shadow detection (NSF START)

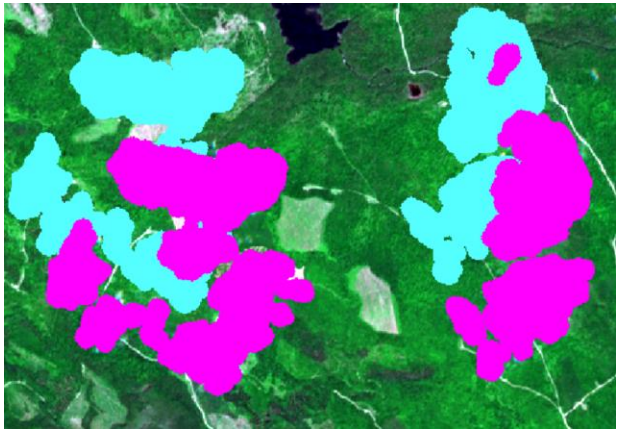
Winter/Spring 2023 – MCC GIST capstone project to develop open-source QGIS plugins for Sentinel-2 cloud/shadow masking

Summer 2023 – MCC GIST summer internship to continue the development of QGIS plugins based on Python FMask plus our ML approach (NSF ATE)

Python FMask

ML-based masking





Major Findings

Student involvement:

Summer 2022 – Two student interns from Monroe Community College Geospatial Information Technology (GIST) program, working at UMaine on machine learning cloud/shadow detection (NSF START)

Winter/Spring 2023 – MCC GIST capstone project to develop open-source QGIS plugins for Sentinel-2 cloud/shadow masking

Summer 2023 – MCC GIST summer internship to continue the development of QGIS plugins based on Python FMask plus our ML approach (NSF ATE)

Summer/Fall 2023 – NSF CAREERS Cyberteam undergraduate research project, funded through an NSF REU award, to build and deploy a cloud-hosted geospatial database application to enable borderless, un-tiled raster processing and machine learning

Company Benefits

Extracting greater value from low-cost remote sensing and geospatial data using multi-objective ML and new data processing approaches

- 10-meter species, forest type, disturbance, and biomass mapping workflows using free or low-cost data
- Maine state land cover and forest type data, fall 2023
- Maine state biomass/carbon data, fall 2023
- High-accuracy, low-bias harvest and disturbance history
- Efficient production software; low-cost data production
- New, extremely efficient methods for point cloud processing
- Reduced time and cost for inventory and mapping

Summary

- Maine state mapping projects were a primary focus over this past year, and will wrap up this fall
- State projects shifted focus from LiDAR to 3D NAIP processing, but methods apply to both
- Originally planned to use NASA CMS plots for independent validation, but too few plots, and too many cutover in recent years
- Validation remains challenging
  - Cross-validation methods using FIA plots
  - Looking for opportunities to work with state and private organizations
- Software is now generalized for use outside of Maine
  - Have FIA data for the northeast; seeking access in other regions
  - Interested in testing the use of other inventory or calibration plot data
  - Motivated to test outside of Maine, later this year