# Progress 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

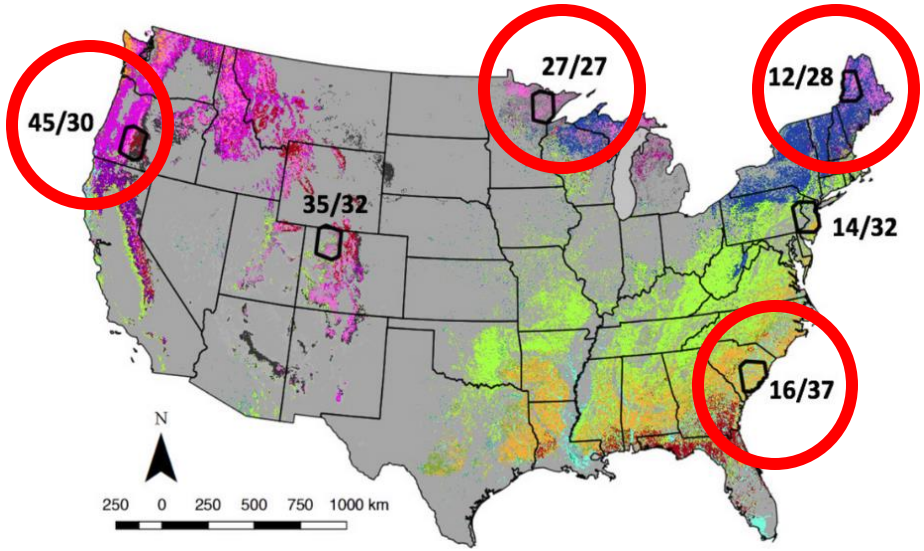
Presenter: Kasey Legaard

# Overview

**Goals:**

Multi-regional validation of automated machine learning and remote sensing methods developed at UMaine

- 10 m species, overstory composition and forest type 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.)

# Progress

**Recent Progress:**

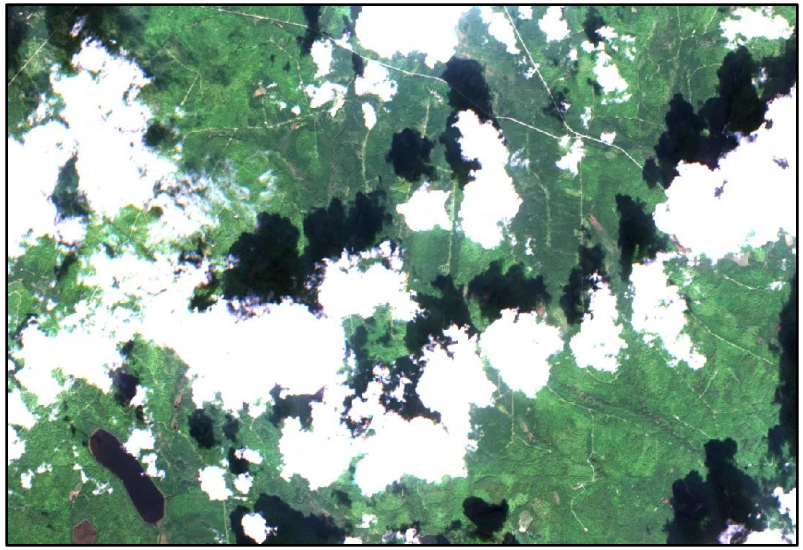
1) Continued improvements to large-area species mapping
   - Improved Sentinel-2 haze correction and cloud masking
   - Improved machine learning in cloud-affected areas
   - Improved prediction adjacent to edges (e.g., recent disturbance, roads, plantations)
2) Nearing completion of workflows and software for time series harvest and disturbance mapping using multi-objective machine learning
   - Efficient reference data handling to eliminate prediction bias
   - Workflows to integrate harvest intensity estimation (biomass removal or harvest classification)
3) Completion of large-area 3D NAIP and LiDAR point cloud processing
   - On-demand and distributed point cloud processing
   - Currently testing software with the 2021 NAIP DSM

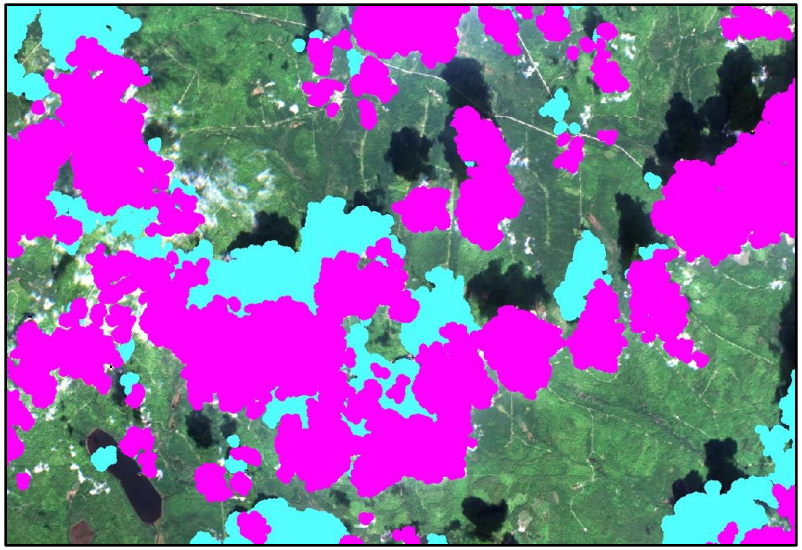
# Progress

**Automated Sentinel-2 cloud/shadow masking:**

Initiated as a CAFS collaboration between the Monroe Community College and UMaine, supported by an NSF START supplemental funding grant

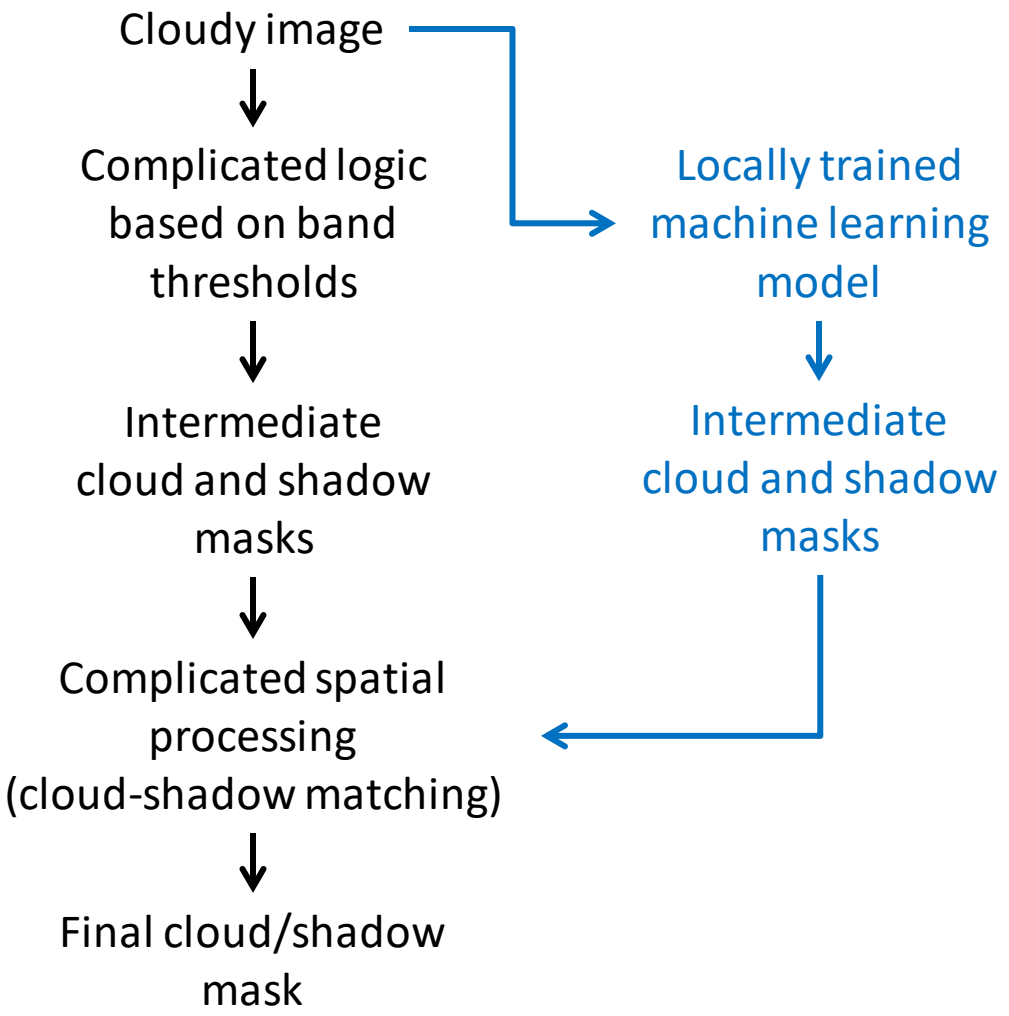
**Project goal:** Improve Sentinel-2 (and Landsat) cloud and shadow masking by integrating machine learning with the Python FMask software (www.pythonfmask.org)

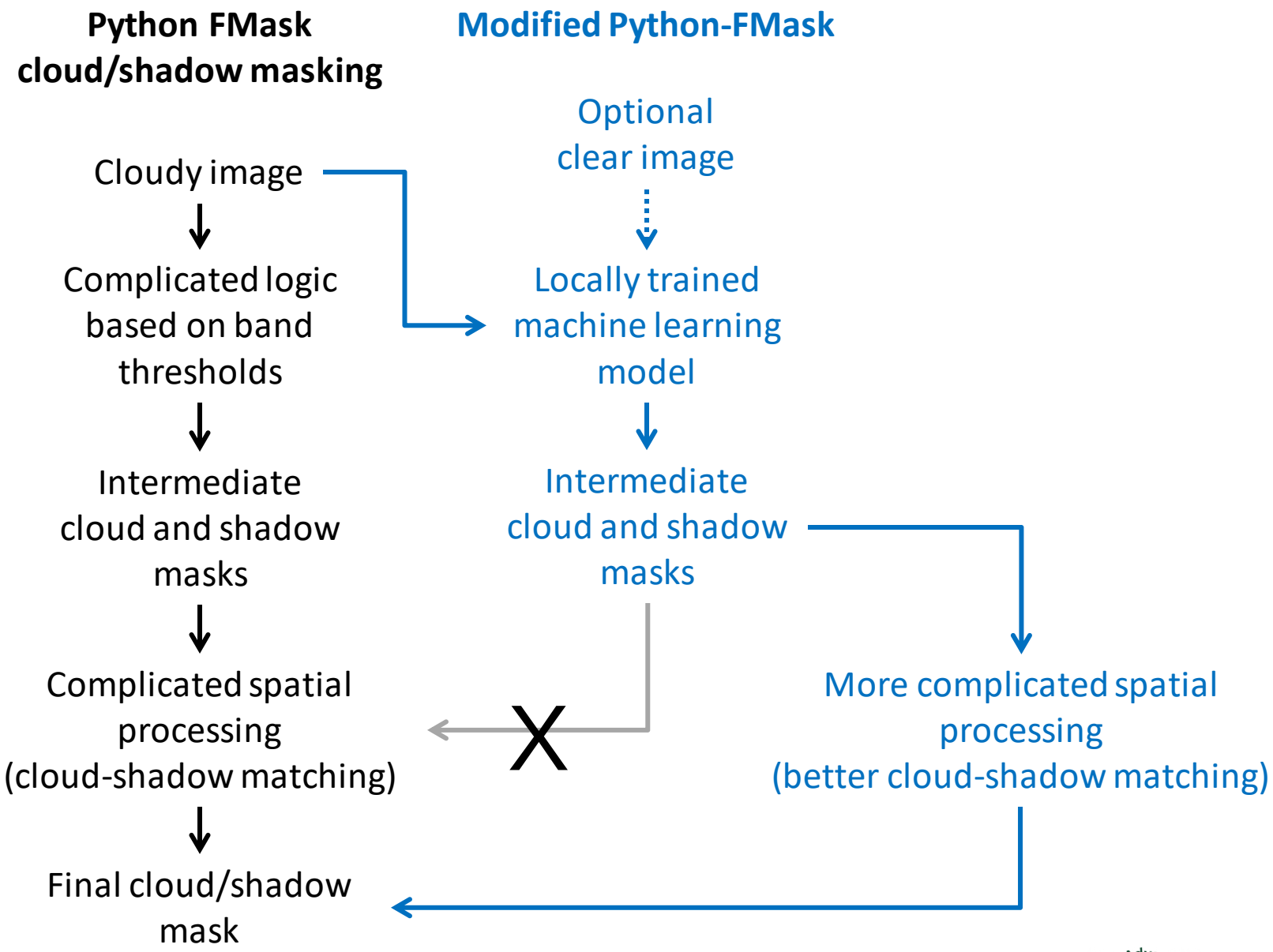




**Python FMask cloud/shadow masking**

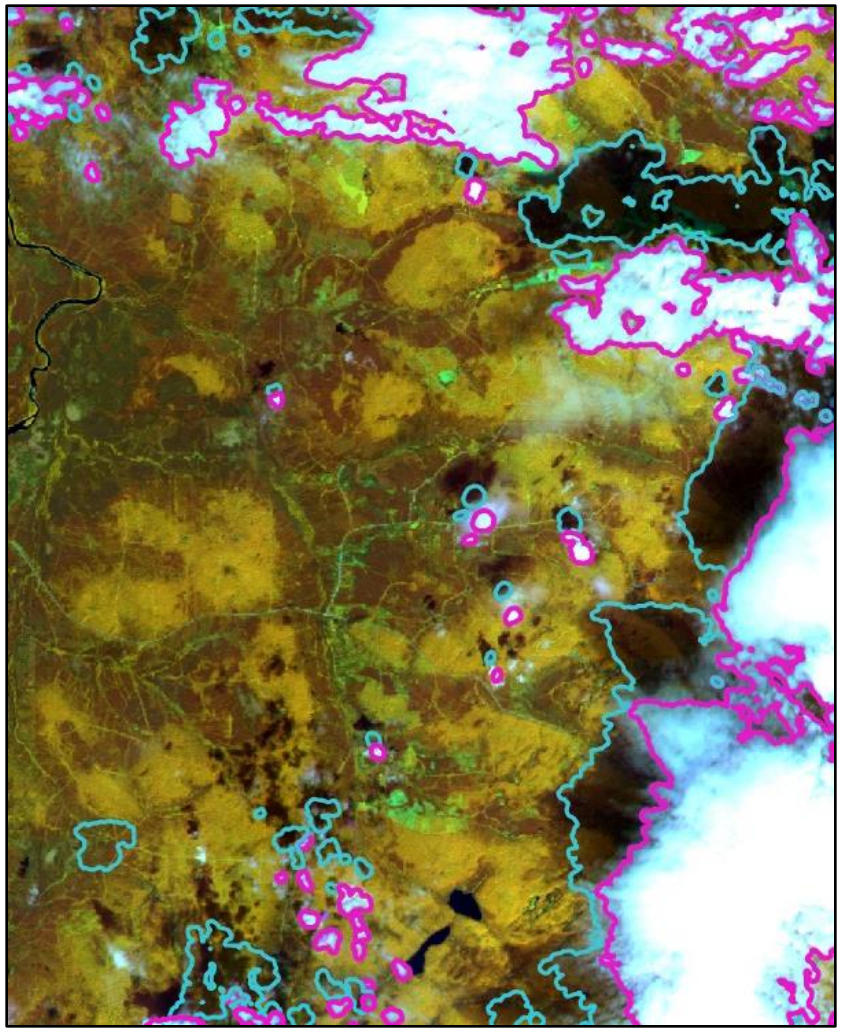
**Modified Python-FMask**



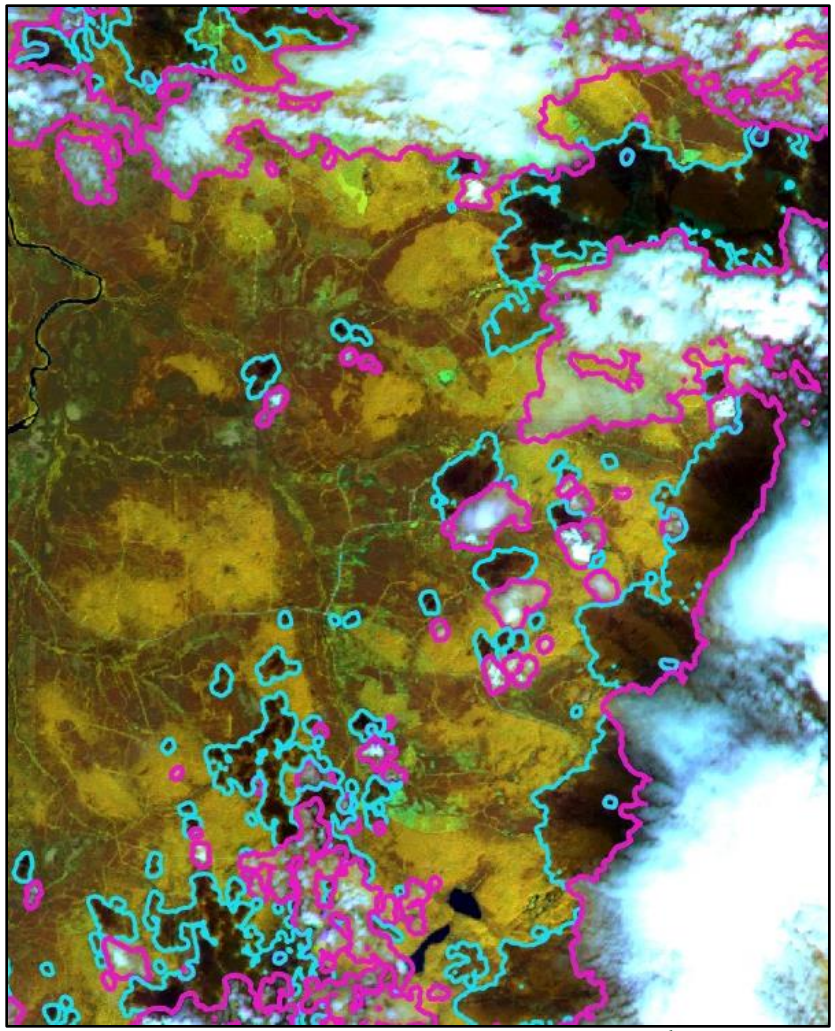

Python FMask cloud/shadow masking
Modified Python-FMask
Optional clear image
Cloudy image
Locally trained machine learning model
Complicated logic based on band thresholds
Intermediate cloud and shadow masks
Intermediate cloud and shadow masks
Complicated spatial processing (cloud-shadow matching)
More complicated spatial processing (better cloud-shadow matching)
Final cloud/shadow mask

# Progress

## Python FMask



## Our modified Python-FMask



# Summary

- Continued improvements to multi-objective ML and large-area mapping workflows and software
- Disturbance mapping and NAIP/LiDAR processing approaching the point that we can integrate multi-source data for large projects
- Limited utility of NASA CMS plot data for validation; relying on FIA for both training and validation
- Working on access to FIA location data in other regions