New Project Proposal

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

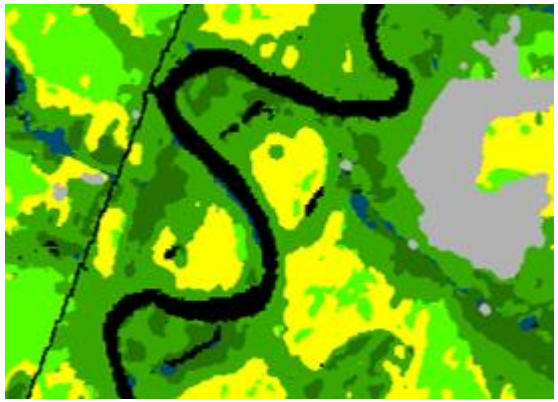
UMaine Intelligent GeoSolutions

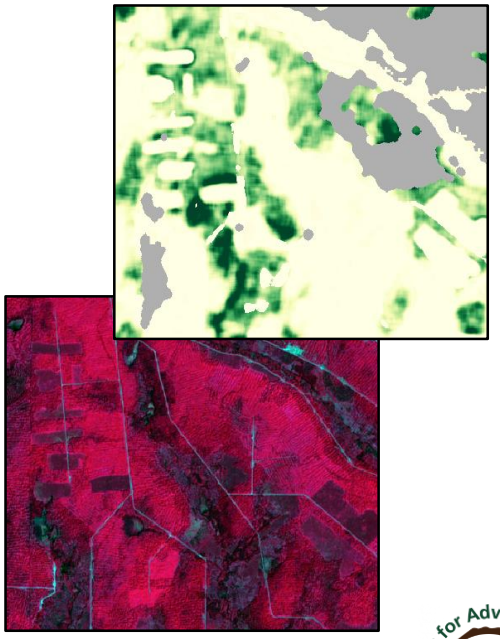
Applied geospatial research and technology transfer initiative

Goal: Support forest management through the provision of high-quality, low-cost spatial data

Forest disturbance

Forest type





Tree species abundance (American beech)

UMaine Intelligent GeoSolutions

Extract maximum value from remote sensing data using improved machine learning methods

Effects of Mismatches of Scale and Location between Predictor and Response Variables on Forest Structure Mapping

Yaguang Xu, Brett G. Dickson, Haydee M. Hampton, Thomas D. Sisk, Jean A. Palumbo, and John W. Prather

Location mismatch between field plots and pixels

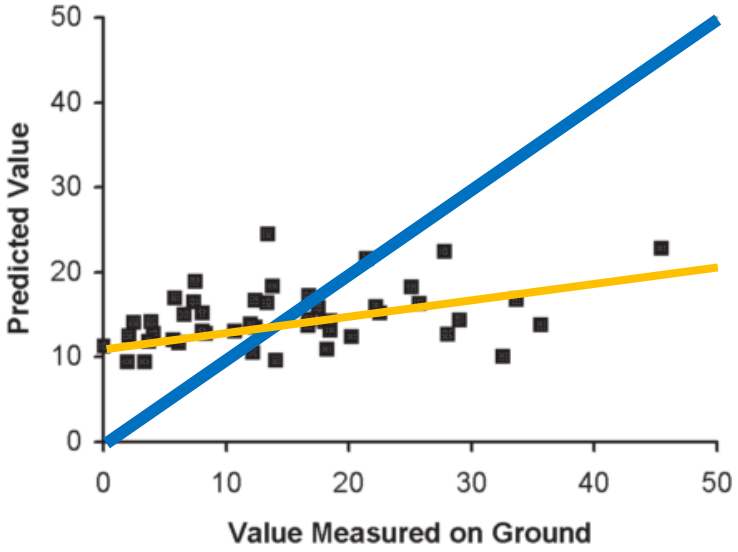


Figure 9. Plot of predicted basal area against ground measured basal area (units: m^2/ha).

Scale mismatch between field plots and pixels

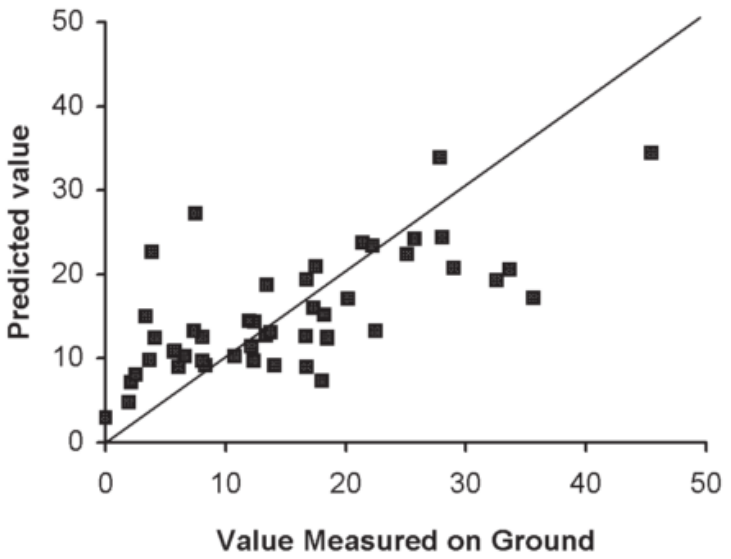
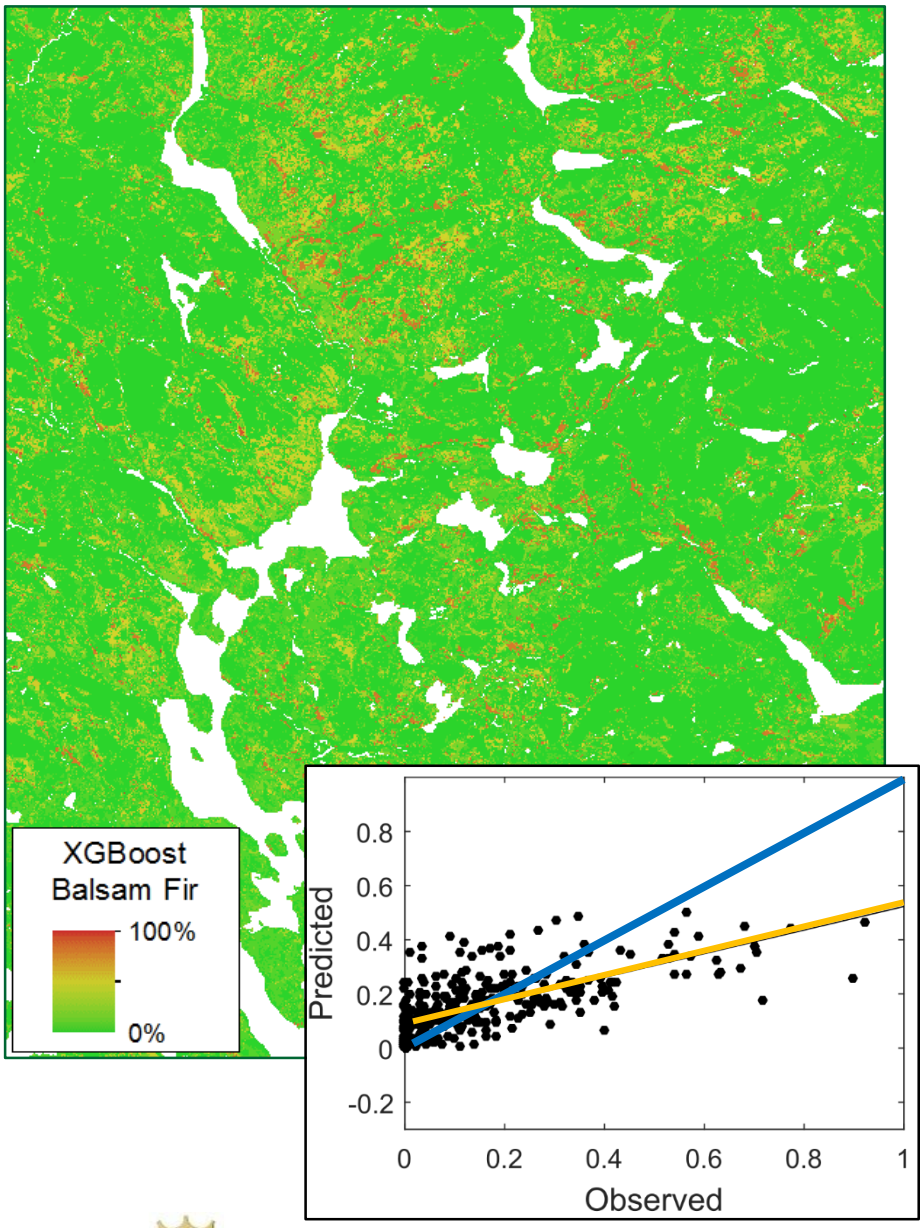


Figure 4. Plot of predicted basal area against basal area measured in ground subplot (units: m^2/ha).



Balsam fir relative abundance (% aboveground live biomass)

XGBoost gradient boosting algorithm

Minimization of total error in the presence of predictor uncertainty causes **attenuation bias** (regression to the mean)

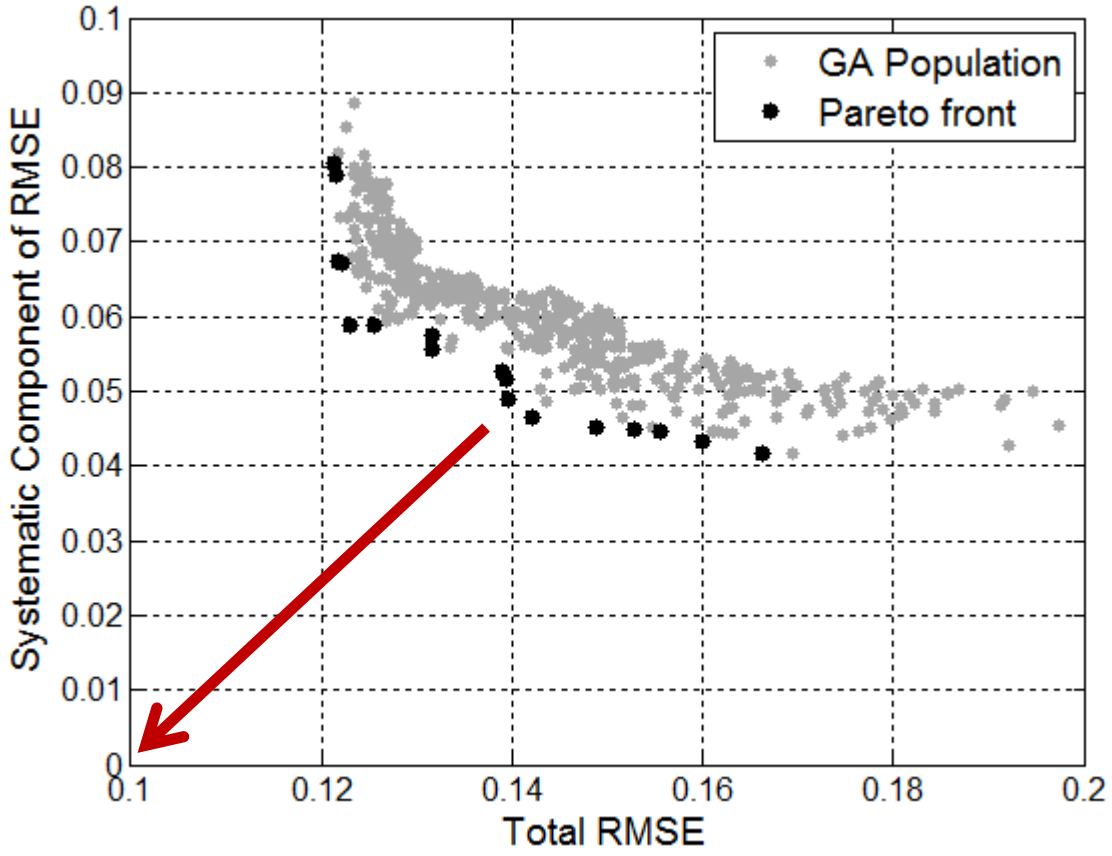
Trade-off between total prediction error and systematic error:

We would prefer low error and low bias

Simultaneous minimization of total and systematic error:

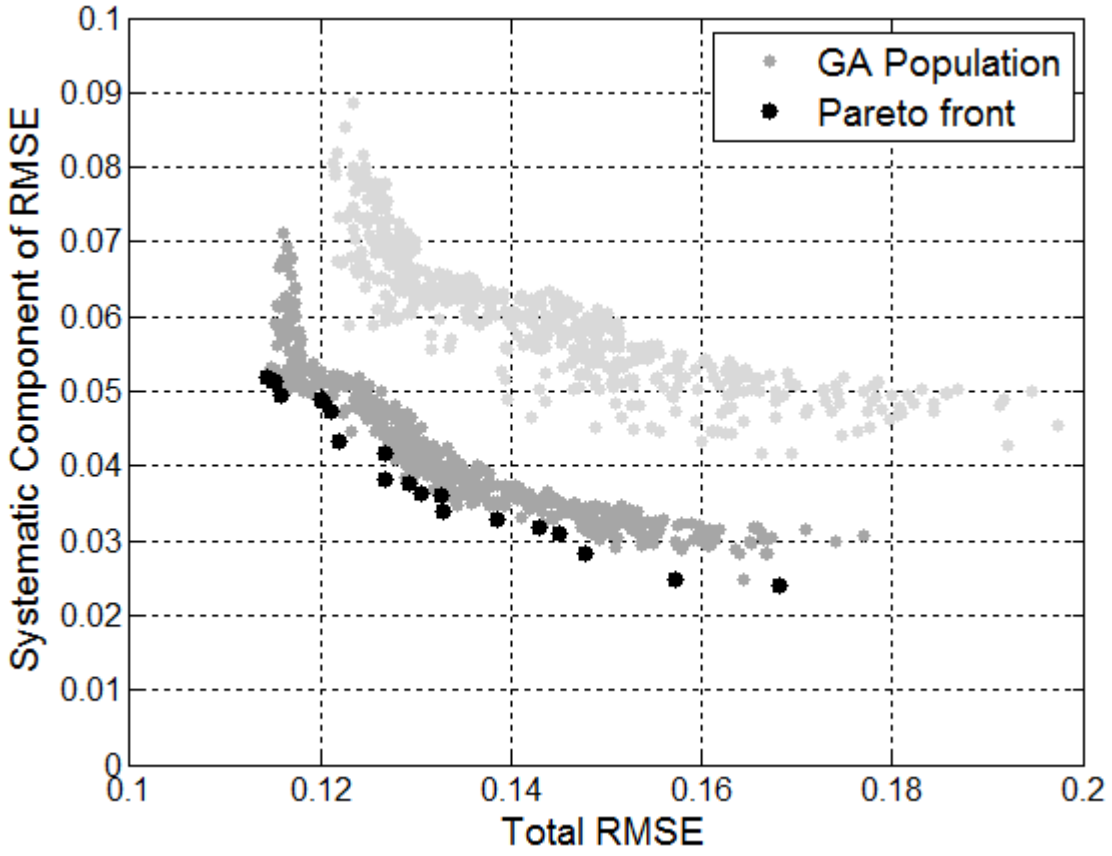
- Support Vector Machines (SVMs) as predictive models
- Multi-objective Genetic Algorithms (GAs) to fit SVMs

Balsam fir % biomass on FIA plots



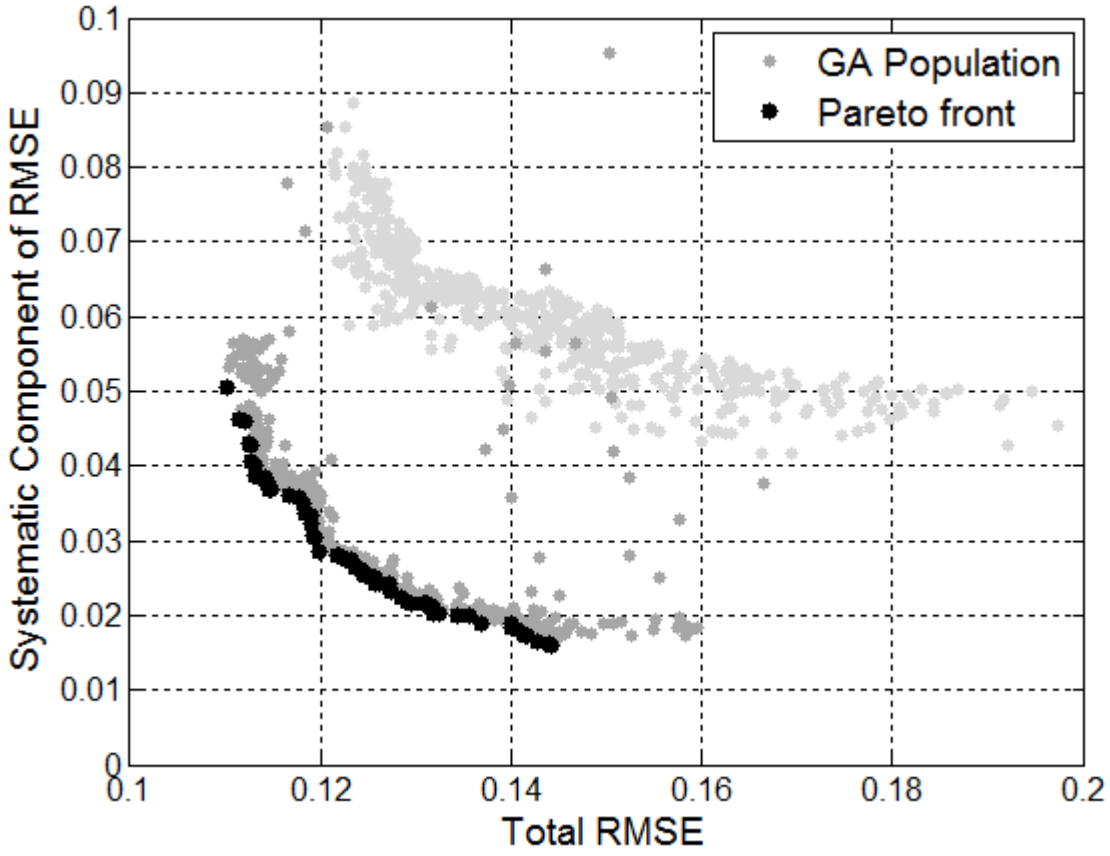
After **10** iterations

Balsam fir % biomass on FIA plots



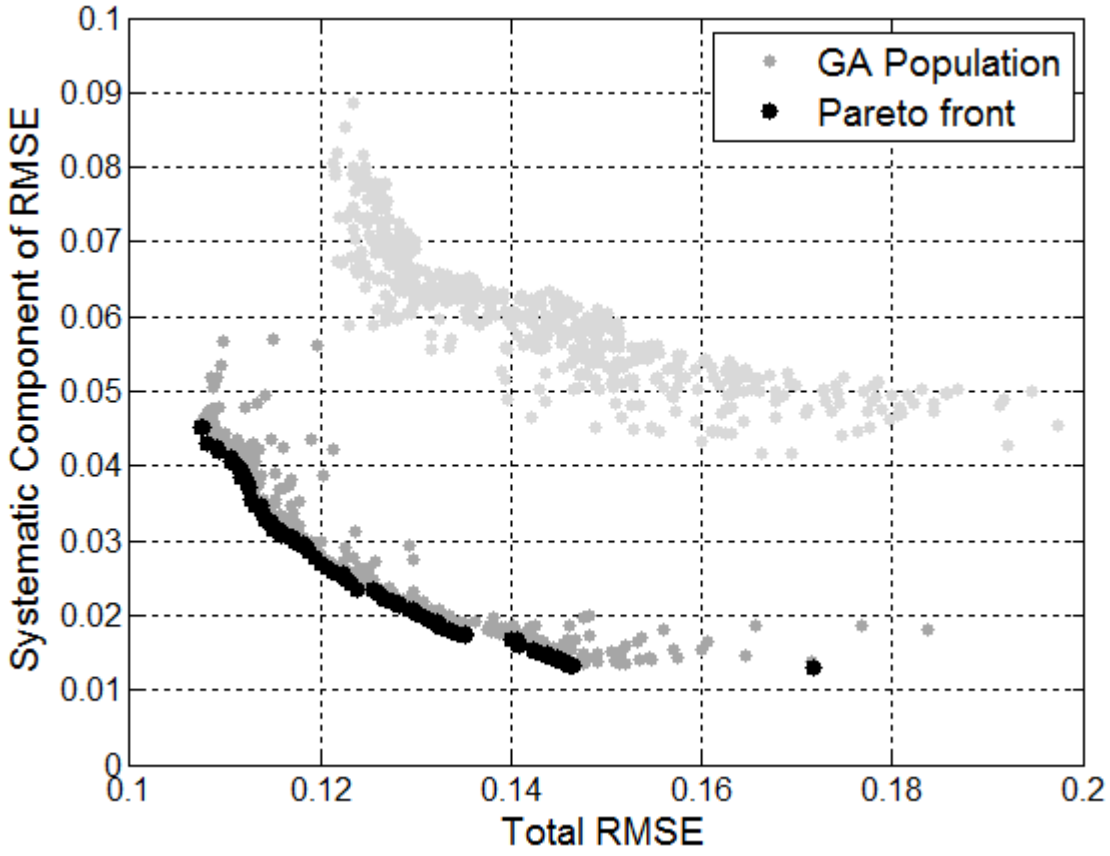
After **20** iterations

Balsam fir % biomass on FIA plots

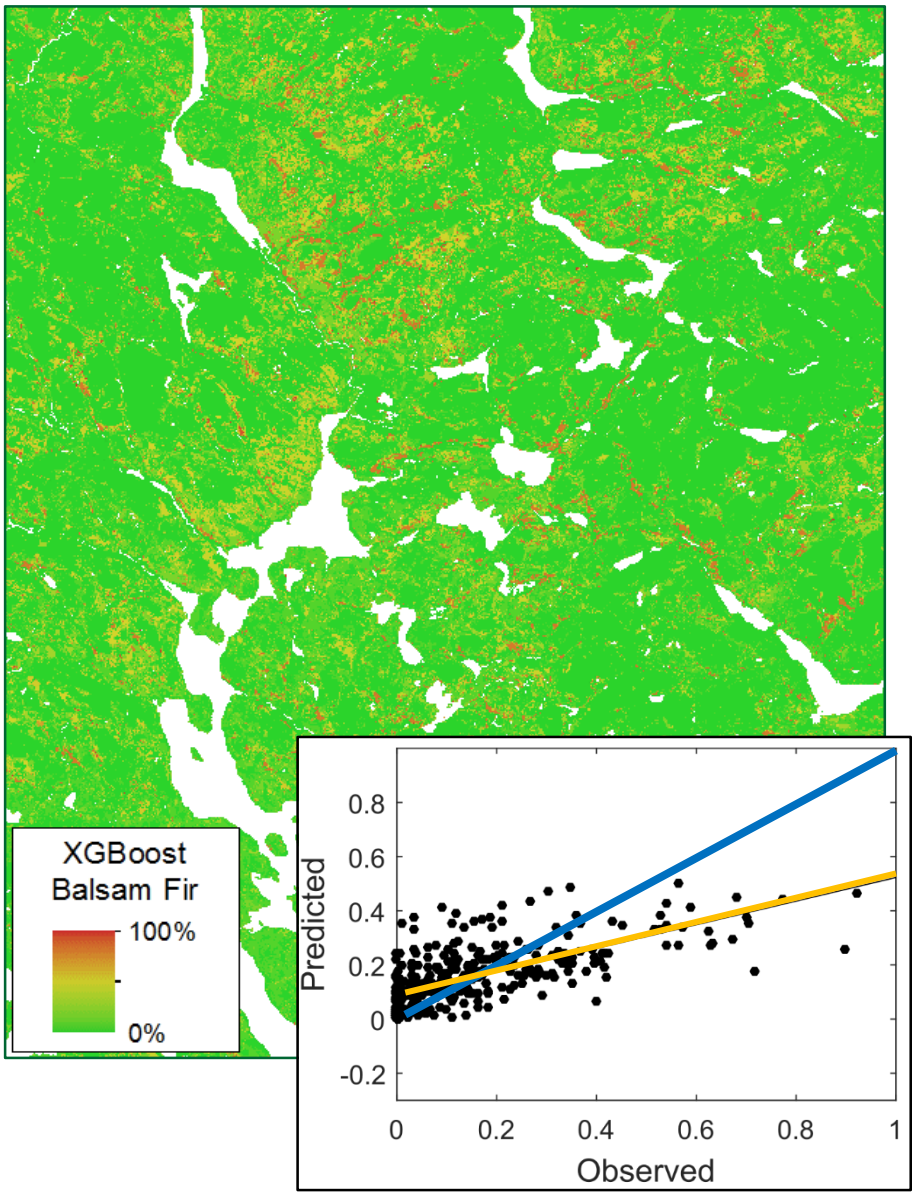


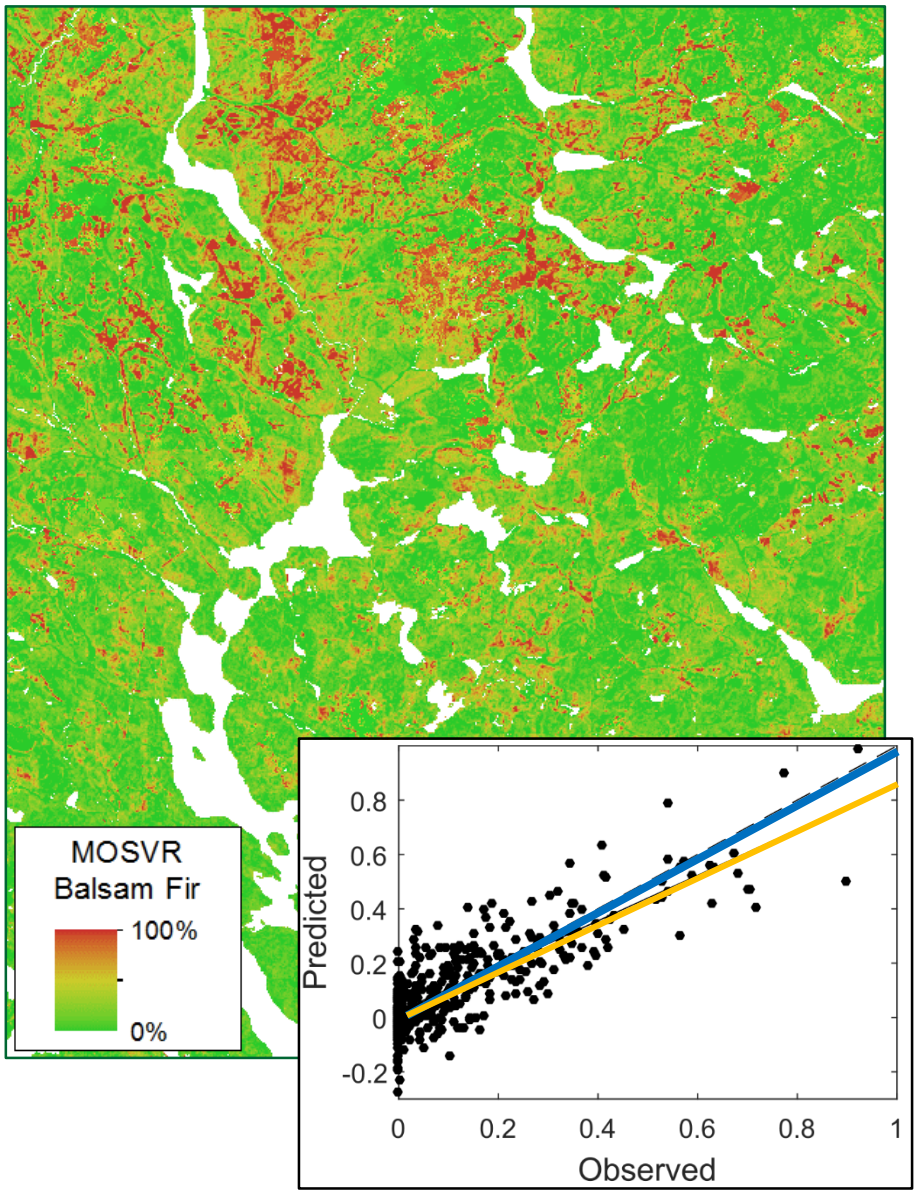
After **40** iterations

Balsam fir % biomass on FIA plots

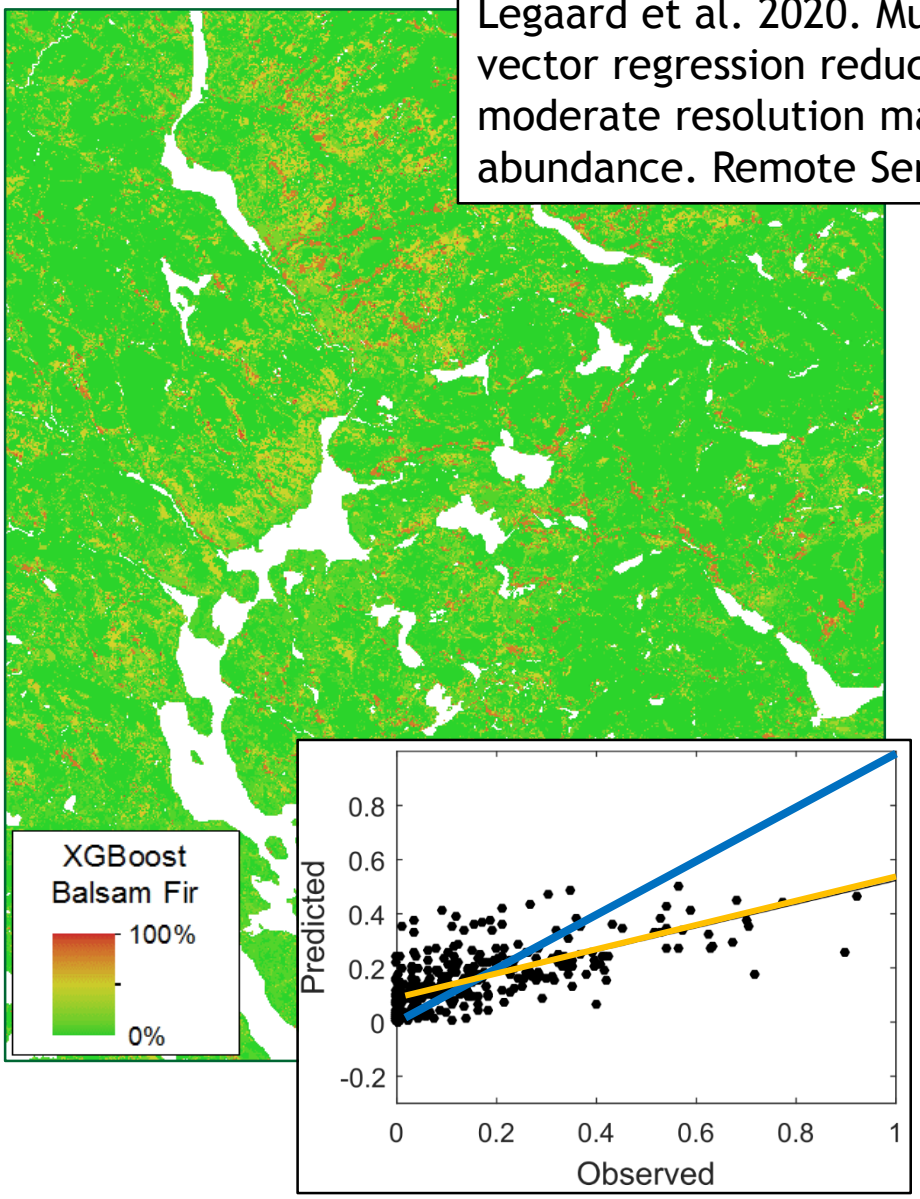


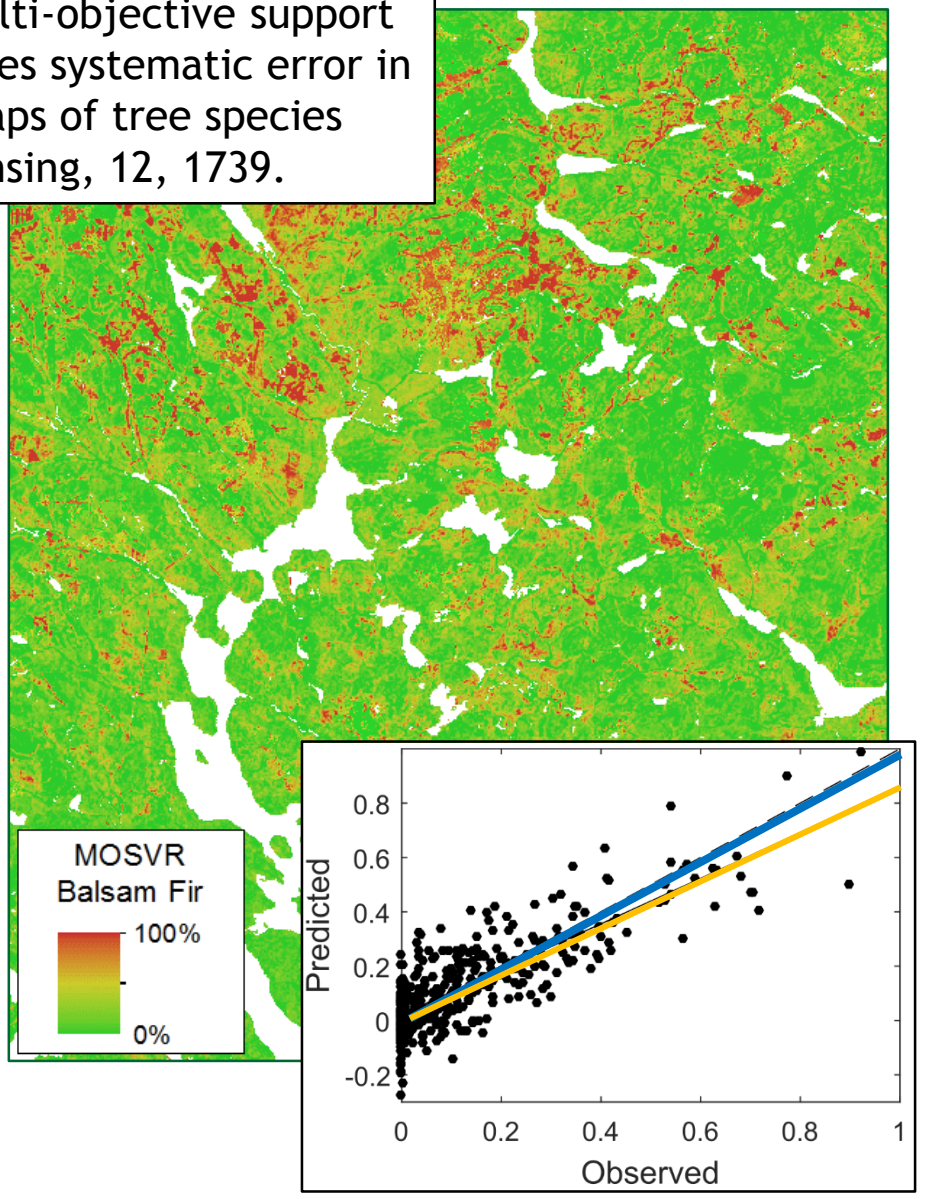
After **80** iterations


XGBoost
Balsam Fir
100%
0%
Predicted
Observed
0
0.2
0.4
0.6
0.8
1
-0.2


MOSVR
Balsam Fir
100%
0%
Predicted
Observed
0
0.2
0.4
0.6
0.8
1
-0.2

Legaard et al. 2020. Multi-objective support vector regression reduces systematic error in moderate resolution maps of tree species abundance. Remote Sensing, 12, 1739.

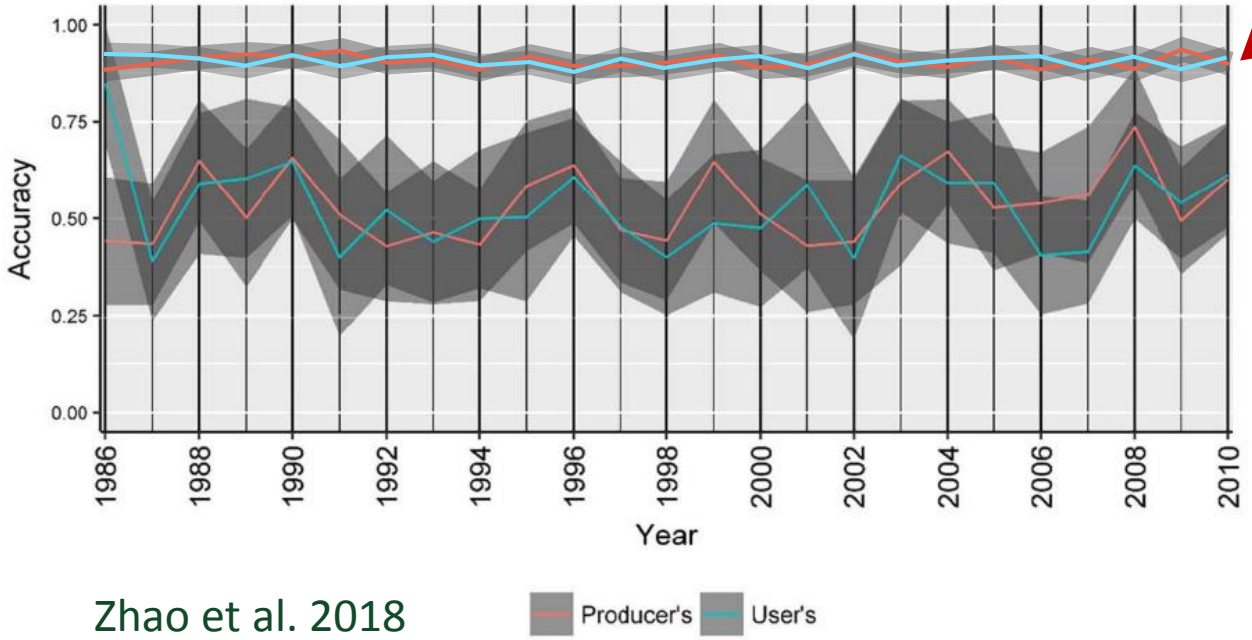




Minimization of bias in change detection:

- Simultaneous minimization of omission and commission error

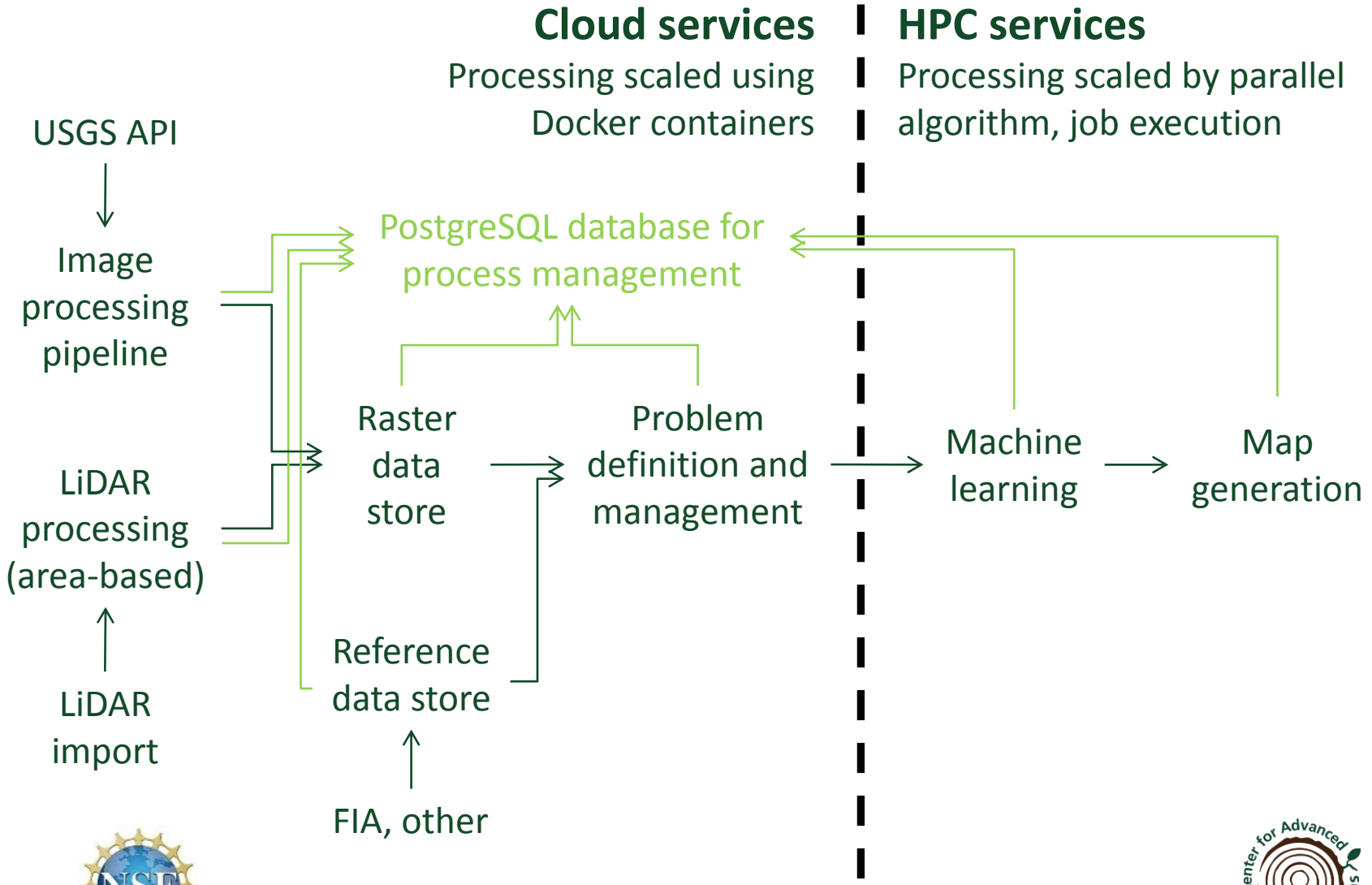
Vegetation Change Tracker, annual Landsat:



Zhao et al. 2018

Our target:
UA = PA = 90+%

Production software (UMaine ACG)



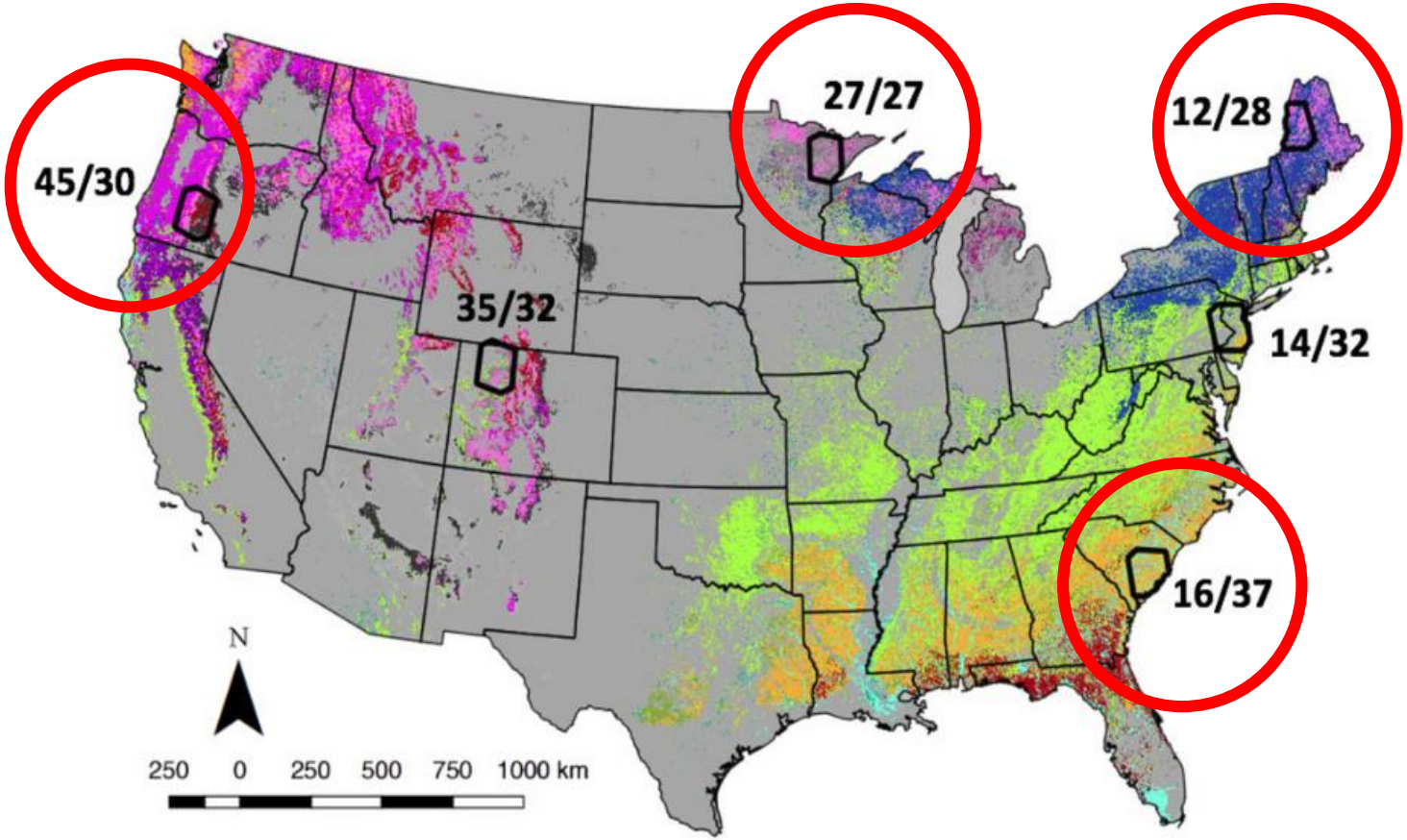
Objectives, year 1:

1) Develop and validate maps of tree species relative abundance at 10 m pixel resolution;
    a) Sentinel 2 imagery processed at 10 m pixel resolution
    b) Integrated use of Sentinel 2 and Landsat 8 imagery (10 m) to improve resolution of vegetation phenology
    c) Two-stage species occurrence/abundance modeling
2) Implement and validate alternative approaches to pooling species-level predictions into a satellite-derived canopy composition map;
3) Develop and validate annual time series of forest disturbance, 1985-present;

Objectives, years 2-3:

4) Implement and validate alternative approaches to aboveground biomass estimation:
   a) Multi-objective ML and Sentinel/Landsat plus disturbance (e.g., timing/intensity of disturbance)
   b) NN imputation of plot data based on multi-objective ML predictions of species/disturbance
   c) Multi-objective ML and commercial LiDAR, Sentinel/Landsat
5) Apply and validate all methods across four large study areas in the northwest, northcentral, northeast, and southeast regions.

Multiregional application and validation, using NASA Carbon Monitoring System data for benchmarking (CMS 2013, Cohen et al.)



Budget (annual, over 3 years):

1)	Salary + fringe:	$28,968
	• 2 months, K. Legaard	
	• 1 month, L. Whitsel	
2)	High-performance computing expenses	$5,000
3)	Travel (CAFS IAB meetings)	$1,500
	Annual total	$35,468
	Project total	$106,406

Deliverables / Company benefits:

1) Validation and refinement of multi-objective ML and remote sensing workflows across regions, forest conditions:
    a) Tree species
    b) Canopy composition / forest type
    c) Timing, intensity of disturbance
    d) Aboveground biomass, by species
2) Full integration of workflows into production software, reducing production time and cost
3) Annual, final reports/presentations
4) Scientific publications (multiregional validation, ML algorithm refinement and comparisons, fusion of LiDAR / multispectral)