UMaine Intelligent GeoSolutions

Kasey Legaard

UM, Research Professor, Geospatial Analytics and Machine Learning

Erin Simons-Legaard

UM, Research Professor, Forest Landscape Modeling

Aaron Weiskittel

UM, Director, Center for Research on Sustainable Forests

IGS overview

Applied geospatial research and technology transfer initiative

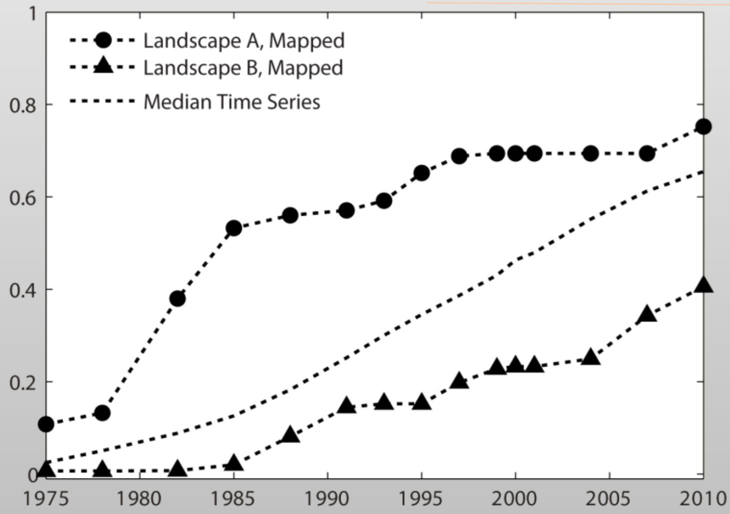
Purpose: Bring forest maps to market to support R&D and data production targeted to user needs

- License of map products to end users
    - One time fee
    - Through service or research agreements
    - Managed by UM Dept. of Industrial Cooperation

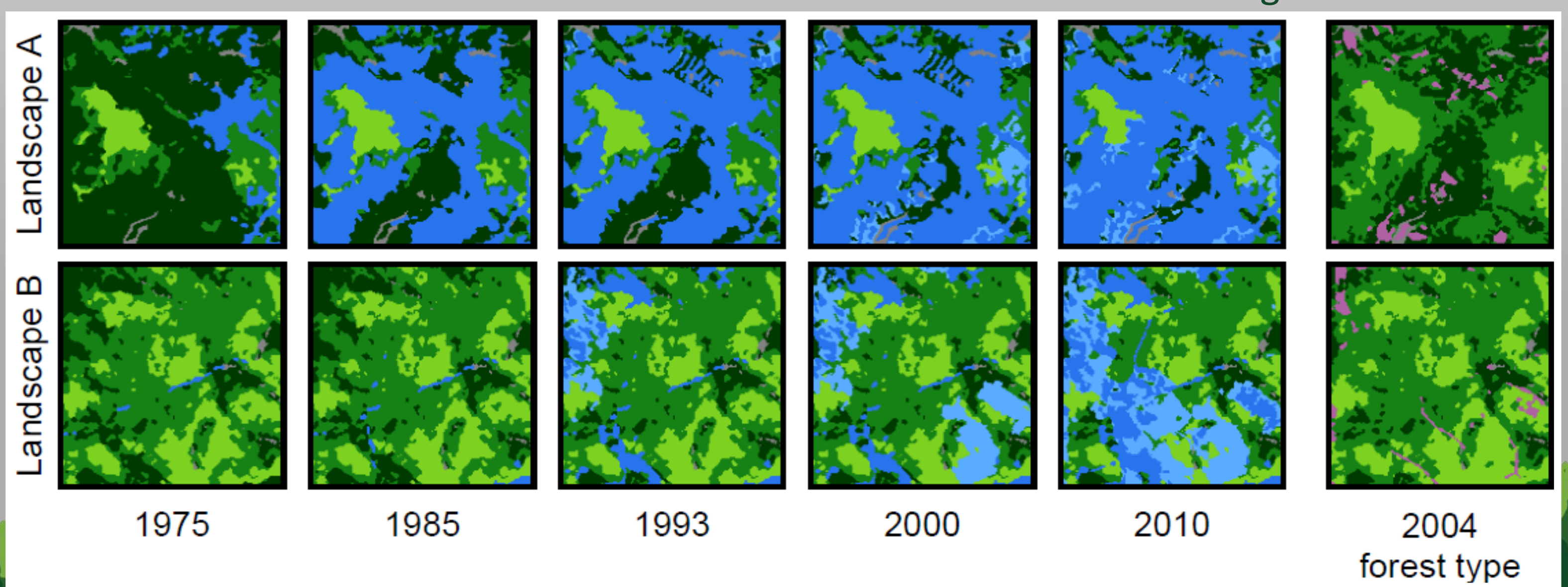
IGS origins: Maine landscape change

Photo: Maine Forest Service

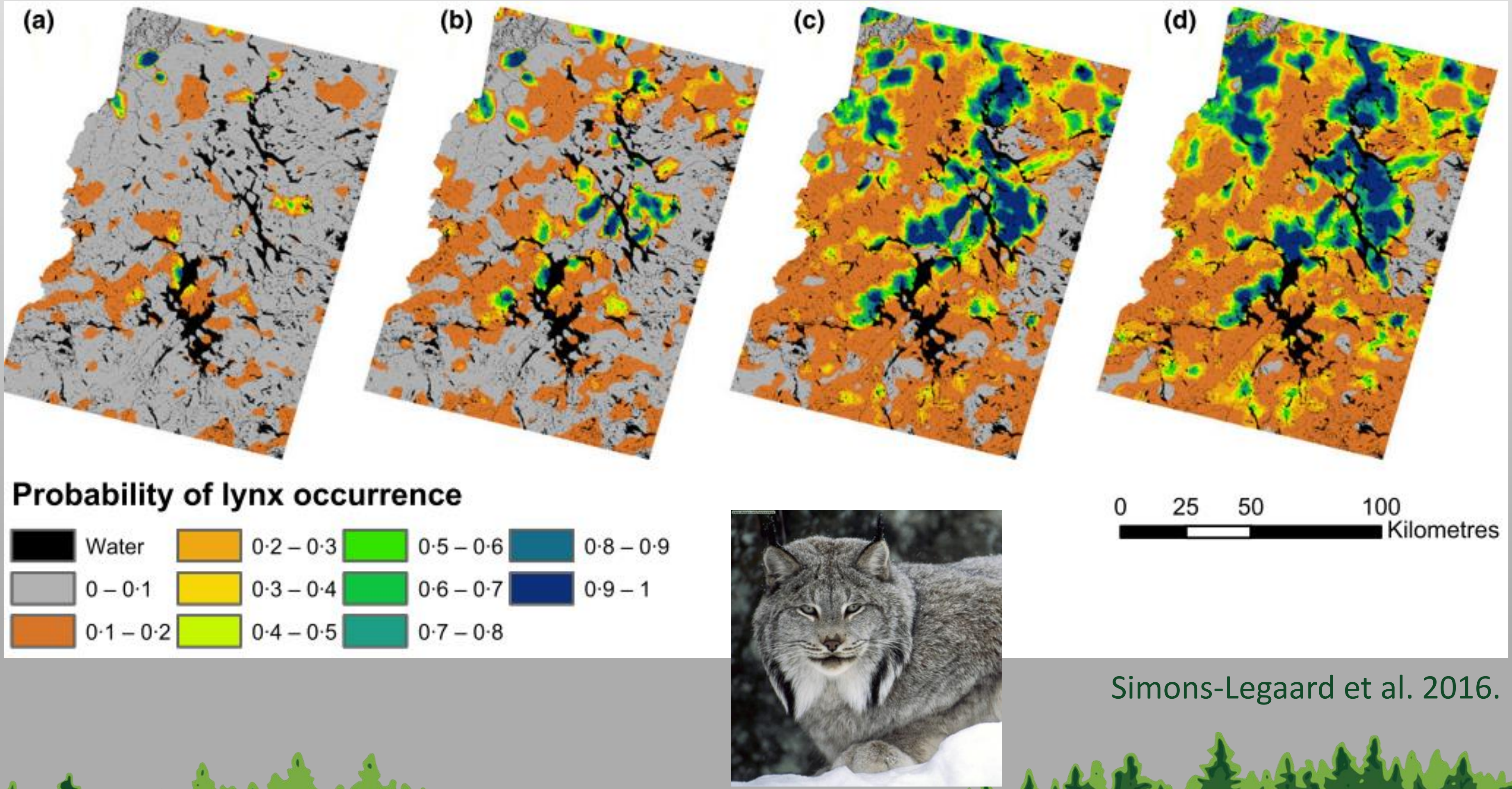
IGS origins: Maine landscape change



Legaard et al. 2015.



IGS origins: Maine landscape change

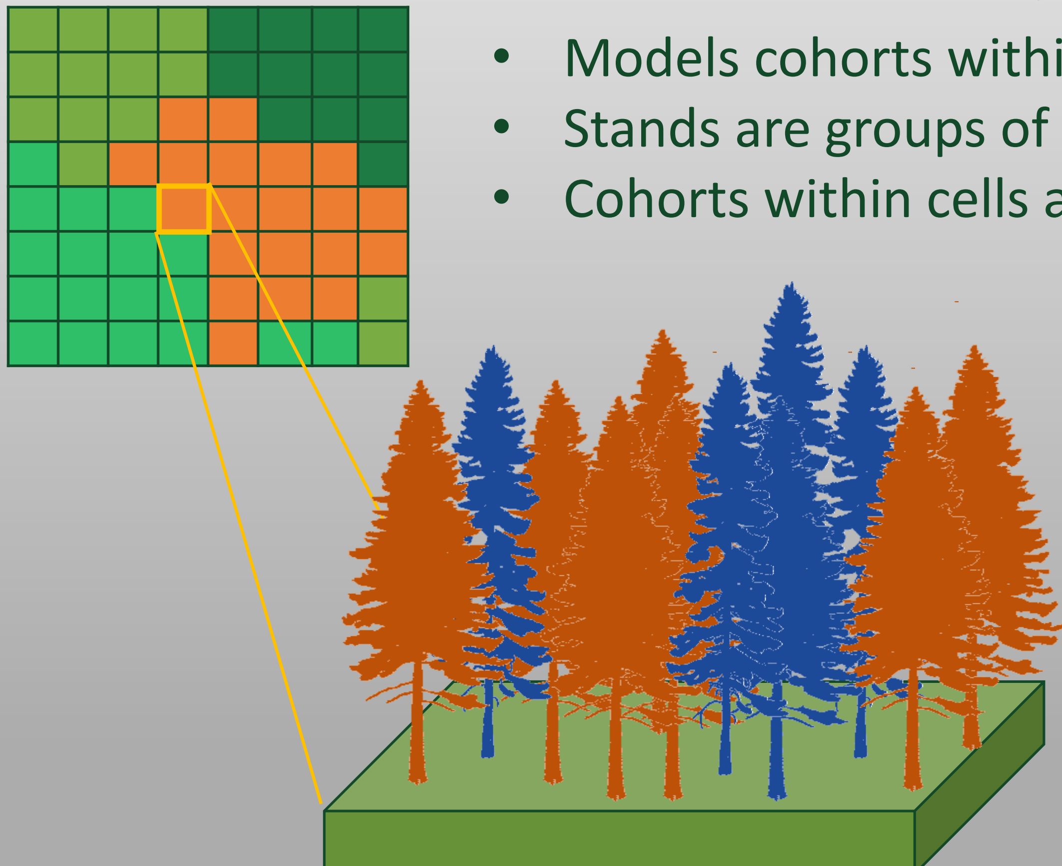


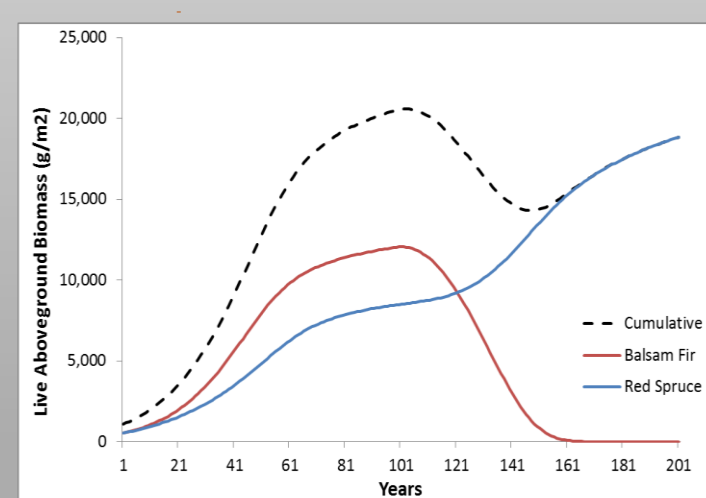
Simons-Legaard et al. 2016.

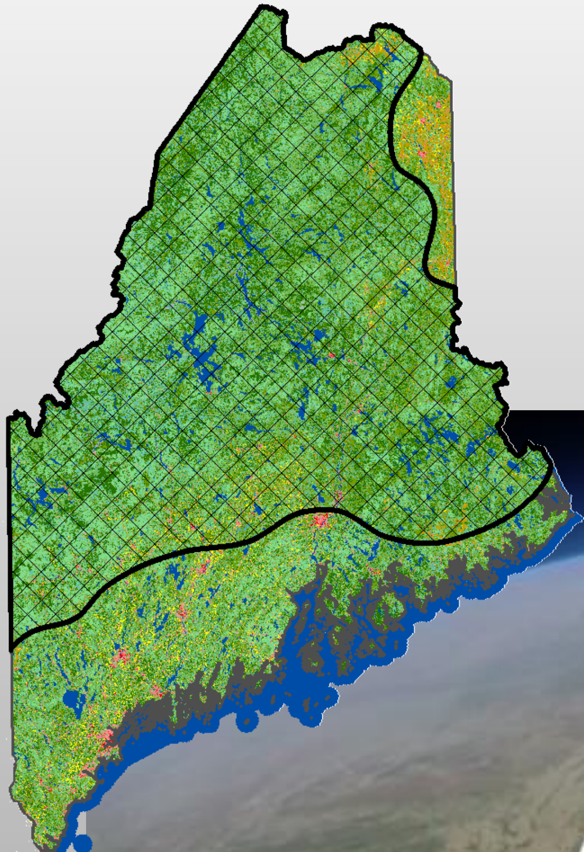
IGS origins: Maine landscape change

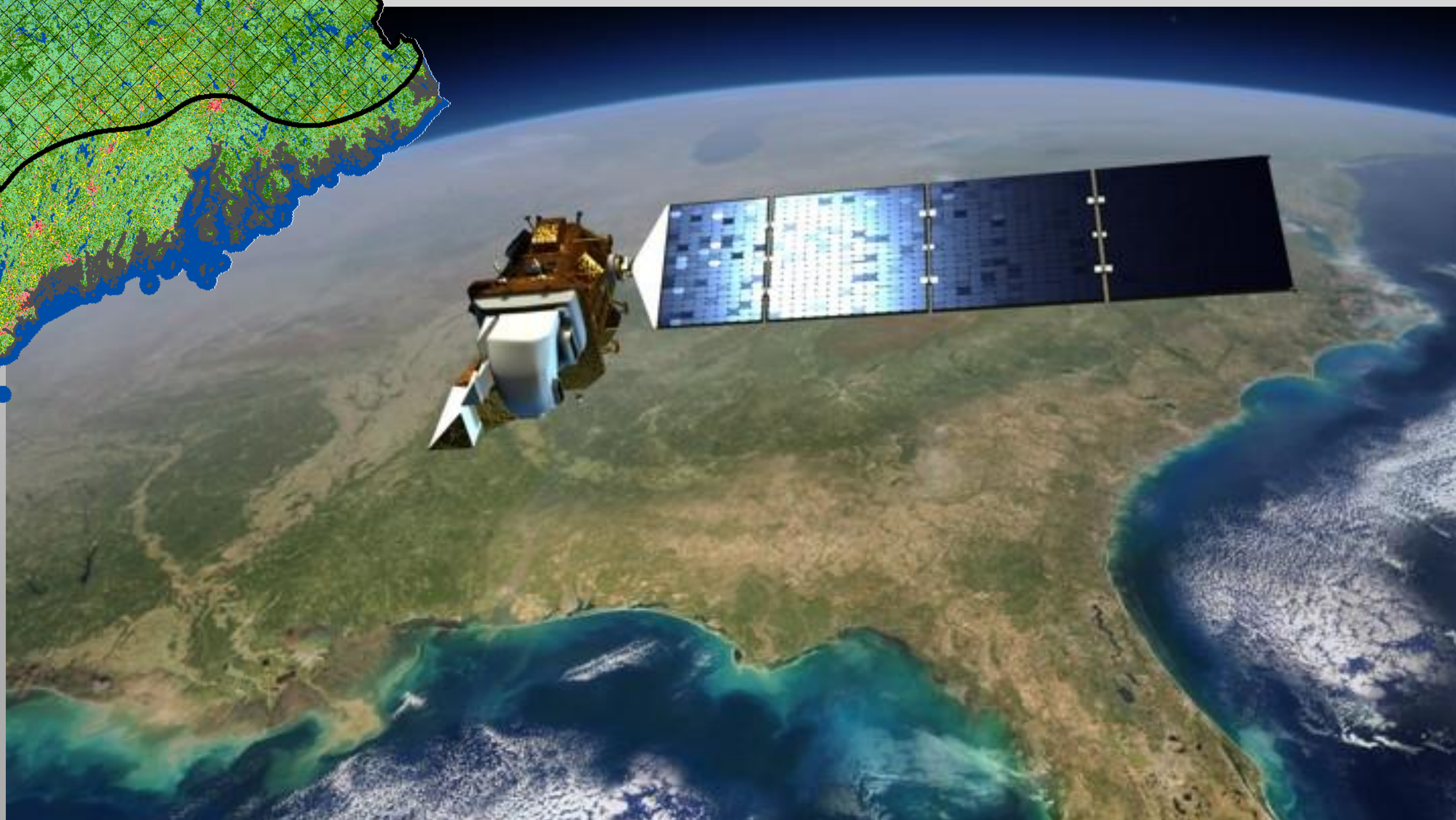
Landis-II forest landscape model:

- Models cohorts within stands
- Stands are groups of like cells
- Cohorts within cells are defined by species and age









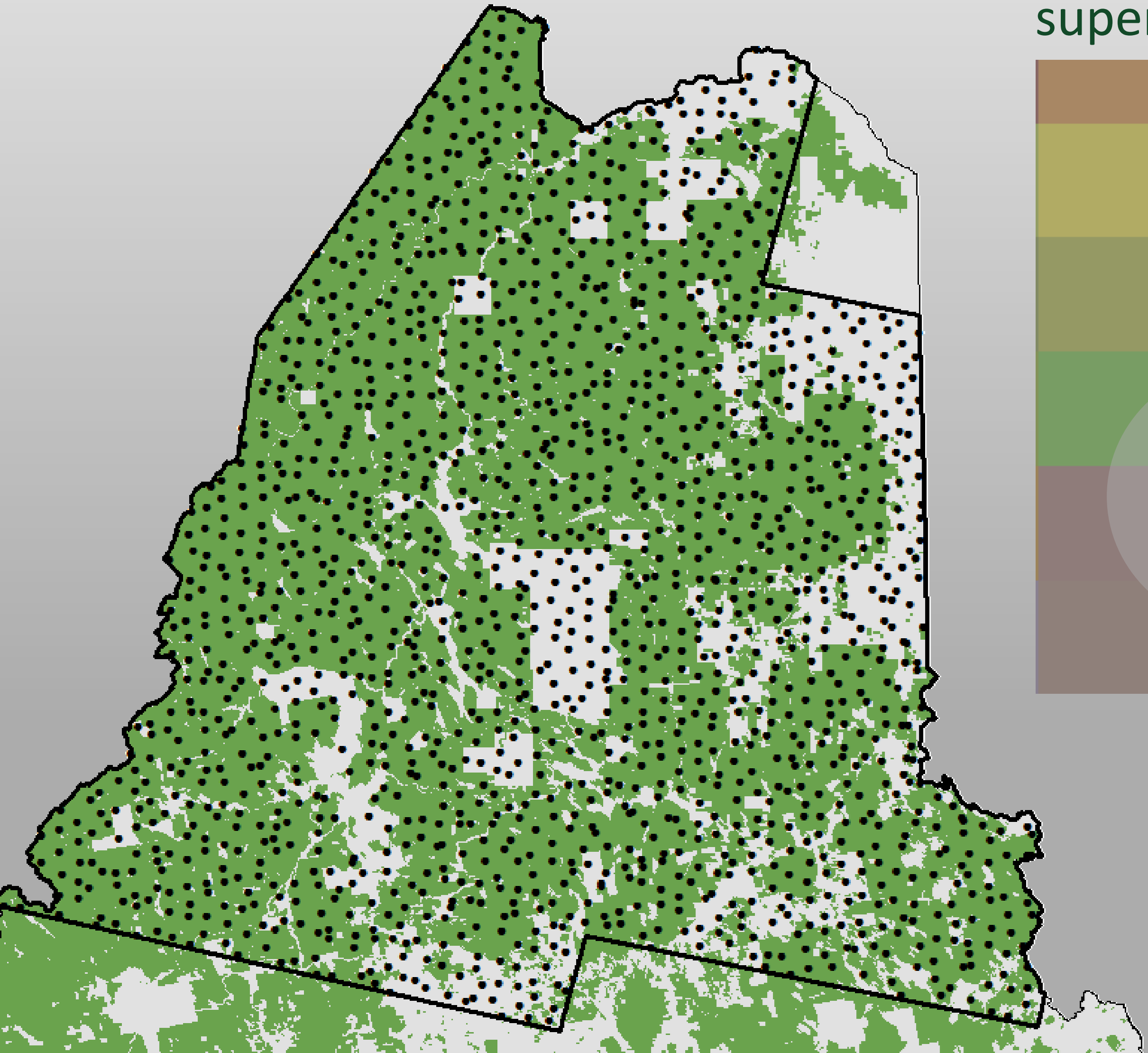
Tree species mapping

Species abundance (% biomass):

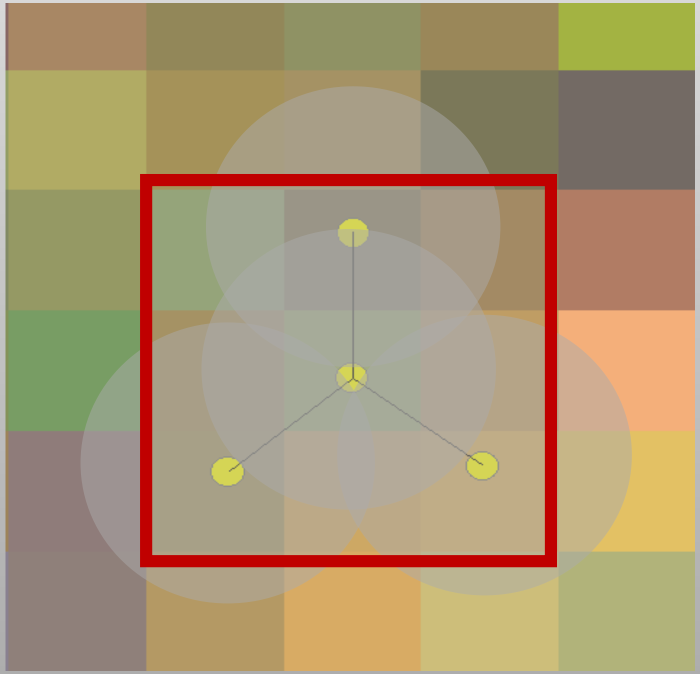
Balsam fir
Red, white, black spruce
E. White pine
N. White cedar
E. Hemlock
Sugar and Red maple
American beech
Yellow birch
Paper birch
Green ash

- 8 Landsat images, acquired late April through mid-October
- Bioclimatic variables
  (USFS RMRS, Moscow Forestry Sciences Laboratory)
- Terrain attributes
  (from National Elevation and Hydrography Datasets)

Tree species mapping



FIA plot (+ location error), superimposed over 30 m pixels:



Public plot locations shown. True plot coordinates provided through collaborative agreement with the USFS NRS FIA Program.

Tree species mapping

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 plots and pixels

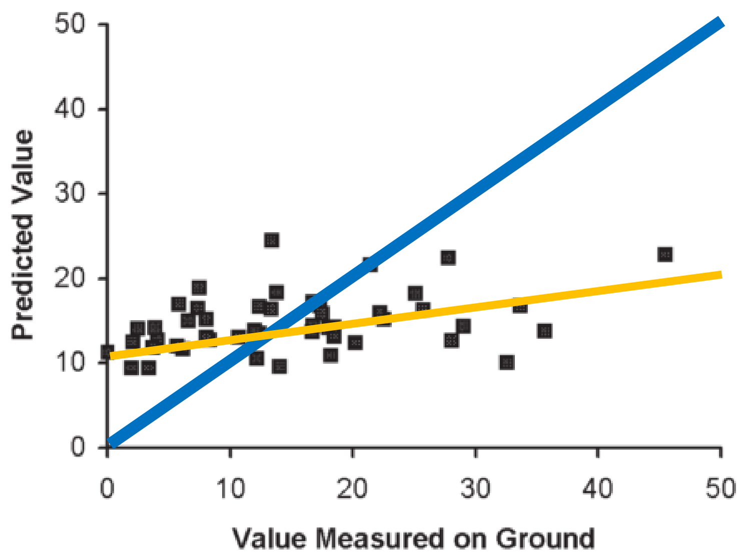


Figure 9. Plot of predicted basal area against ground measured basal area (units: m^2/ha).

Scale mismatch between plots and pixels

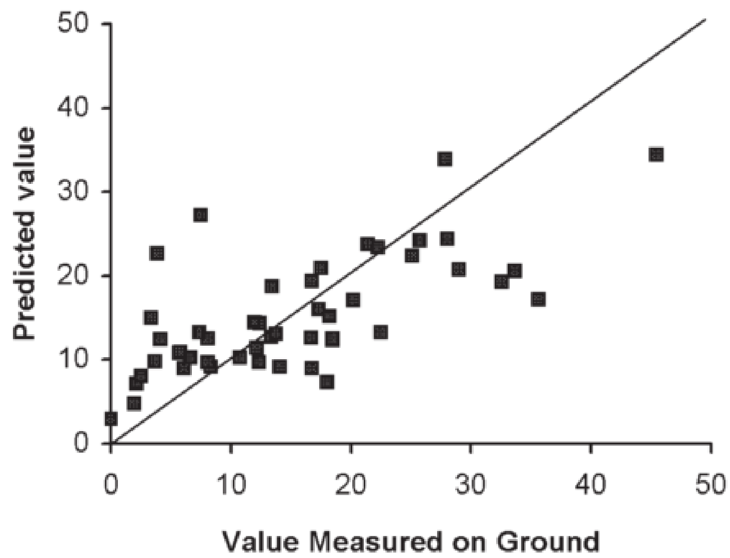
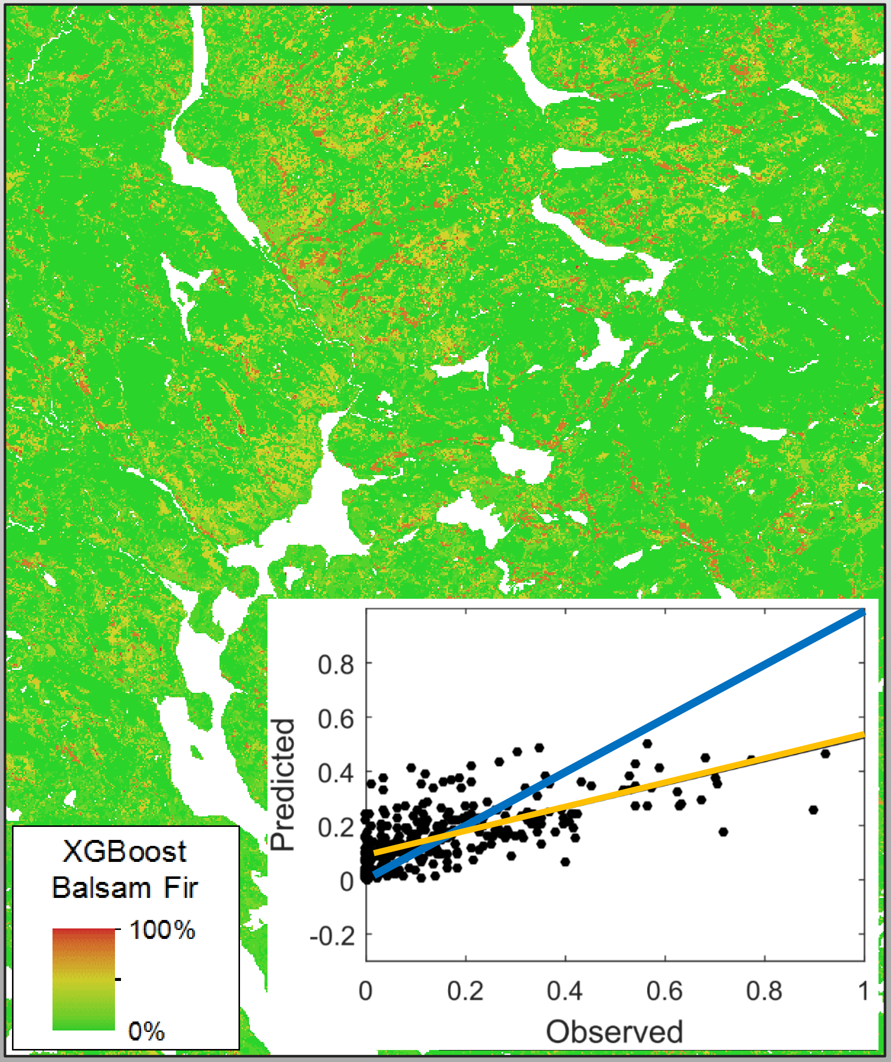


Figure 4. Plot of predicted basal area against basal area measured in ground subplot (units: m^2/ha).

Balsam fir prediction



XGBoost gradient boosting algorithm

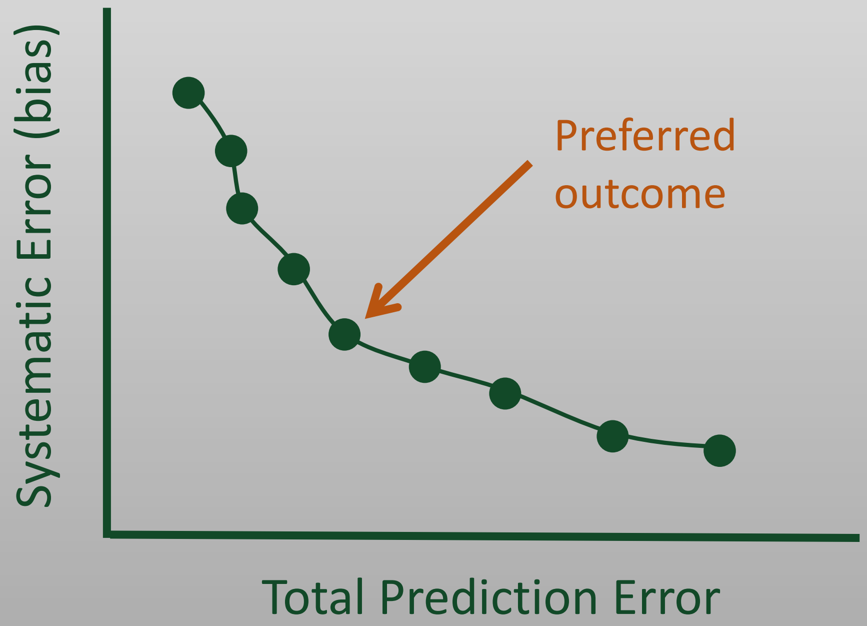
- Minimizes total error, in part by predicting values nearer to the mean

= attenuation bias

Multi-objective machine learning

Minimization of total prediction error leads to high bias when fitting models subject to uncertainty in predictor variables

We would prefer low error *and* low bias



Multi-objective machine learning

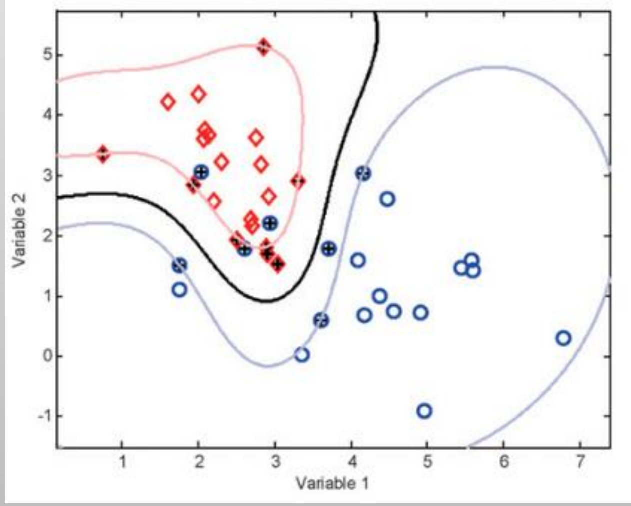
Simultaneous minimization of total prediction error and bias:

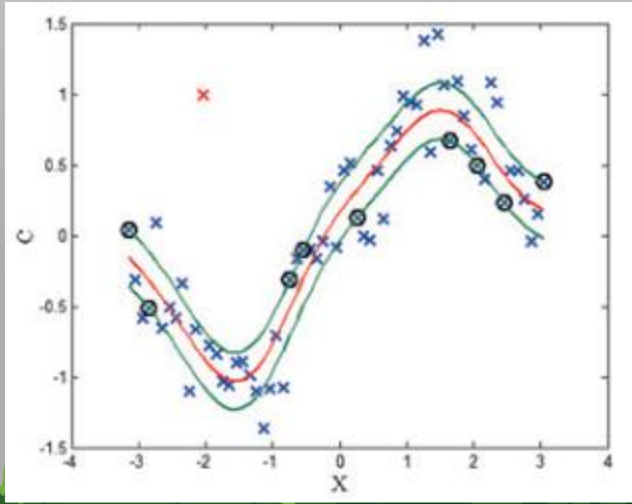
Use a flexible machine learning algorithm capable of fitting complex relationships

- Support Vector Machines (SVMs)

Use a training process capable of evaluating a lot of different SVMs

- Genetic Algorithms (GAs)





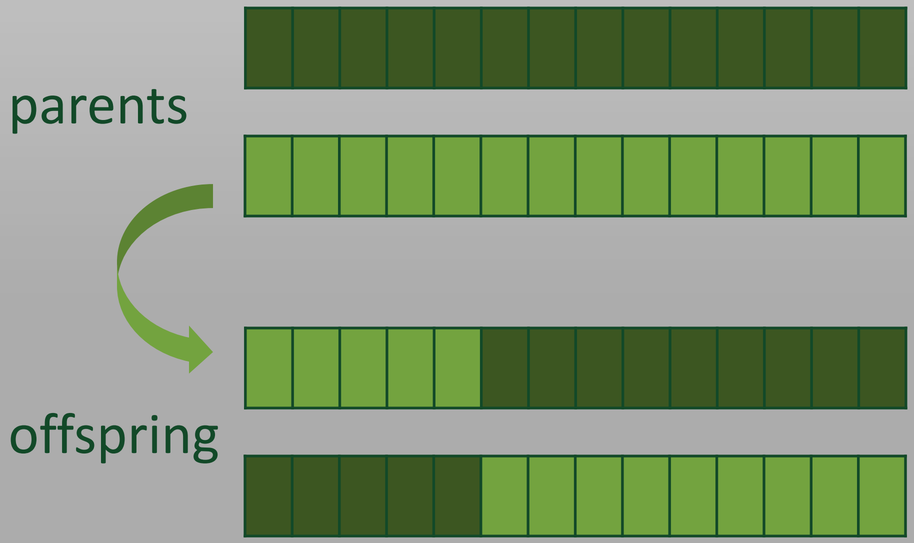
Brereton and Lloyd. 2010.

Multi-objective machine learning

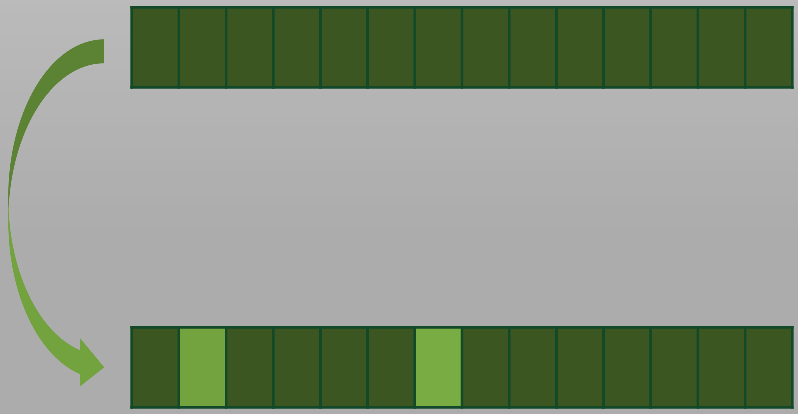
GA operates on a population of SVMs

- e.g., 1000 individual models, each competing to reproduce
- Survival of the fittest, where fitness is determined by overall prediction error and bias

Genetic recombination:



Genetic mutation:



Multi-objective machine learning

Train current generation of models

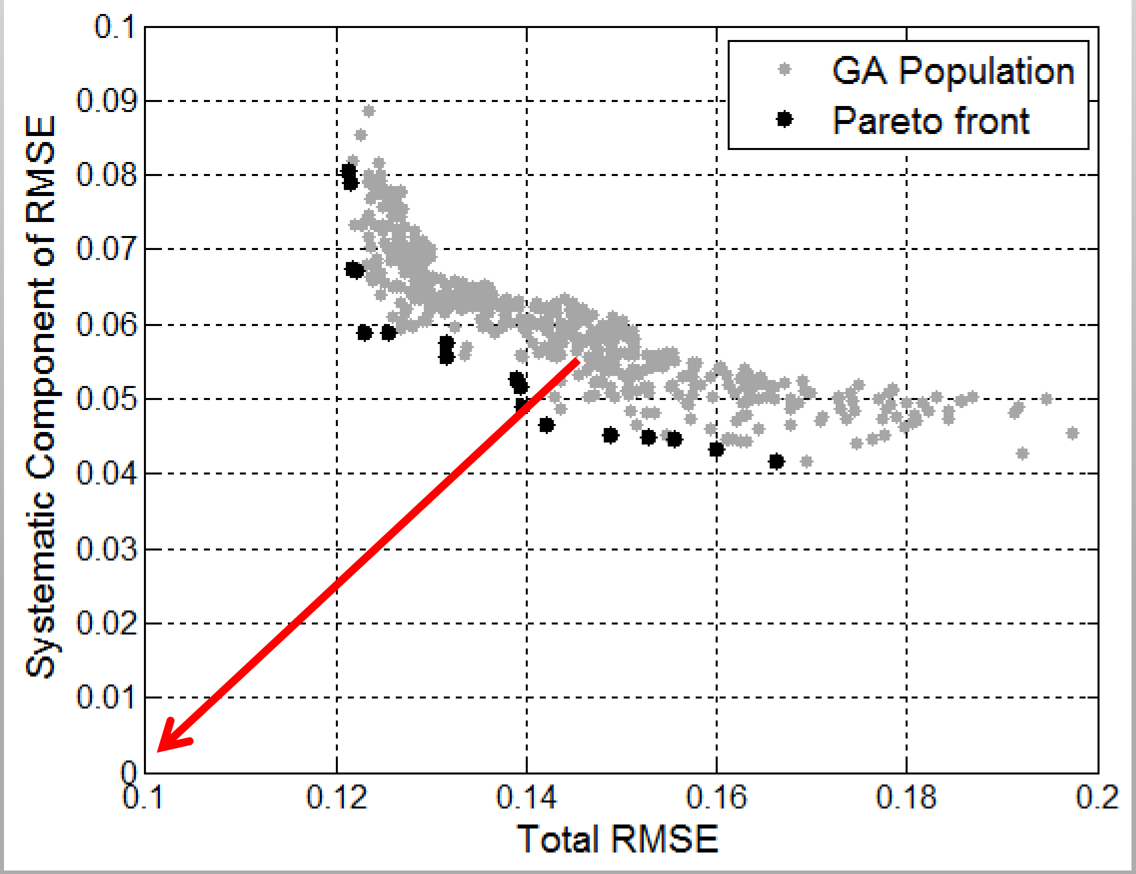
Evaluate fitness of current generation

Select parents of next generation

Apply genetic operations to obtain next generation

Multi-objective machine learning

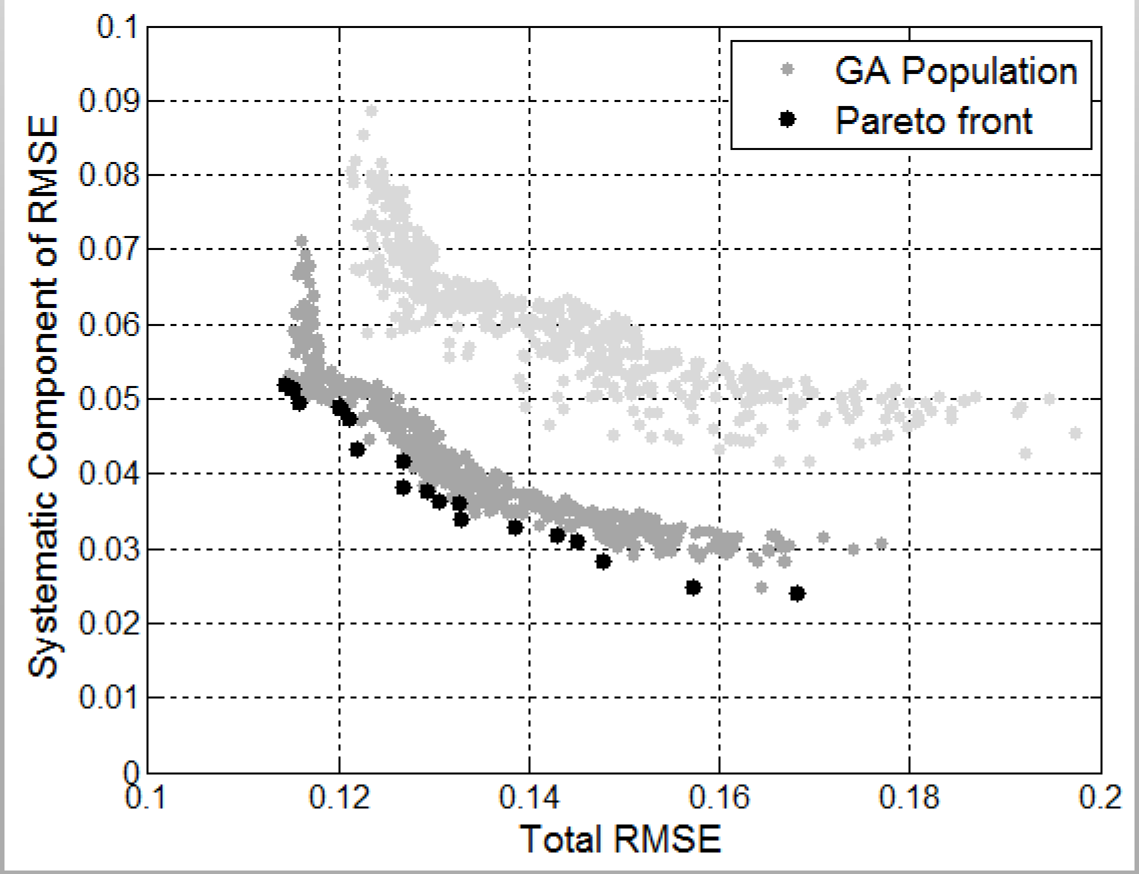
Multi-objective support vector regression –
simultaneous minimization of total and systematic RMSE:



Population status
After **10** generations

Multi-objective machine learning

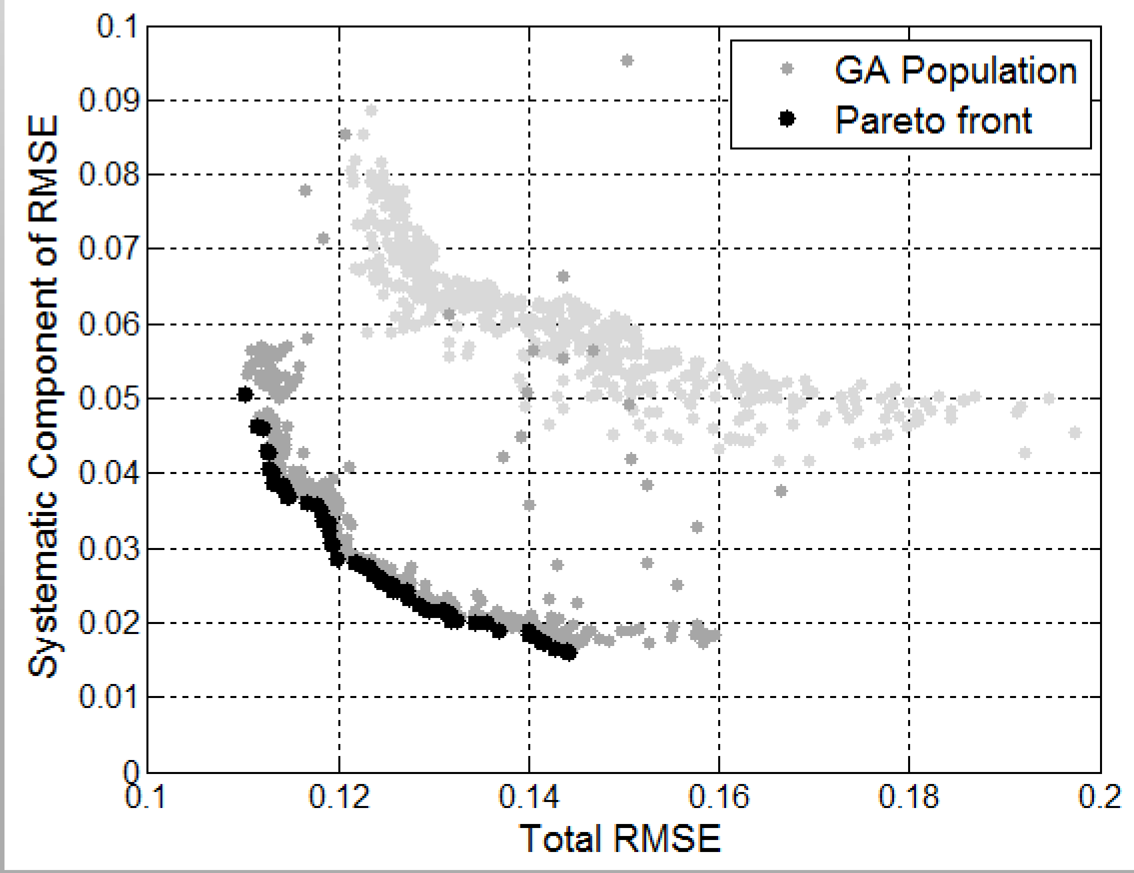
Multi-objective support vector regression –
simultaneous minimization of total and systematic RMSE:



Population status
After **20** generations

Multi-objective machine learning

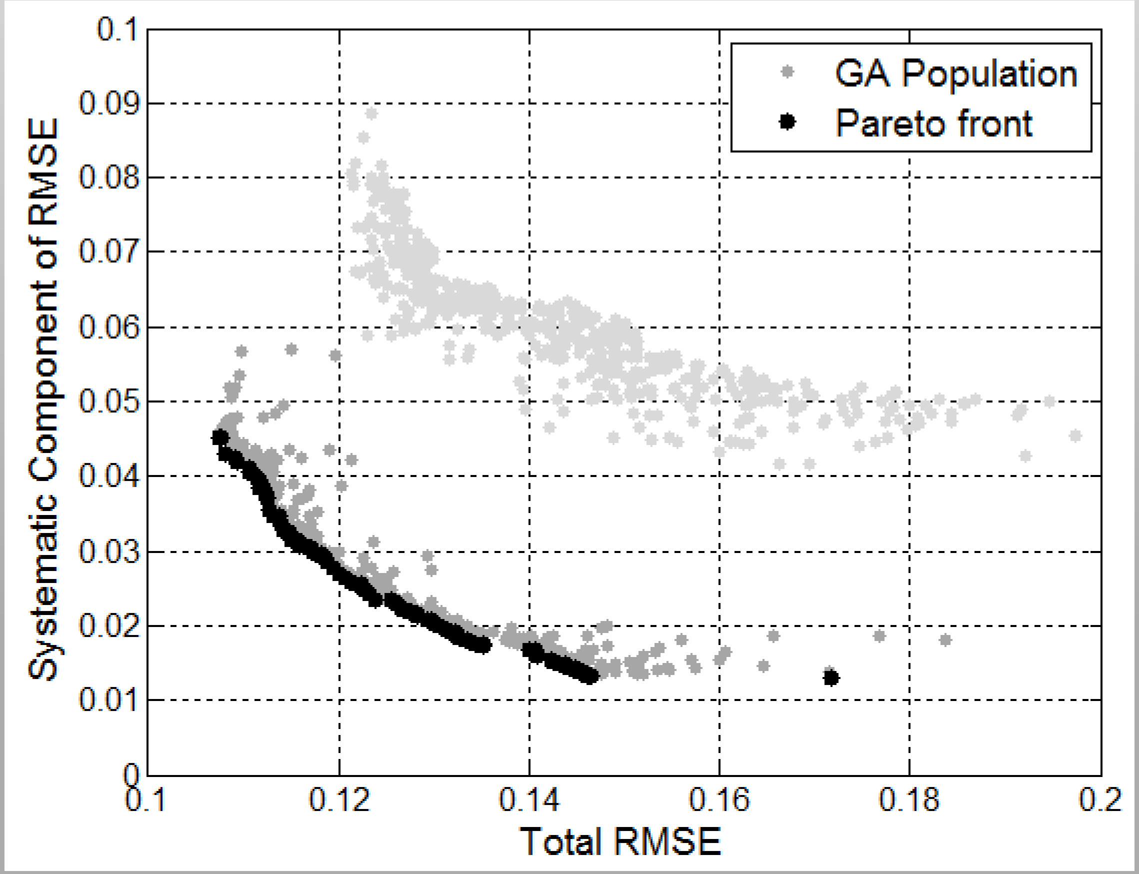
Multi-objective support vector regression –
simultaneous minimization of total and systematic RMSE:



Population status
After **40** generations

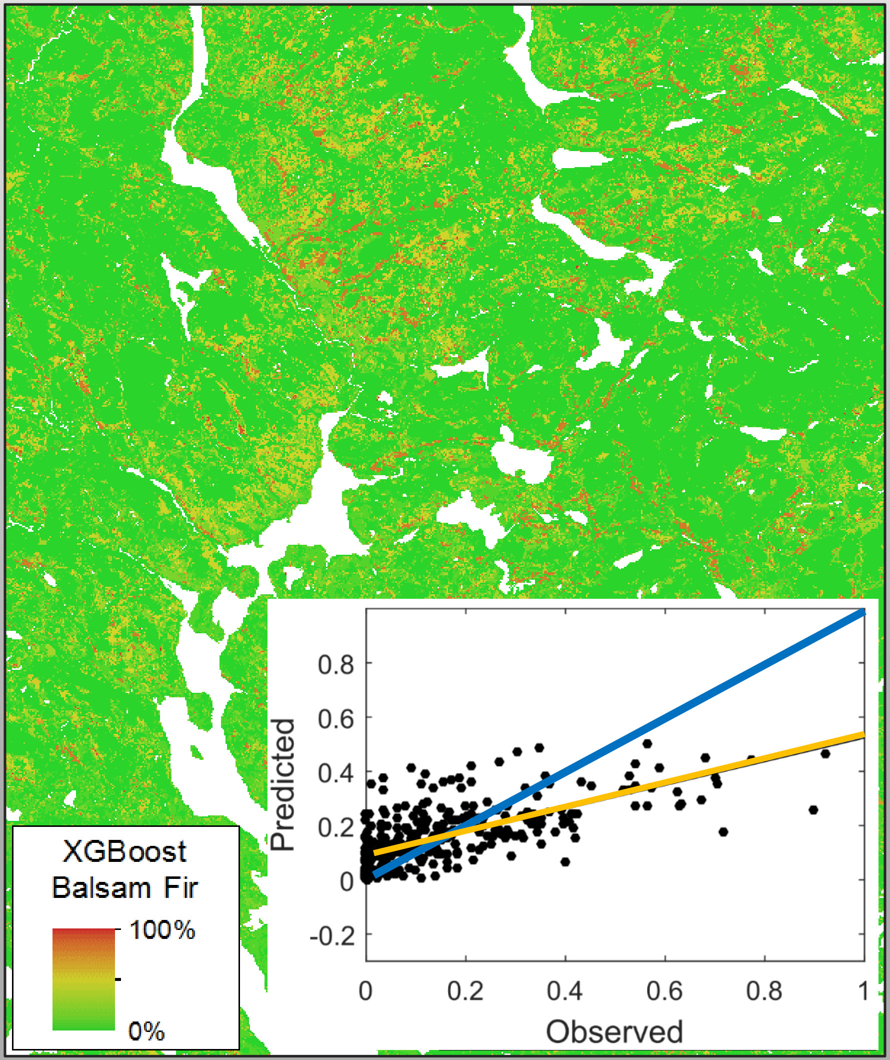
Multi-objective machine learning

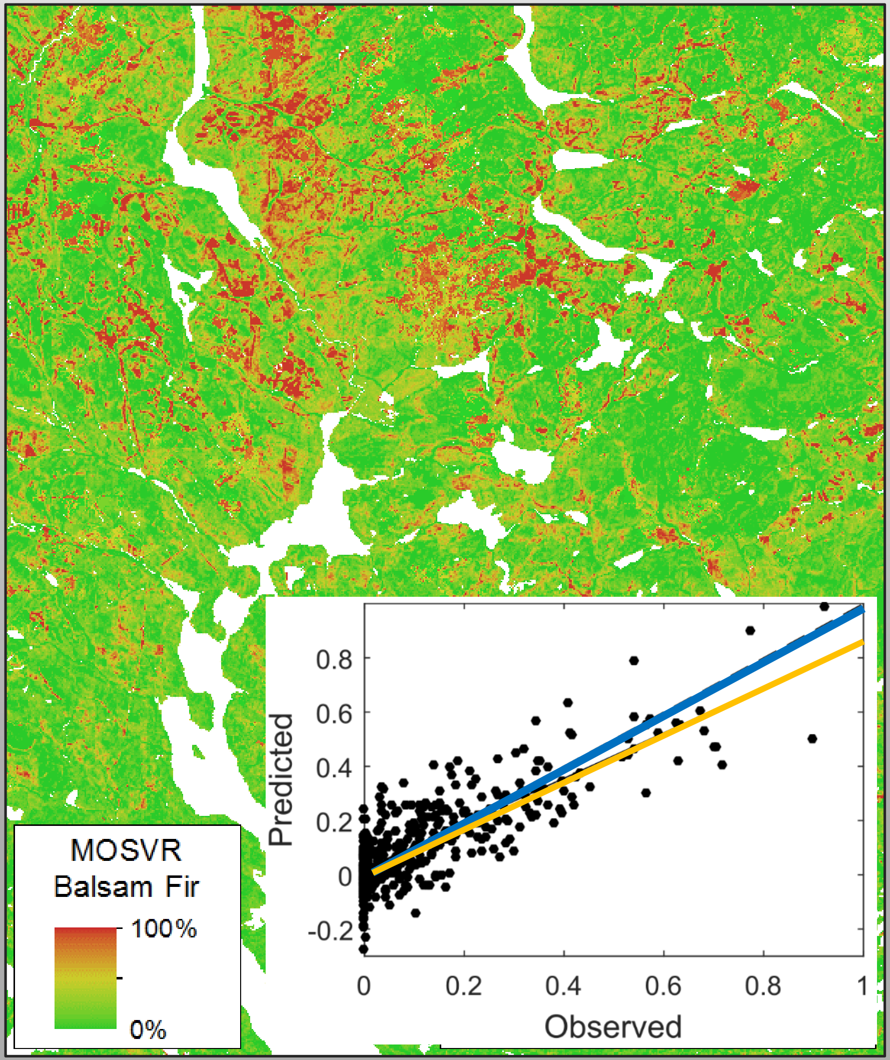
Multi-objective support vector regression –
simultaneous minimization of total and systematic RMSE:



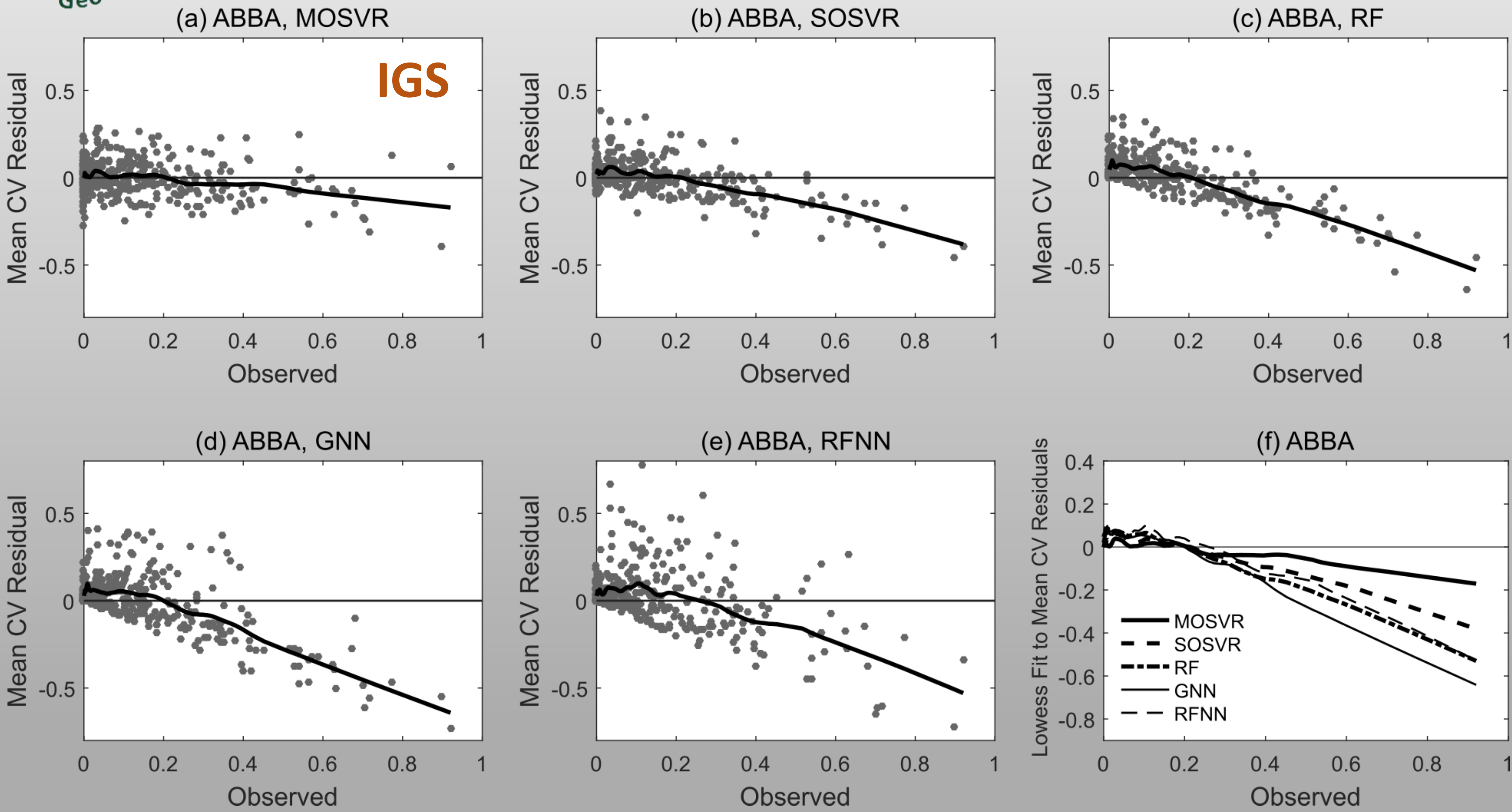
Population status
After **80** generations

Balsam fir model prediction



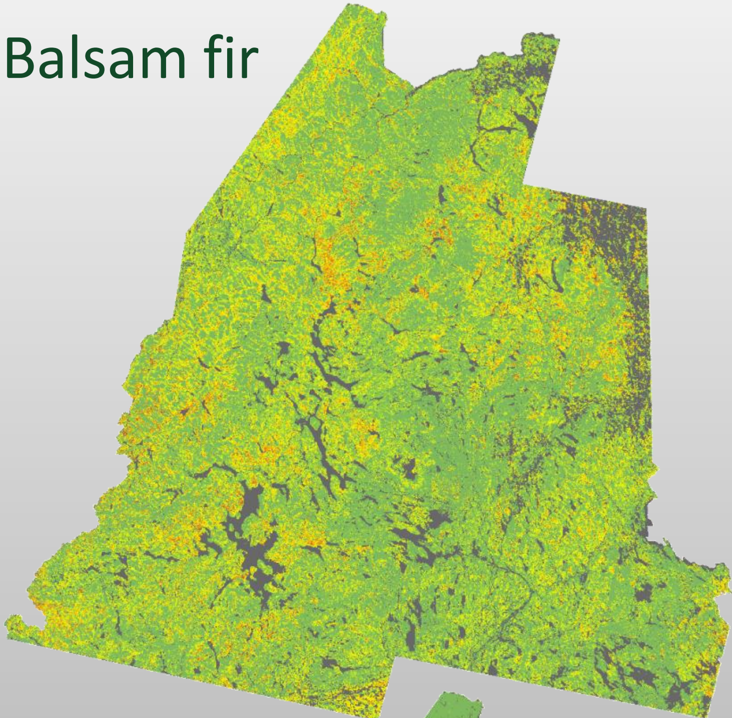


Balsam fir model comparisons

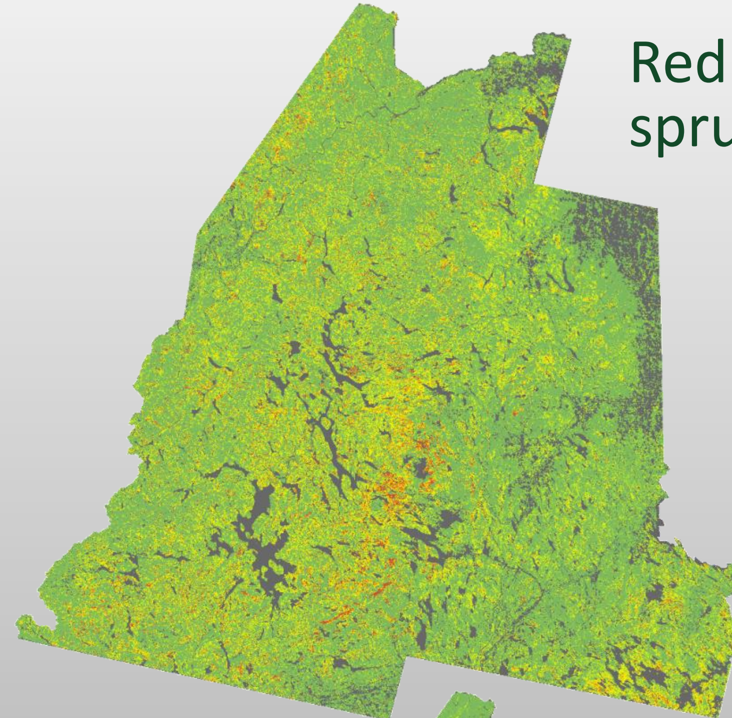


Multi-objective support vector regression maintains accuracy while reducing bias

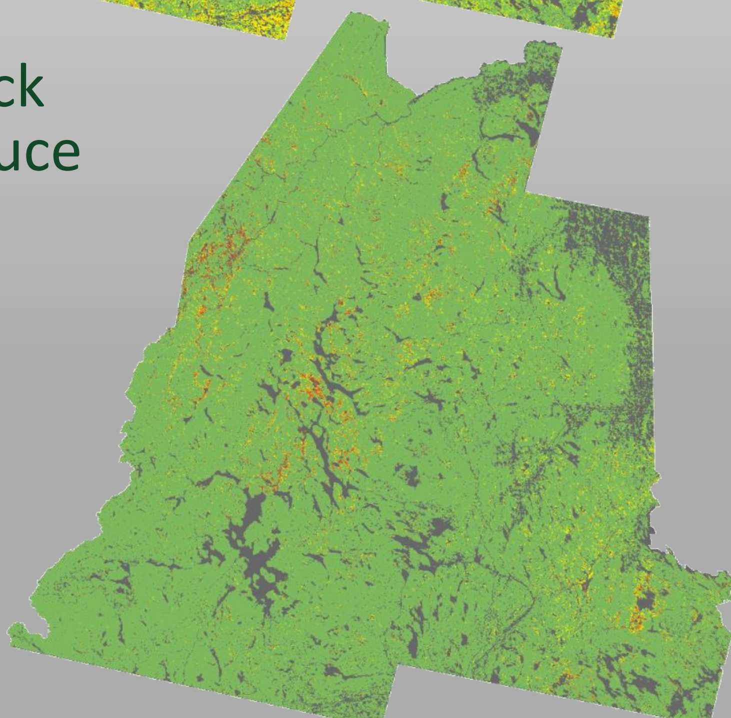
Balsam fir



Red spruce



Black spruce



White spruce

Forest disturbance classification

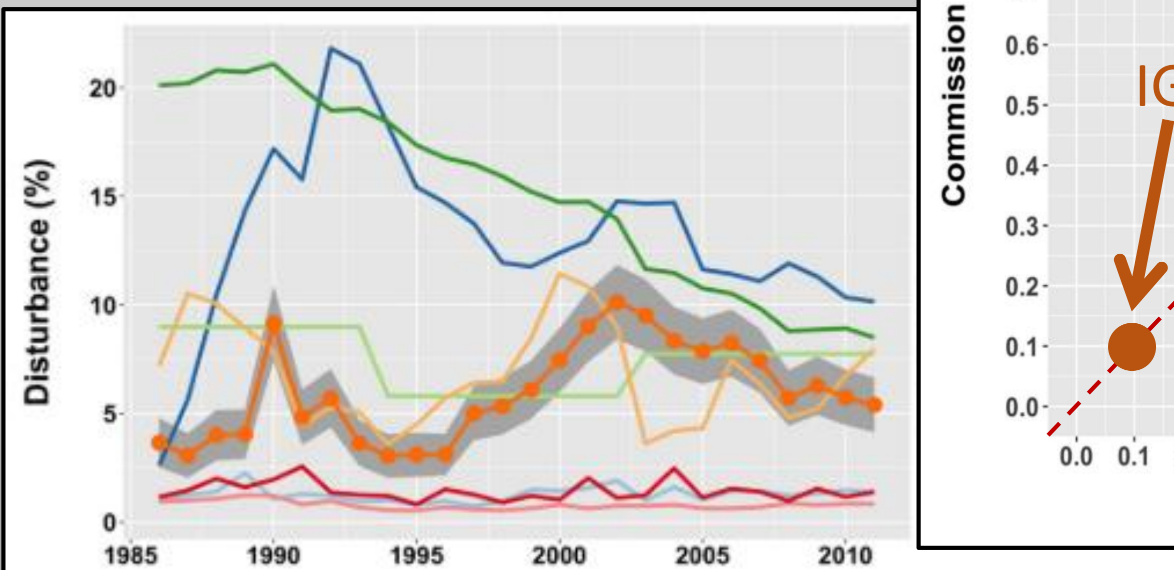
Accuracy achieved (Landsat imagery):

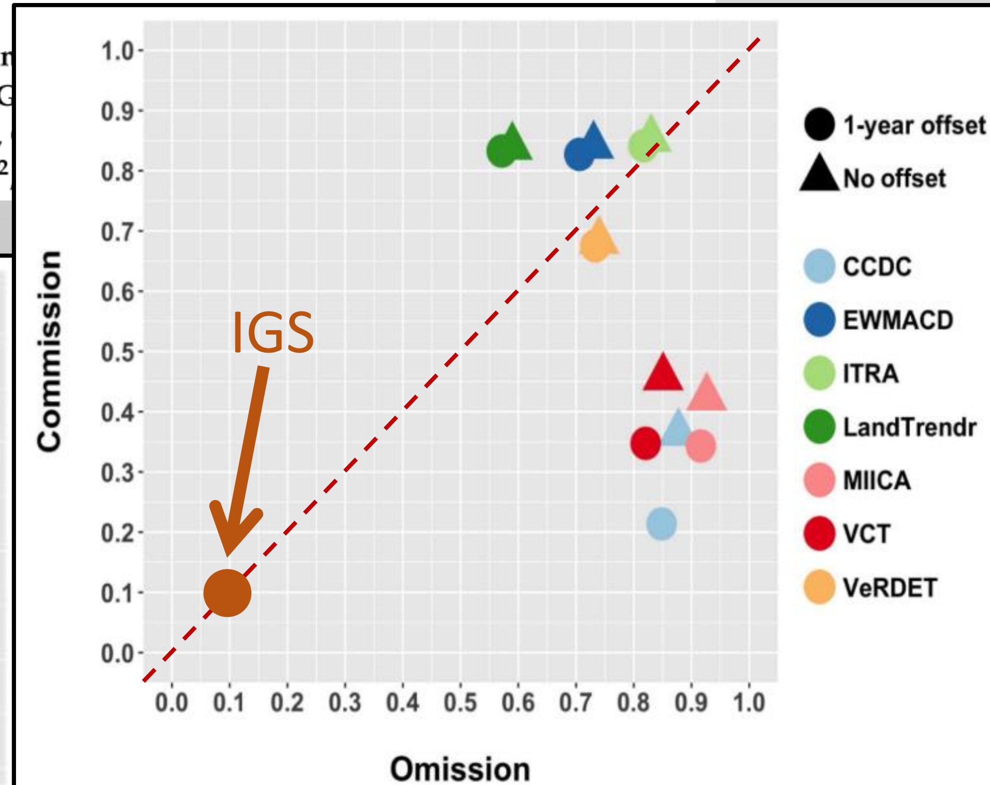
- Overall accuracy >97%
- Change class accuracy >90%, with no prediction bias

Forest disturbance classification error

How Similar Are Forest Disturbance Maps Derived from Different Landsat Time Series Algorithms?

Warren B. Cohen [1,*], Sean P. Healey [2], Zhiqian
C. Kenneth Brewer [5], Evan B. Brooks [6], Noel G
Robert E. Kennedy [9], Thomas R. Loveland [10],
James E. Vogelmann [10], Curtis E. Woodcock [12]





Forest disturbance classification

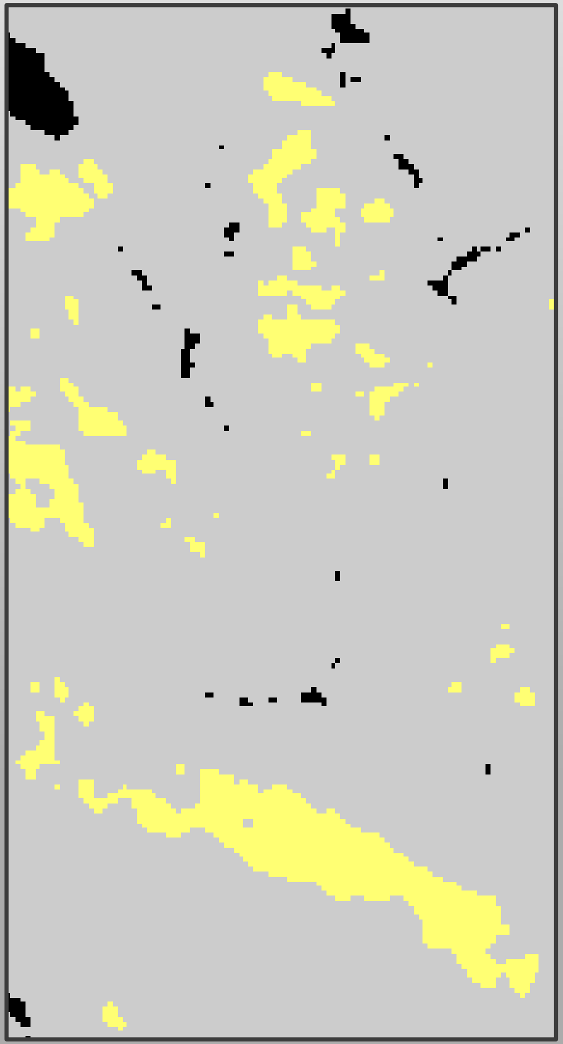
Training/validation samples

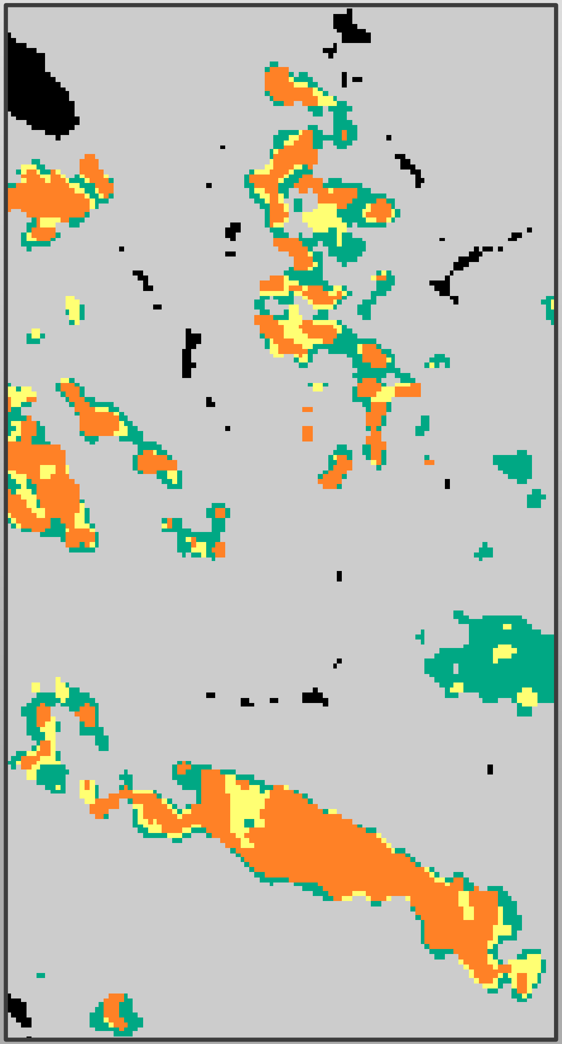
High accuracy and control of omission and commission error requires a large number of reference samples

Traditional approach relies on visual image interpretation

Multi-objective outcomes enable a semi-supervised approach where the machines do most of the work

Forest disturbance classification





User's accuracy = 98%
Commission error = 2%

User's accuracy = 90%
Producer's accuracy = 90%

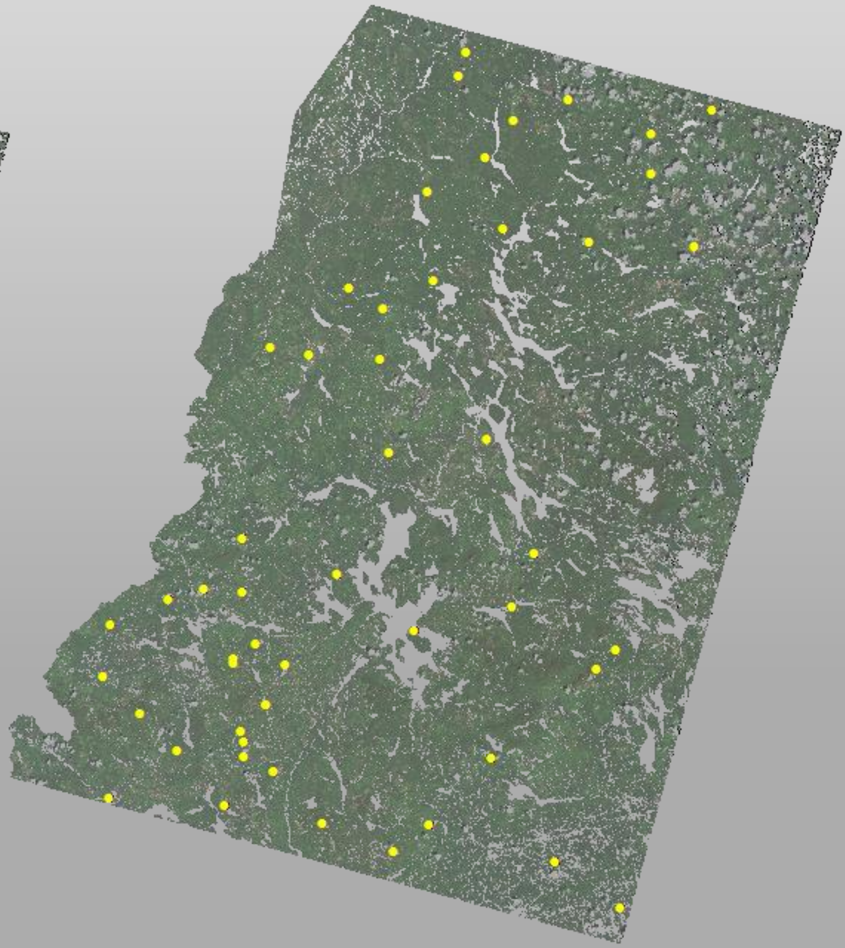
Producer's accuracy = 98%
Omission error = 2%

Compare multiple Pareto-optimal maps, and focus attention where maps disagree

Forest disturbance classification

Training/validation samples

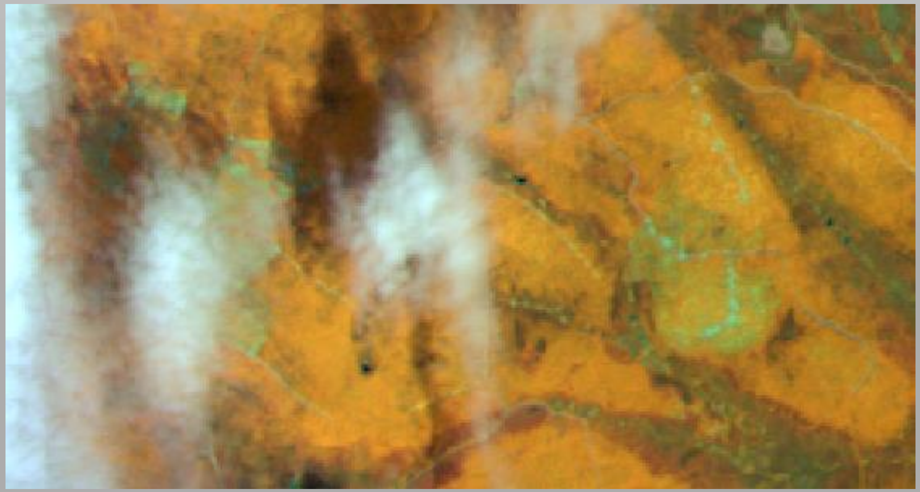
Human interpreted training/validation samples

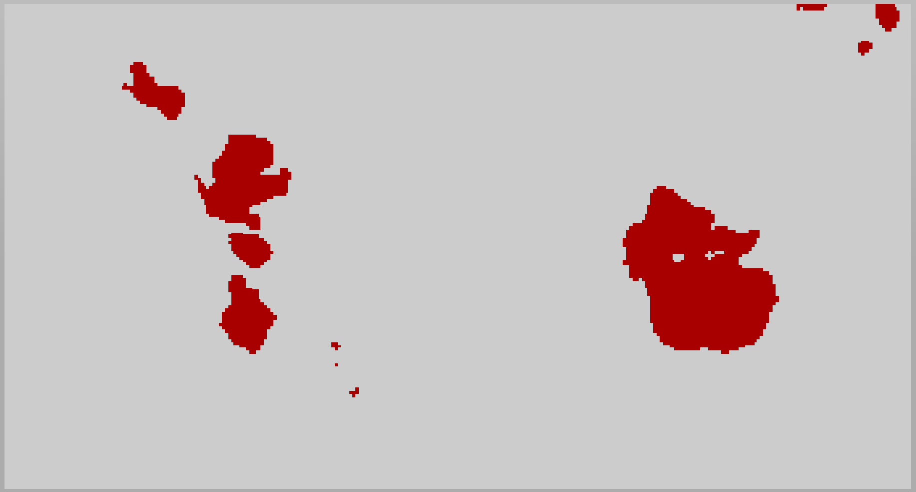


Forest disturbance classification

High accuracy with no cloud masking

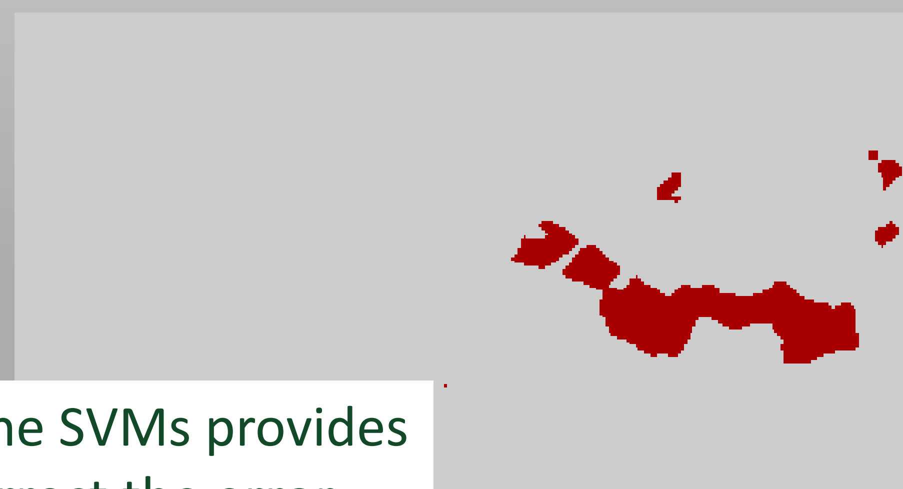
SVMs are trained to distinguish cloud from canopy disturbance



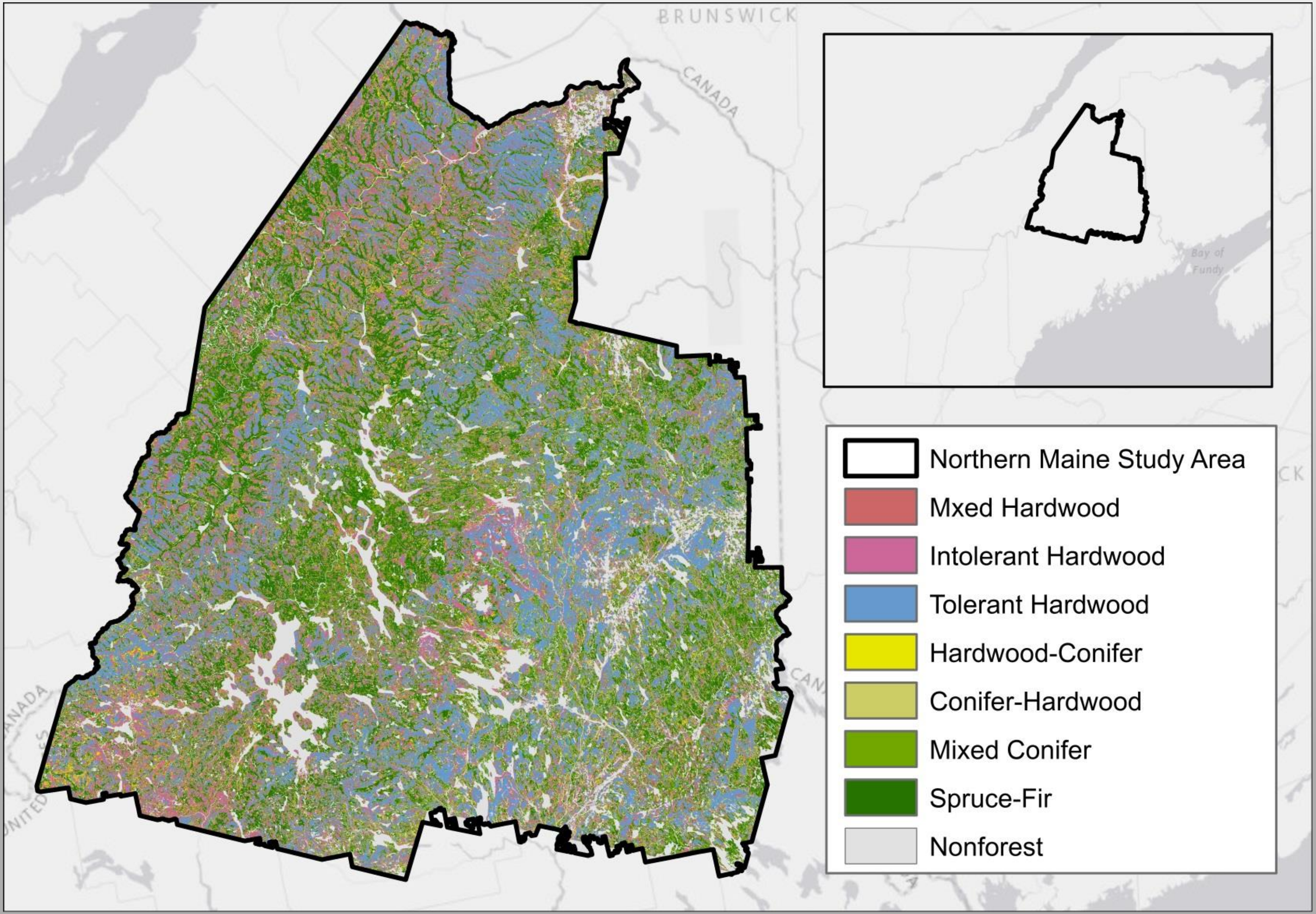


Forest disturbance classification

No examples of time 1 cloud shadow in the initial training data



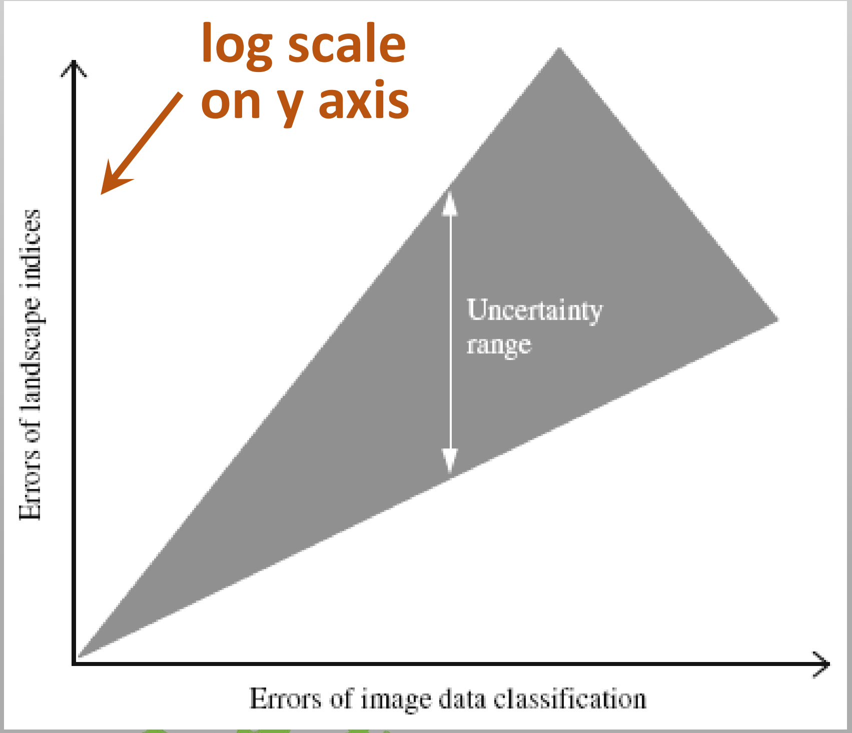
GA used to train the SVMs provides simple way to correct the error – no masks, no editing


BRUNSWICK
CANADA
Bay of Fundy
Northern Maine Study Area
Mxed Hardwood
Intolerant Hardwood
Tolerant Hardwood
Hardwood-Conifer
Conifer-Hardwood
Mixed Conifer
Spruce-Fir
Nonforest

Classification bias and fragmentation metrics

On the accuracy of landscape pattern analysis using remote sensing data

Guofan Shao · Jianguo Wu



<u>Exponential</u> increase in metric error (and error uncertainty) with increasing map error

Recommendations include…

- Accurate maps
- **Balanced classification error**

IGS products & applications

Low-cost, high-value forest map products:

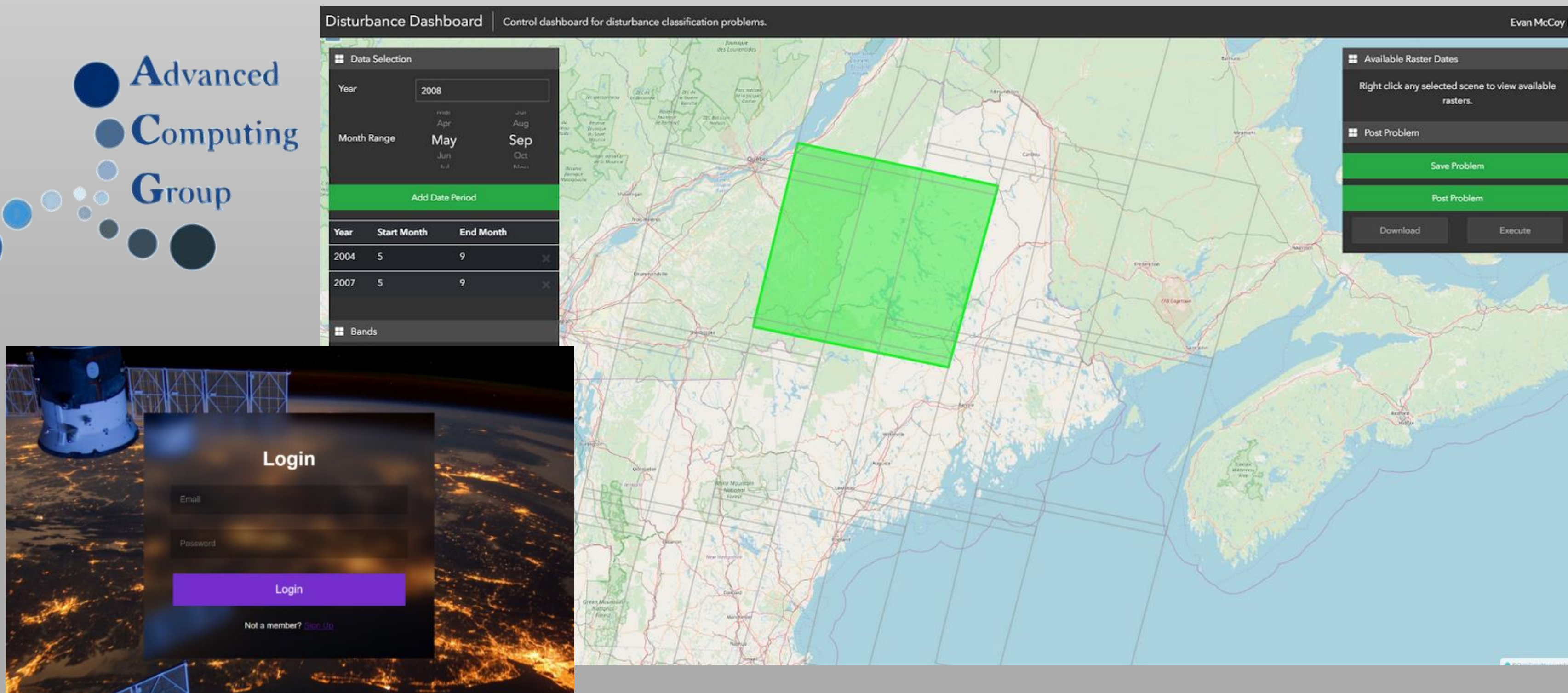
- Tree species distribution and abundance
- Forest types
- Disturbance history, monitoring
- Regeneration status
- Susceptibility, vulnerability to forest pests
- Wildlife habitat and ecosystem services
- Landscape change (retrospective and prospective)
- Regular updates

IGS products & applications

Multi-year collaboration with the UMaine Advanced Computing Group

- ML code base run on UMaine supercomputing cluster
- Highly automated workflows run on UMaine cloud resources

Advanced Computing Group

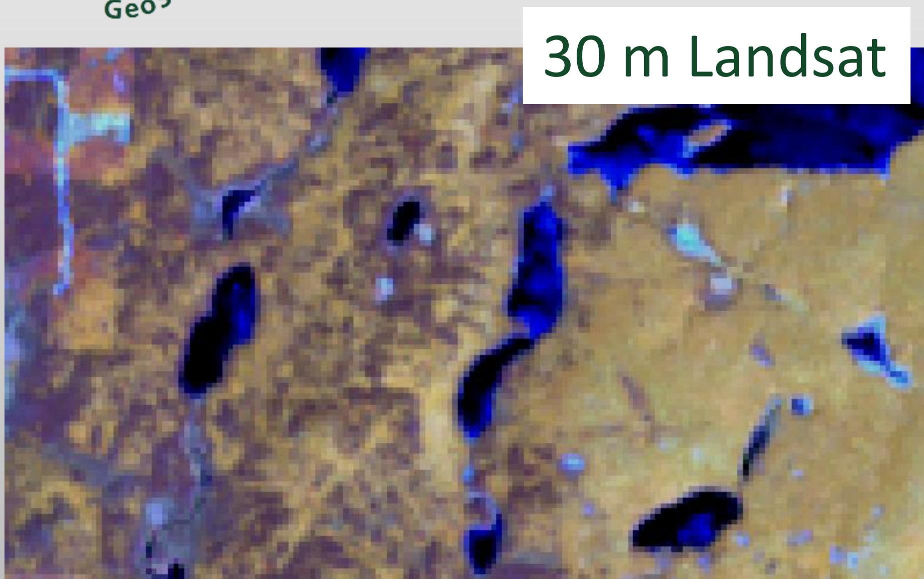


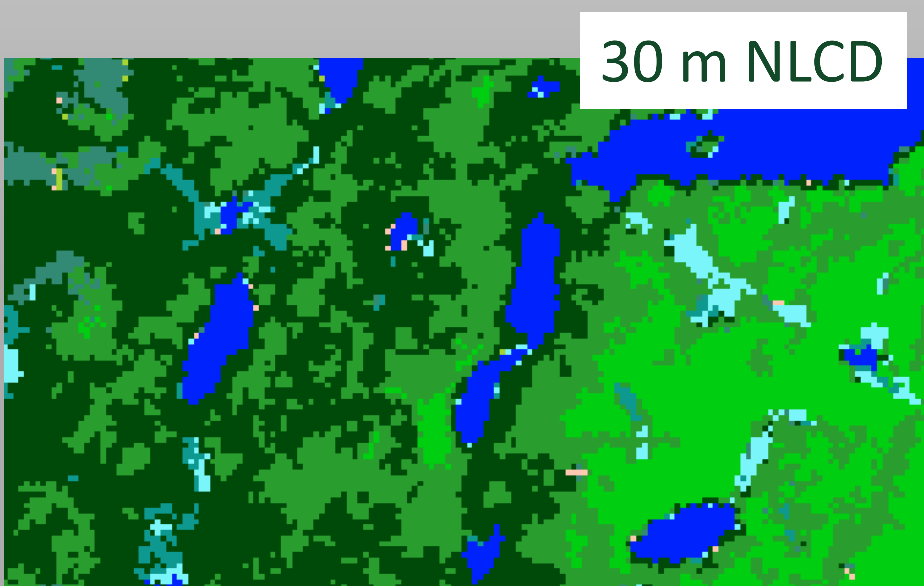
IGS products & applications

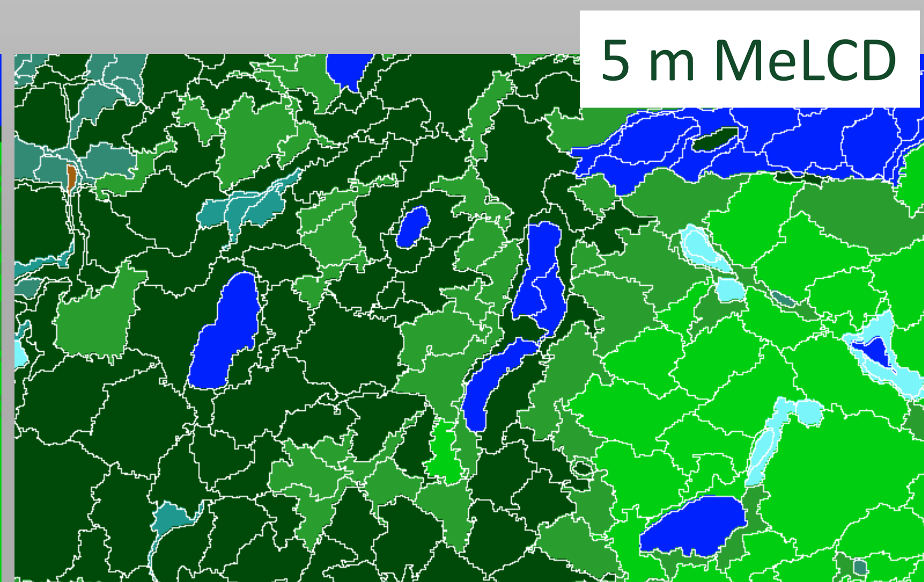
Forest type mapping from advanced ML and 10 m Sentinel 2 imagery (+ bioclimatic and terrain variables)

- Heavily automated where automation makes sense
- High accuracy; control of error distributions
- Efficient use of available training data; ability to target acquisition of additional training data to high value locations
- Need for post-processing, product editing is minimal to none

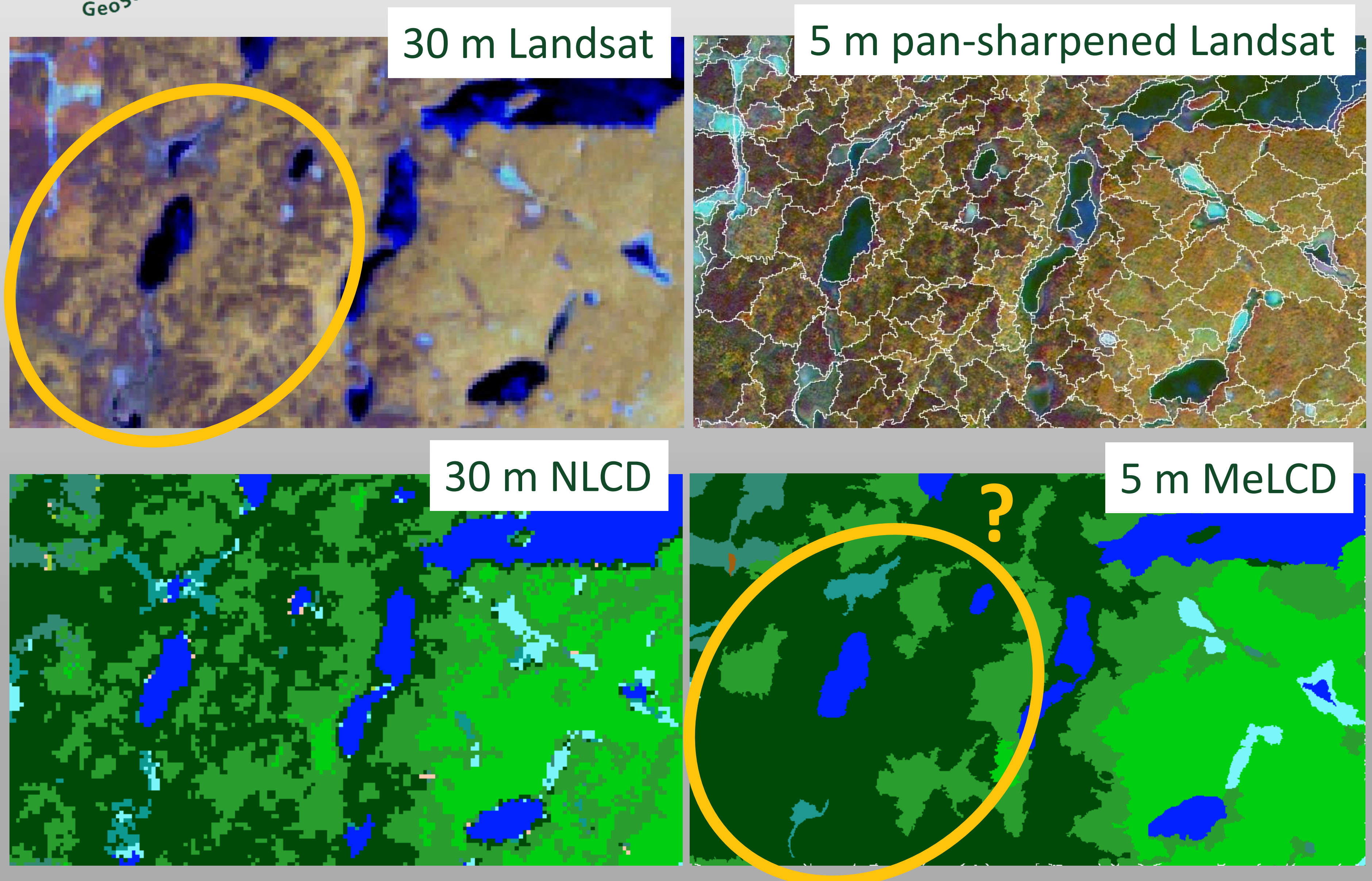
A digression on pixel resolution…





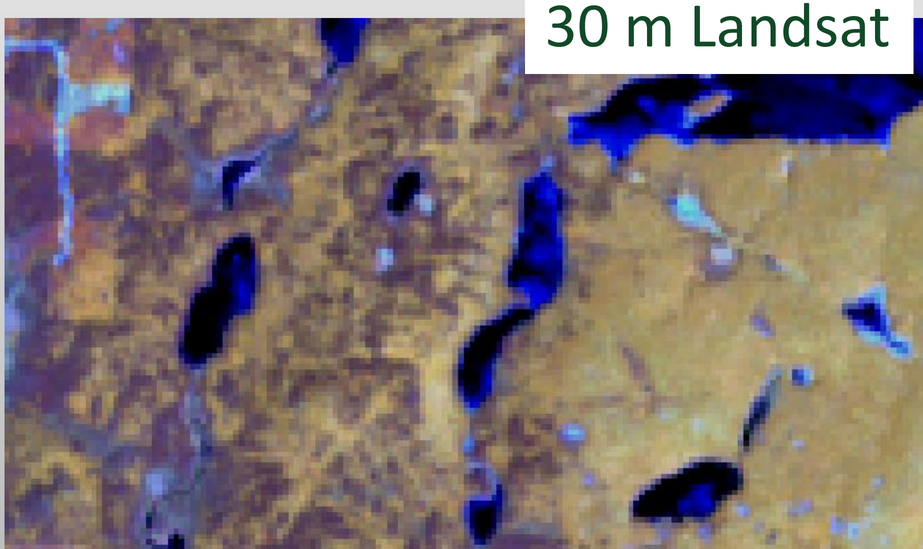


A digression on pixel resolution…



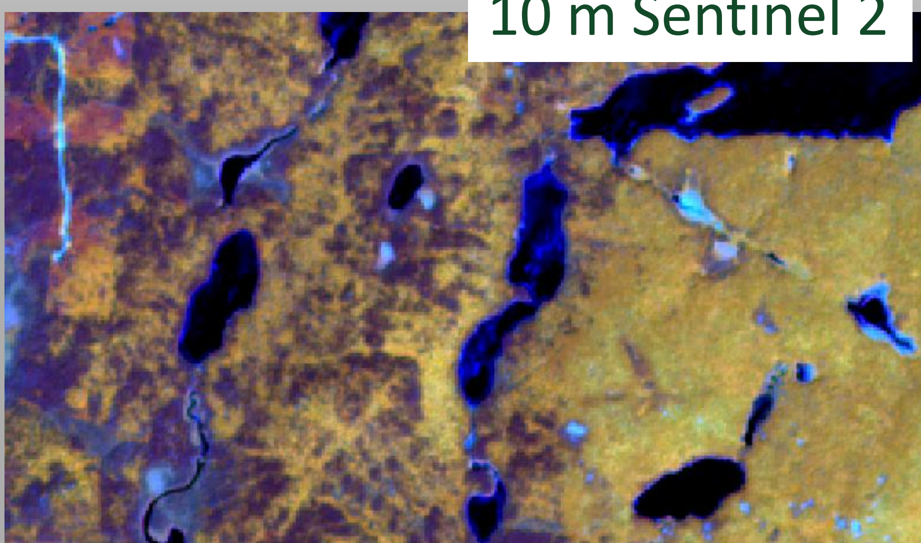
A digression on pixel resolution…

30 m Landsat



5 m pan-sharpened Landsat

10 m Sentinel 2



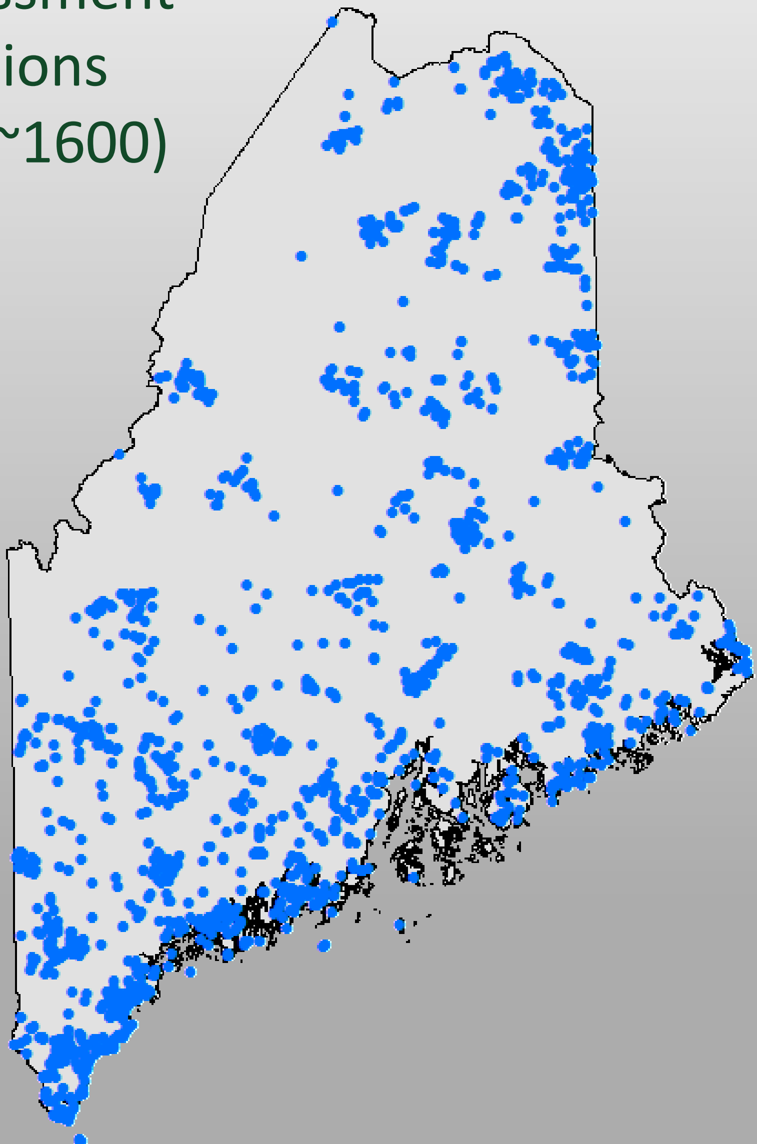
<1 m NAIP

IGS products & applications

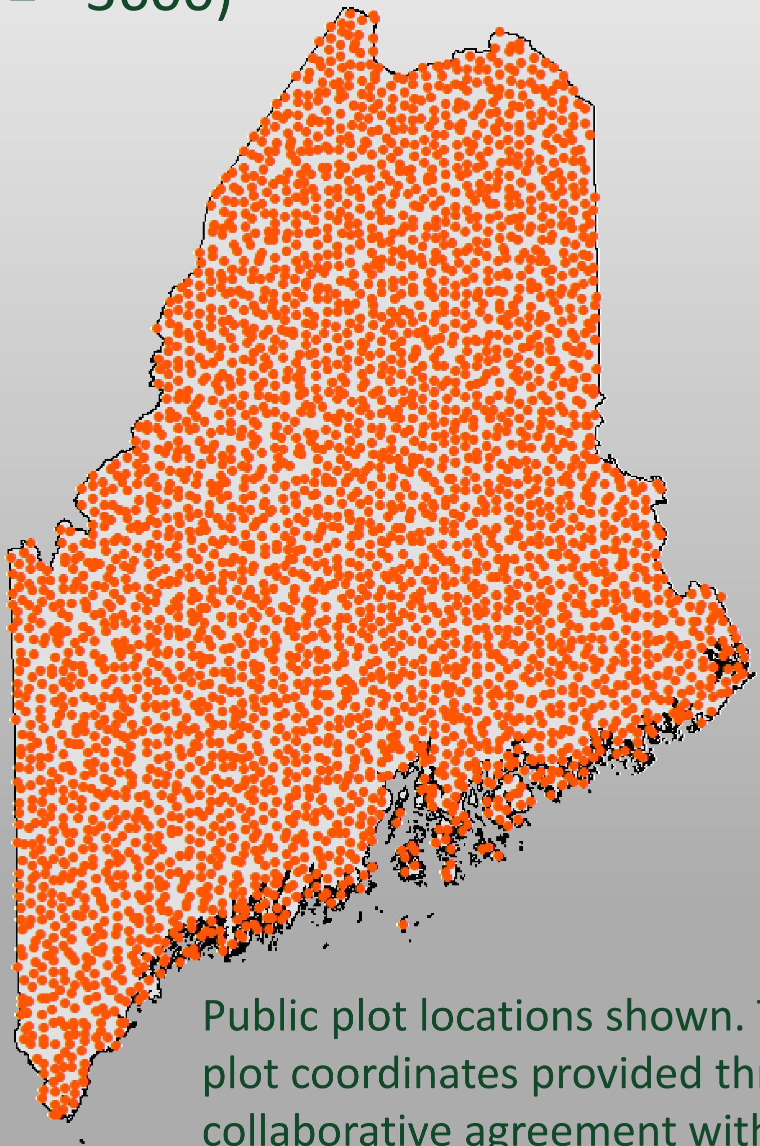
Use of USFS FIA forest plot data (i.e., confidential plot locations) for model training and validation

- Permits classification by FIA forest type:
  - Maple/beech/birch
  - Spruce/fir
  - Aspen/birch
  - White/red/jack pine
  - Oak/hickory
  - Elm/ash/cottonwood
  - Oak/pine

MeLCD accuracy assessment locations (n = ~1600)



USFS FIA plots (n = ~3600)



Public plot locations shown. True plot coordinates provided through collaborative agreement with the USFS NRS FIA Program.