Feb. 6, 2019

Recognizing forests using publicly-available data

Delineating and characterizing Pacific Northwest forest conditions with machine learning models

David Diaz

MOTIVATION

Automate accurate forest stand mapping to facilitate management planning

OUR TASK IS TO SEGMENT AND LABEL FOREST STANDS

"A contiguous community of trees sufficiently uniform… to distinguish it from adjacent communities."

Nyland (2007)

Related efforts all use pixel-level classification, aim to provide a broader set of predicted variables, and are usually designed for regional-scale analysis

CSIRO PUBLISHING
www.publish.csiro.au/journals/ijwf *International Journal of Wildland Fire* **2009**, *18*, 235–249

LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment

Matthew G. Rollins

US Geological Survey, Center for Earth Resources Observation and Science (EROS), Sioux Falls, SD 57198, USA. Email: mrollins@usgs.gov

252 IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 4, NO. 2, JUNE 2011

Monitoring Landscape Change for LANDFIRE Using Multi-Temporal Satellite Imagery and Ancillary Data

James E. Vogelmann, Jay R. Kost, Brian Tolk, Stephen Howard, Karen Short, Xuexia Chen, *Associate Member, IEEE*, Chengquan Huang, Kari Pabst, and Matthew G. Rollins

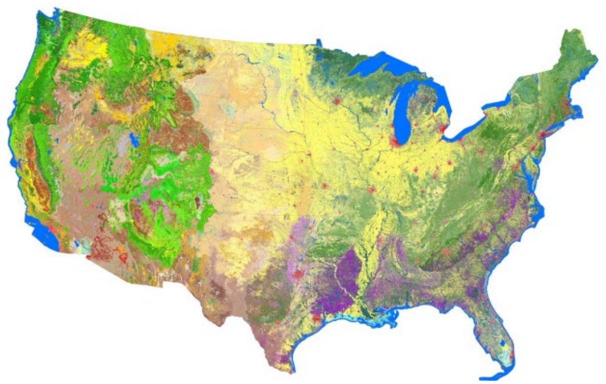


Fig. 1. LANDFIRE existing vegetation data set developed for the conterminous United States. Different shades of green and purple represent different types of forest cover. Shades of tan, brown and orange represent rangeland and grassland ecosystems. Light yellow represents agriculture, red represents urban, and blue represents water.

Predictive mapping of forest composition and structure with direct gradient analysis and nearest- neighbor imputation in coastal Oregon, U.S.A.

Janet L Ohmann and , Matthew J Gregory

Canadian Journal of Forest Research, 2002, 32(4): 725-741, https://doi.org/10.1139/x02-011

ELSEVIER
Forest Ecology and Management
Volume 358, 15 December 2015, Pages 154-164

Imputed forest structure uncertainty varies across elevational and longitudinal gradients in the western Cascade Mountains, Oregon, USA

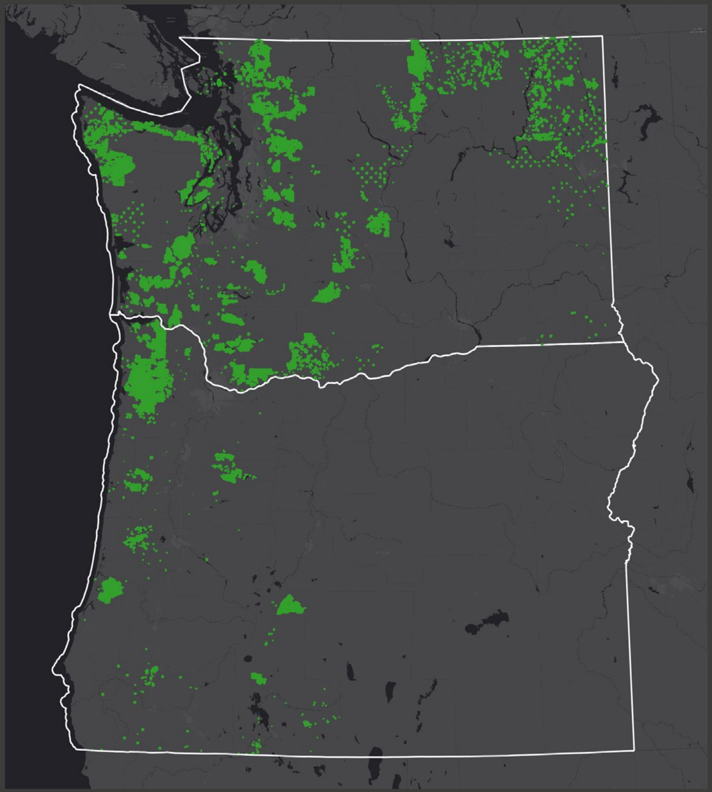
David M. Bell [a], Matthew J. Gregory [b], Janet L. Ohmann [b]

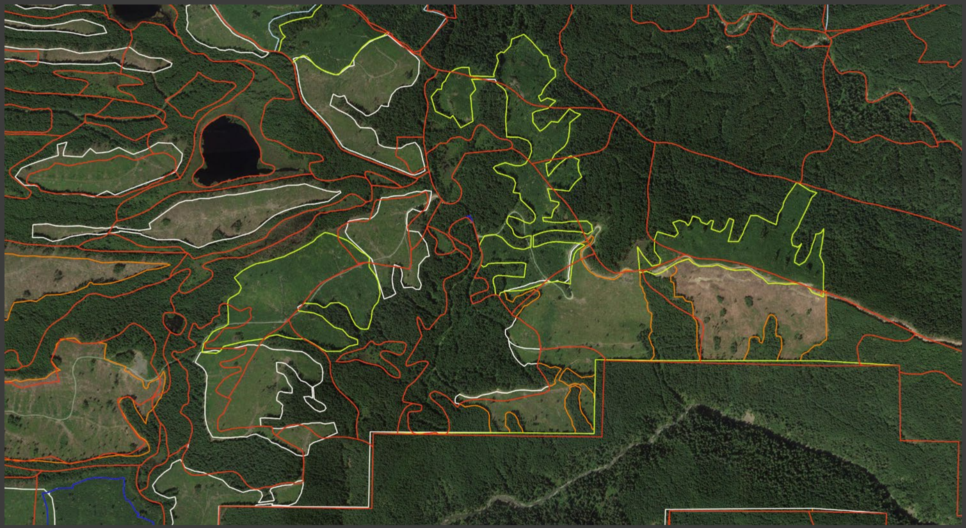
ELSEVIER
Remote Sensing of Environment
Volume 112, Issue 5, 15 May 2008, Pages 2232-2245

Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data

Andrew T. Hudak [a], Nicholas L. Crookston [a], Jeffrey S. Evans [a], David E. Hall [a], Michael J. Falkowski [b]

FOREST INVENTORY DATA ACQUIRED THROUGH PUBLIC RECORDS REQUESTS

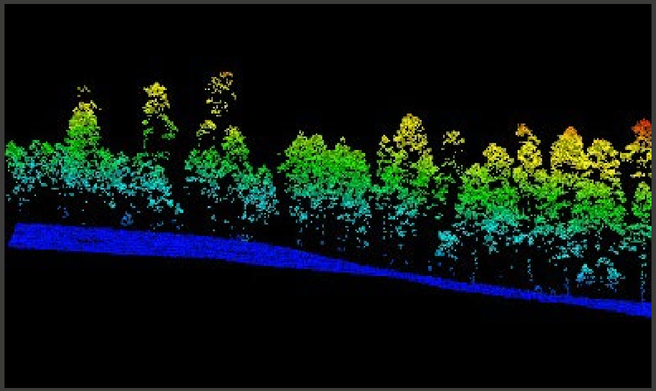




- WA DNR: 369,000 stand polygons over six different years (2004-2017)
- ODF: 15,000 current stand polygons

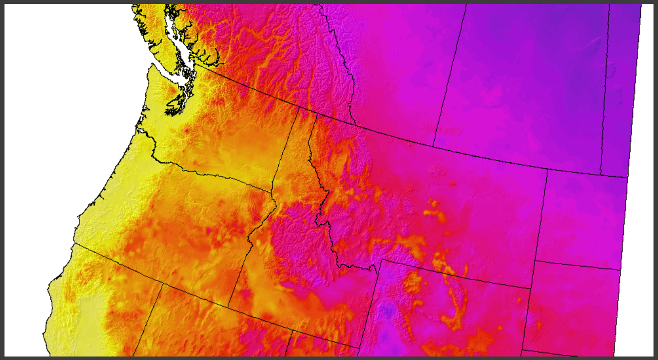
450,000 plot measurements have occurred in these stands, containing tree-level observations (species, diameter, height, etc.)

GENERATING FEATURES



LIDAR

Publicly-available lidar point clouds covering several million acres are being processed into 20+ rasters characterizing terrain and canopy. 0.5-1m resolution surface and intensity rasters, 10m resolution canopy metrics.



CLIMATE

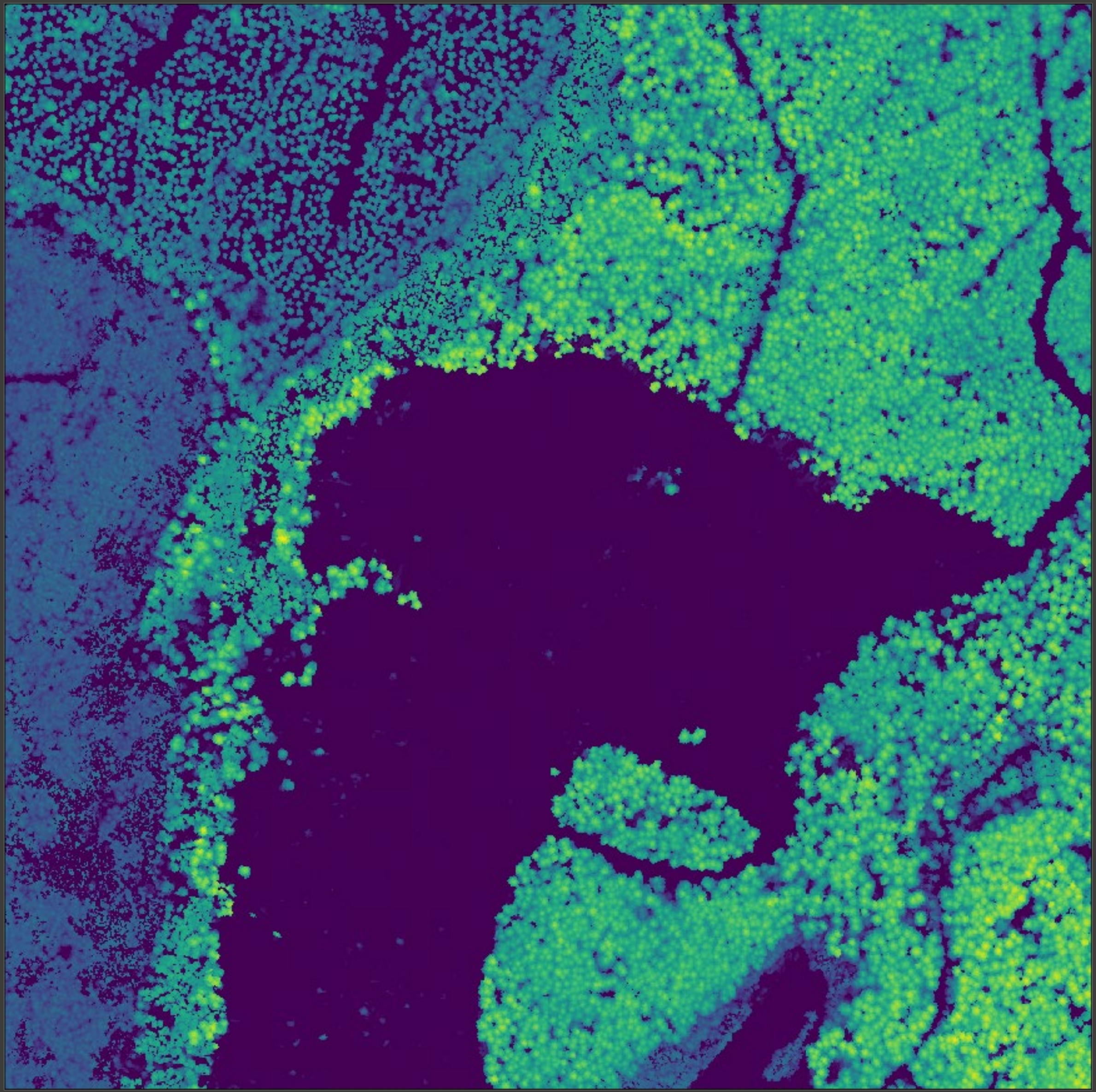
Down-scaled monthly, seasonal, and annual climatic data and derived metrics relevant for vegetation modeling extracted from Climate WNA (Wang 2012).

IMAGERY

Aerial (NAIP) and satellite imagery extracted using Google Earth Engine. Time series of several images per year collected from Sentinel and/or Landsat to facilitate species identification.

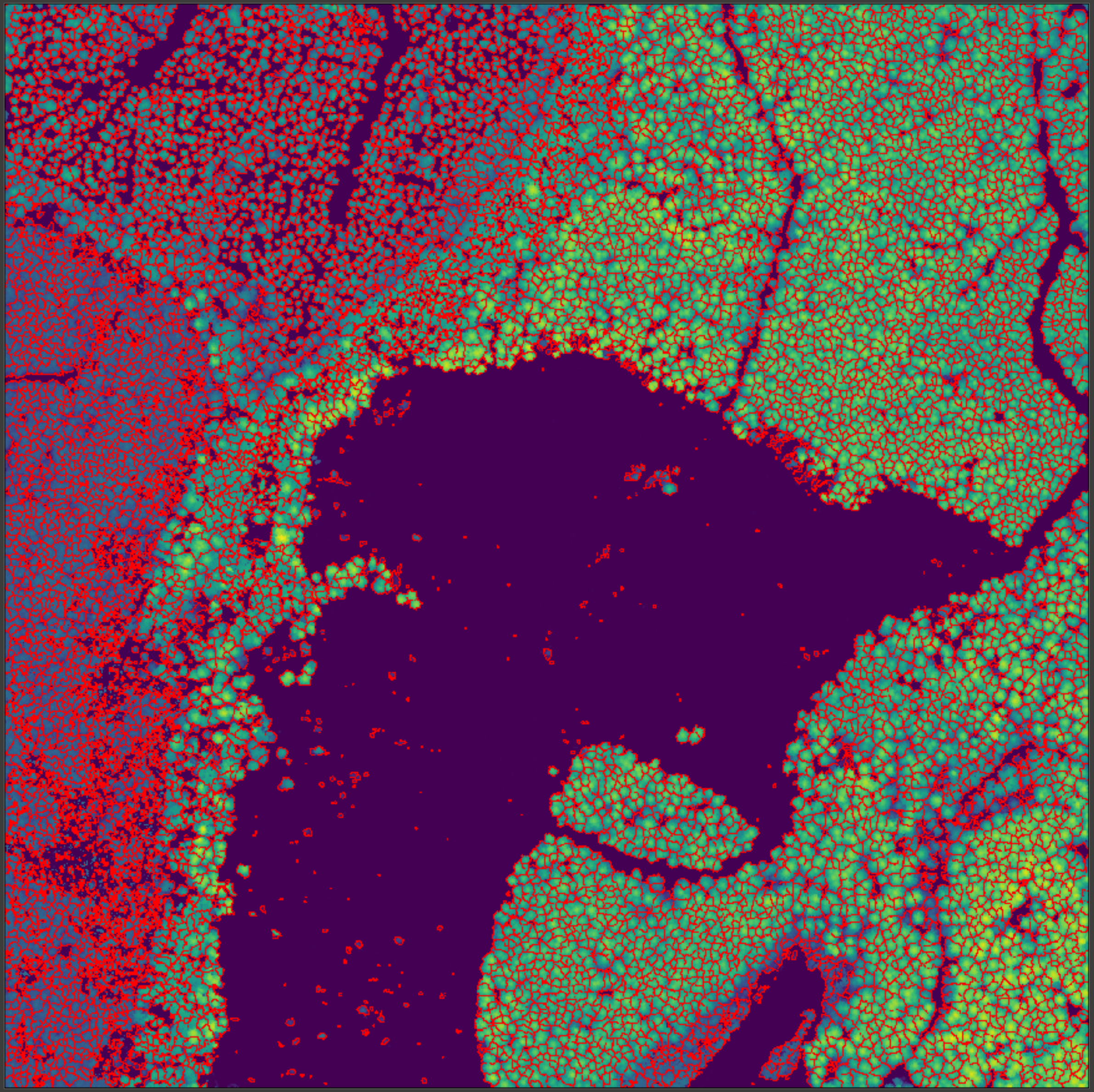
Generating Superpixels from Tree Associated Objects

Generating Superpixels from Tree Associated Objects



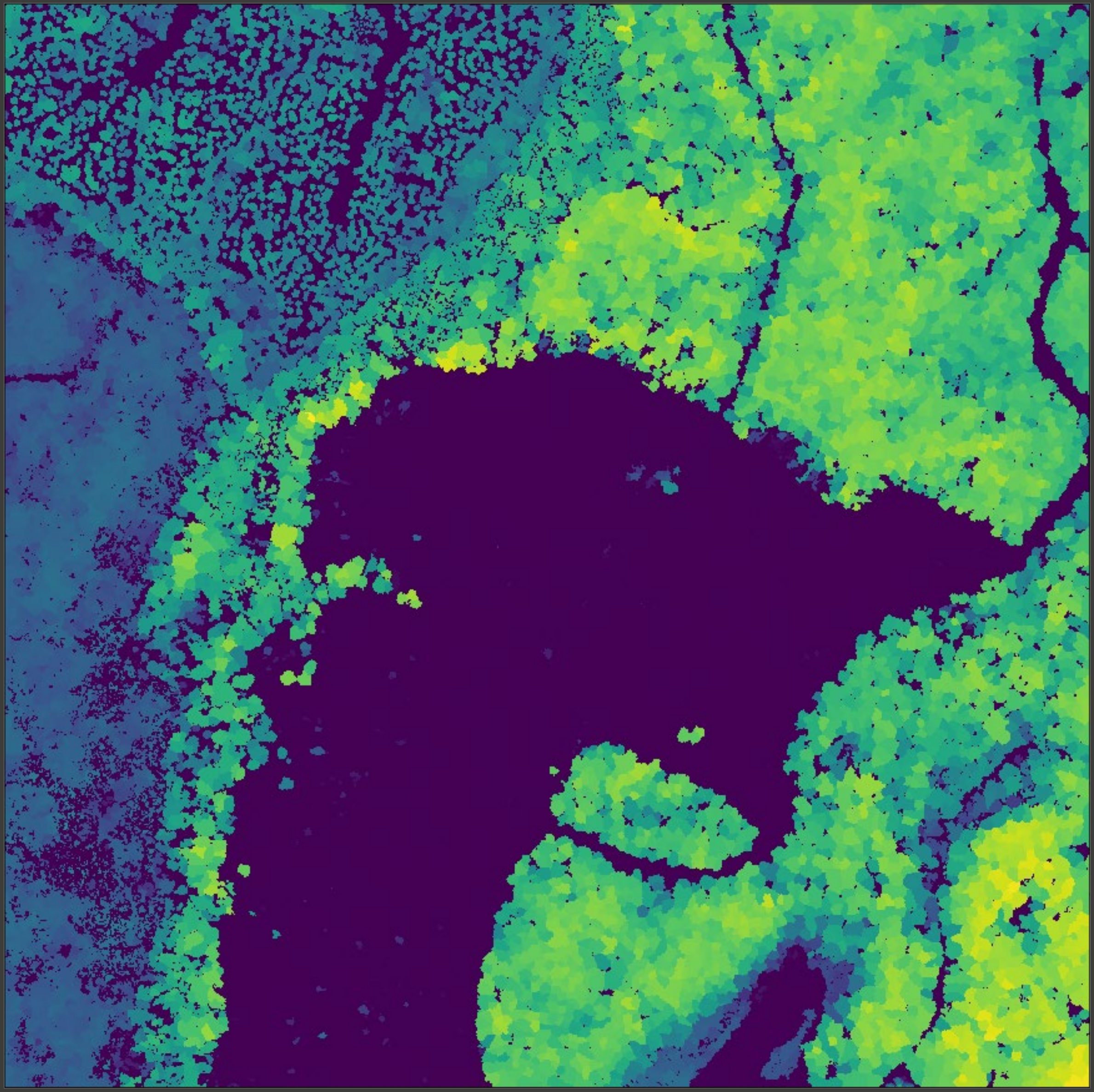
Perform watershed segmentation of the canopy height model.

Homogenizing Tree Objects to Generate “Superpixels”



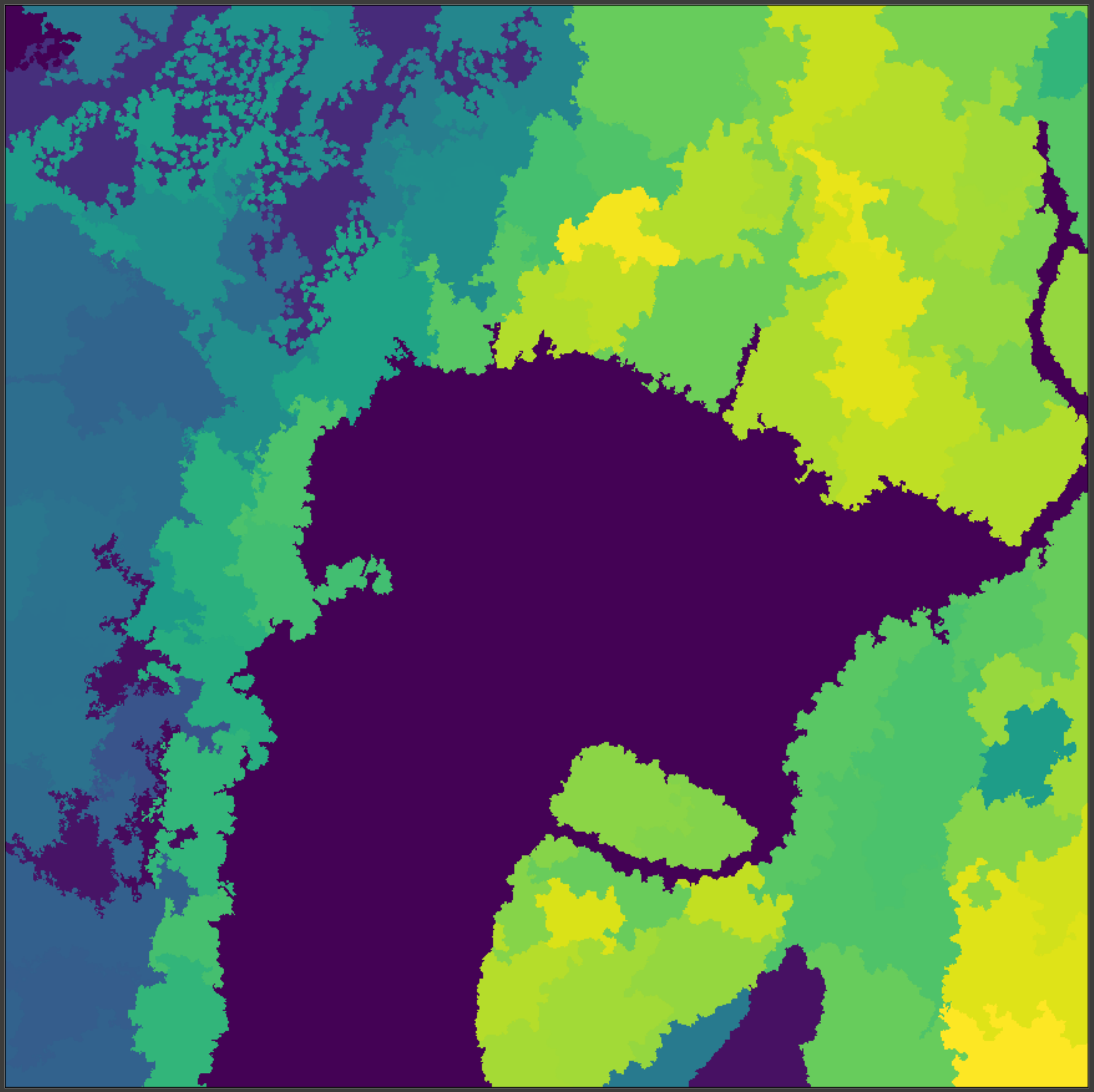
Perform watershed segmentation of the canopy height model.

Homogenizing Tree Objects to Generate “Superpixels”



Then reassign values for each “basin” using a mean or max of each raster used for stand segmentation.

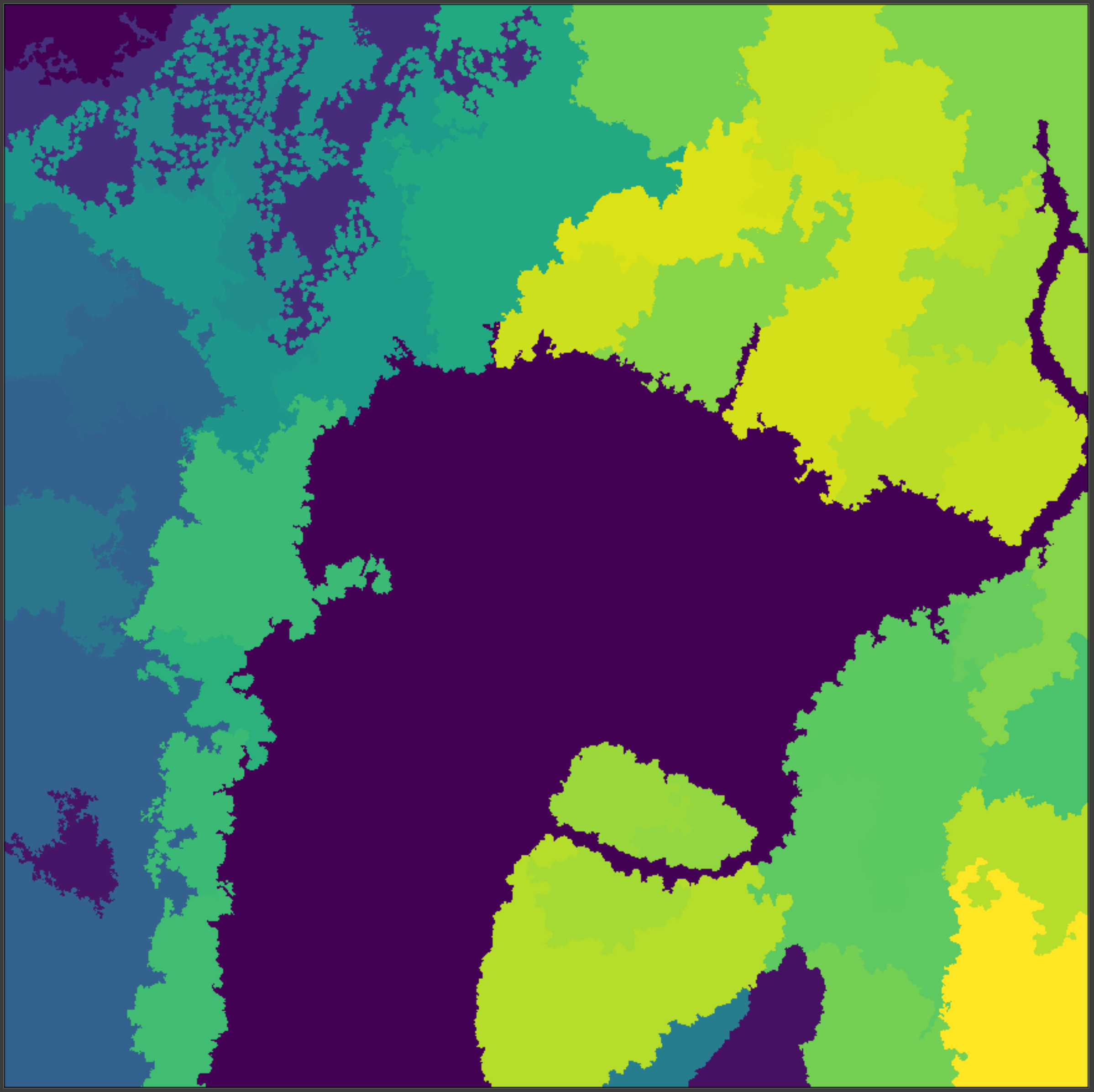
Homogenizing Tree Objects to Generate "Superpixels"



Then, use graph-based image segmentation to merge these objects into "stands" > 0.5 acre, or > 1.0 acre.

These are the objects we'll learn to classify.

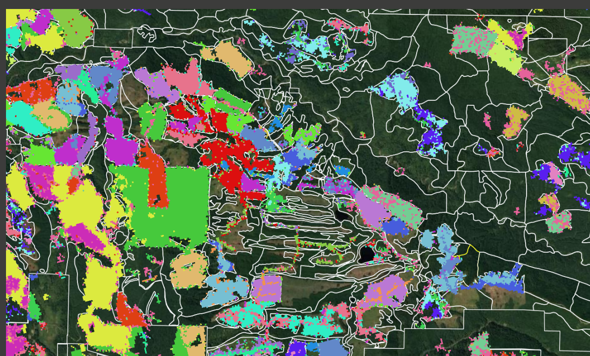
Homogenizing Tree Objects to Generate "Superpixels"



Then, use graph-based image segmentation to merge these objects into "stands" > 0.5 acre, or > 1.0 acre.

These are the objects we'll learn to classify.

GENERATING LABELS



Year of disturbance based on LandTrendr (Kennedy et al. 2012).

Tree measurements are grown forward from sampling date using a growth-and-yield model (FVS) on an annual timestep until any detected disturbance.

Generating annual stand conditions significantly expands the number of samples we can learn from.

Primary focus for label attributes:

- Forest Type: *dominant species*
- Size Class: *avg. tree diameter*
- Stocking Level: *canopy cover, stand density index, basal area*
- # Canopy Layers

Data Transformation

Unsupervised segmentation of existing stand areas. New segments contained within existing stands are assigned the forest inventory attributes of that stand.

Histograms or distributions of raster values within each segment are generated.

Linear Discriminant Analysis (supervised dimensionality reduction) is applied to transform histogram values to maximize separation between known stand types.

Features

Transformed histograms of raster data that occur within rasterized “stand” polygons.

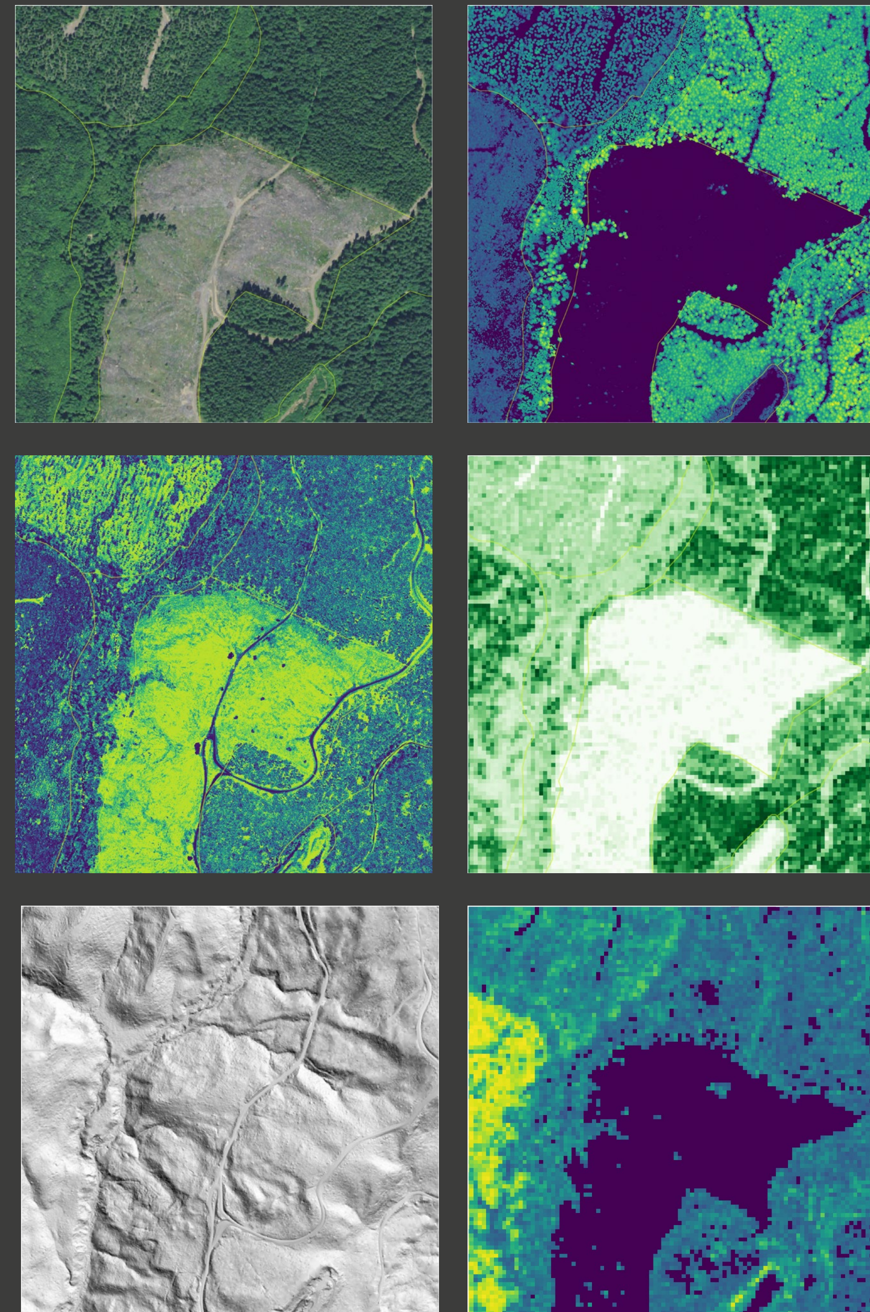
Labels

Several indicators (multi-output prediction) of forest type, size, stocking, and structure.

BENEFITS OF TWO-PHASE APPROACH

Segment, then Classify

- Follows stand-paradigm widespread in professional forestry (as opposed to pixel-based labeling)
- Effectively leverages multi-resolution raster data
- Avoids misfit between plots and pixels
- Re-segmentation of stands prior to model training may resolve imprecise or inaccurate original stand boundaries
- A wide variety of stand-level and plot-level metrics can be imputed to segmented stands, as point estimates or as distributions



MODEL-TUNING

- Perform gridsearch of RF model hyperparameters and number of LDA-transformed features used for model training to minimize cross-validation loss (MSE for quantitative labels, cross-entropy for categorical labels)

ACCURACY & BENCHMARKING

- Compare probability of predicted forest attributes to field-measured (and FVS-modeled) inventory attributes:
- Summaries for different regions and forest types
- Summaries at different spatial scales (stand, parcel, watershed, ecoregion)
- Performance of this approach will be benchmarked against overlapping areas and years with LANDFIRE and GNN datasets.

POST-PROCESSING

Automated stand boundaries are very detailed. Need to simplify and smooth.

Develop and apply additional RF imputation models to generate more detailed information about each stand:

- Plots, treelists that can be incorporated directly into growth-and-yield models
- Distributions of timber volume, biomass, carbon storage