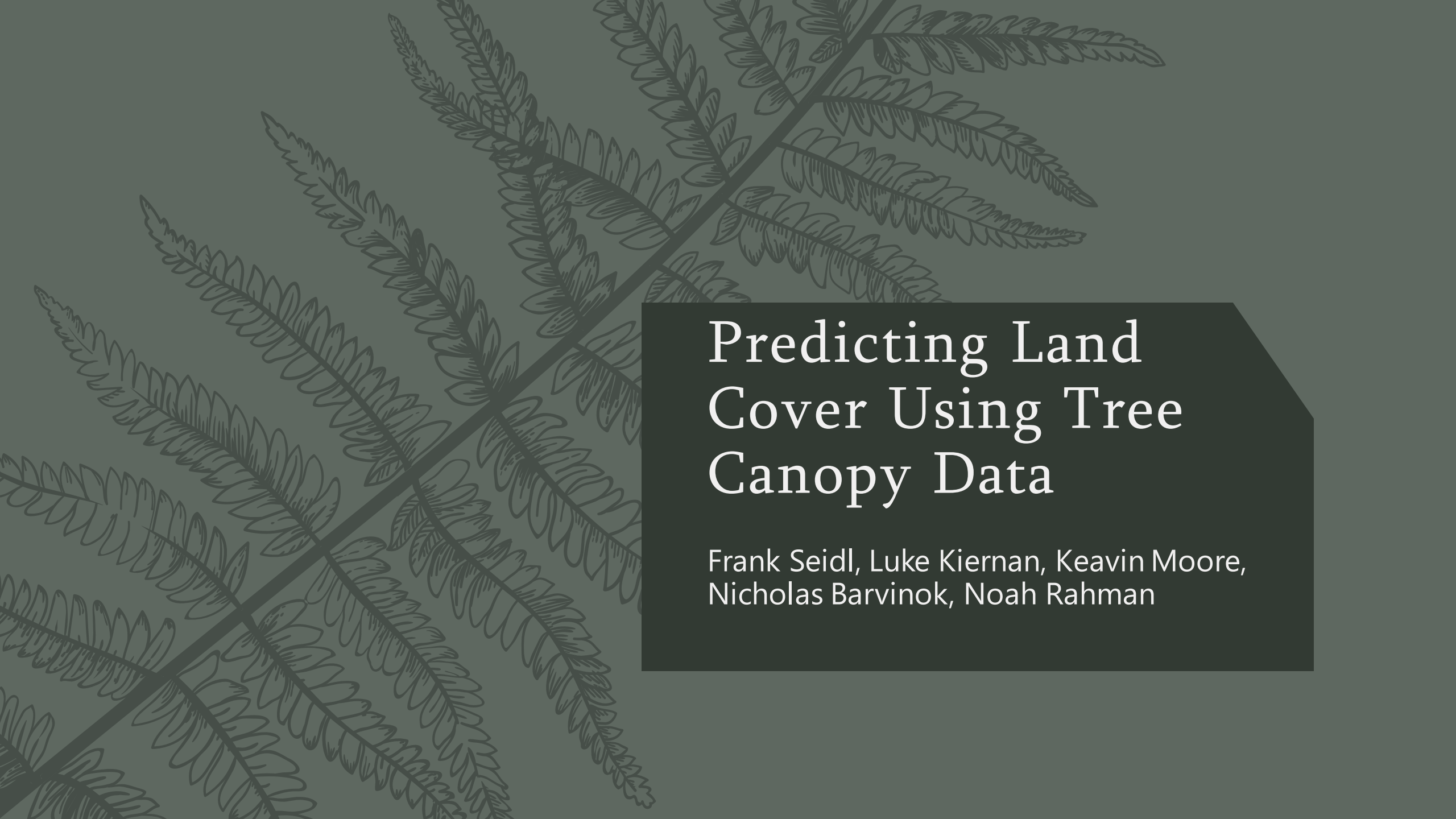


Predicting Land Cover Using Tree Canopy Data

Frank Seidl, Luke Kiernan, Keavin Moore, Nicholas Barvinok, Noah Rahman

Research Question

Is there a correlation between land cover and tree canopy height and structure?

Tree canopy height data can help...

Wildfire Prevention Researchers

predict the most at risk areas for wildfire

Vegetation & Groundwater Researchers

conveniently predict vegetation distribution

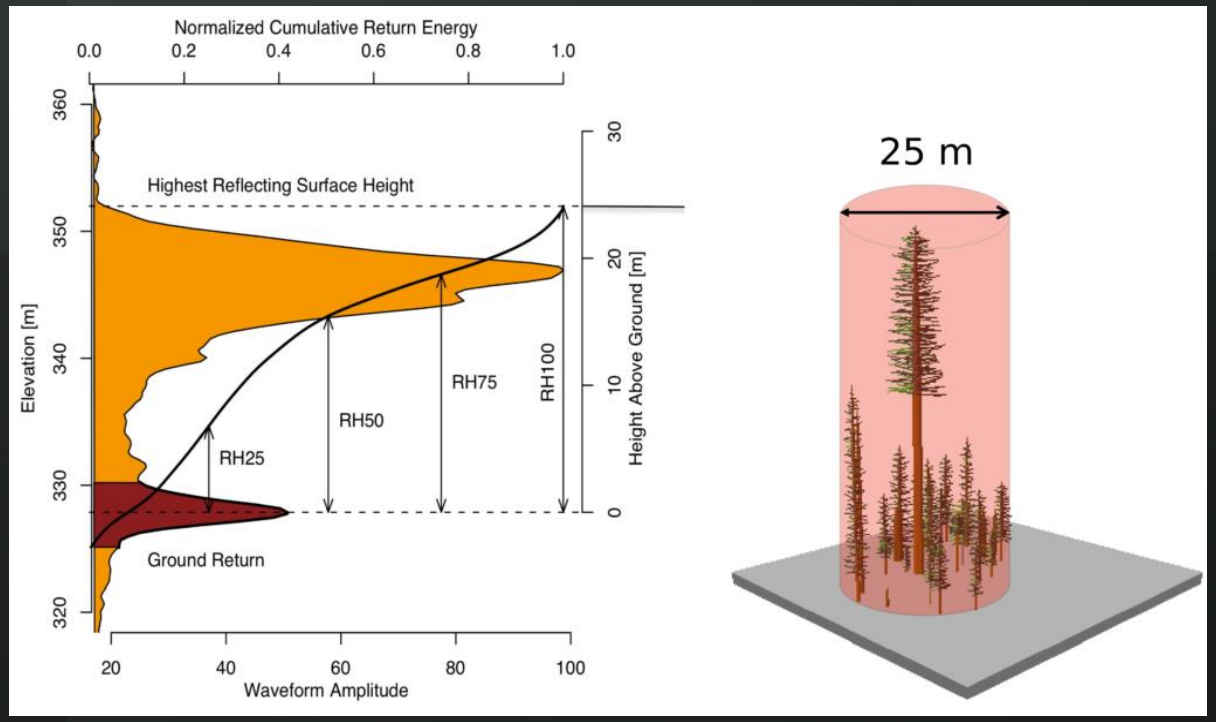
Real Estate Developers

determine the best region for new housing developments

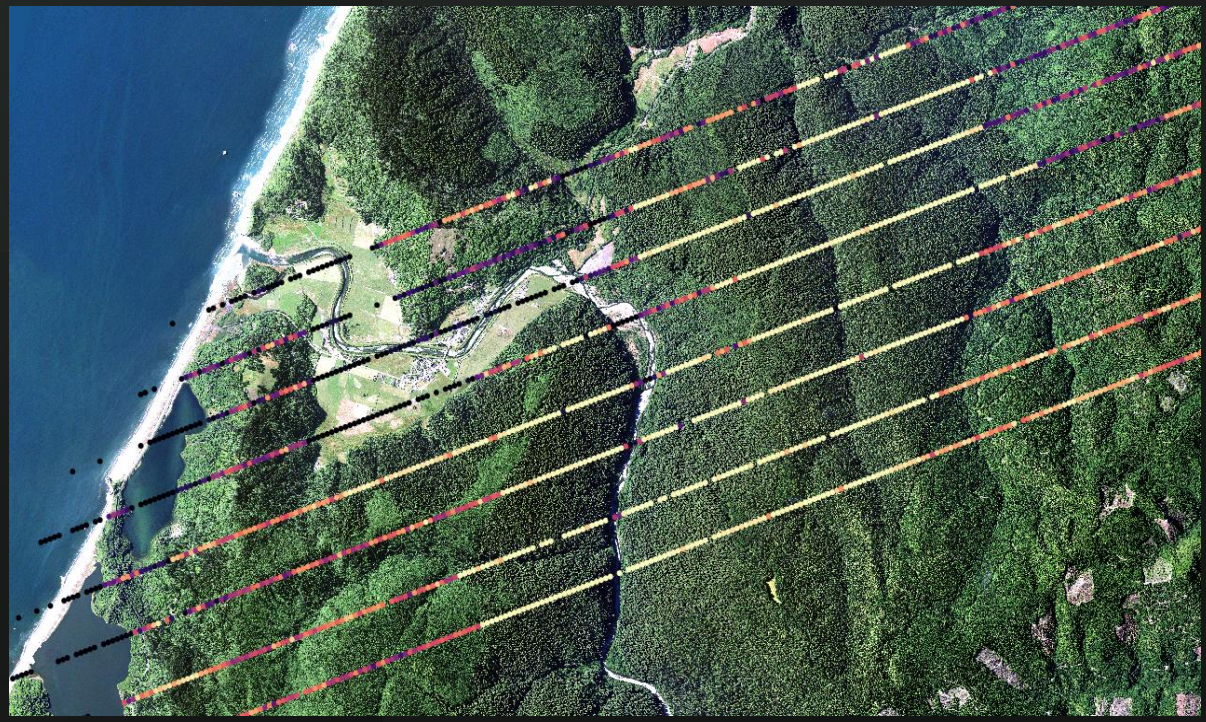
Solar Energy Companies

improve shadow prediction

Global Ecosystem Dynamics Investigation (GEDI) Data



Source: UMD GEDI Products Overview

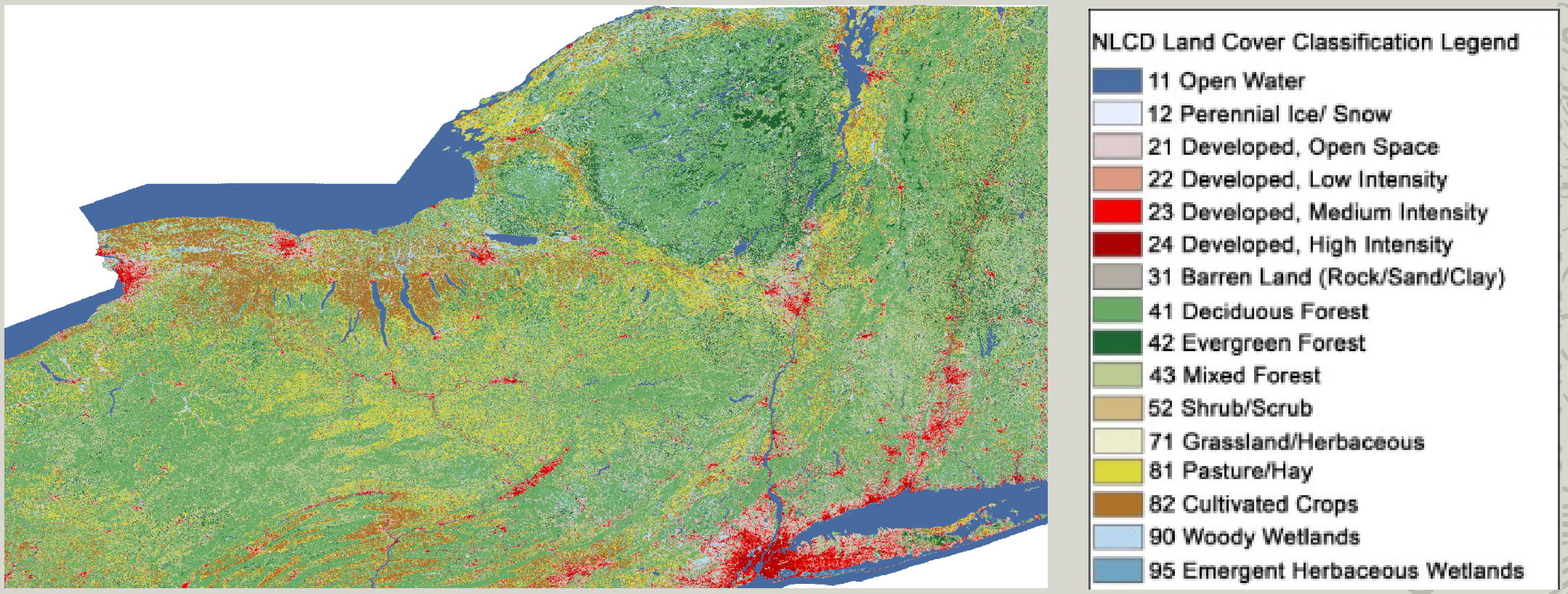


Source: LPDAAC USGS, Redwood National Forest

GEDI Data pre-processing/cleaning

- Geographic sub-setting (New York).
- Filter bad data using quality/degradation flags.
- Downsample relative height metrics, *rh_n* --- these are our training features.

Multi-Resolution Land Characteristics (MRLC) Data



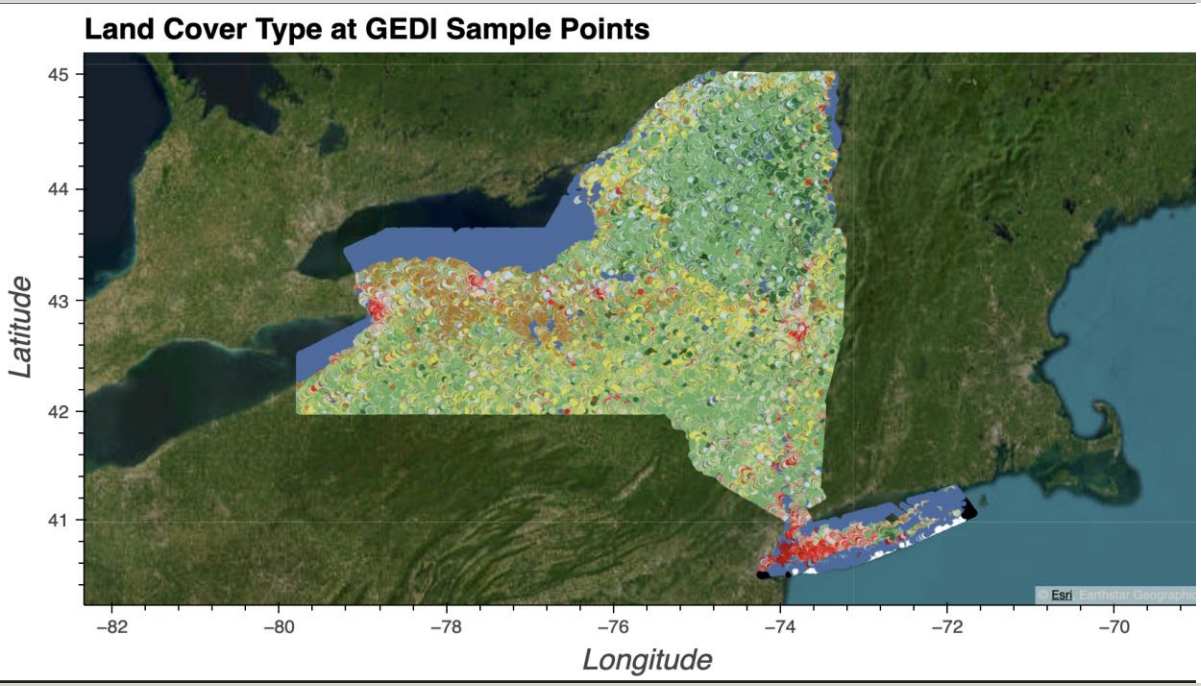
MRLC Data pre-processing/cleaning

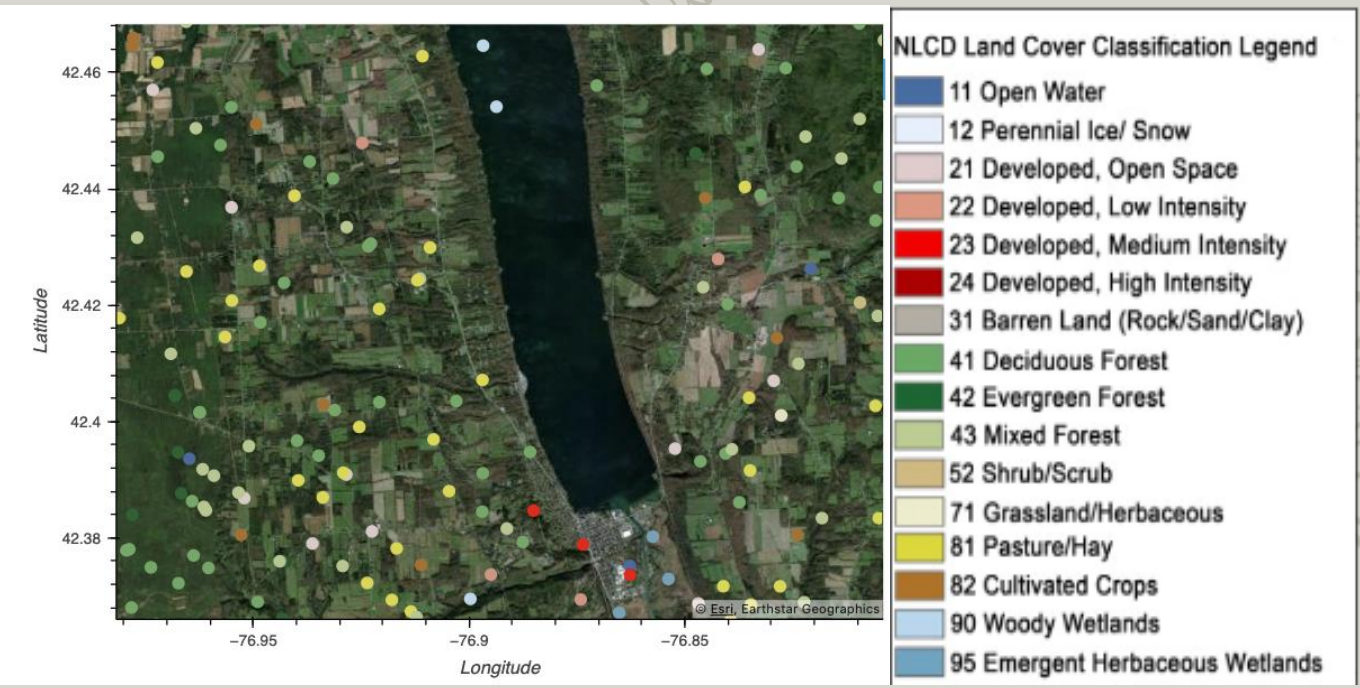
- Download as geo-marked raster data
- Combine with GEDI observations:
  convert (lat, long) → pixel → cover type.
- Combine cover types into groups --- these are our prediction labels.

GEDI-L2A Vector Canopy Top Height dataset

← Predictors & Targets →

MRLC land cover data for New York State



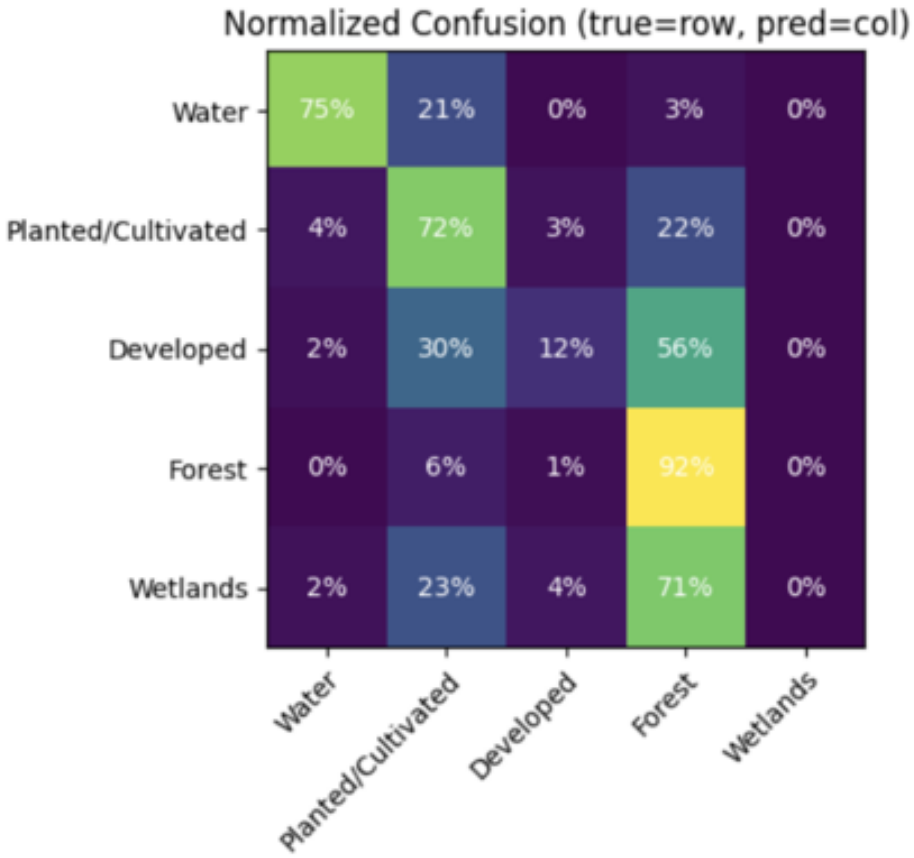


k-Nearest Neighbors Classification

- Classify point as majority label among k nearest training points.
- Neighborhood size hyperparameter k has large near-optimal range, [~25, ~100].
- Curse of dimensionality: 101-dimensional rh profiles require low-dimensional embedding.
- Downsampling, downsampling followed by robust scaling, and PCA yield similar results.
- Following results downsample to only the 0th, 10th, 20th, …, 100th rh percentiles.

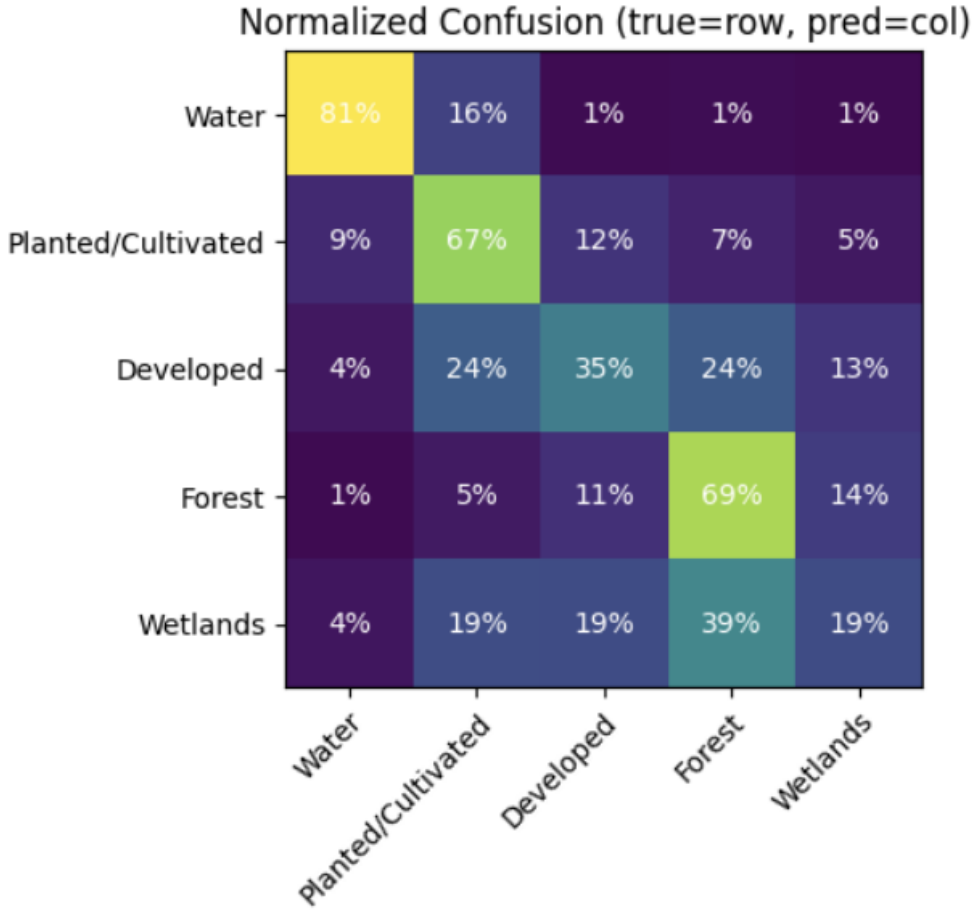
Representative training data

kNN strongly favors "Forest" label, as most of NYS is forest. Wetlands are (almost) never predicted.

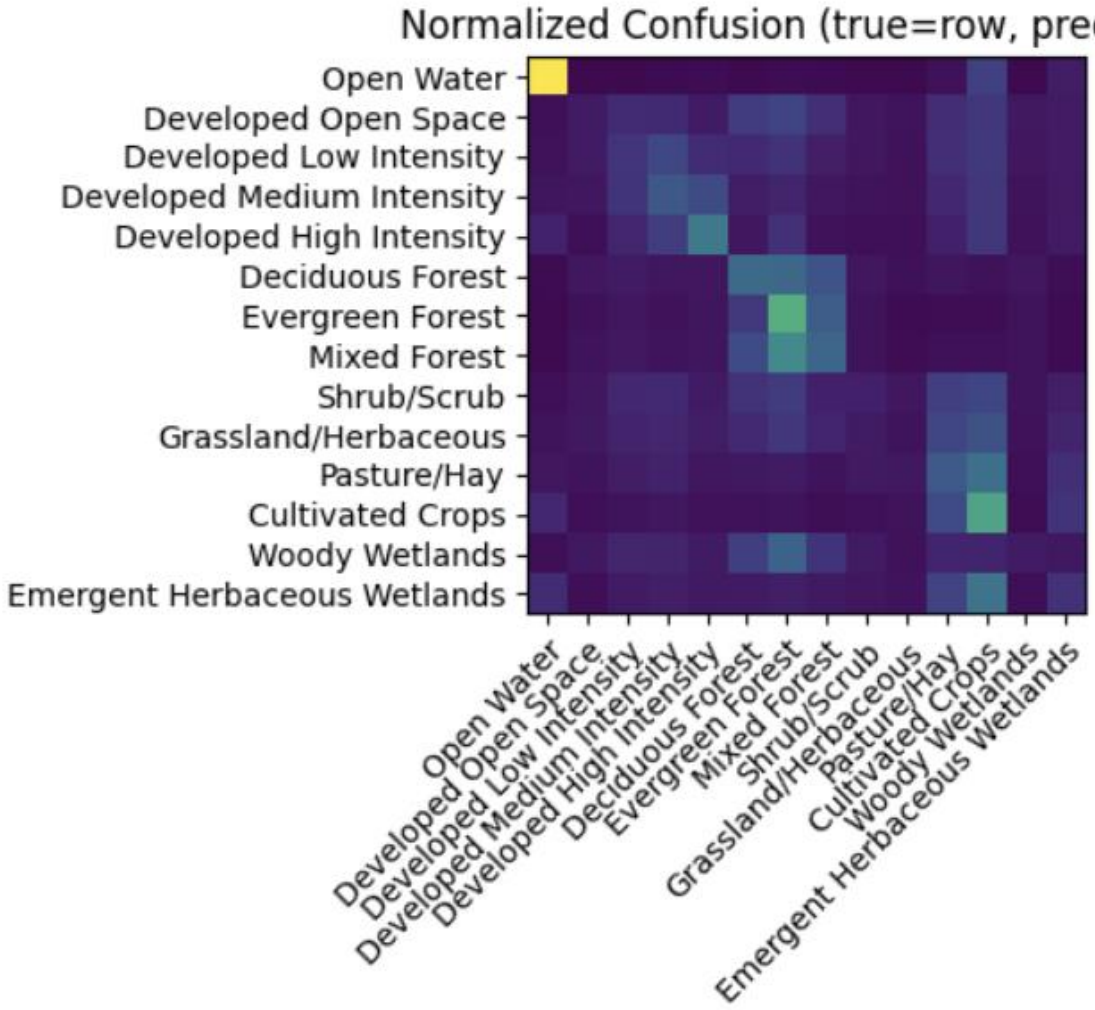


Balanced training data

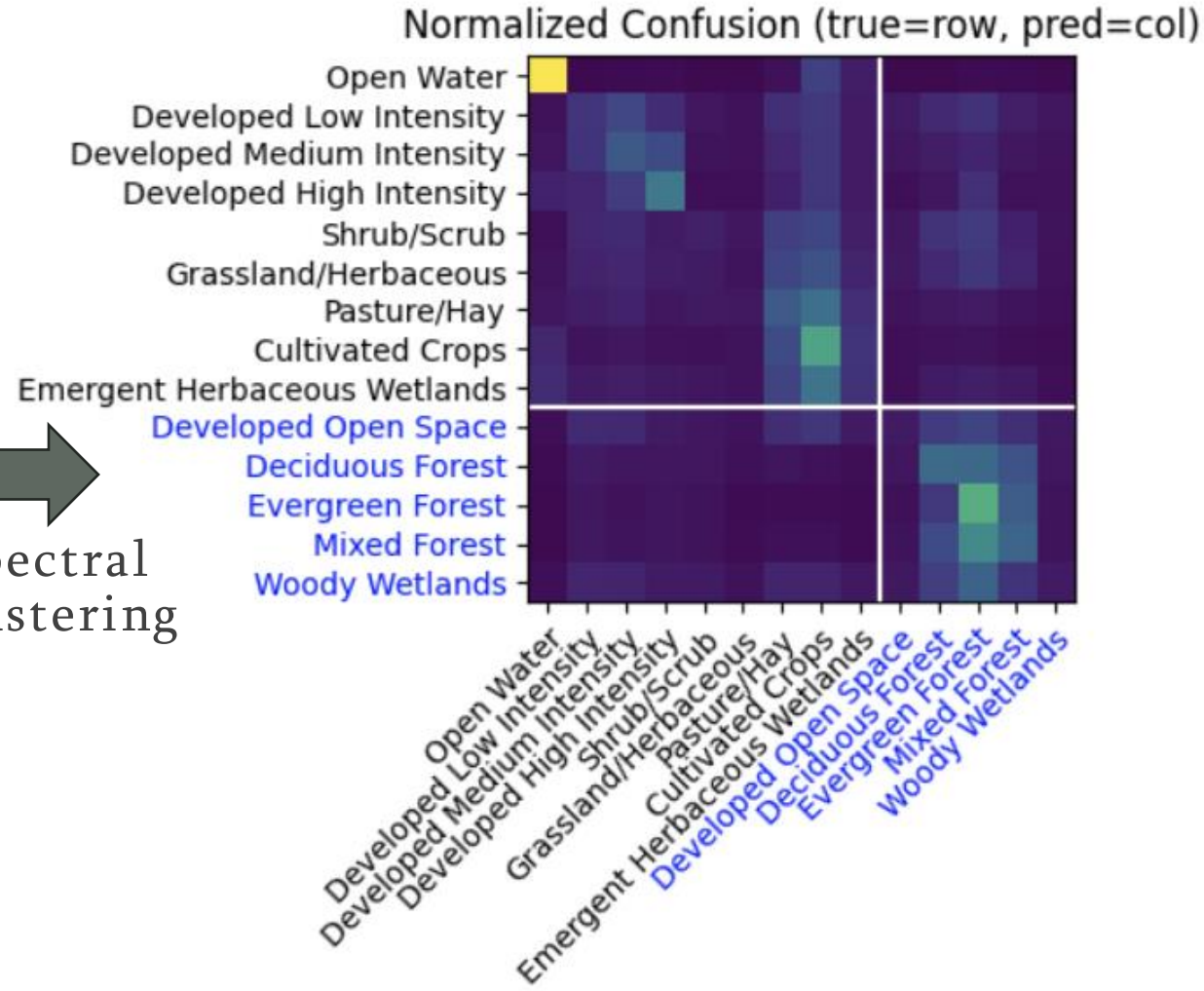
No "Forest" bias, wetlands sometimes predicted. Raw accuracy diminished slightly.



Each tile is rounded to the nearest whole percent.



Spectral Clustering



Next Steps

Compare with other New York data

e.g., real estate data, energy usage data

Compare with data outside New York

e.g., historical wildfire data in California

Improve UI

Make a website, app, etc.

Repeat the experiment

Pick a location other than New York and compare

Thanks

To the Erdos Institute, including Roman Holowinsky, Steven Gubkin, Alec Clott, Zach Hafen-Saavedra

The Global Ecosystem Dynamics Investigation (GEDI)

The Multi-Resolution Land Characteristics (MRLC) Consortium