Machine learning for natural resource management

Jenny R. Kreiger, PhD
December 3, 2019

Photo by John B. Kalla, via Flickr

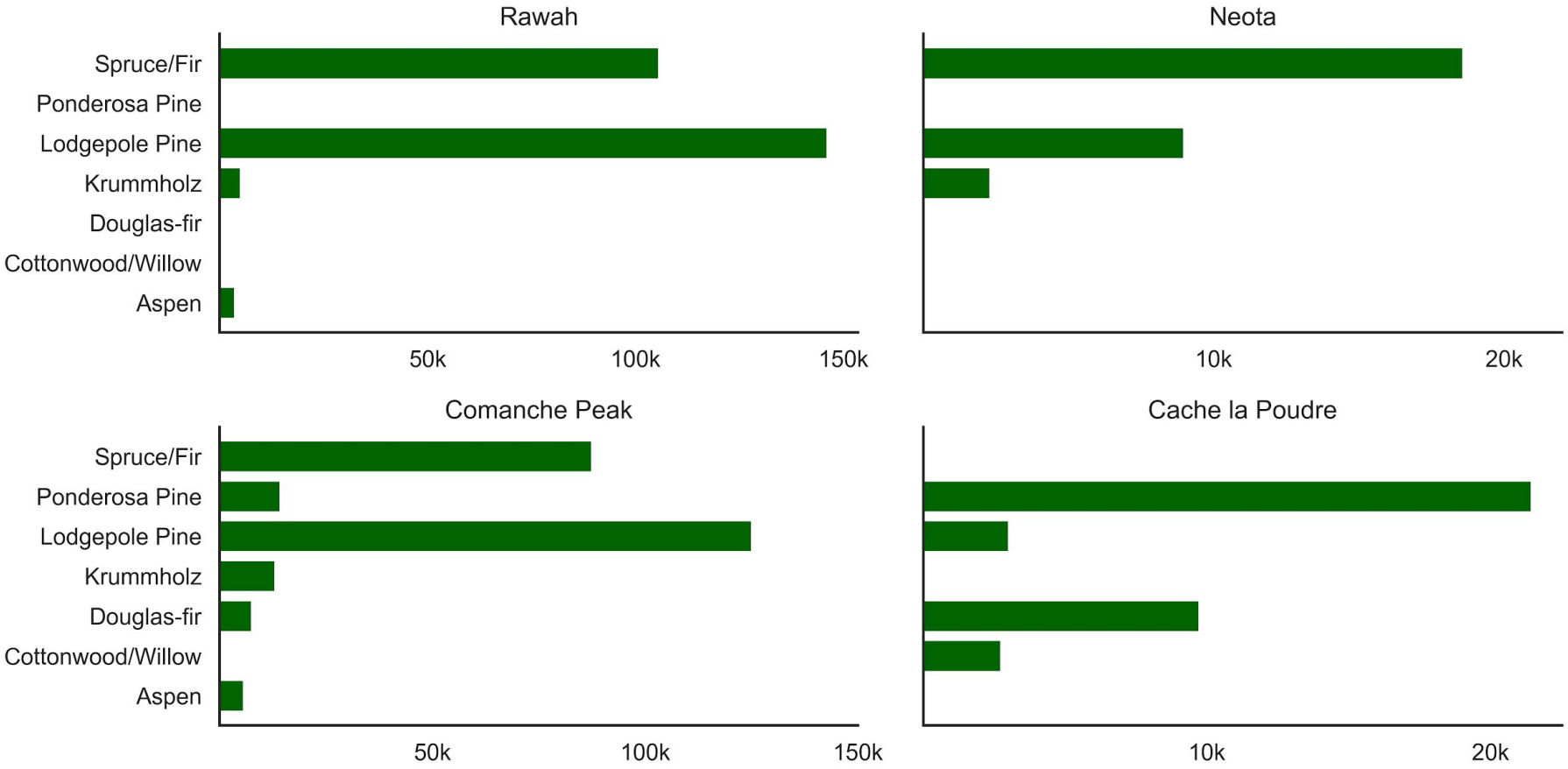
Why use machine learning?

Affordable

Scalable

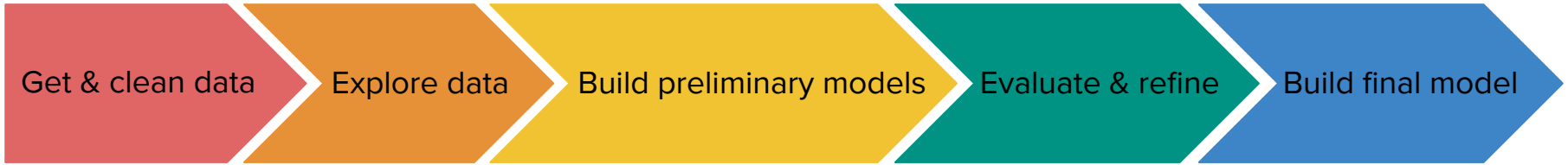
Powerful

Four CO Wilderness Areas



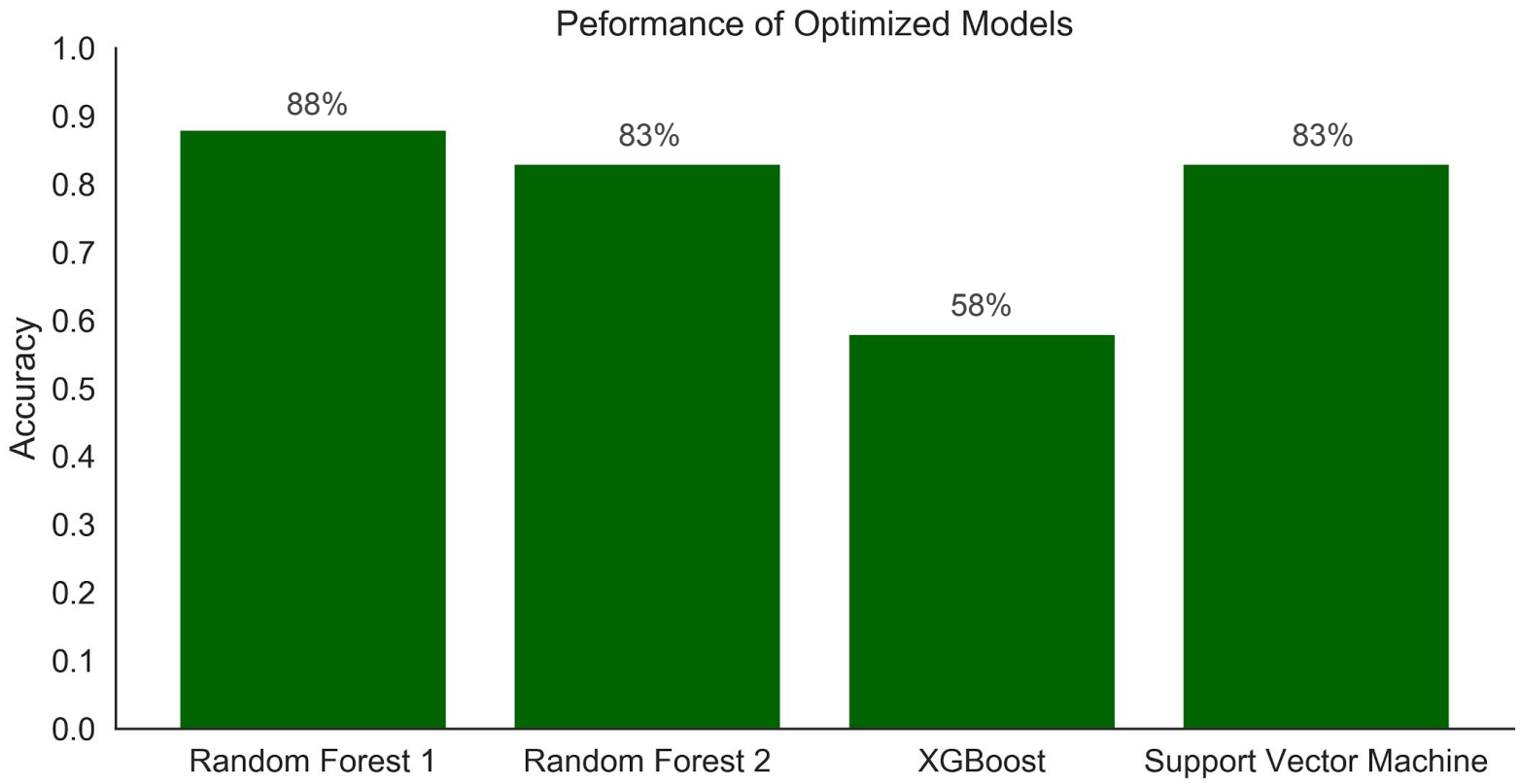
Tracts with various tree cover types

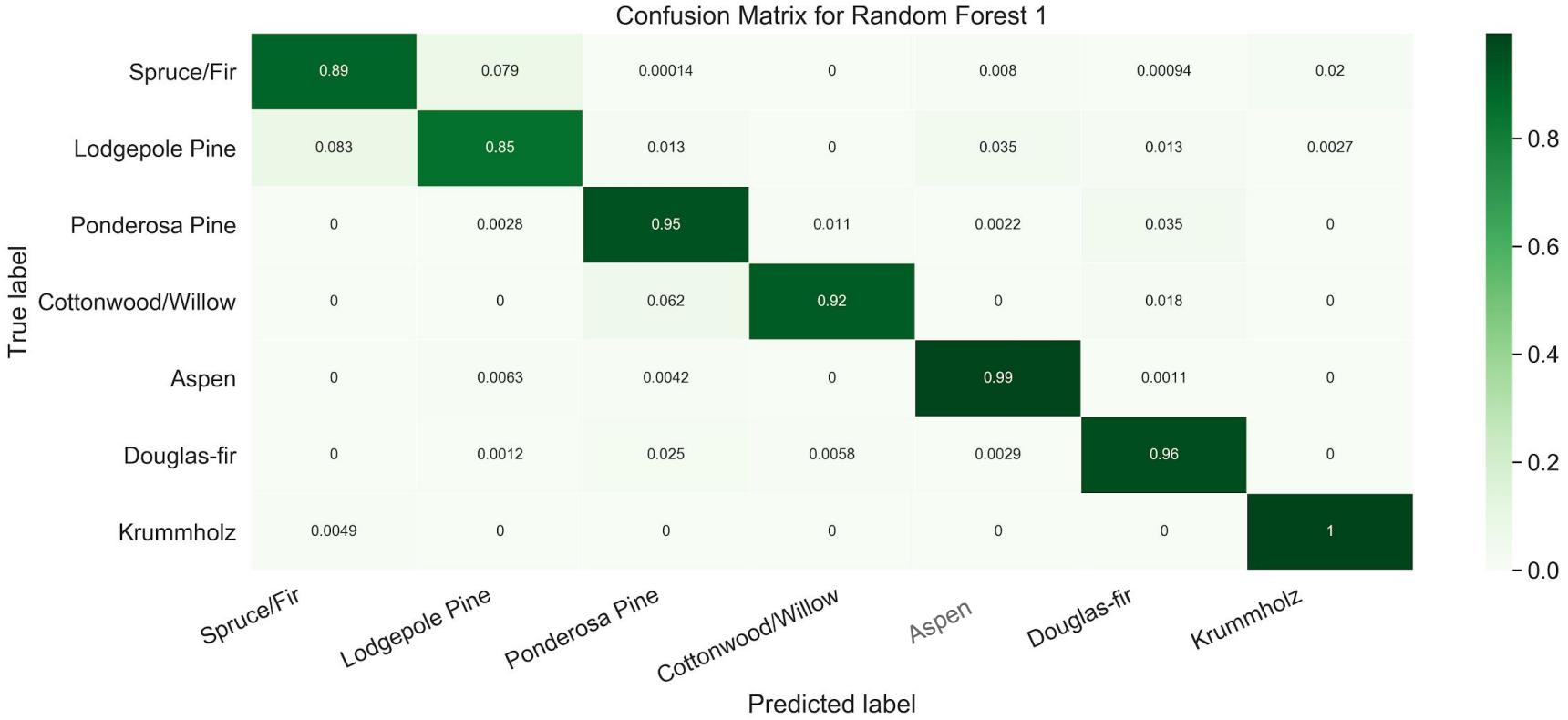
Methodology



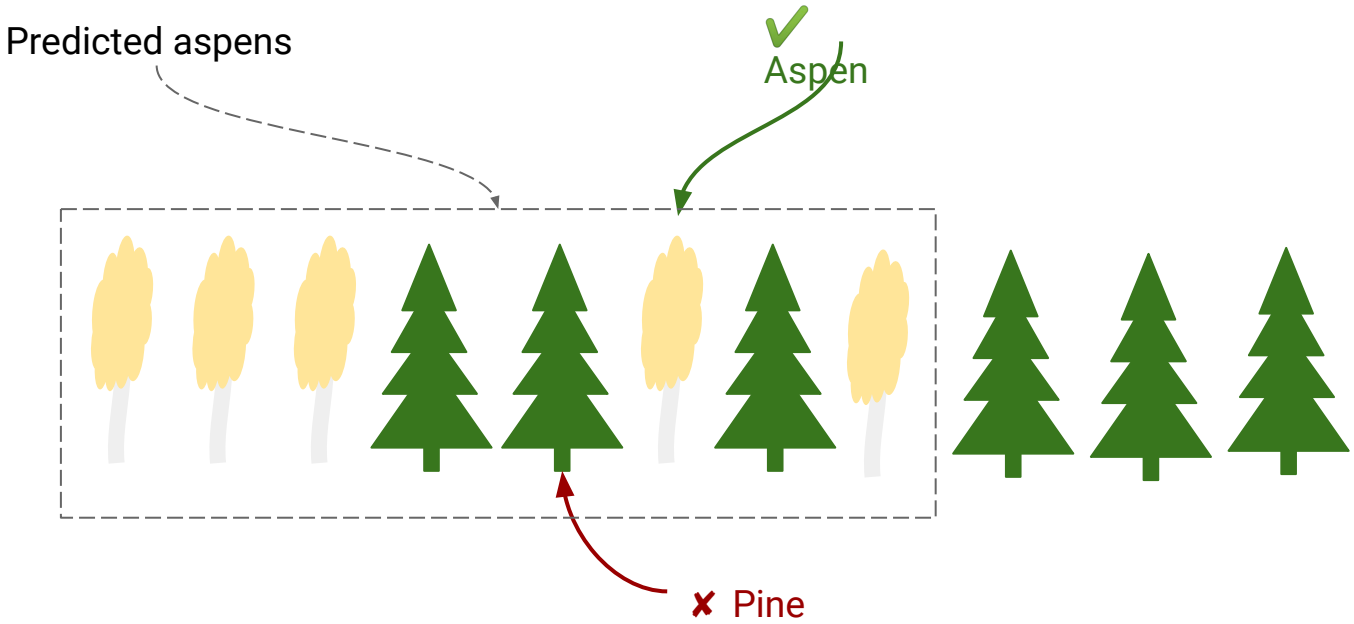
88%

Accuracy when identifying 7 tree cover types from cartographic data alone


Peformance of Optimized Models
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
88%
83%
58%
83%
Random Forest 1
Random Forest 2
XGBoost
Support Vector Machine

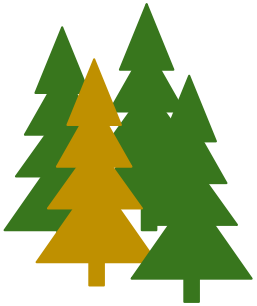

Confusion Matrix for Random Forest 1
Spruce/Fir
Lodgepole Pine
Ponderosa Pine
Cottonwood/Willow
Aspen
Douglas-fir
Krummholz
0.89 0.079 0.00014 0 0.008 0.00094 0.02
0.083 0.85 0.013 0 0.035 0.013 0.0027
0 0.0028 0.95 0.011 0.0022 0.035 0
0 0 0.062 0.92 0 0.018 0
0 0.0063 0.0042 0 0.99 0.0011 0
0 0.0012 0.025 0.0058 0.0029 0.96 0
0.0049 0 0 0 0 0 1
True label
Predicted label
0.8
0.6
0.4
0.2
0.0

Model weaknesses

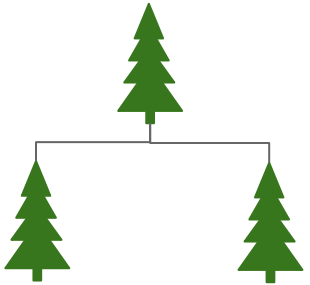


Recommendations

Multiple models for comparison



Choose the best model for the problem



Random Forest is quick and easy

Future work

Refine site-specific models

Incorporate data from other regions

Compare predictions to satellite images

Focus on areas critically affected by climate change

Thanks!

GitHub: github.com/jrkreiger/random-forest-trees

Twitter: @j_re

LinkedIn: linkedin.com/in/jrkreiger