Predicting fire line effectiveness with machine learning

Dennis W. Hallema[*], Christopher D. O'Connor, Matthew P. Thompson, Ge Sun, Steven G. McNulty, David E. Calkin & Katherine L. Martin

[*]North Carolina State University, Dept. Forestry and Environmental Resources
USDA Forest Service Rocky Mountain Research Station
USDA Forest Service Southern Research Station

Challenges in fire planning and operations management

- Assigning resources where they can be most effective
- Avoiding locations with extreme fire behavior + unacceptable risk for fire response team safety
- Existing models: limited resolution, predictive power
- Objective: Build a fast PCL model with better performance

Predicting potential fire control locations (PCLs)

Response variables

- High-risk zones (HRZs): Locations of increased fire hazard to be avoided (improve fire safety)
- Locations of control opportunities with high probability of success (PCLs)

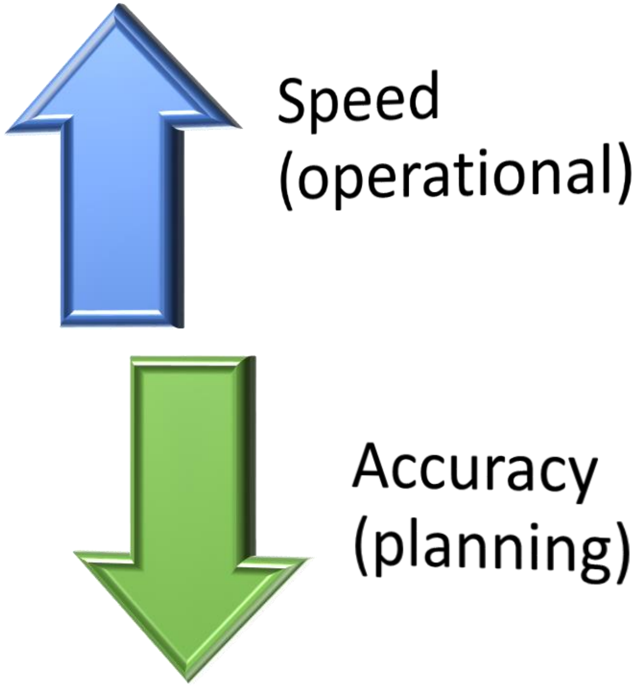
Predictors of PCLs

- Fire weather
- SDI suppression difficulty (snag density, slope steepness)
- Safety conditions (distance to safe zone, accessibility)
- Other topography (ridge distance)

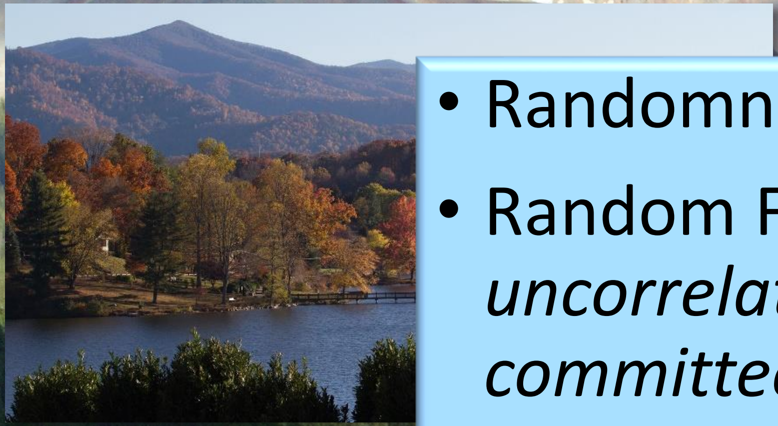
Thompson and others

Supervised learning classifiers

- Logistic Regression
- Support Vector Machine
- Decision Tree
- <u>Random Forest: a trade-off</u>
- Gradient Boosting Machine
- Artificial Neural Network

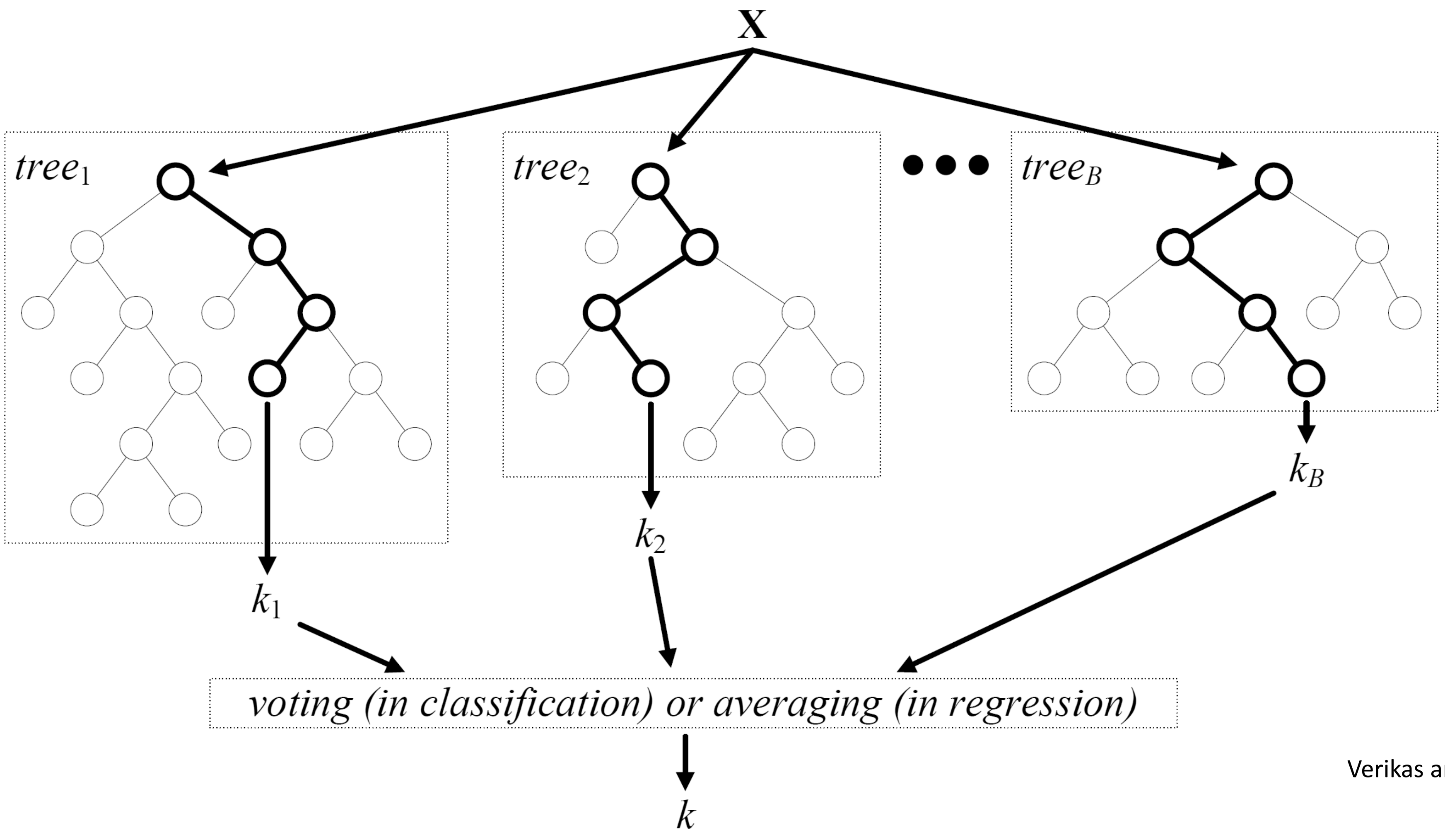


Random Forest



- Randomness ≠ “*We don’t know*”
- Random Forest = *“Large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.”*
- Ensemble method analogous to voting
- Fast parallel computation of trees

Random Forest

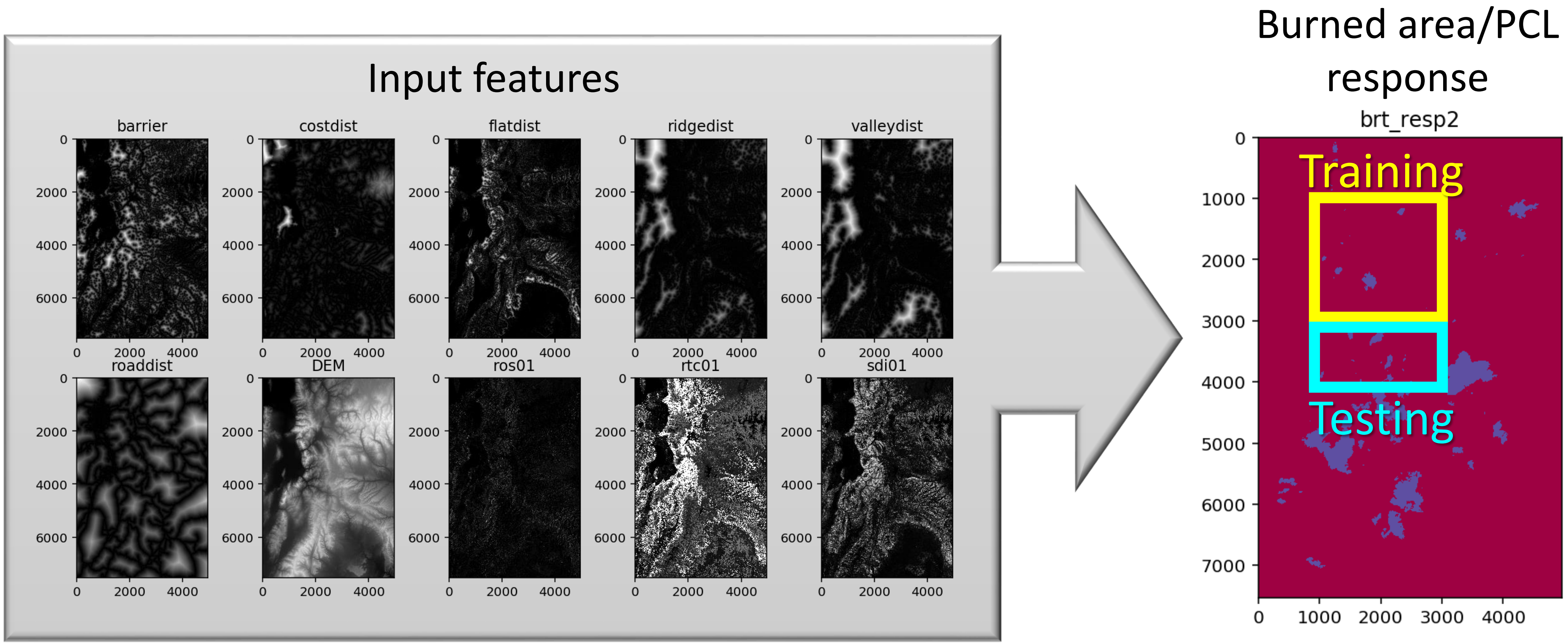


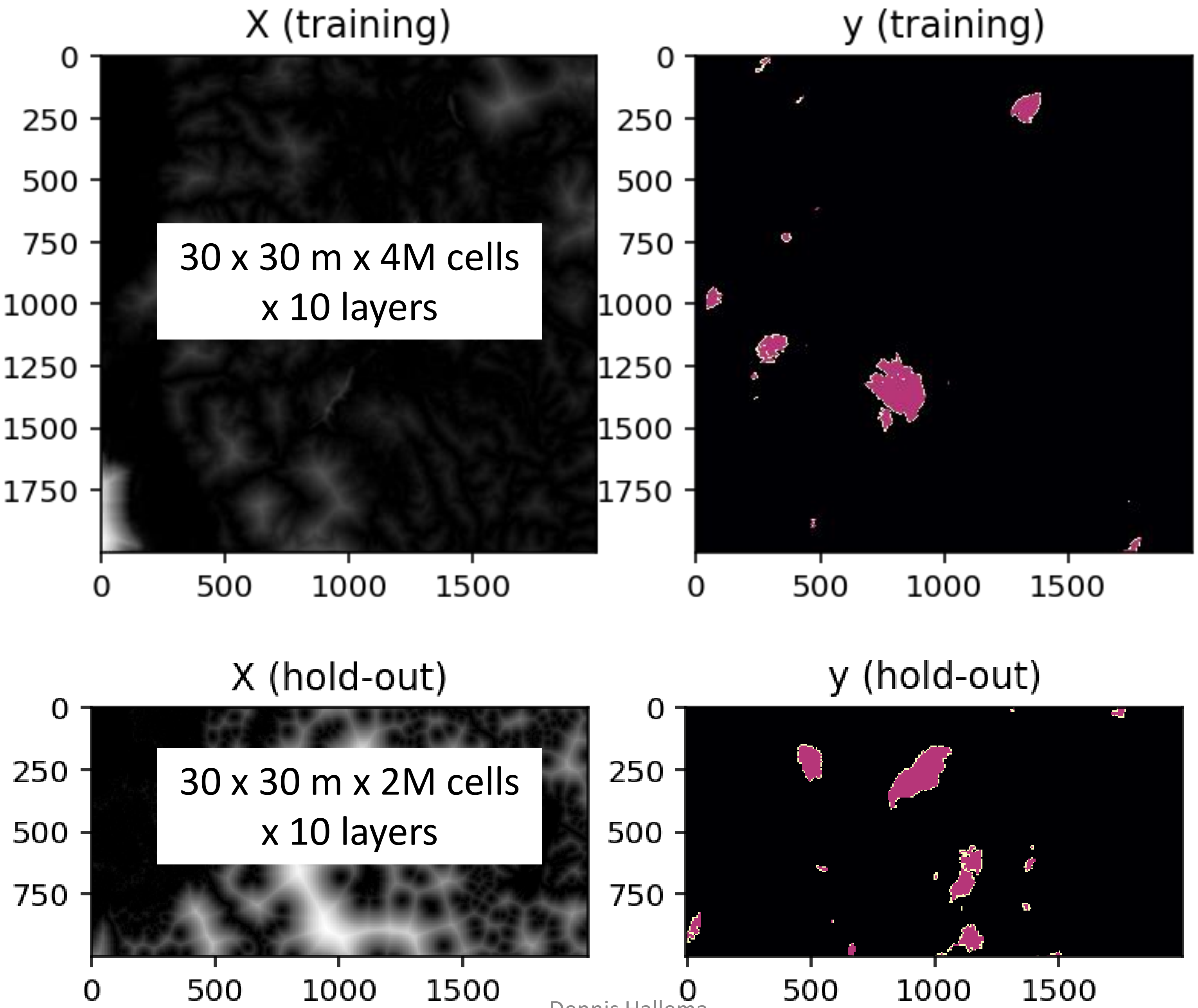
Verikas and others (2016)

Building a Random Forest in scikit-learn

1. Read 10 feature rasters and 1 burned area/PCL response raster
2. Subset area of interest
3. Split into contiguous training and testing sets
4. Fit Random Forest to training set
5. Predict testing set

2018 Pole Creek Fire in Central Utah

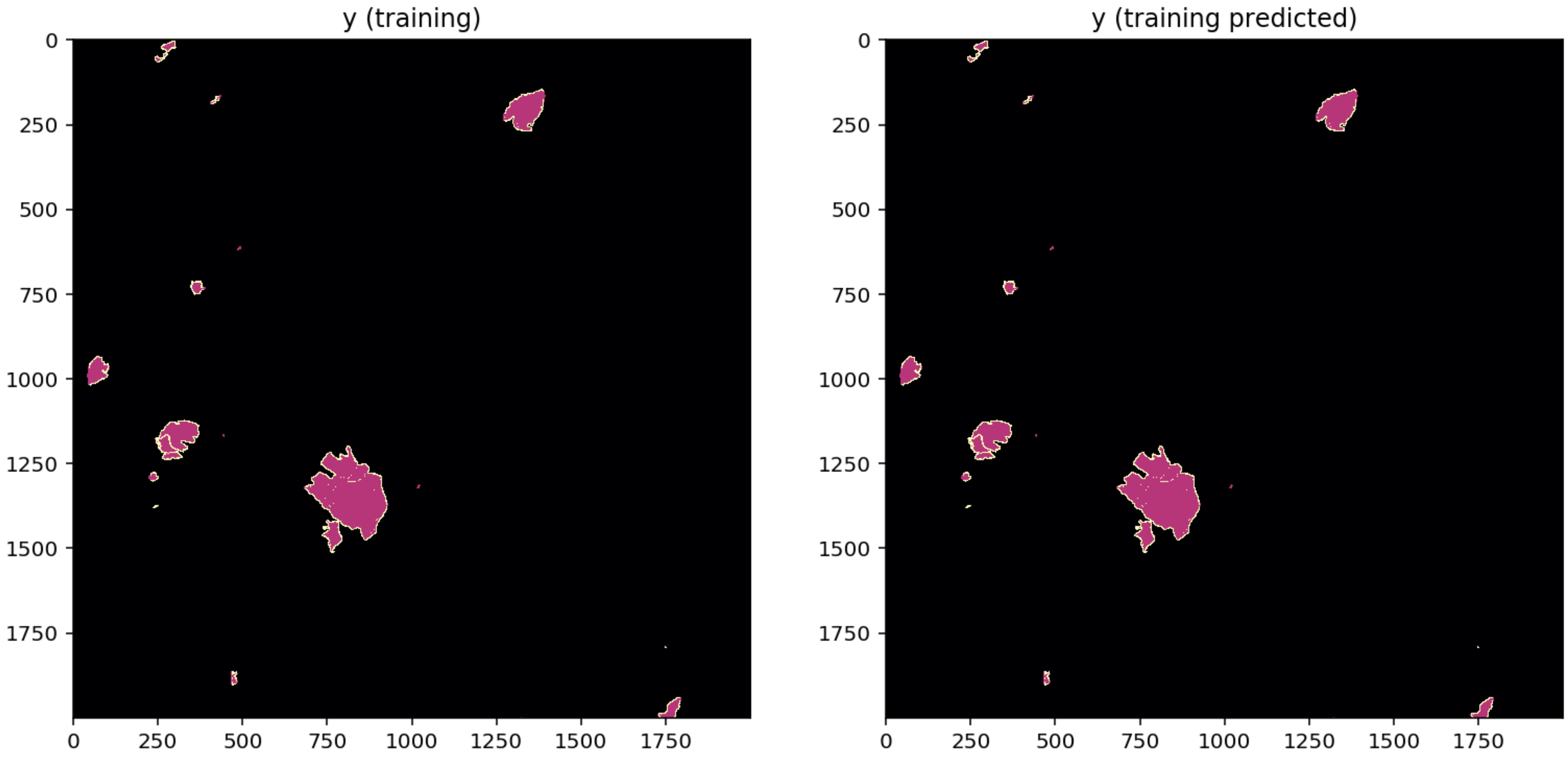



X (training)
30 x 30 m x 4M cells
x 10 layers
y (training)
X (hold-out)
30 x 30 m x 2M cells
x 10 layers
y (hold-out)
0
250
500
750
1000
1250
1500
1750
0
500
1000
1500

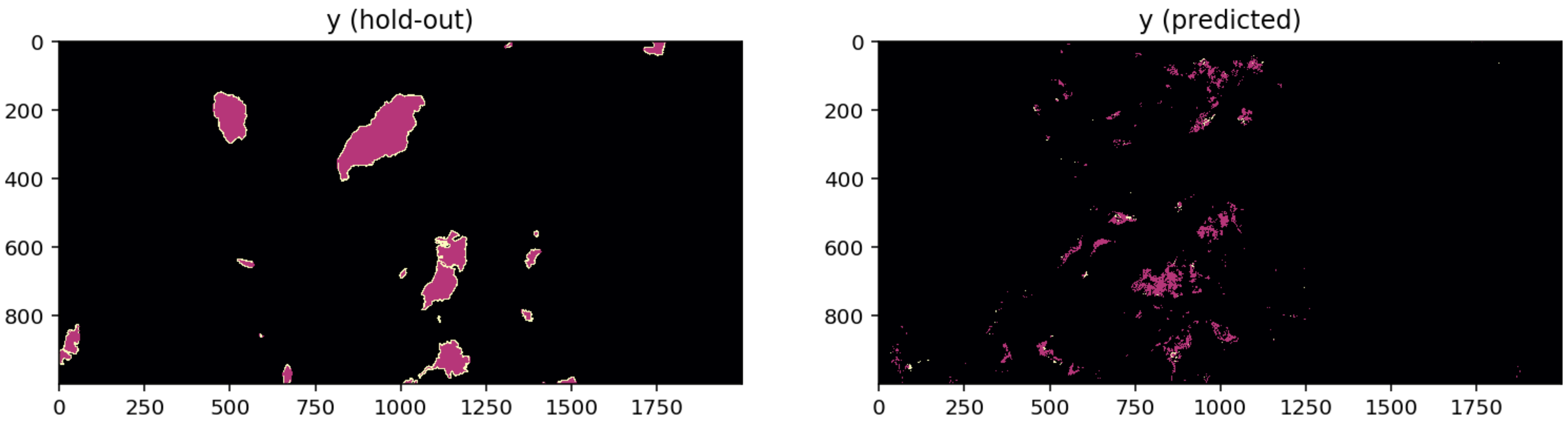
Training the HRZ/PCL Random Forest

RF Parameter	Values
Number of trees (“voters”) in forest	50
Criterion to measure the quality of a split	Gini
Max depth of tree	No maximum
Min samples required to split internal node	2
Min samples required to be at leaf node	1

100% accuracy for training

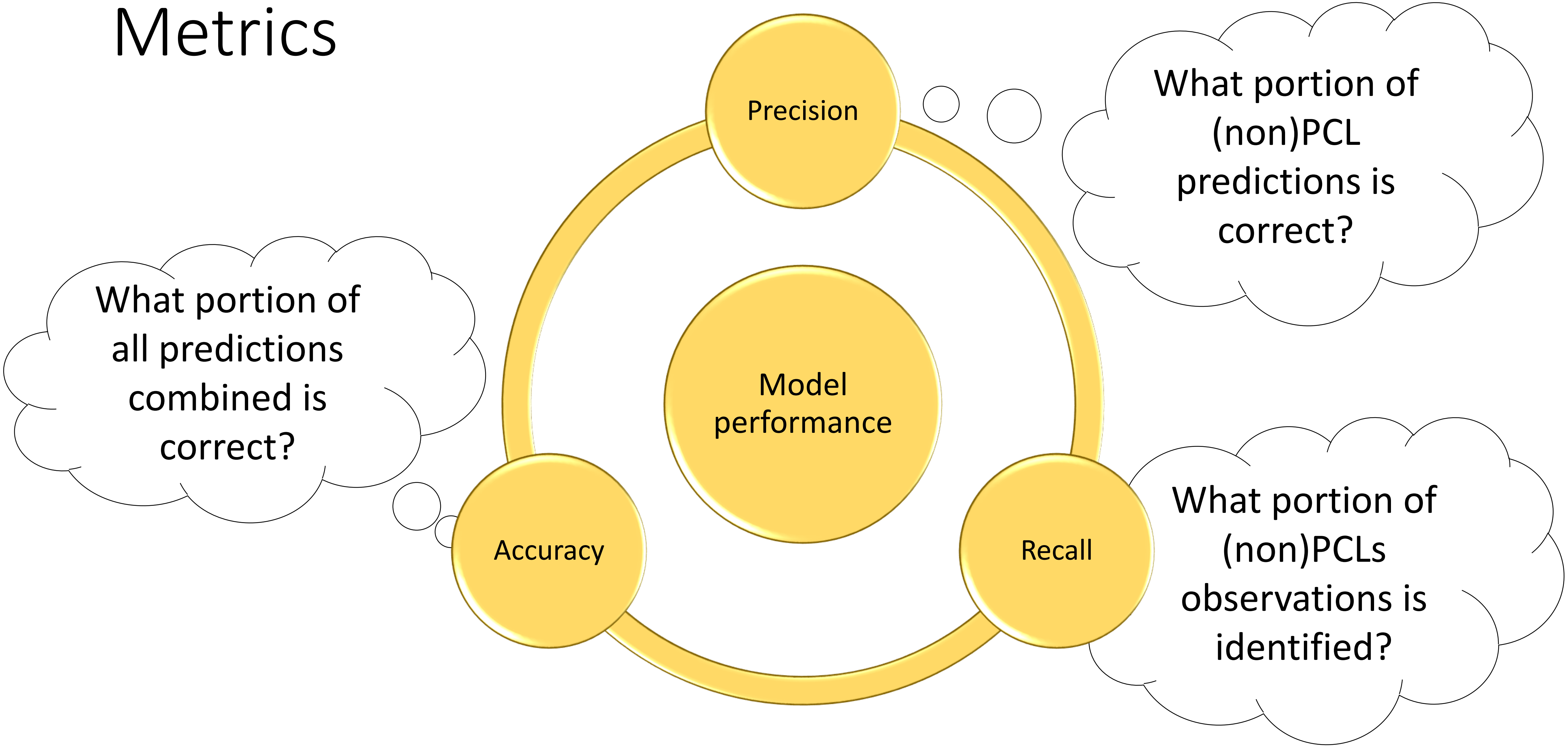


Fast hold-out prediction with Random Forest



Red – High-risk fire zone (HRZ)
Yellow – Potential fire control location (PCL)
Black – Unaffected by fire

Metrics


Precision
Model performance
Accuracy
Recall
What portion of (non)PCL predictions is correct?
What portion of all predictions combined is correct?
What portion of (non)PCLs observations is identified?

Training performance

- Mean squared error = 0.0000
- <u>Accuracy for all predictions = 100%</u>

	precision	recall	f1-score	support
Unaffect (0)	1.00	1.00	1.00	3927024
HRZ (1)	1.00	1.00	1.00	58898
PCL (2)	1.00	1.00	1.00	14078

Testing performance

- Mean squared error = 0.0861
- Accuracy for all predictions = 94%

	precision	recall	f1-score	support
Unaffect(0)	0.96	0.98	0.97	1919476
HRZ(1)	0.08	0.05	0.06	63665
PCL (2)	0.01	0.00	0.00	16859

Take-away

- Predicted HRZs near observed burn areas, PCLs highly localized
- Unaffected areas located correctly
- Power of fast parallel computation with Random Forests

Funding: USDA Forest Service Rocky Mountain Research Station. Acknowledgement: Yu Wei, Colorado State University. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the policies and views of the U.S. Government. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.