Machine learning to predict final fire size at the time of ignition

Shane Coffield

PhD Candidate in Earth System Science at UC Irvine

Collaborators:

Jim Randerson and Yang Chen *(Earth System Science)*

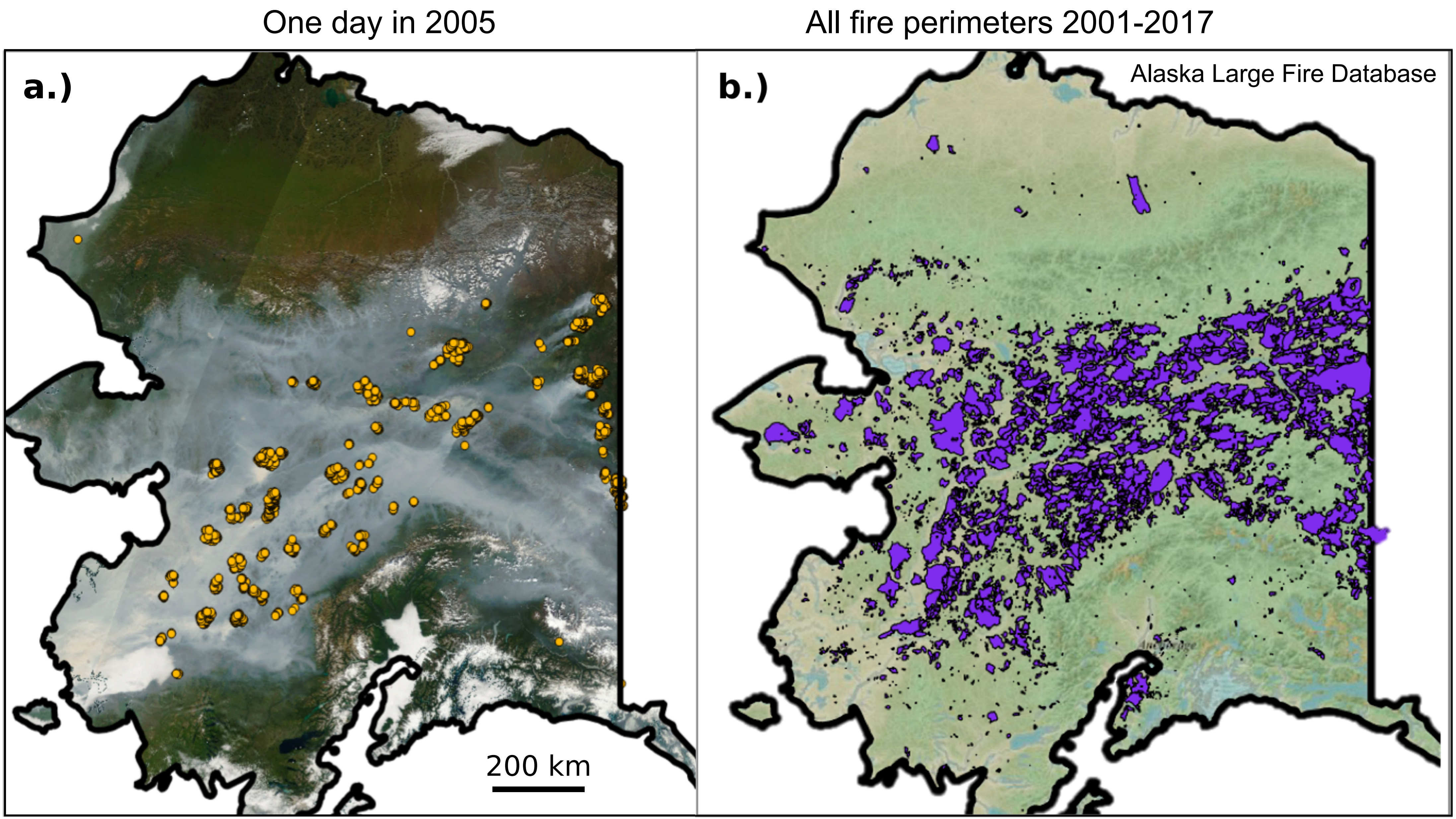
Casey Graff and Padhraic Smyth (*Information and Computer Science)*

Efi Foufoula-Georgiou *(Civil and Environmental Engineering)*

March 25, 2020

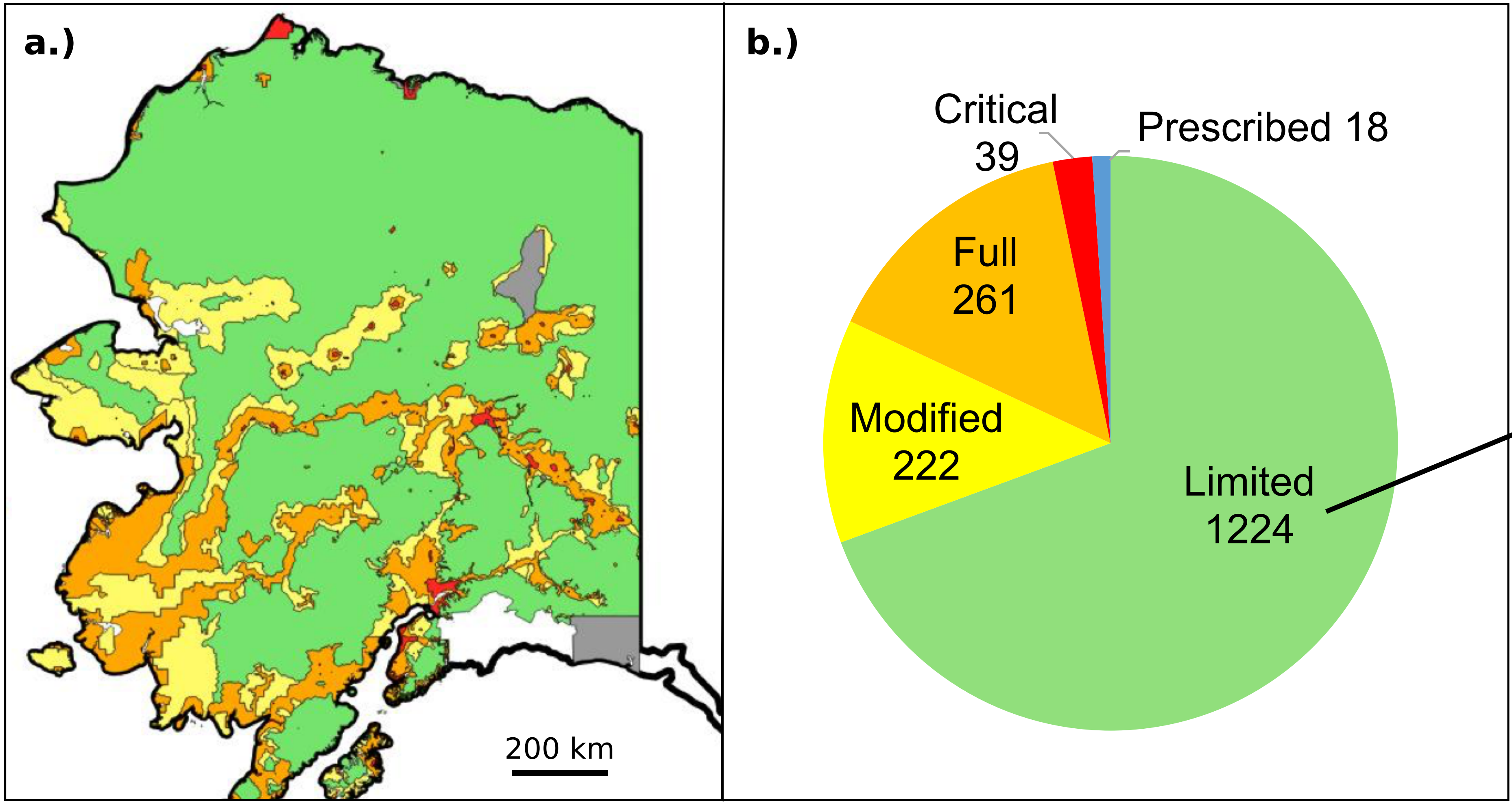
Alaska Fire Science Consortium Spring Workshop

Study area & data



Study area & data

Fire management zones



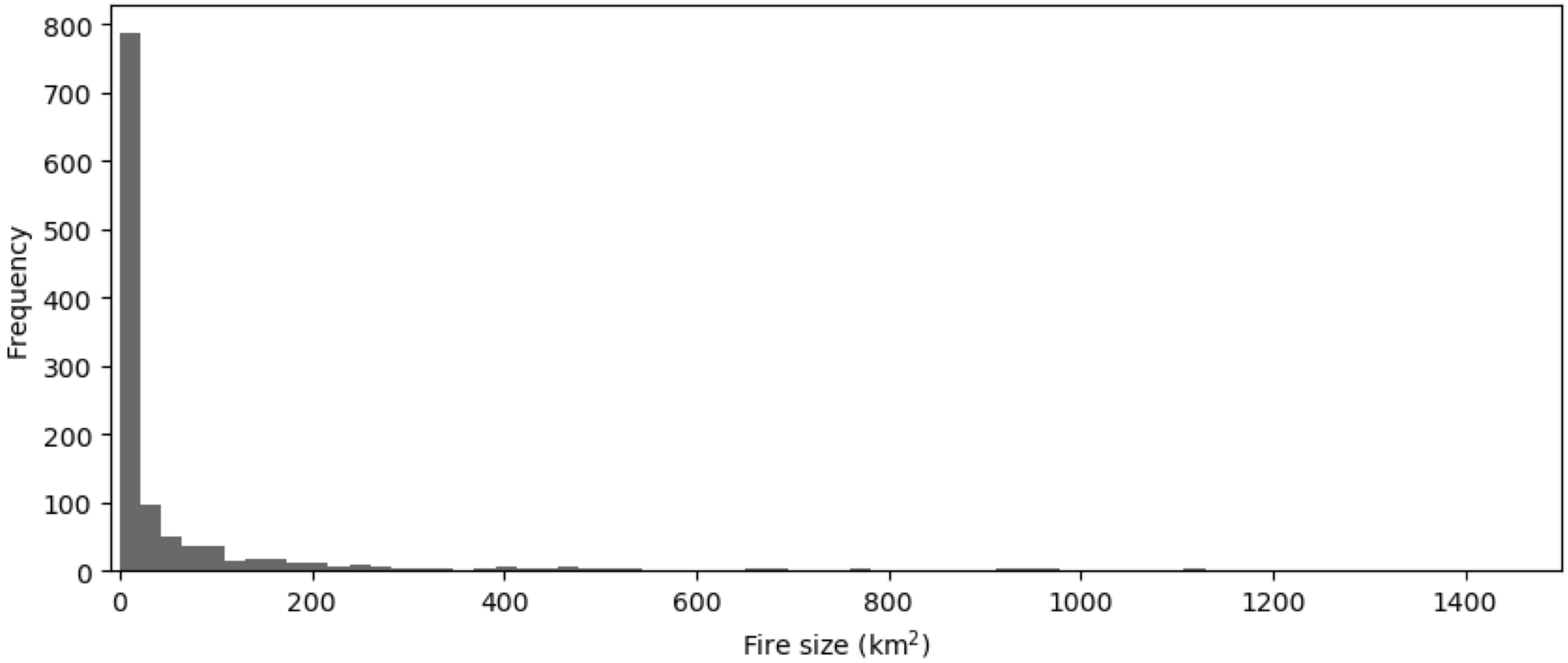
Further filtered by agreement to MODIS fire detections

↓

1168 fires in our dataset

Motivation

- Many fires occur (1168 fires over 17 years)
- Only a few become very large
- Ecosystems are vulnerable to climate-driven increases in fire activity
- Fire fighting resources are limited



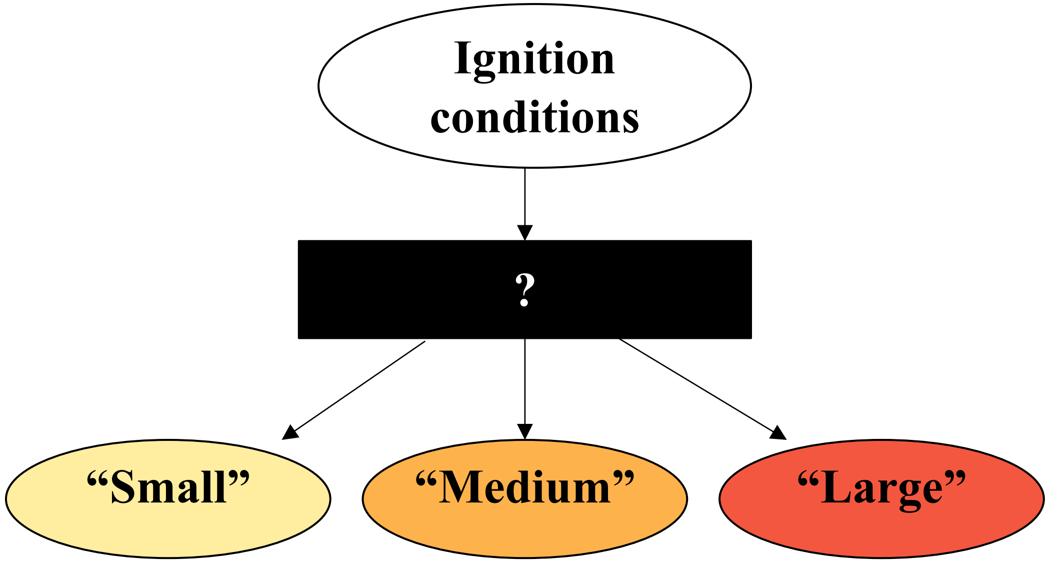
Approach

Science question

What variables control fire size from the time of ignition?

Topography, Weather, Vegetation at different scales

Application: Triaging approach to fire prediction; Using only information available at the time of ignition, predict whether a fire will be **small**, **medium**, or **large** by the time it has run its course.



Approach

Data

	Source	Variables	Resolution
Fires	Alaska Large Fire Database (1168 fires in ”limited” mgmt.)	• Ignition location • Ignition date • Final burned area	
Weather	ECMWF ERA-5 reanalysis	• 2-m air temperature • Relative humidity • Precipitation • 10-m wind speed • Surface pressure • Vapor pressure deficit (derived)	0.25º; hourly (aggregated to daily)
Topography	USGS GTOPO30	• Slope • Aspect • Elevation	1 km
Vegetation	LANDFIRE Existing Vegetation Type	• Fraction hardwood (aspen/birch) • Fraction softwood (black or white spruce) • Fraction shrub • Fraction grass	30 m; for specific years (2001, 2008, 2010, 2012, 2014)

Modeling

- 1168 fires divided into terciles
  - “Small” (< 1.2 km^2)
  - ”Medium”
  - ”Large” (> 19.8 km^2)
- Decision Tree classifier
  - 10-fold cross-validation
  - Performance metric: mean outgroup classification accuracy
- Tune for optimality
  - Spatial window around ignition for vegetation cover
  - Time window around ignition for weather data
  - Tree size
  - Predictor variables to include
- Compare to other ML algorithms
  - Random forest
  - K-nearest neighbors
  - Gradient boosting
  - Multi-layer perceptron (MLP)

Approach

Data

	Source	Variables	Resolution
Fires	Alaska Large Fire Database (1168 fires in ”limited” mgmt.)	• Ignition location • Ignition date • Final burned area	
Weather	ECMWF ERA-5 reanalysis	• 2-m air temperature • Relative humidity • Precipitation • 10-m wind speed • Surface pressure • Vapor pressure deficit (derived)	0.25º; hourly (aggregated to daily)
Topography	USGS GTOPO30	• Slope • Aspect • Elevation	1 km
Vegetation	LANDFIRE Existing Vegetation Type	• Fraction hardwood (aspen/birch) • Fraction softwood (black or white spruce) • Fraction shrub • Fraction grass	30 m; for specific years (2001, 2008, 2010, 2012, 2014)

Modeling

- 1168 fires divided into terciles
  - “Small” (< 1.2 km^2)
  - ”Medium”
  - ”Large” (> 19.8 km^2)
- Decision Tree classifier
  - 10-fold cross-validation
  - Performance metric: mean outgroup classification accuracy
- Tune for optimality
  - Spatial window around ignition for vegetation cover
  - Time window around ignition for weather data
  - Tree size
  - Predictor variables to include
- Compare to other ML algorithms
  - Random forest
  - K-nearest neighbors
  - Gradient boosting
  - Multi-layer perceptron (MLP)

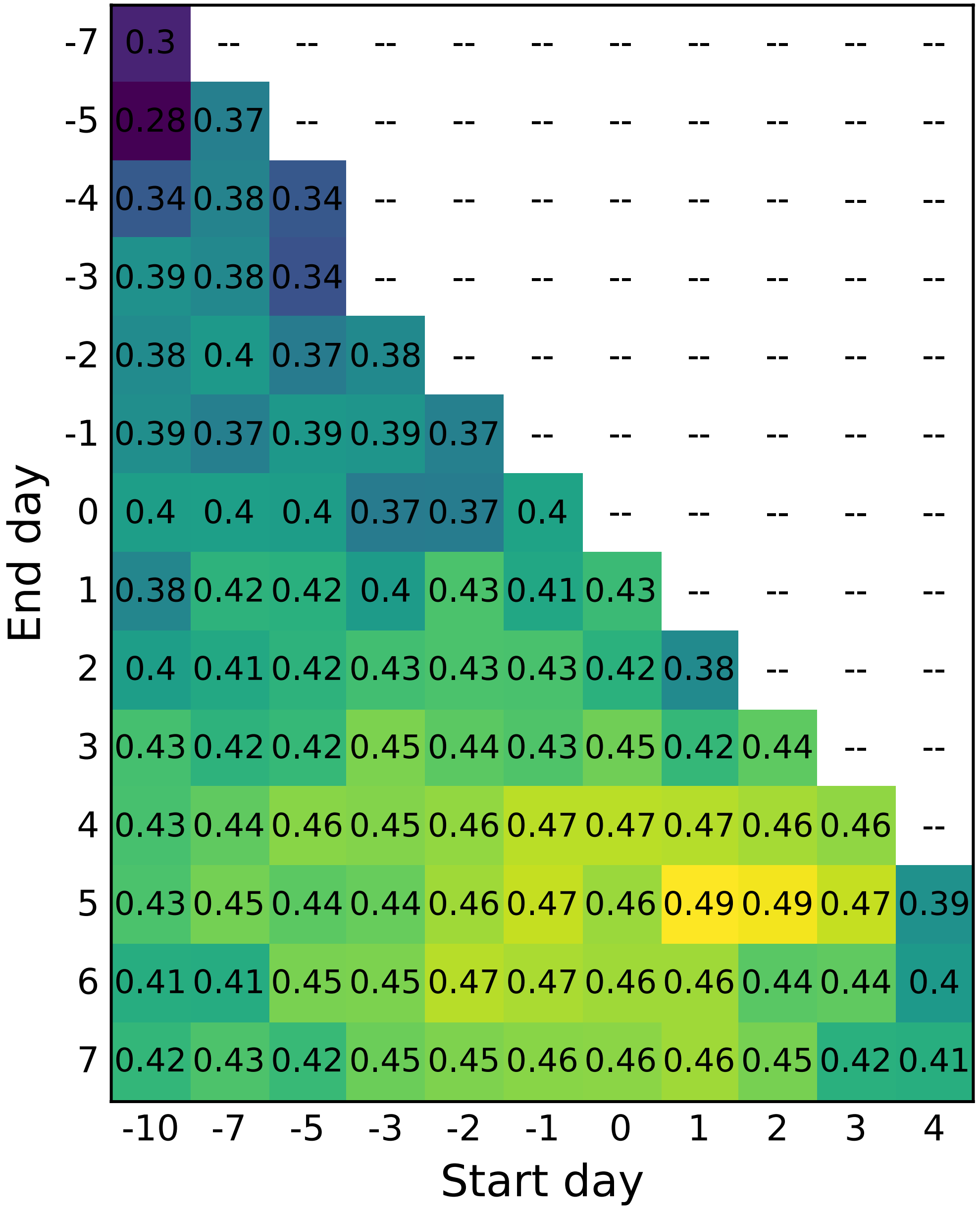
Results

Optimal radius
around ignitions to average veg data
- 4km

Optimal time window
around ignitions to average weather data
- 1 day after ignition → 5 days after ignition

Optimal tree size
- 4 nodes

Most important predictors
- VPD (1 to 5 days)
- Spruce fraction (4 km around ignition)



Results

Accuracy including different variables

Added note about VPD:

Vapor pressure deficit is a derived metric combining T and RH. VPD is the *difference* between the saturation vapor pressure (VP_{sat}) and the observed vapor pressured (VP), whereas RH is the ratio of VP/VP_{sat}. VPD scales more linearly with ET and increases faster at high temperatures than RH decreases

Variables included	Accuracy of best model	p-value
Random classification		
None	33.3 ± 4.4%	-
One-variable models		
RH	47.2 ± 4.9%	<0.001*
T	39.4 ± 6.4%	0.013*
P	45.7 ± 5.0%	<0.001*
VPD	**49.2 ± 4.7%**	<0.001*
W	29.6 ± 9.0%	0.868
SP	31.6 ± 9.7%	0.689
T_{anom}	37.6 ± 6.7%	0.055
Two-variable models		
VPD, T	49.2 ± 4.7%	<0.001*
VPD, P	48.8 ± 5.5%	<0.001*
VPD, RH	47.8 ± 3.8%	<0.001*
T, P	44.4 ± 5.6%	<0.001*
T, RH	44.0 ± 5.6%	<0.001*
P, RH	45.2 ± 4.3%	<0.001*
Three-variable models		
VPD, T, P	48.8 ± 5.5%	<0.001*
VPD, T, RH	45.7 ± 5.8%	<0.001*
VPD, P, RH	47.8 ± 3.8%	<0.001*
T, P, RH	43.1 ± 5.3%	<0.001*
Four-variable model		
VPD, T, P, RH	45.5 ± 5.9%	<0.001*

Variables included	Accuracy of best model	p-value
Random classification		
None	33.3 ± 4.4%	-
One-variable models		
Spruce fraction	40.7 ± 7.1%	0.007*
Birch/aspen fraction	29.4 ± 4.8%	0.962
Day of year	39.1 ± 7.3%	0.025*
Slope	36.2 ± 7.1%	0.145
Aspect	26.4 ± 7.6%	0.989
Elevation	34.7 ± 6.0%	0.280
Combination models		
Spruce, birch/aspen	40.4 ± 7.2%	0.009*
VPD, spruce	**50.4 ± 5.2%**	<0.001*
VPD, birch/aspen	49.2 ± 4.7%	<0.001*

Results

Decision Tree Classification (small, medium, or large fire)

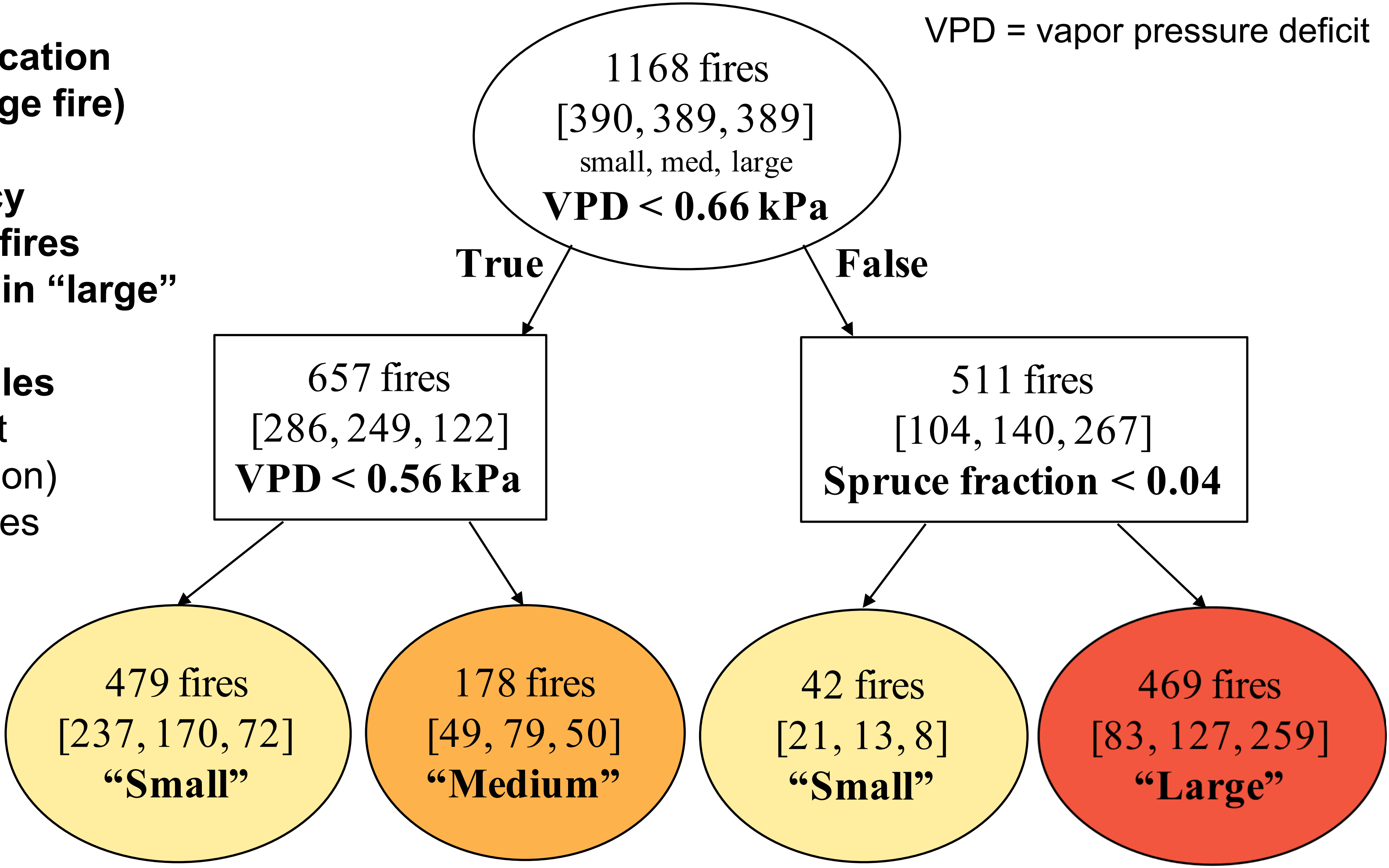
- **50% overall accuracy**
- **65% recall for large fires**
- **75% of burned area in "large"**

Most important variables

- Vapor pressure deficit
  (1-5 days after ignition)
- Fraction of spruce trees
  (In a 4-km radius)

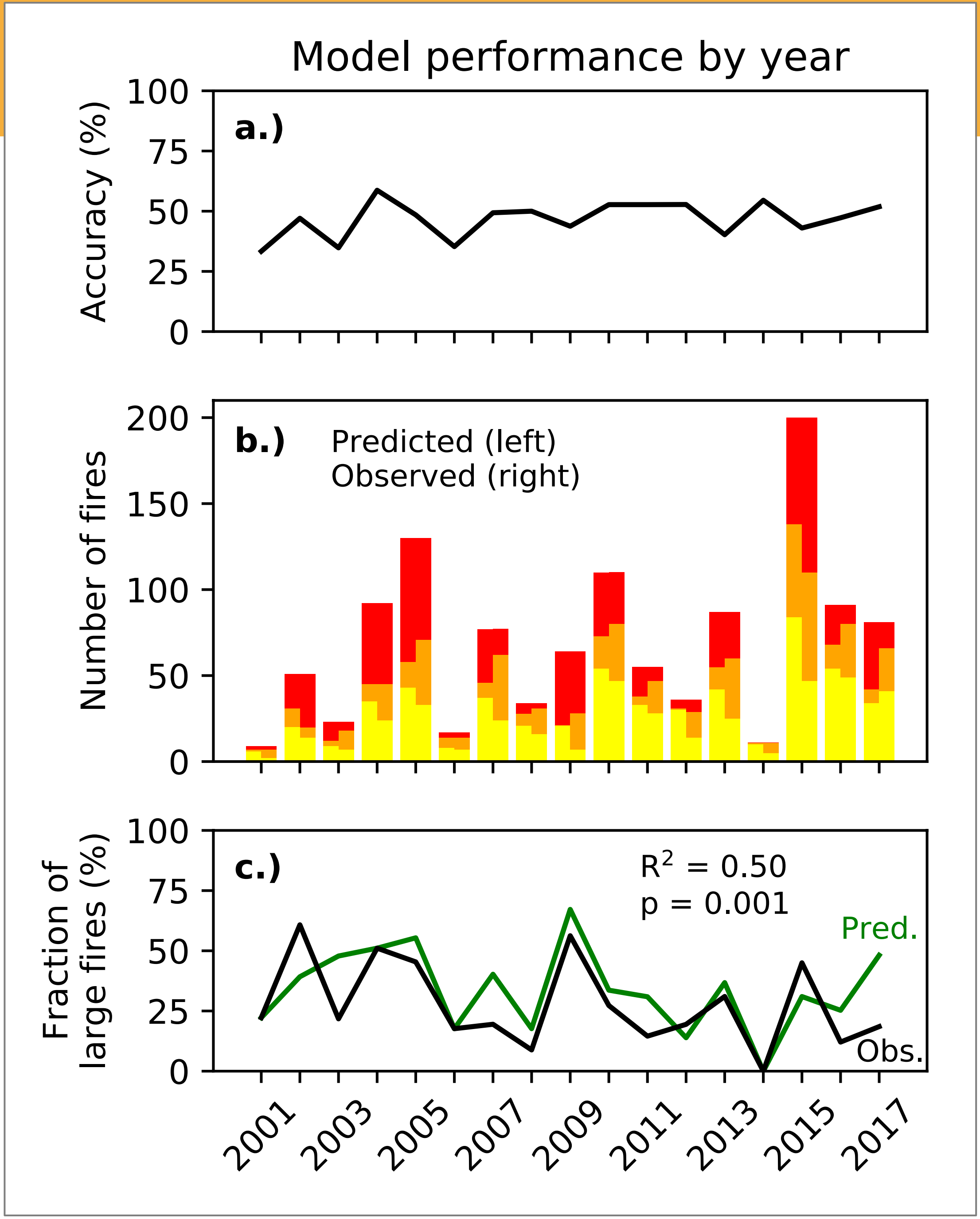
More complex ML algorithms did not outperform decision trees

VPD = vapor pressure deficit



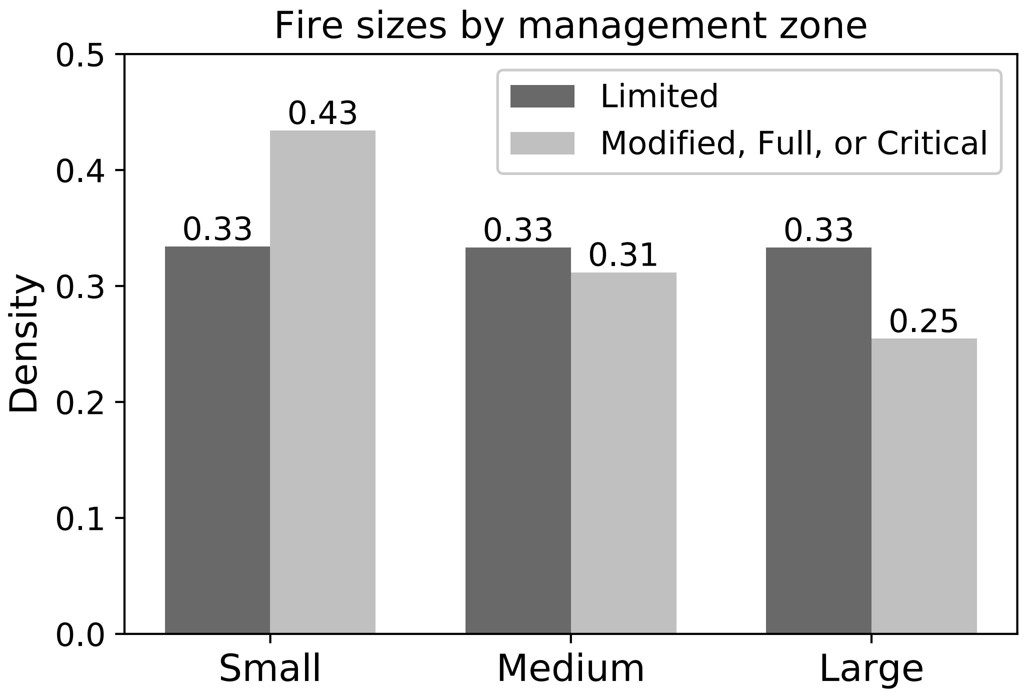
Results

Simple decision tree model generally captures the interannual variability of fire size



Results

Considering fires in more managed zones….



	Management option		
	Critical, full, or modified	Limited	All
Mean fire size (km^2)	52.0	79.9	71.8
Burned area/year (km^2/yr)	1183	4449	5631
Fires/year	22.0	55.0	78.0
Area (km^2)	170,543	463,038	633,581
Burned area/year/area (km^2/yr/km^2)	6.94×10^{-3}	9.61×10^{-3}	8.89×10^{-3}
Fires/year/area (#/yr/km^2)	1.29×10^{-4}	1.19×10^{-4}	1.23×10^{-4}

Fires in more managed zones are

- Smaller
- Burn less total area
- Higher spatial density (More frequent when normalized by zone area)

Inferred human footprint is to

- Increase the number of fires by 3.4%
- Decrease total annual burned area by 7.5%

Results

When applied to areas more managed fire zones, we find that….

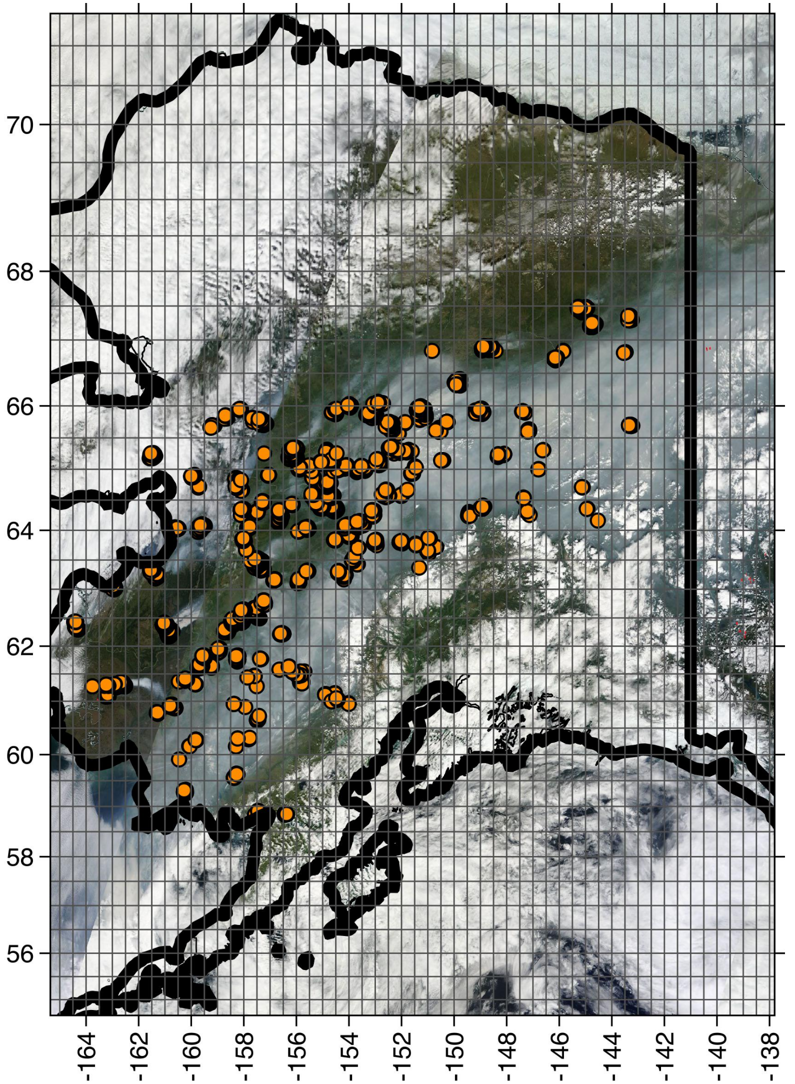
- Accuracy decreases (50% → 43%) but recall does not
- Without suppression, those fires likely would have been disproportionately larger
  - ~1.8 times larger than observed
  - Humans tend to ignite fires on drier days than lightning
  - Suppression efforts decreased burned area by about 44% in those areas

Phillip Spor, UAF

Other research

Forecasting daily wildfire activity using Poisson regression
Casey Graff et al. in *IEEE Transactions on Geoscience and Remote Sensing*

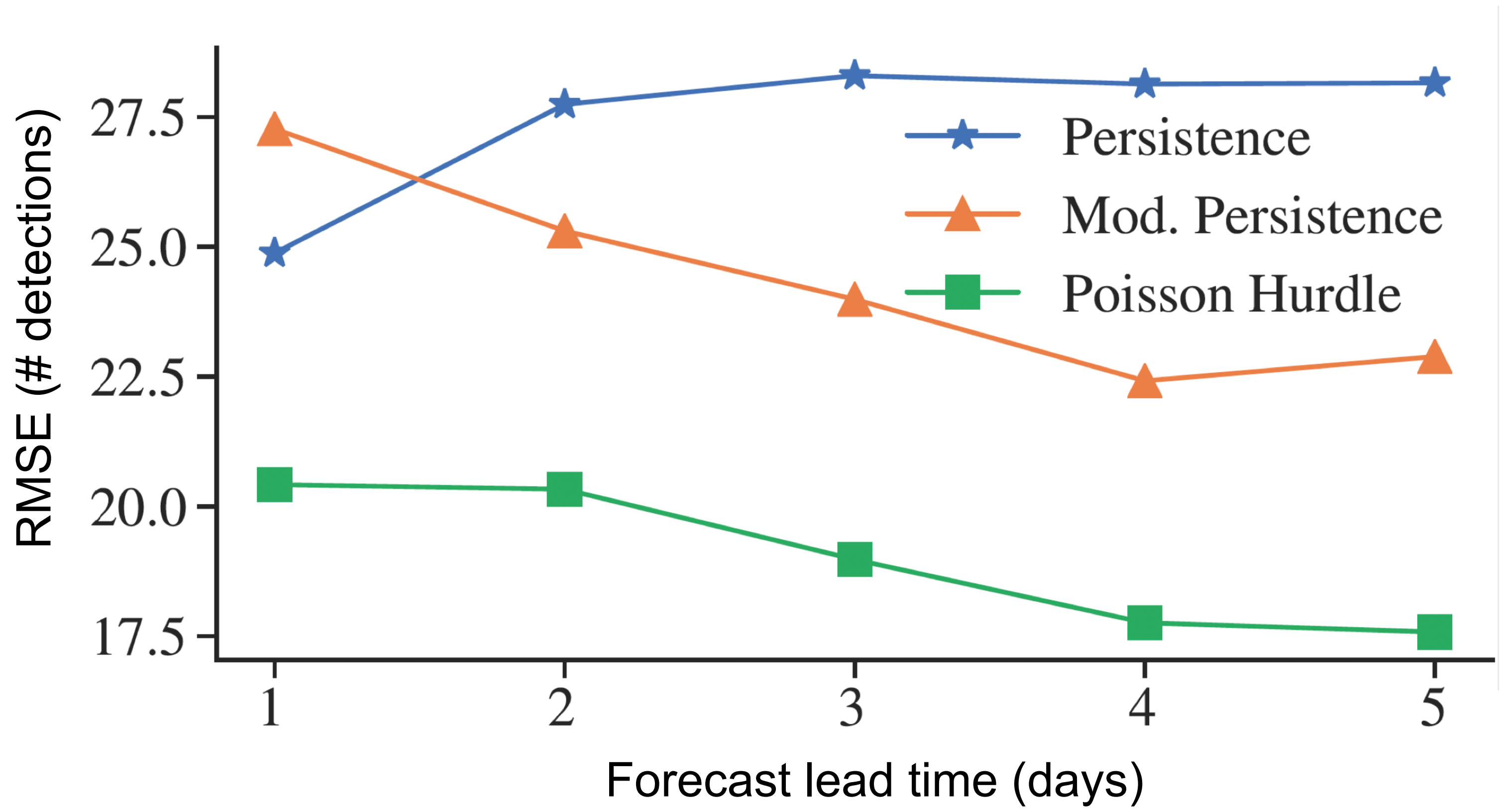
- **Goal**: Offer basis for improved smoke forecasting over current persistence models
- **Approach**:
  - Leverage weather information and learned patterns of fire evolution
  - Use MODIS fire detections as a proxy for smoke
  - Predict number of detections in each ½º gridcell 1-5 days in advance

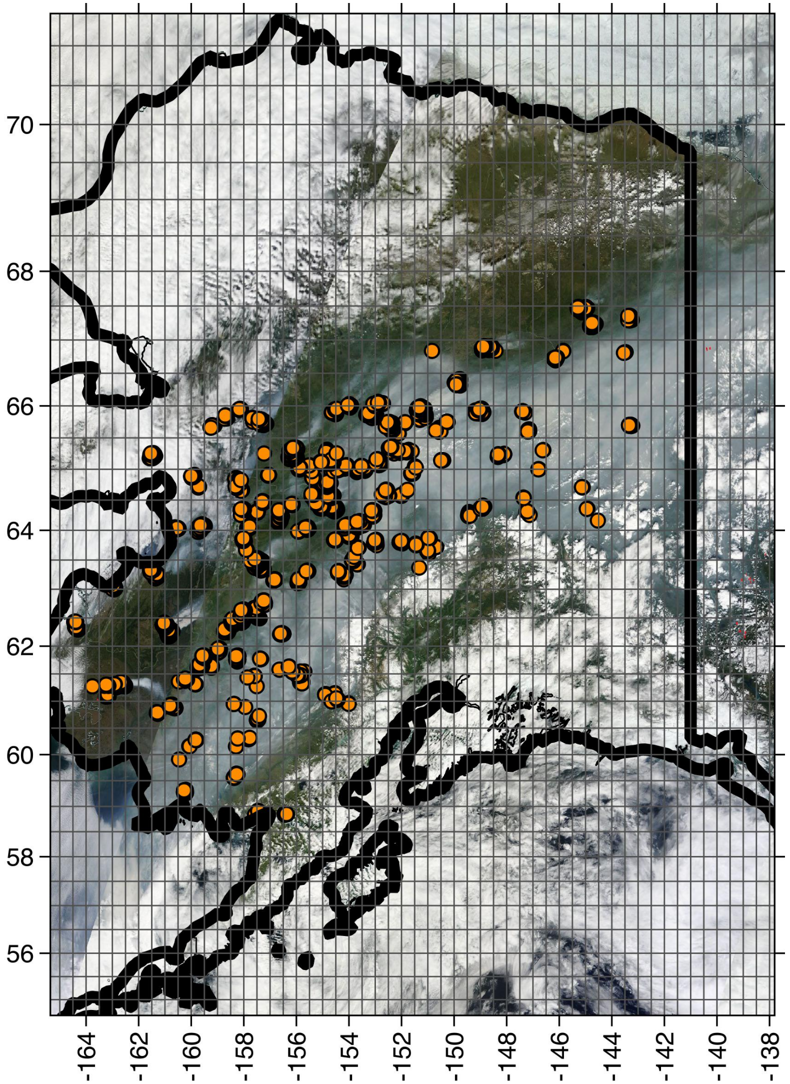


Other research

Forecasting daily wildfire activity using Poisson regression
Casey Graff et al. in *IEEE Transactions on Geoscience and Remote Sensing*

- **Results**





Conclusions

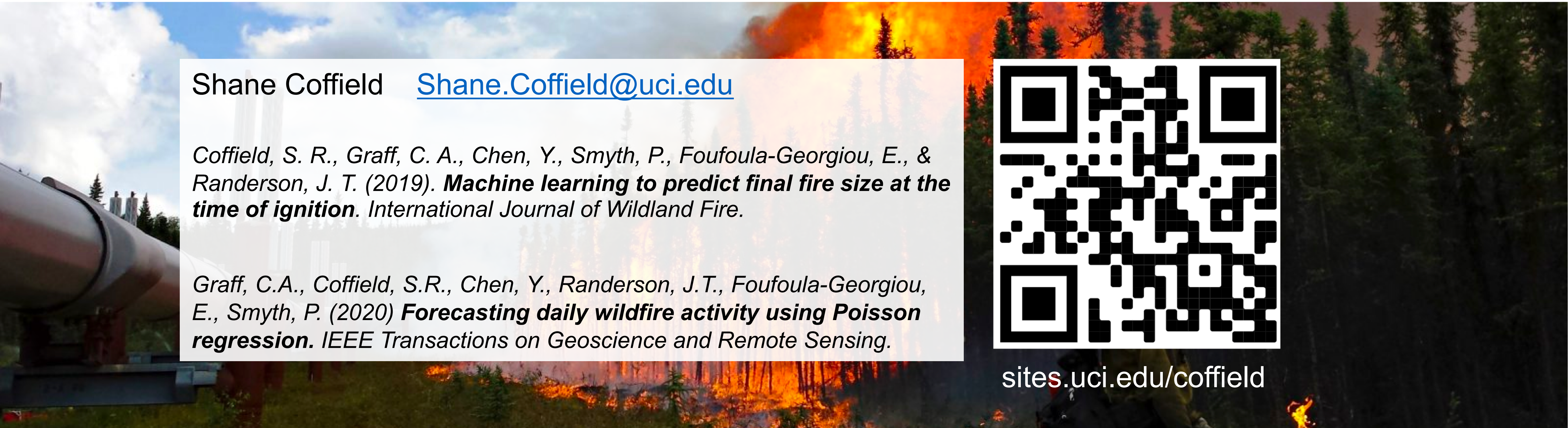
- Our simple machine learning approach can triage fires into 3 size groups with 50% accuracy
- VPD in the first 5 days of a fire is the best predictor of final fire size
- Identifying role of management: In the context of climate change, new approaches may be needed to maintain the current fire regime (and slow the impact of new extreme fire seasons).
    - These simple statistical techniques may help guide suppression efforts to protect vulnerable ecosystems and carbon

Phillip Spor, UAF

Thank you!

Discussion points

- Future approaches to fire management given climate change
- Usefulness of this triaging approach in application? Worth developing further?
- Collaborations at intersection of ML and fire prediction in Alaska or California



Shane Coffield   Shane.Coffield@uci.edu

Coffield, S. R., Graff, C. A., Chen, Y., Smyth, P., Foufoula-Georgiou, E., & Randerson, J. T. (2019). ***Machine learning to predict final fire size at the time of ignition****. International Journal of Wildland Fire.*

Graff, C.A., Coffield, S.R., Chen, Y., Randerson, J.T., Foufoula-Georgiou, E., Smyth, P. (2020) ***Forecasting daily wildfire activity using Poisson regression.*** *IEEE Transactions on Geoscience and Remote Sensing.*

sites.uci.edu/coffield