Poster #GC11F-1164

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

Shane R. Coffield[1], Casey A. Graff[2], Yang Chen[1], Padhraic Smyth[2], Efi Foufoula-Georgiou[1,3], James T. Randerson[1]

UC Irvine Depts. of [1]Earth System Science, [2]Computer Science, and [3]Civil & Environmental Engineering

Key Points

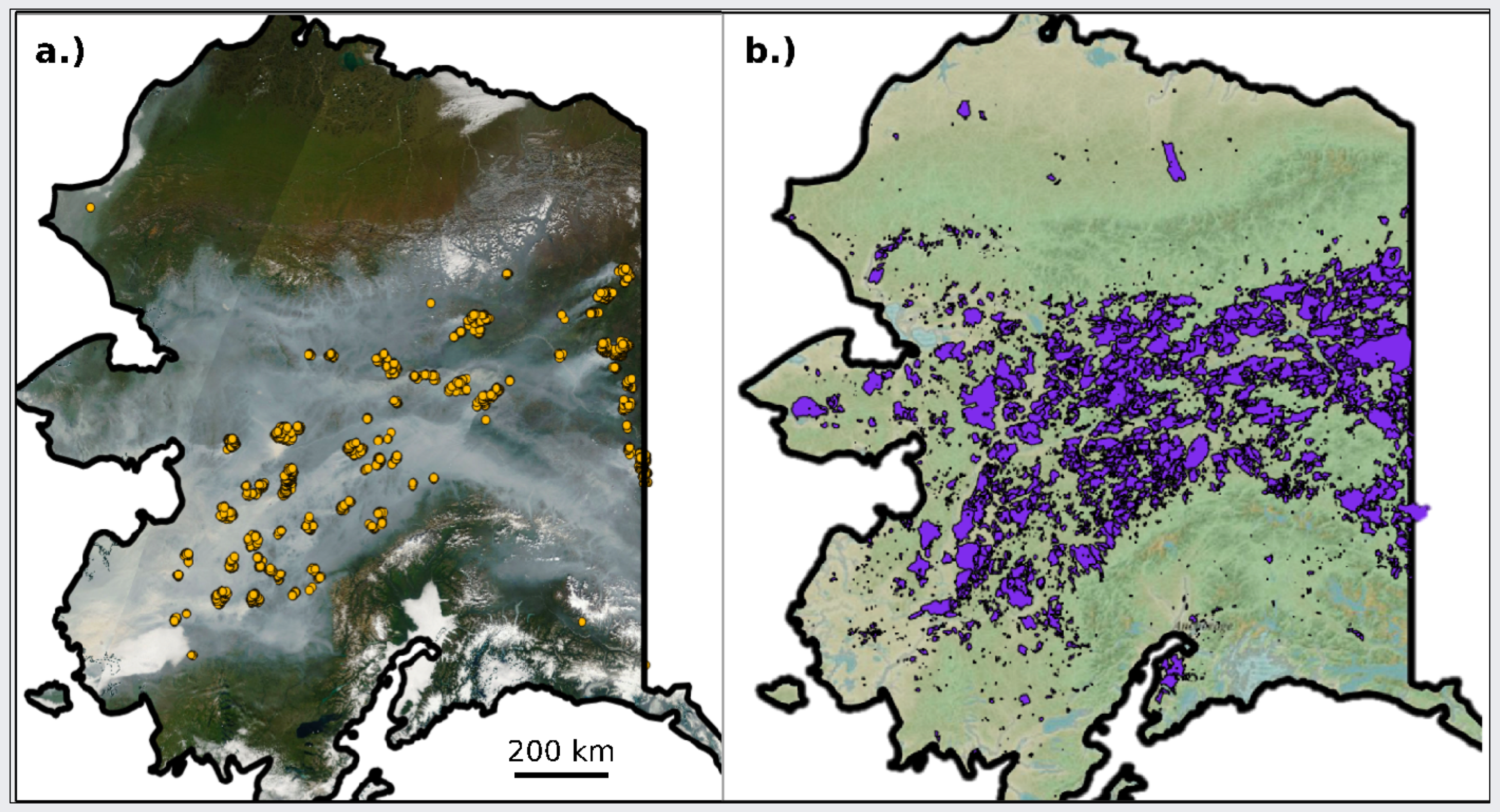
- **Climate change** may require new approaches for **fire management**
- **Decision trees** can be used to classify ignitions as leading to **small, medium, or large fires** with **50% accuracy**
- Decision trees were as accurate as more complex machine learning methods
- Ignitions identified as "large" by our model ultimately accounted for 75% of burned area

Objective - Develop and validate a new framework for wildfire prediction, to triage fires using only information available at the time of ignition

Research questions

1. What environmental variables can explain final fire size from the time of ignition?
2. Which machine learning approaches perform with highest accuracy while maintaining interpretability?

Methods



Study area – Alaska

Wildfires in Alaska on one day (a) and over 17 years (b), exacerbated by climate change

Datasets

Wildfires: Alaska Large Fire Database (2001-2017); only fires in the "limited" management zone which are not actively suppressed by humans; sorted into terciles:

"**small**" < 1.2 km² | "**medium**" 1.2-19.8 km² | "**large**" > 19.8 km²

Weather: ECMWF ERA5 temperature, precip., wind speed, surface pressure, relative humidity, and vapor pressure deficit (derived)

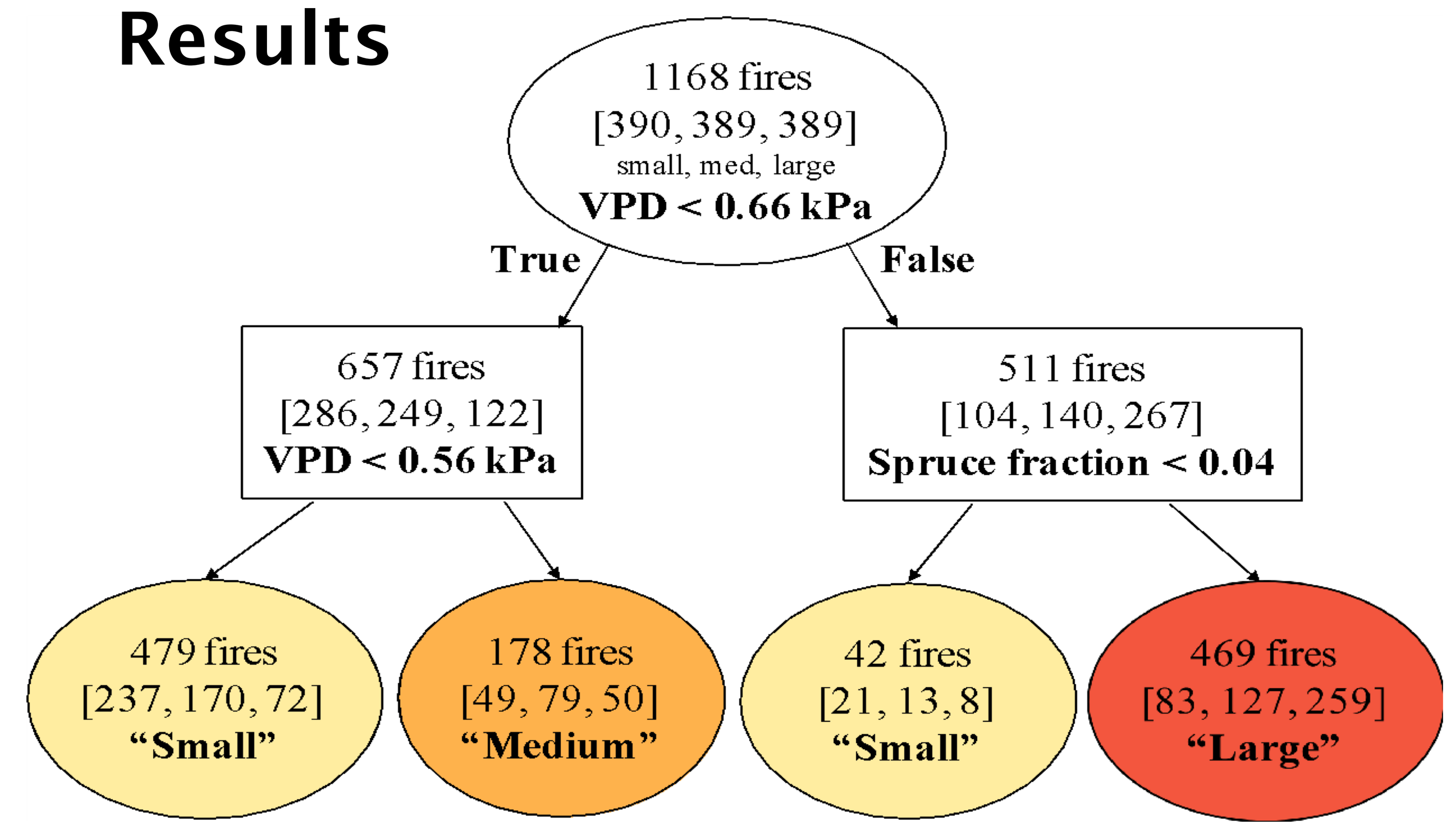
Topography: USGS GTOPO30 global DEM

Vegetation: LANDFIRE Existing Vegetation Type

Modeling approach

- ➢ Tune **time window** to average weather data (figure →)
- ➢ Tune **spatial window** to average vegetation data
- ➢ Tune model **architecture** (number of nodes for decision tree)
- ➢ Compare combinations of **input variables**
- ➢ Compare **algorithms**: decision tree, random forest, k-nearest neighbors, gradient boosting, multi-layer perceptron
- ➢ Select optimal model based on highest **accuracy** in 10-fold cross validation

Results



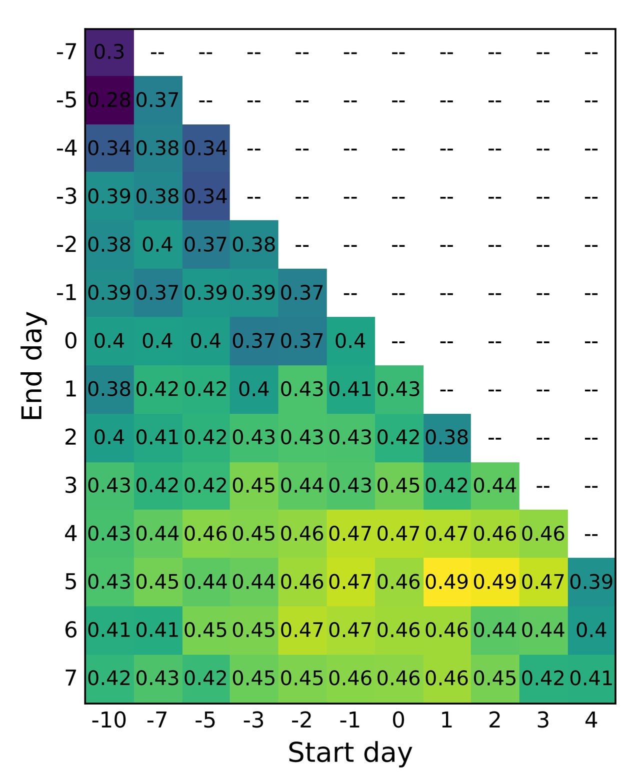
Final decision tree (summary visualization from fitting a single tree to all data)

- Optimal predictive window for weather: 1-5 days after ignition
- Optimal predictive window for vegetation: within 4 km of ignition
- Optimal input variables: vapor pressure deficit (VPD) and fraction of spruce trees
- Classification accuracy (validation): 50.4%
- Decision trees performed similarly to more complex algorithms

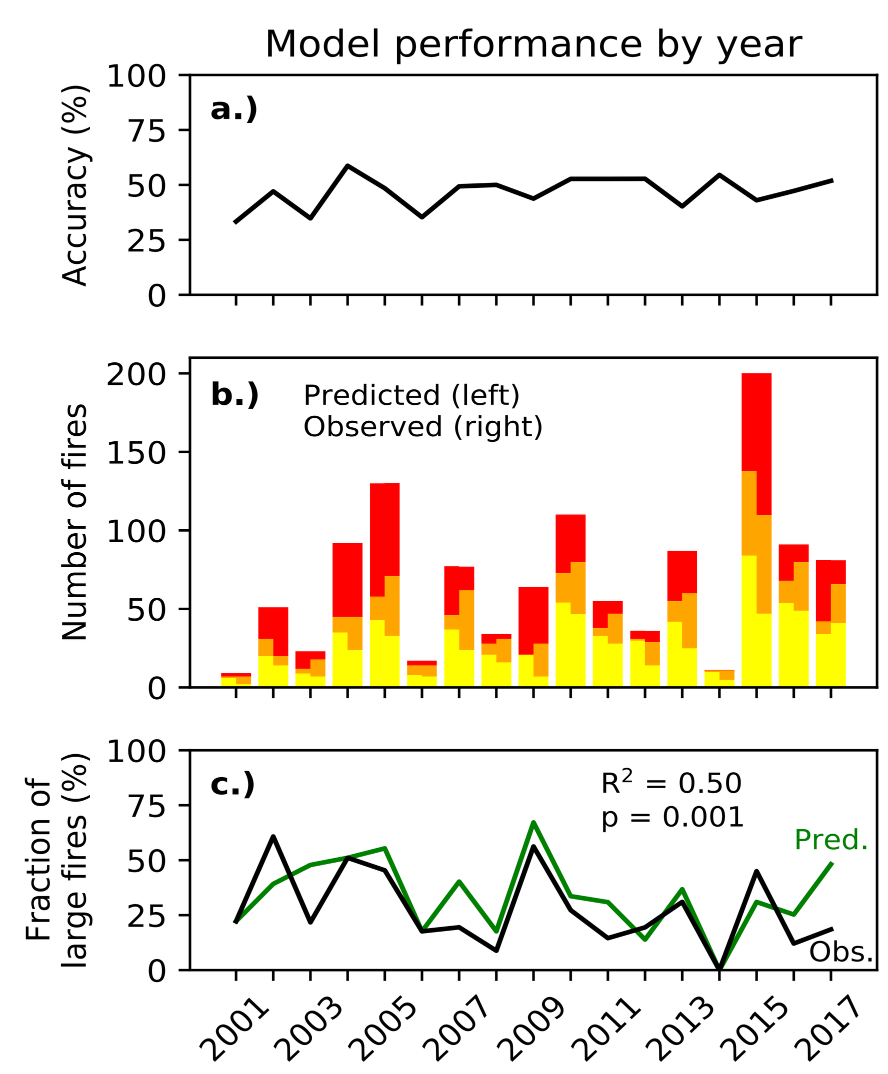
Summary statistics for best decision tree model

- Best performance for largest size class
- 40% of fires predicted as "large" accounted for a disproportionate amount (75%) of total burned area.

Accuracy	50.4 ± 5.2%
Recall for large fires	65.2 ± 8.4%
Precision for large fires	52.5 ± 11.8%
Burned area accounted for by fires classified as large	74.9 ± 12.6%
Improvement in weighted error over a null model	36.3 ± 5.9%



Tuning for optimal time window of weather data

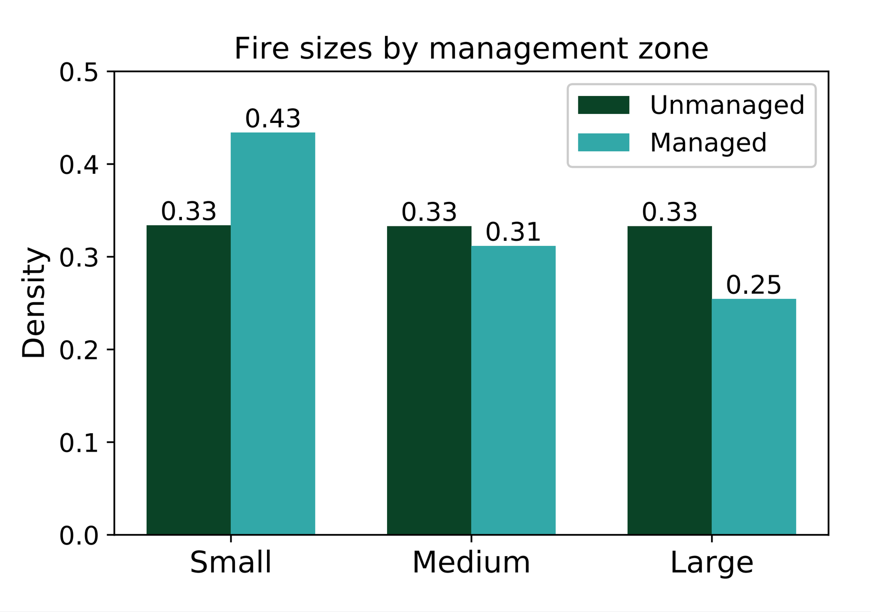


Model performance by year, capturing the interannual variability of fires

AGU100 December 2019

Discussion

Application to areas with active fire management/suppression

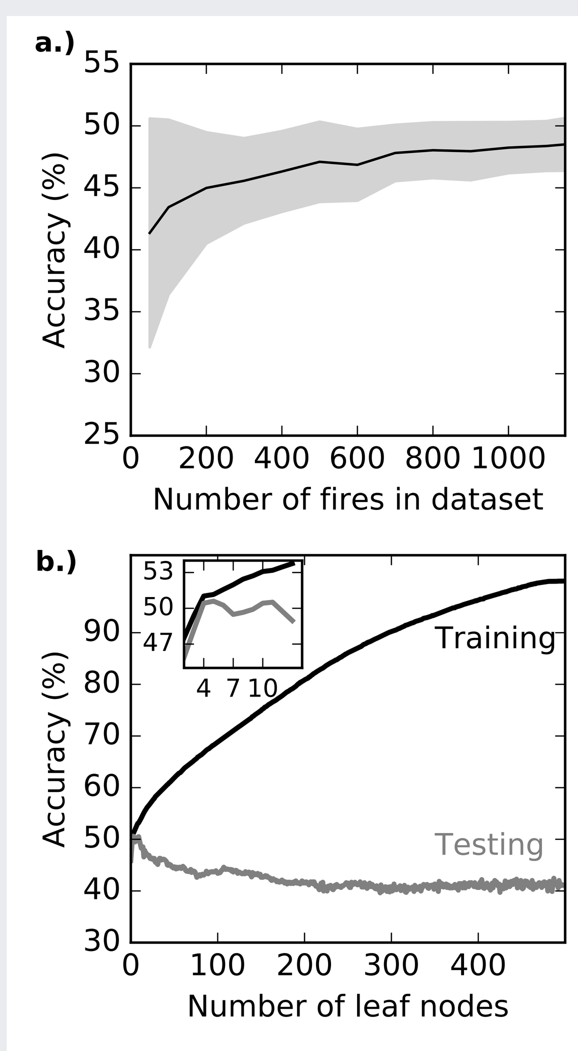


- Fires in more managed zones are smaller but 8% more frequent.
- We estimate that the effect of humans on Alaska's fire regime is to increase total fire frequency by 3.4% but to decrease total burned area by 7.5%.

Summary statistics for model applied to managed zones

- Decrease in total accuracy and precision
- Similar recall for large fires: the model can still "catch" the fires that do become large.
- Disproportionate overprediction of large fires (48% vs. 40%) due to higher VPD during human-ignited fires in populated/managed zones

Accuracy	43.0%
Recall for large fires	64.3%
Precision for large fires	34.0%
Burned area accounted for by fires classified as large	70.6%
Improvement in weighted error over a null model	22.2%



Limiting factors

- Not limited by size of dataset or overfitting →
- Likely limited to 50% accuracy due to incomplete characterization of fuels and loss of information in constructing simple input variables.

Our results show promise for early identification of large fires, and future research should continue applying machine learning with more complex input parameters.

Acknowledgements

NSF Graduate Research Fellowship DGE-1839285; NSF 1633631 UCI NRT Machine Learning and Physical Sciences (MAPS) Training Program; NASA NNX15AQ06A CSULA/UCI DIRECT-STEM project; NSF CNS-1730158; DOE Office of Science RUBISCO Science Focus Area and NASA's SMAP, IDS, and CMS programs; NSF DMS-1839336 TRIPODS program; and NASA NNX16AO56G GPM program.

Link to full paper in *International Journal of Wildland Fire*

DOI 10.1071/WF19023

Contact

shane.coffield@uci.edu