Developing Wildfire Risk and Burn Area Prediction Models Using Comprehensive Environmental Variables in California with Machine Learning

Introduction and Background

- Around 100,000 wildfires occur in the U.S. per year, destroying thousands of acres while costing precious lives, structures, and over 1.5 billion dollars annually.
- California ranks as the state with the most extreme wildfire damages in the U.S.

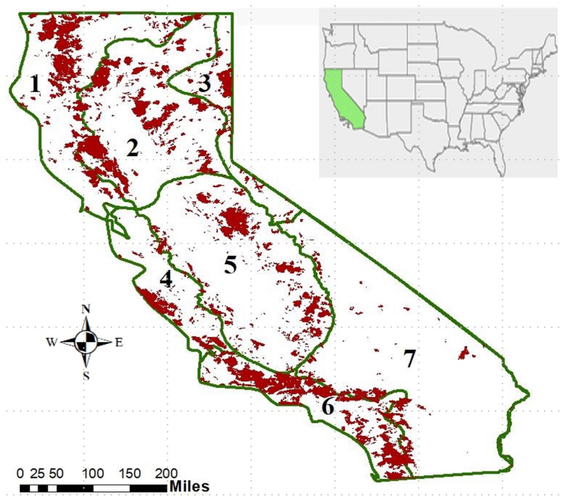

Fig. Areas burned by wildfires in the 7 climatic divisions of California from 1984-2019. (Angela Chen)

Previous Research Issues:

- A limited focus on only one aspect of wildfire prediction: either risk or burn area, but not both.
- Examining relationships between only three to four environmental variables with wildfires over a short time period of only 5-10 years.

Research Objectives

- Analyzing the relationships between wildfires and nine environmental variables over a 36-year study period.
- Creating a robust wildfire risk prediction model that can accurately predict the probability of a fire occurring.
- Creating a robust wildfire burn area prediction model that can predict the number of acres a wildfire is projected to burn.

Methods

Wavelet Analysis

(Torrence and Compo, 1998)

- Continuous wavelet transform (CWT)
- Cross Wavelet transform (XWT)
- Wavelet transform coherence (WTC)

Random Forest

(Breiman et al, 1984; Breiman, 2001)

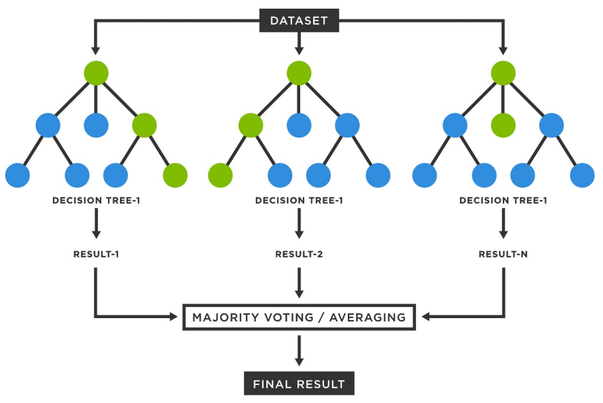

Fig. Visual of the random forest algorithm with decision trees. (tibco.com)

Analyzation of Environmental Variables

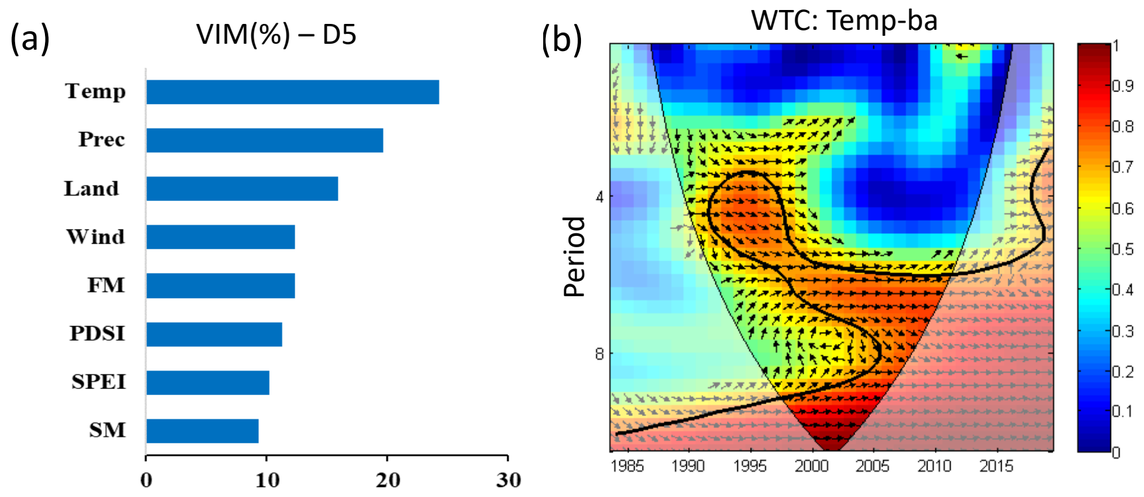

Fig. The (a) Relative Importance of Variables and (b) WTC for temperature for the San Joaquin Division. Air temperature is the most important variable across all divisions. (Angela Chen)

Wildfire Burn Area Prediction Model

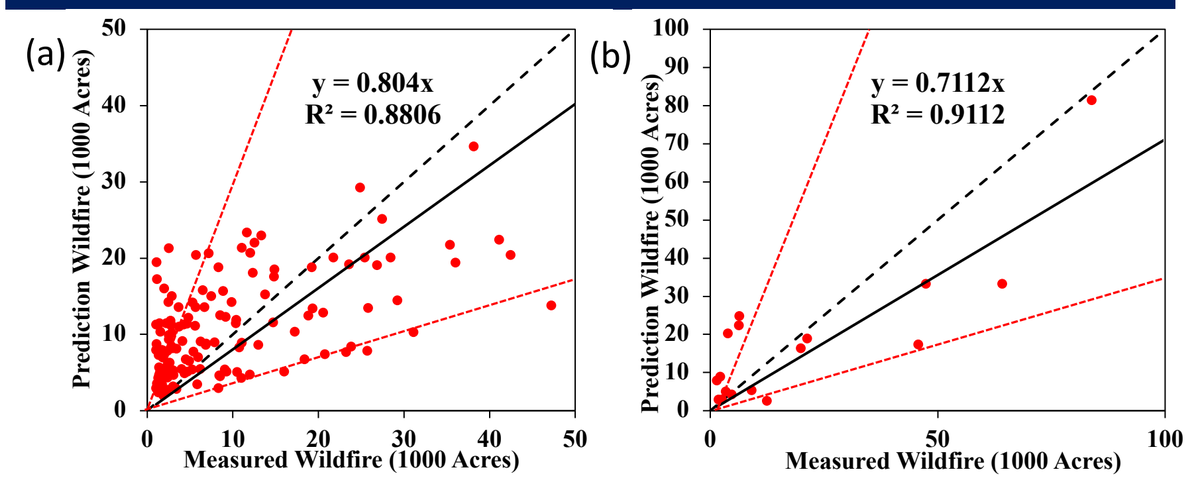

Fig. (a) Random Forest training results (b) validating predicted results with observation in San Joaquin Drainage. The average R^2 across all 7 divisions is 0.91. (Angela Chen)

Data and Variables

Analyzing Relationships between Wildfire Burn Area and:

- Temperature
- Precipitation
- Wind Speed
- Evapotranspiration
- Drought
- Topography
- Land Cover
- Soil Moisture
- Fuel Moisture

Wildfire Risk Prediction Model

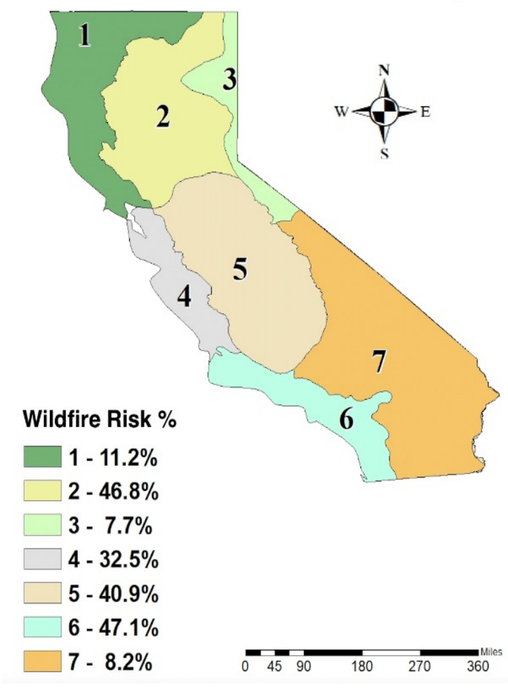

Fig. Wildfire risk probability in the 7 climate divisions of California in 2023 on annual time-scale (Angela Chen)

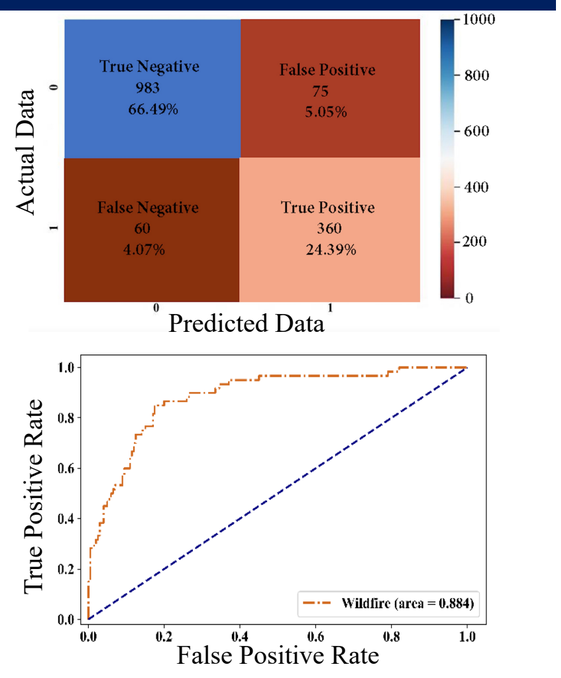

Fig. The generated confusion matrix and ROC curve, showing the F1-accuracy score of the wildfire risk model is 90.88%.

Conclusion

- Analyzed wildfire relationships with nine comprehensive environmental variables, instead of the usual three to four.
- Temperature and precipitation are top two drivers for CA wildfires across all seven divisions.
- Robust prediction models for both wildfire risk and burn area were developed. They can be used in the prevention and alleviation of wildfires, leading to less damages on the environment and human health.

References (Selected)

1.A. R. Keyser and A. L. Westerling. Predicting increasing high severity area burned for three forested regions in the western United States using extreme value theory. Forest Ecology and Management, 432:694–706, Jan. 2019.

2.A. Mull. Wildfires Have Made Smoke Days California's Snow Days - The Atlantic, 2018.

3.S. E. Finlay, A. Moffat, R. Gazzard, D. Baker, and V. Murray. Health Impacts of Wildfires. PLoS Currents, 2012.