Satellite-Based Prediction of Fire Risk in Northern California

Caroline Famiglietti, Natan Holtzman, & Jake Campolo

cfamigli@stanford.edu (CS229)
nholtzma@stanford.edu (CS229)
jcampolo@stanford.edu (CS229A)

Abstract

In recent decades, **climate change** has drastically influenced key characteristics and **patterns of wildfire** across the global land surface[1]. **In California, wildfires are increasing in magnitude, scale, frequency, and duration**, thus heightening risks to human populations and ecosystems[2]. A clearer understanding of the patterns of fire development and spread in California, including the spatial distribution of vulnerability to disturbance, is **vital in adaptively managing land and resources**. To model the development and spread of wildfire in Northern California, we have applied **logistic regression** with forward stepwise selection, **decision trees** with gradient boosting, and a **multilayer perceptron** to a robust dataset of **ecological, hydrological, and meteorological variables derived largely from remote sensing**[3]. While we were able to predict non-fire more accurately than fire, predictions for the latter retained **75 - 80% overall accuracy**. Our work demonstrates the strong potential of using remote sensing assets to preemptively **identify fire risk** and **inform prevention efforts** in the coming decades.

Data & Features

Fire occurrence is partially controlled by the **local vegetation and climate conditions**, *e.g.* an area experiencing dryness/drought is more fire prone. We selected a variety of climate and reflectance variables to characterize **pre-disturbance conditions** impacting the probability of fire.

Climate variables

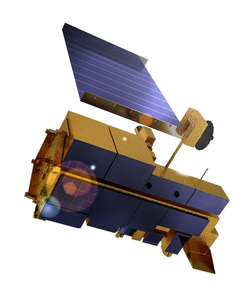
- *Land surface temperature*
- Evapotranspiration (ET)
- Precipitation
- Drought severity index
- Soil moisture (4 depths)
- Wind speed

Reflectance variables

- *Reflectance (R, G, B, near-infrared, shortwave infrared 1 & 2)*
- *Normalized difference vegetation index*
- *Normalized difference moisture index*
- *Normalized difference wetness index*
- *Tasseled cap greenness*
- *Tasseled cap wetness*

Data sources:

- MODIS instrument aboard the Terra satellite *(italicized variables)*
- GLDAS and CHIRPS assimilation datasets (non-italicized variables)



For each variable, we computed at most 4 variations:

- 1-day lagged average
- 1-week lagged average
- 1-month lagged average
- 3-month lagged average

Sampling strategy:

- On each day from 2001 to 2017, we extracted all variables at **all "fire" pixels** and at **an equal number of randomly selected "non-fire" pixels**.
- "Fire" / "non-fire" labels were derived from the **Terra MODIS *presence of fire* product**, which maps fires based on their pre- and post-fire infrared signals ("burn index").

$\Longrightarrow$ **Dimensions of full dataset:** 903,921 (*pixel-days*) × 112 (*features*)

Models

Logistic regression with forward stepwise selection

Let $y \mid x; \theta \sim \text{Bernoulli}(\phi)$. We have

$$\phi = p(y=1|x;\theta) = \frac{1}{1+\exp(-\theta^T x)} = g(\theta^T x)$$

where $g(z) = \frac{1}{1+e^{-z}} \quad \forall\, z \in \mathbb{R}$.

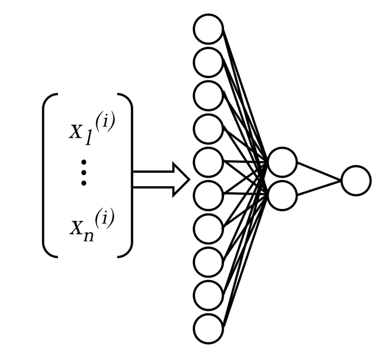
Trees with gradient boosting

1. Set $\hat{f}(x) = 0$ and $r_i = y_i$ for all i in X.
2. For bootstrap samples $b = 1, 2, ..., B$, repeat:
   a) Fit tree $\hat{f}_b$ to the training data (X, r)
   b) Update $\hat{f}(x) = \hat{f}(x) + \lambda \hat{f}_b(x)$
   c) Update $r_i = r_i - \lambda \hat{f}_b(x_i)$
3. Return $\hat{f}(x)$

Multi-layer perceptron (MLP)

Activation: ReLU
f(x) = max(0, x)

Optimization: Stochastic gradient descent

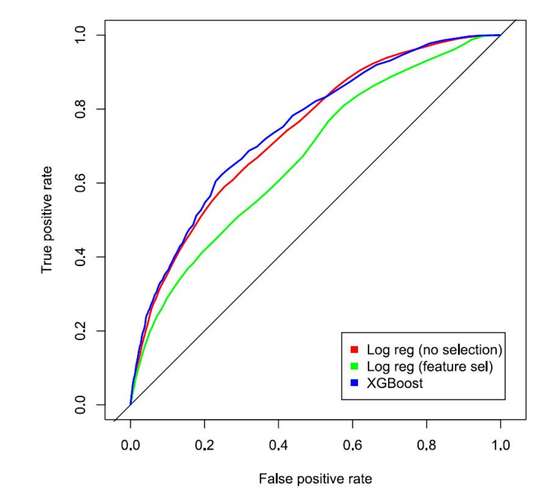


Results

Model	Train Error (%)	Test Error (%)
Logistic regression *(C)*	20.62	22.71
Logistic regression *(R)*	20.93	23.24
Logistic regression *(C,R)*	21.33	22.71
Boosted trees *(C)*	21.07	23.11
Boosted trees *(R)*	19.57	22.62
Boosted trees *(C,R)*	20.38	22.37
MLP *(C)*	25.20	24.69
MLP *(R)*	20.51	22.72
MLP *(C,R)*	18.97	22.21

✳ (C): climate variables only; (R): reflectance variables only; (C,R): climate and reflectance variables. m_{train} = 499,475 data points from 2001–2016; m_{test} = 37,687 data points from 2017.

ROC curves



*✳ Forward selection of **4 features** reduces AUC on test set from 0.74 to 0.68.*

Confusion matrix

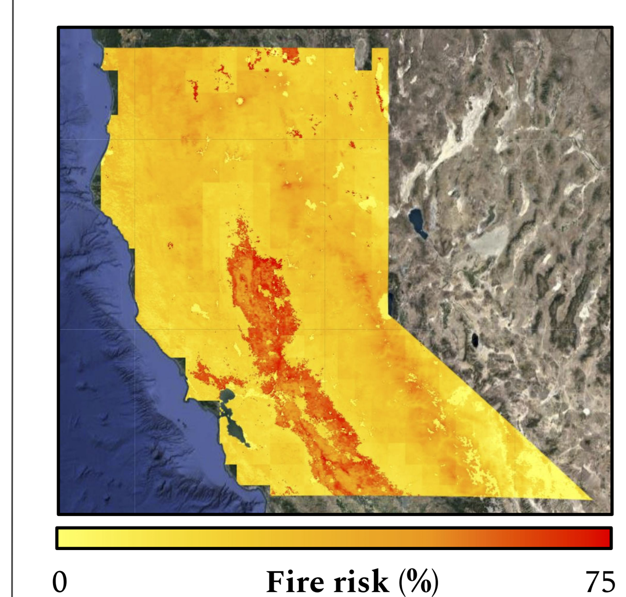
	(Predicted) No Fire	*(Predicted)* Fire
(Actual) No Fire	26,453	1,930
(Actual) Fire	6,502	2,802

Fire:
Precision: 0.592
Recall: 0.301
F1: 0.399

No Fire:
Precision: 0.803
Recall: 0.932
F1: 0.863

*✳ For conciseness, only statistics for **boosted trees** are shown above. Results for the other models are similar.*

Fire risk map
(Average for June–August 2017)



*✳ Risk derived using coefficients from **logistic regression (C, R)**.*

Discussion

Model hyperparameter choices

We used **cross-validation techniques** to determine "optimal":

- Logistic regression feature selection
- Number of trees in ensemble
- Neural network architecture

Feature importances

- Variations in error between *(C)*, *(R)*, and *(C,R)* subsets were **minimal**, suggesting our dataset had **more variables than were needed.**
- *Logistic regression:* LC, GCVI, SWIR, ET
- *Boosted trees:* LC, precipitation, LST, WS

Prediction accuracy

Unmeasured causes of fire ignition are **potential sources of error** in our models.

- *Anthropogenic*: crop residue burning, smoking, fireworks, electrical failure, campfires, arson
- *Natural*: lightning

References

[1]Pechony O, Shindell DT (2010). Driving forces of global wildfires over the past millennium and the forthcoming century. *PNAS*, http://www.pnas.org/content/107/45/19167.

[2]Westerling AL, Hidalgo HG, Cayan DR, Swetnam TW (2006). Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity. *Science*, http://science.sciencemag.org/content/313/5789/940.full.

[3]Hawbaker TJ, Vanderhoof MK, Beal Y-J, Takacs JD, Schmidt GL, Falgout JT, Williams B, Fairaux NM, Caldwell MK, Picotte JJ, Howard SM, Stitt S, Dwyer JL (2017). Mapping burned areas using dense time-series of Landsat data. *Remote Sensing of Environment*, https://doi.org/10.1016/j.rse.2017.06.027.

Future

1. Can we use **projections of future climate** to **infer expected wildfire dynamics** in the region in the coming decades? $\Longrightarrow$ *Using climate model outputs such as **RCP 4.5 / 8.5**, which forecast climate change from now to 2100, we could apply our models to these predicted data to assess **future changes in fire risk** over the next century.*

2. Can we predict **post-disturbance effects** on impacted areas? (*e.g.* economic impacts/costs of damages) $\Longrightarrow$ *We could compile **spatially-explicit population, infrastructure, and economic data** in order to extrapolate potential costs and damages from fires predicted by our model, which would be useful for policymakers in conducting cost benefit analysis of fire prevention.*