Spread of Wildfire Pollutants in California

Courtney Moran, Akwasi Owusu-Akyaw, Richard Li CS 229 Fall 2019

Stanford University

Abstract

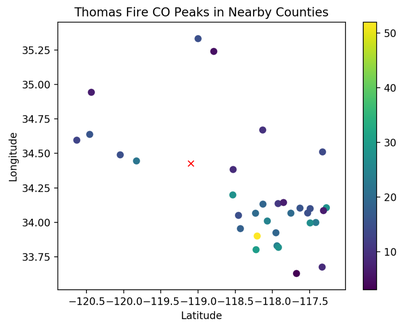
Our goal for this project was to create a tool which cautions civilians of poor air quality in the event of a wildfire. To accomplish this, we created a model that predicts the CO concentration of nearby California counties for the duration of a large wildfire. To train our model, we utilized previous Californian large scaled wildfires and utilized the following input features:

1. Fire size (Acres)
2. Direction relative to fire (degrees)
3. Distance from fire (meters)

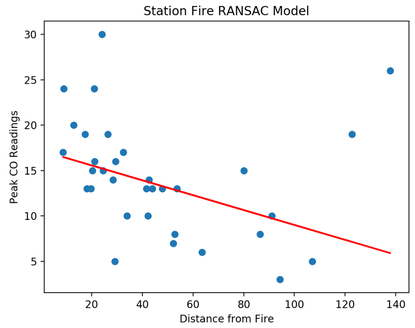
Using higher order polynomial regressions, we tested our model using a wildfire excluded from our training set, obtaining NMSE's of < 0.3 for 65 % of the fires we tested.

Initial Findings

We acquired our CO2 concentration data from the EPA website, which details the daily CO2 concentration. In addition, we obtained the center location of past wildfires in California by looking into wildfire history. From this, we created our initial plots below:



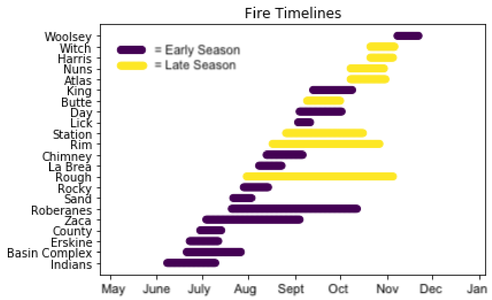
Initial Model

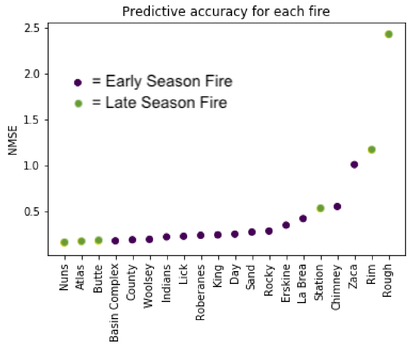


We started with a simple linear regression model and a RANSAC model using only one feature, distance from the fire.

We trained our linear model on 24 of the major wildfires in California between 2006 and 2018. From there, we tested our model on one more fire that we excluded from our training set. Our RANSAC model had a Means Square error (MSE) of 38.14 compared to 48.22 for the linear model.

Current Model





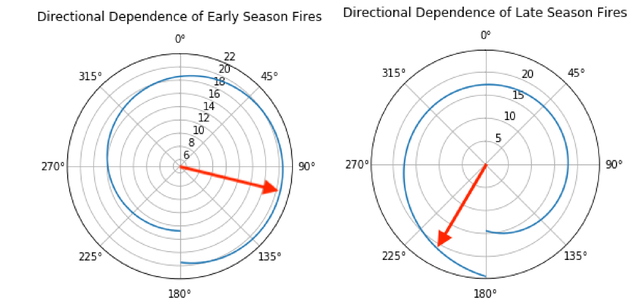
The 24 fires were categorized into early or late season fires, to capture the characteristic wind patterns at each time of year. The model was trained separately for the two groups, which improved model accuracy compared to no segmentation.

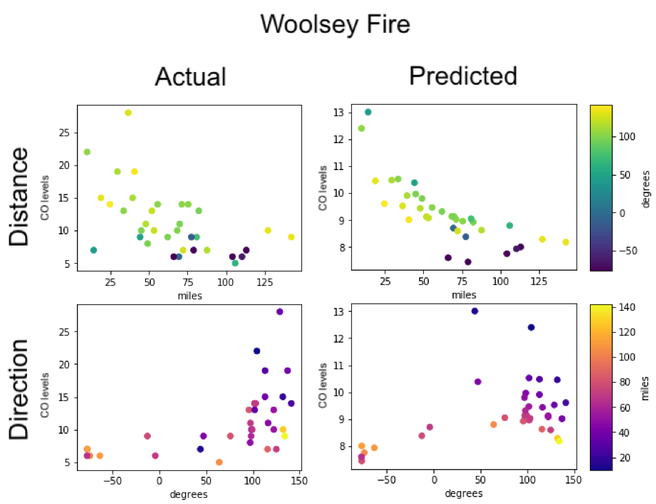
Sensor direction relative to the fire was added as a feature to account for the seasonality of wind patterns.

We expanded the linear model to include 3rd order polynomial features, which we observed describes our training data well without overfitting or sacrificing generalizability on the test set.

The predictive accuracy of the model was <0.3 NMSE for 65% of the fires. Late season fires had disproportionately higher error, likely due to the limited size of the dataset.

Results





The directional dependence of early season vs late season fires is consistent with characteristic wind patterns in California.

For most of the year, coastal air from the ocean blows inland (towards the East). From October-November, however, warm winds towards the Southwest (Diablo winds) can dominate.

The CO intensity falls off with distance, which is modeled with an R^2-like relationship. The purple dots, which represent sensors west of the fire, have disproportionately low readings, as expected.

Average CO levels increase with fire size (not shown) with a nearly linear relationship.

References

EPA, "AirData website File Download page," EPA, 13-2019. [Online]. Available: https://aqs.epa.gov/aqsweb/airdata/

Future Research

As another option, we could have explored using decision tree regressions. We believe that this option could have helped us section the data by seasons, and create a more accurate piece-wise model to describe CO concentration.

Contact Information

moranc@stanford.edu
akwasio@stanford.edu
rli2014@stanford.edu