Are California Wildfires Getting more Intense in the North or South?

Felicia Cruz

2021-12-2

Introduction

Growing up in the Antelope Valley, located in the Mojave Desert at the northern edge of LA County, I experienced wildfires at a very young age. As I grew up, orange haze outside, flying ash, and evacuation announcements on the news were not out of the ordinary during the later summer months. I will never forget taking a day trip to the beach with my family one summer only to come home to barricades outside our neighborhood because a wildfire had jumped the nearby ridge deeming it unsafe to enter. While our house, and my dog inside, was thankfully intact, we were reminded of the event for months by the charred hill next to our house.

The statistics describing the frequency, intensity, and destructiveness of California wildfires are alarming to say the least, and are all likely to be further exacerbated by the climate crisis in the years to come. As temperature and droughts increase with elevated greenhouse gas emissions, it is probable that we can expect more wildfires in the future, especially with the fire seasons getting longer.

In the past couple decades, California's wildfire season has progressively gotten worse. A recent study by researchers at UC Irvine found that the annual burn season has lengthened since 2000 and that the yearly peak has shifted from August to July. In recent years especially, fires are getting both more intense and more frequent. For example, the 2017 season was the most destructive wildfire season on record at the time, with 1.2 million acres burned. The largest fire during that season was the Thomas Fire in Santa Barbara County, which was California's largest modern wildfire at the time. In 2020, 4.2 million acres were burned, amounting to more than 4% of the state's total land. This resulted in 2020 being the largest wildfire season recorded in California's modern history. The August 2020 Complex Fire alone burned more than 1 million acres, making it the first "gigafire" on record.
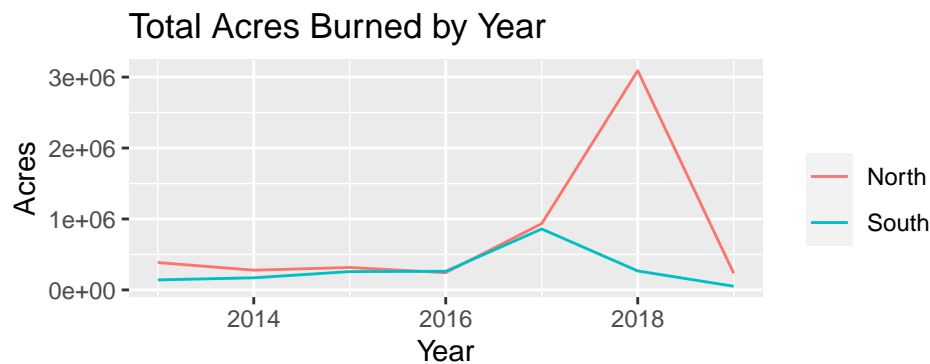

Figure 1: Total Acres Burned (2013-2019). Acres burned in the North spiked in 2018, followed by a large drop-off in 2019.

I am interested in how the intensity of California wildfires has changed over time, and if this change is different in the north compared to the south. Are California wildfires getting more intense in the northern part of the state or in the south? Is there a difference in trends between start date and acres burned for fires in the north versus fires in the South?

From my preliminary research, there does not seem to be existing evidence on this question. While many articles discuss trends in wildfire season as a whole and factors that contribute to the growing wildfires in California, there is not much to be said about how geographic location is potentially correlated with changes in wildfires over time.

Data Description

To explore changes in California's wildfire season and compare trends in the North to the South, I will be using a dataset which contains over 1,600 wildfire events in the state between 2013 and 2020. This dataset is made available on Kaggle.com and was originally scraped from records on the CAL FIRE website. For each fire, this dataset contains 40 variables; for the purposes of this analysis, the most relevant variables include latitude and longitude, start date, and acres burned. A potential limitation of this dataset is that it only includes wildfires responded to by CAL FIRE; many wildfire events that contribute to the overall trends in the north and the south may not be included in this dataset.

The dataset can be found here: https://www.kaggle.com/ananthu017/california-wildfire-incidents-20132020

Analysis Plan

Before looking at differences in geographic trends, I will first explore how start date affects wildfire intensity by running the following regression:

$$acres\_burned_i = \beta_0 + \beta_1 \cdot started_i + \varepsilon_i$$

Next, in order to assess the effects of wildfire start date and location on the number of acres burned, I will be using the following interaction model:

$$acres\_burned_i = \beta_0 + \beta_1 \cdot started_i + \beta_2 \cdot north\_south_i + \beta_3 \cdot started_i \cdot north\_south_i + \varepsilon_i$$

The variable `north_south` takes on a value of "North" for all wildfires occurring above 36 degrees latitude, and everything below this is assigned "South". I have chosen this interaction model in order to see if wildfire trends differ in the north and south.

After running this regression, I will determine if the slope coefficients are statistically significant by looking at the associated p-values.

Lastly, due to many extreme values resulting from record-breaking fires in recent years, I will winsorize my data at the 99th percentile to try to eliminate some noise and see if my regression results change.

Results

Linear regression (acres_burned~started)

From the first linear regression exploring the effects of start date on wildfire intensity, we can see that for every one day increase in the start date, there is a 1.08 increase in the number of acres burned. While this shows a positive trend, these results are not significant, as the p-values for both the intercept and the slope are quite high.

Table 1: Linear Regression of Acres Burned on Start Date

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-13959.94	17060.52	-0.82	0.41
started	1.08	0.99	1.09	0.28

Interaction model

After including an interaction term for the California region a wildfire occurred in, we see a difference in slopes for the two regions. From Table 2, we can see that the intercept for fires that occur in the South is 24,345 acres above the intercept for fires in the South. Despite the higher intercept for fires in the South, the slope of the regression line is 1.56 less than that of the regression line for fires in the North. In other words, the trend line for northern fires is steeper despite a lower intercept. Just like our first regression results, the p-values are very high. While this model does show a difference in slopes, these results are not statistically significant and, therefore, there is not enough evidence to reject the hypothesis that there is a difference in the trends of wildfire intensity over time due to geographic location.

Table 2: Multiple Linear Regression

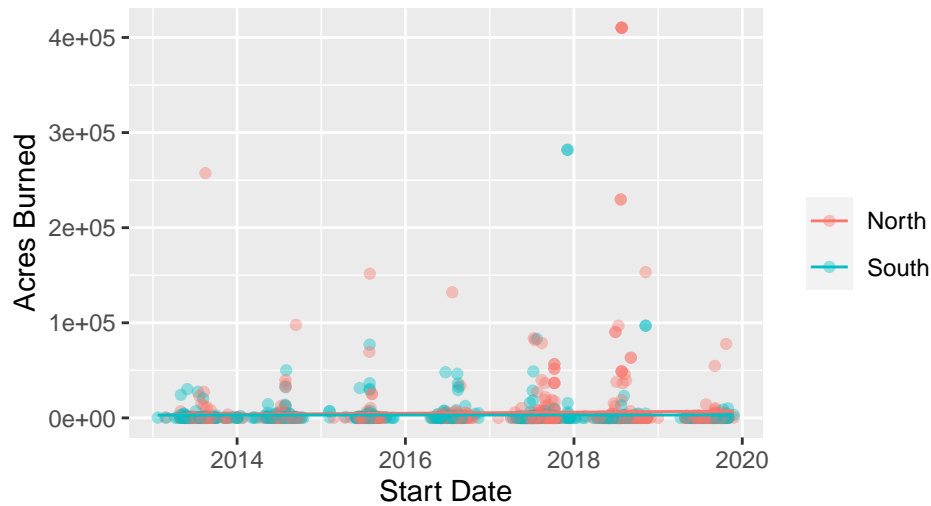|	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-21405.07	22705.70	-0.94	0.35
started	1.57	1.31	1.19	0.23
north_southSouth	24344.65	34590.61	0.70	0.48
started:north_southSouth	-1.56	2.01	-0.78	0.44

Figure 2: Acres Burned vs. Start Date and Region

Winsorization at the 99th percentile

In an attempt to eliminate some of the noise caused by this wildfire event data, I decided to winsorize my `acres_burned` values at the 99th percentile (90,288 acres).

3

Comparing these results to the previous regression before winsorization, β_1, the slope of the line for fires occurring in northern California, is smaller. Additionally, the slope for the South regression line after winsorization is -0.45, compared to a barely positive slope previously.

The p-value for the coefficient on the interaction term decreases after winsorization from 0.44 to 0.31, but it is still far too high to reject the hypothesis that there are different trends in California wildfire intensity based on region.

Table 3: Multiple linear regression after winsorization at the 99th percentile

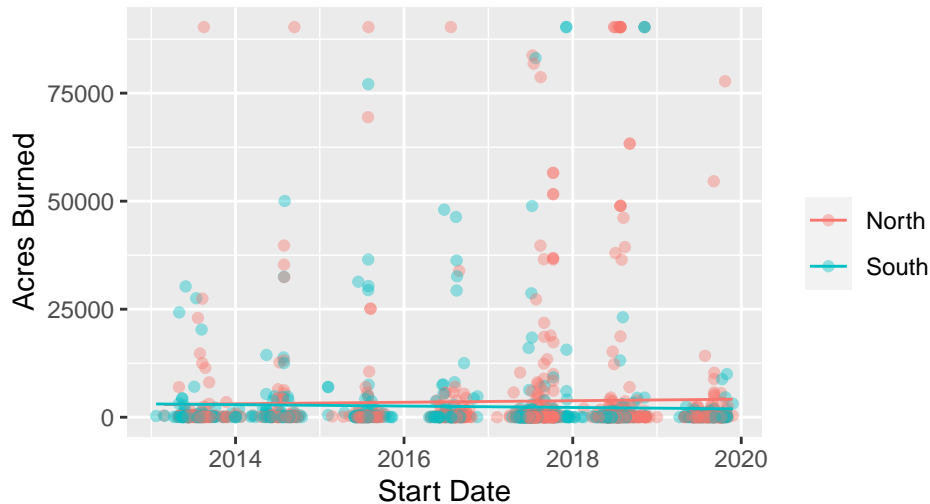|	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	-4783.42	10434.59	-0.46	0.65
started	0.49	0.60	0.81	0.42
north_southSouth	14905.85	15896.39	0.94	0.35
started:north_southSouth	-0.94	0.92	-1.02	0.31

Figure 3: Acres Burned vs. Start Date and Region after Winsorization

Lastly, after winsorizing `acres_burned` at the 99th percentile, I decided to drop the observations above the 99th percentile entirely as a sensitivity check for my model. After re-running my interaction model, based on the results in Table 4 it does seem that my original model was sensitive to the outliers. Now, the p-value for the interaction term is 0.03 which is so much lower than the original results and the winsorized results. While this would suggest significance at the 5% level, I think more analysis would need to be done to see if dropping outliers is the appropriate route to take.

Table 4: Multiple linear regression after removing observations above the 99th percentile

	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	-4174.55	7611.38	-0.55	0.58
started	0.39	0.44	0.89	0.37
north_southSouth	23747.00	11568.90	2.05	0.04
started:north_southSouth	-1.42	0.67	-2.12	0.03

Further Analysis

Overall, more data would be needed to perform a more robust and comprehensive analysis to accurately assess how regional location affects the trends in wildfire intensity in California over time. This dataset is vry limited in that it only includes wildfire events during a seven year span from 2013-2019. To get a more representative picture of the trends in acres burned for northern and southern wildfires in California over time, a much larger temporal range is needed. Additionally, 2020 data was not included which means observations from the largest fire season on record are not taken into account here. With more data points spanning a 20-30 year period, I suspect much more interesting and informative results.

Beyond doing a regression analysis to explore differences in geographic trends in wildfire intensity over time, doing time series analyses to produce predictions of wildfire intensity would also be useful. This type of analysis would be of interest to government organizations, first responders in California, and the general public living in fire-prone areas of the state.

Lastly, looking at seasonal trends over time can help identify if and how the wildfire season itself is changing. This type of analysis could help answer other questions such as: Is the wildfire season in California getting longer? Are more intense wildfires happening at a certain point in the season?

Link to my GitHub repository: https://github.com/fmcruz23/fmcruz23.github.io/tree/main/_posts/2021-11-29-eds-222-final-blog

References

California Department of Forestry and Fire Protection (CAL FIRE). "2018 Incident Archive." Cal Fire Department of Forestry and Fire Protection, https://www.fire.ca.gov/incidents/2018/.

California Department of Forestry and Fire Protection (CAL FIRE). "2020 Incident Archive." Cal Fire Department of Forestry and Fire Protection, https://www.fire.ca.gov/incidents/2020/.

"California's Wildfire Season Has Lengthened, and Its Peak Is Now Earlier in the Year." (GSTDTAP): California's Wildfire Season Has Lengthened, and Its Peak Is Now Earlier in the Year, 22 Apr. 2021, http://resp.llas.ac.cn/C666/handle/2XK7JSWQ/323227.

"One Reason for Northern California's Terrible Fire Season: Less Rain than Southern California." Los Angeles Times, Los Angeles Times, 1 Oct. 2020, https://www.latimes.com/california/story/2020-10-01/northern-california-fire-season-less-rain-than-southern-california.

Sadegh, Mojtaba, et al. "The Year the West Was Burning: How the 2020 Wildfire Season Got so Extreme." ScholarWorks, https://scholarworks.boisestate.edu/civileng_facpubs/149/.