Analyzing Wildfire Prone Areas in California: A Multifactorial Approach - CS249 Winter 2024

Anthony Holmes anthonyjholmes1@g.ucla.edu
Department of Computer Science, UCLA, Los Angeles, CA, 90024.

Abstract

This study explores the interrelation between weather conditions and wildfire occurrences in California through comprehensive data analysis. By utilizing records from the National Interagency Fire Center (NIFC) and daily weather reports from NOAA's National Centers for Environmental Information (NCEI), the research aims to identify factors contributing to the susceptibility of certain areas to wildfires. Employing a range of analytical methods, including linear regression, T-tests, correlation analysis, ANOVA, and Chi-square tests, the study examines the impact of temperature, humidity, and geographical factors on wildfire characteristics.

Key findings reveal that while weather station data near wildfire perimeters may not reliably indicate temperature due to potential skewing, a general trend of rising temperatures and decreasing rainfall preceding the wildfire season is evident, contributing to increased wildfire risk. The study also highlights the absence of statistically significant differences in relative humidity between dry and wet periods and the lack of significant impact of weather station proximity on wildfire characteristics. Additionally, no significant differences were found in wildfire sizes across different causes, and no significant association was observed between fire cause and management complexity.

These results underscore the complex dynamics governing wildfire behavior and emphasize the importance of comprehensive strategies for wildfire prediction, management, and mitigation that consider a broad spectrum of environmental, temporal, and geographical factors. For reference, code is available at : `https://github.com/anthonyjholmes1/com-sci-249-Winter24`

1 Introduction

California's diverse landscapes, ranging from dense forests to sprawling urban areas, have been increasingly threatened by wildfires. These devastating events not only pose a risk to human life and property but also have profound ecological impacts, making the study of their causative factors an urgent concern. The susceptibility of areas to wildfires is influenced by a complex interplay of factors, including but not limited to vegetation types, weather patterns, agricultural practices, and economic conditions. Gaining a deeper understanding of how these elements interact is key to coming up with plans to reduce the risk of wildfires and safeguard communities at risk. This project aims to analyze the factors that make certain areas in California more prone to wildfires than others.

1.1 Motivation

My idea behind this study originates from my own experiences growing up in Ireland, a country that has year-round mild climate, where extreme weather conditions are a rarity. The winter of 2009 to early 2010, however, differed from this norm; Ireland was gripped by one of its most severe cold spells in history. Ireland is not equipped to handle sub-zero temperatures. My family and I faced

Preprint. Under review.

considerable hardship when the plummeting temperatures transformed roads into treacherous ice paths, and we lived on top of a steep hill which effectively isolated us. The freezing conditions led to our water pipes bursting and the central heating system failing, leaving us without water and reliant on wood logs for warmth. We walked nearly 2 miles every day to gather basic goods to survive this spell. This period of unforeseen hardship, lasting two weeks, deeply impacted me and ignited a commitment to contributing positively to global society.

This commitment strengthened when I moved to California in 2018, where I was captivated by the United States' breathtaking natural landscapes, from its vast national parks and diverse wildlife to its picturesque coastlines. The destruction brought by the August Complex wildfire in 2020, with its toll on life and nature, was a turning point for me. It prompted the urgency of taking action to preserve our environment.

In my professional role today, where I focus on identifying and rectifying inefficiencies in manufacturing processes, I have already begun to contribute to environmental sustainability. These efforts have led to reductions in carbon emissions and minimized raw material usage. Pursuing a graduate degree represents a continuation of this journey, driven by my belief in the power of data to effect change. With the data seemingly within reach, my goal is to deepen my understanding of how to leverage this information effectively.

This project marks a significant step towards realizing my long-term ambition of dedicating my career to the environmental sector. It is an opportunity to apply the education that I have received so far at UCLA and to embark on a path that I hope will allow me to make a meaningful and lasting impact on our world.

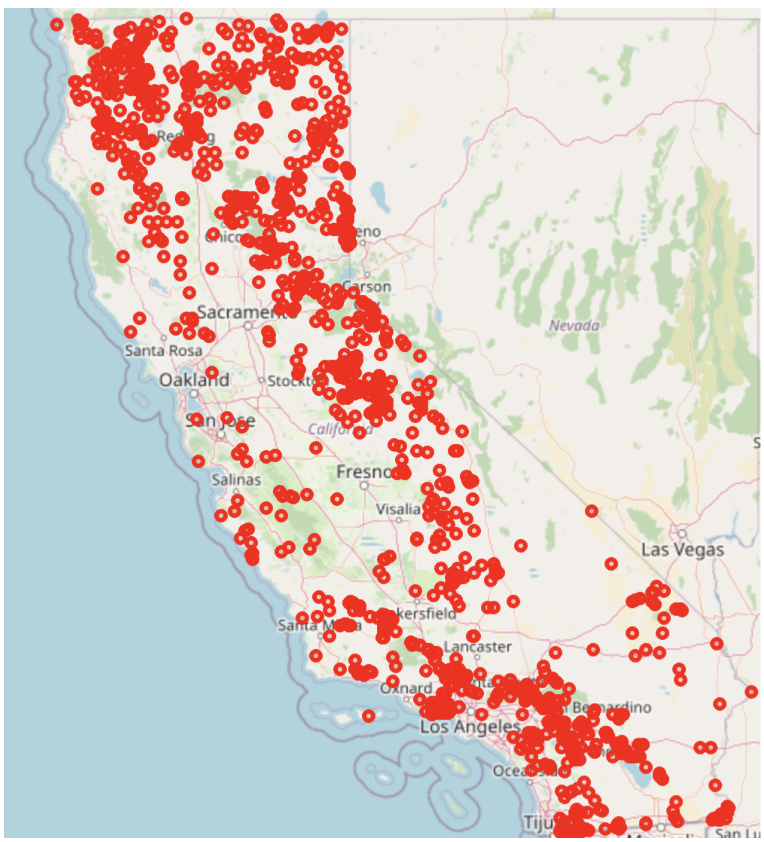

Figure 1: Image of the distribution for wildfires in California between 2020-2024

1.2 Research

The study from Keeley and Syphard [2021] posits that the wildfire scenario of 2020, marked by an array of large-scale fires, is not an anomaly but part of a longstanding historical trend in California.

Although there's a noticeable rise in the frequency and intensity of wildfires in recent times, instances of similar scope have been documented in the past. The research underscores the critical role of historical wildfire data in comprehending and preparing for the contemporary and future realities of wildfire management, especially against the backdrop of evolving climate conditions and expanding human development in susceptible regions.

Wang [2020] underscore the extensive, layered economic impacts of wildfires, with a significant portion stemming from indirect and remote economic disruptions. This comprehensive assessment serves as a clarion call for policymakers and stakeholders to fortify wildfire prevention, preparedness, and economic resilience strategies, especially as wildfire risks loom larger in the shadow of climate change, economic expansion, and population growth in susceptible areas. This study not only enriches the discourse on the economic dimensions of wildfires but also pioneers a methodological blueprint for evaluating the full economic spectra of natural disasters.

2 Data

In exploring the relationship between wildfire occurrences and weather patterns in California, I leveraged two primary datasets: wildfire records from the National Interagency Fire Center (NIFC) and daily weather reports from NOAA's National Centers for Environmental Information (NCEI).

2.1 Wildfire Data

The wildfire dataset, obtained from NIFC's website, provides comprehensive records of wildfire incidents across the United States from 2020 onwards. While earlier records were available, they were hosted on a different platform, and I focused on the more detailed and complete dataset from 2020 to present. This dataset includes information on the location, size, cause, and duration of each wildfire, among other attributes.

2.2 Weather Data

To complement the wildfire data, I utilized daily weather reports from NOAA, available at NCEI's archive. This dataset encompasses weather observations from stations worldwide, including a wide array of weather elements such as temperature, precipitation, wind speed, and humidity. Additionally, geographical information for each station was incorporated to accurately match weather observations to the corresponding locations in California. This was facilitated by a California geographic boundary shapefile sourced from California's open data portalCA.gov, allowing for the exclusion of stations outside California's borders.

2.3 Data Transformation and Analysis

A key step in my analysis was determining the proximity of weather stations to wildfire locations, for which I employed the haversine formula to calculate distances. This analysis unveiled anomalies, particularly in temperature readings near wildfires, with some gauges indicating extreme temperatures exceeding 350 degrees Fahrenheit. Such findings hinted at the direct impact of wildfires on local weather station readings.

Given the voluminous nature of the weather data, with millions of data points annually, I conducted a selective filtering process to streamline the dataset for relevance to wildfires. Initial filters excluded elements like snow, under the premise that they would not significantly influence wildfire dynamics. However, this decision might have overlooked the indirect effects of snowmelt on soil moisture and wildfire susceptibility. To manage the complexity of the weather data, I transformed it into a more accessible format using pivot tables, enabling a day-by-day comparison of weather conditions across different stations.

2.4 Adjustments and Considerations

Throughout this process, data cleaning was essential to ensure the accuracy and relevance of the analysis. This included removing outliers and addressing missing or erroneous data points, such as

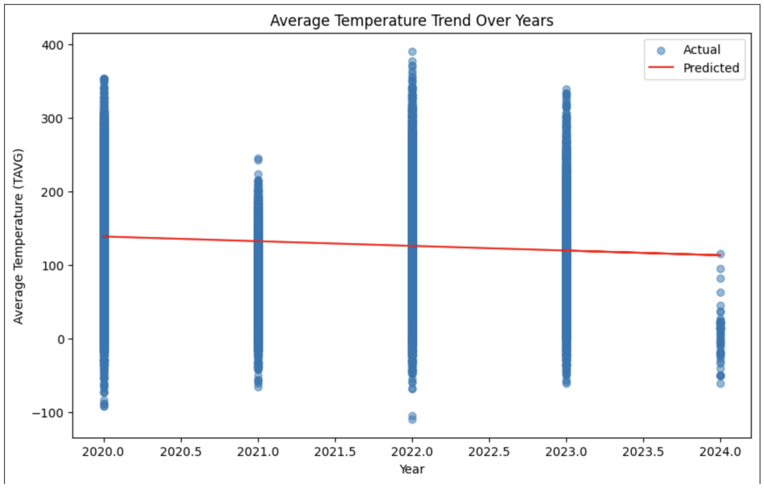

Figure 2: Image of the average distribution of temperature for wildfires in California between 2020-2024

the aforementioned extreme temperature readings. By refining the data, I aimed to derive meaningful insights into how weather patterns influence wildfire behavior and risk in California.

The exploratory analysis revealed critical insights into the interaction between weather conditions and wildfire occurrences. It underscored the importance of considering a broad range of weather elements in wildfire research and highlighted potential areas for further investigation, such as the role of soil moisture and the indirect effects of various weather phenomena on wildfire dynamics.

In summary, this analysis represents a comprehensive effort to understand the multifaceted relationship between wildfires and weather in California, leveraging detailed records from NIFC and NOAA. Through meticulous data transformation and cleaning, I aimed to uncover patterns that could inform wildfire prediction, management, and mitigation strategies.

3 Methodology

The choice of methods was guided by the hypothesis that certain weather conditions could significantly influence the occurrence, duration, and severity of wildfires. Below is an overview of the techniques used and the rationale behind them.

3.1 Linear Regression Analysis

I employed a linear regression model to examine how average temperature affects wildfire characteristics over the years. This method was selected because it allows for quantifying the relationship between a continuous predictor (average temperature) and continuous outcome variables (e.g., wildfire size, duration). Linear regression is appropriate for this analysis due to its ability to handle continuous data and provide insights into trends over time.

3.2 T-test on relative humidity

I conducted a T-test to compare the relative humidity levels during dry and wet months of the year. The T-test is suitable here as it tests the hypothesis that there are significant differences in mean values between two groups (dry vs. wet months) under the assumption that the data follows a normal distribution. This comparison helps to understand how seasonal changes in humidity could impact wildfire risk.

3.3 Wildfire frequency distribution

To visualize the frequency of wildfires, I generated a heatmap. This method is particularly useful for identifying patterns and hotspots of wildfire occurrences within a geographical area. The heatmap enables quick identification of areas with higher wildfire frequencies, which could suggest a correlation with specific weather conditions or geographical features.

3.4 Correlation Analysis

I assessed the correlation between wildfire duration and size using Pearson correlation coefficient and further examined their relationship using Spearman correlation. Pearson's correlation was chosen to measure the strength and direction of the linear relationship between two continuous variables. Spearman's correlation was used as a non-parametric measure to assess the monotonic relationship between these variables, offering insights even when the relationship is not linear.

3.5 T-test on Distance from Station

To explore how the proximity of weather stations to wildfire locations affects various metrics (e.g., detection time, temperature readings), I conducted a T-test comparing different distance categories. This test was applied under the hypothesis that closer stations might provide more accurate or relevant weather data related to wildfires.

3.6 ANOVA Test on Fire Size and Cause

An ANOVA test was utilized to investigate if there are statistically significant differences in wildfire sizes across different causes. Given that fire cause is a categorical variable and fire size is continuous, ANOVA is an appropriate choice for comparing means across multiple groups.

3.7 Chi-square Test

To explore the relationship between fire cause and management complexity, a Chi-square test was applied. This test is appropriate for categorical data, assessing whether there is a significant association between two categorical variables.

3.8 Methodology Justification

Each selected method aligns with the type of data analyzed and the specific research questions posed. By employing a combination of statistical tests and visualizations, I aimed to capture both the quantitative relationships between weather conditions and wildfires and the spatial-temporal patterns of wildfire occurrences. These methods were chosen for their applicability to the dataset's nature and the hypotheses being tested, ensuring a comprehensive understanding of the factors contributing to wildfire dynamics in California.

4 Results

4.1 Average Distribution of Temperature and Wildfire Frequency

The analysis reveals that temperatures from weather stations situated close to wildfire perimeters are not reliable indicators due to potential skewing. However, a general trend of rising temperatures and decreased rainfall leading up to the wildfire season (primarily July to September) was identified. This condition contributes to an abundance of dry foliage, increasing wildfire risk. This finding underscores the importance of considering spatial relationships between weather stations and wildfires for accurate data interpretation.

4.2 T-test for Relative Humidity

The T-test comparing relative humidity during dry and wet periods indicated no statistically significant difference in mean relative humidity between these periods (P-value: 0.322). This result suggests that the criteria used to define "dry" and "wet" may not capture the nuances affecting relative humidity's impact on wildfire occurrence. .

Threshold in Kilometers	T-statistic	P-value
0	nan	nan
1	-0.34741332738699404	0.728351115772618
2	-0.8118627545734829	0.41705616526746114
3	0.7343785743784186	0.46288358060048906
4	0.20267855156853556	0.8394259313813381
5	-0.3056136346429608	0.7599599969964965
6	-0.6602207312336403	0.5092583841337037
7	-1.1487924375559941	0.2509057210591464
8	-1.5650184582471722	0.11788279137622282
9	-0.2906080995582357	0.7714090037279535
10	-0.610528494707311	0.541645165644099
11	-1.0025876489905503	0.31629306672845814
12	-1.173736711715816	0.24076922069719078
13	-0.39233327725713874	0.6948923677436738
14	-0.3994036874109522	0.689677782072164
15	-0.7538865569254503	0.4510879892979436
16	-0.7342706852712716	0.46294929161362997
17	-1.0889157981348345	0.2764432796704451
18	-0.571908040719506	0.5675078643063625
19	-0.8126397569094396	0.41661060921398707
20	-1.0508798487638857	0.2935578966826053
21	-1.4923310358347142	0.13591565819012819
22	-1.6946294568514557	0.09044511064617497
23	-1.5761386207929957	0.11529797857558527
24	0.43163948292099813	0.6660929216294758
25	0.43163948292099813	0.6660929216294758
26	0.41719777398098307	0.6766198568287021
27	0.3927624475890921	0.6945754295260776
28	0.37086558241799755	0.7108131787357115
29	0.34320043611991774	0.7315170381983496
30	0.27999982853858635	0.7795333171185049
31	0.27999982853858635	0.7795333171185049
32	0.242323108206505	0.8085775863351321
33	0.1977002092195283	0.8433182595124803
34	0.1977002092195283	0.8433182595124803
35	0.1977002092195283	0.8433182595124803
36	0.13964738916270453	0.8889656346307822
37	0.13964738916270453	0.8889656346307822
38	0.13964738916270453	0.8889656346307822
39	0.13964738916270453	0.8889656346307822

Table 1: Table containing the station distance from wildfire, t-statistic, and p-values for the first 40 kilometers

4.3 T-test for Distance from Station

Analysis across various distances showed no significant impact of the proximity of weather stations to wildfires on the examined characteristics, with most P-values exceeding the significance threshold. However, a trend worth investigating further was observed at around 22 kilometers, albeit without strong statistical evidence.

4.4 ANOVA Test on Fire Size and Cause

The ANOVA test found no significant differences in wildfire characteristics across different fire causes (P-value: 0.517616). This implies that the reason for a wildfire's ignition does not significantly affect its size or other examined traits within this dataset. This outcome suggests that factors beyond the initial cause may play a more critical role in determining wildfire behavior and impacts.

4.5 Chi-square Test between Fire Cause and Management Complexity

The Chi-square test yielded a P-value of 0.3746, indicating no significant association between fire cause and management complexity. This finding suggests that the complexity of managing a wildfire is not significantly influenced by its cause, pointing towards a uniform approach in managing wildfires regardless of their origin. It emphasizes the need for broad strategies adaptable to various wildfire scenarios.

4.6 Spearman Correlation between Wildfire Duration and Size

The Spearman correlation coefficient of 0.545 indicates a moderate positive relationship between wildfire duration and size, suggesting that longer-lasting wildfires tend to cover more area. This relationship, while significant, is not strong enough to predict size from duration accurately, indicating the influence of other variables not captured in this simple correlation.

5 Implications and Conclusions

The findings from these analyses contribute to a nuanced understanding of the factors influencing wildfire behavior and management. While some expected relationships, like the impact of relative humidity and proximity to weather stations, showed no significant effects, the moderate correlation between duration and size points to complex dynamics governing wildfire spread. These insights underline the multifaceted nature of wildfire behavior, necessitating comprehensive strategies for prediction, management, and mitigation that account for a wide range of environmental, temporal, and geographical factors.

5.1 New Questions and Future Directions

1. Effect of Snow and Snowmelt: Realizing the overlooked significance of snow and snowmelt in contributing to soil moisture prompts further investigation into how these factors influence wildfire risk and behavior, potentially affecting seasonal and long-term wildfire patterns.
2. Refined Spatial Analysis: The intriguing hint of influence at a 22-kilometer threshold for weather station distance invites a more granular spatial analysis, possibly incorporating topographical features and exploring localized weather patterns.
3. More Complex Models: The moderate correlation between wildfire duration and size suggests the utility of developing predictive models that incorporate a broader range of variables, including detailed fuel models, human activity patterns, and advanced weather metrics.
4. Extended Temporal Scope: With more time, expanding the dataset to include records before 2020 and integrating satellite observations could enrich the analysis, enabling long-term trend analysis and the assessment of climate change impacts on wildfire dynamics.

5.2 Final words

The exploratory and analytical work undertaken has shed light on the intricate relationship between weather conditions and wildfires, challenging some initial assumptions while confirming others. The insights gained underscore the complexity of wildfire ecology and the critical role of comprehensive, context-aware analysis. Looking ahead, the integration of more granular environmental datasets, the application of more sophisticated analytical methods, and a consideration of the wider ecological and climatic factors would significantly enhance our comprehension of wildfire dynamics and inform better prevention and management practices. Interestingly, this study has also heightened my awareness of the nuanced intricacies involved in accurately gathering and interpreting data. It has underscored the myriad factors that can influence wildfire behavior, some of which may boil down to mere chance or human influence. The process underscored the criticality of meticulous data analysis, cleaning, transformation, and the inherent challenges and risks associated with data imputation.

References

CA.gov. California shapefile. URL https://data.ca.gov/dataset/ca-geographic-boundaries. Accessed: 2024-03-05.

Jon E. Keeley and Alexandra D. Syphard. Large california wildfires: 2020 fires in historical context. *Fire Ecology*, 17(22), 2021. doi: 10.1186/s42408-021-00110-7. URL https://link.springer.com/article/10.1186/s42408-021-00110-7.

NIFC. Nifc fire reports. URL https://www.nifc.gov/fire-information/statistics/wildfires. Accessed: 2024-03-05.

NOAA. Noaa daily data. URL https://www.ncei.noaa.gov/pub/data/ghcn/daily/. Accessed: 2024-03-05.

Zhu Mac Kinnon Geng Zhang Zheng Lei Shao Gong Davis Wang, Guan. Economic footprint of california wildfires in 2018, 2020. URL https://drive.google.com/file/d/1a1lLj-46Q1a5INaIXEOFwvd3T43ZWdNk/view. Accessed: 2024-03-05.