Minimization of CA Wildfire Impact

Yuan Chang Hanxiang Jiang Yiqing Li Yuxin Liu

March 2023

Abstract

Extreme fires are a growing threat to public health and safety, to homes, to air quality and climate goals, and to national forests. Due to the climate change, we are seeing fires that burn larger and hotter on average than ever before around the world. While all fire spots need to be taken care of, some emit higher threats than others implicitly. In this project, we wish to use convex optimization to find out the subset of locations that needs to be taken care of if we wish to minimize the financial, social and geological impact of wildfires. Our project focus solely on the state of California, and using individual counties as unit of regions. Our results shown that the level of priority generally have little to do with region size or proximity to large bodies of water, rather its geographical characterises. In particular, counties where the national forests and parks reside are usually prime targets for fire containment.

1 Introduction

1.1 Motivation

In 2022, there were 7396 fire incidents in the State of California. Among those chaotic disaster situations, we are facing the unprecedented devastation of many catastrophic wildfires: 2,569,386 acres of land were burned, 3846 structures were destroyed, and 3 causalities.[1] As California residents, we want to provide potential optimizations of how to allocate limited resources: fire trucks, firefighters, equipment, etc. Our motivation is to minimize the impact of the wildfires with respect to budget constraints that are determined by many factors.

1.2 Previous Work

Previous research focused on risk prediction, evacuation strategies & planning, sheltering, reconstruction, potential pollution, and health effects. Researchers have explored various models using statistical and data-driven machine-learning approaches with diverse focuses, such as the severity & impact of wildfires and the investigation of the probability of ignition or burning.[2] Logistic Regression (LR), Random Forest (RF), Boosting Regression Trees (BRT) models were adopted for fire risk classification based on historical statistics of fire and weather conditions. Researchers indicated that RF and BRT models were more accurate under the evaluation of human-induced wildfires in Spain.[3] Support Vector Machine model was used to predict the fire hazard level of a day given previous weather conditions, such as minimal and maximal temperature, average humidity, wind speed, and solar radiation. [3]

Previous studies with a focus on intensity and effects of wildfires applied statistics-based approach, which involved the numbers to imply or deduce the cause and effect [2]. For simulating wildfires, there were two techniques including vector implementations and raster implementations. The vector implementation makes use of the outer shape constituted by expanding fire points using the spread model and local factors like weather and topography. The raster implementation practices a group of contiguous cells that can be burning or not burning. Correspondingly, Cellular Automata approach was applied to track the movement of the fire through the cells in the simulation domain. [4]

Spatio-temporal Data Mining techniques were also frequently used in fire prediction. Their application included forecasting and trend analysis, association rule mining for prediction of ongoing forest fire development, pattern detection for sequence of fire events, and cluster analysis and identification of fire spots [5].

However, there is only limited research on the optimization of resource allocation. Therefore, we want to explore and contribute how to allocate insufficient resources more efficiently.

1.3 Intended Contributions

This project aims to provide a subset of which counties should be rescued first with limited resources to minimize the financial, social, and geological impact of wildfires, given the fire spots. It is important to note that this prioritization does not mean that we will ignore other fire spots or fail to provide assistance. Instead, it will help us allocate resources efficiently to minimize property loss and save lives. This project can also serve as a reference for whether the government or related departments should raise the budget for wildfire prevention and mitigation.

1.4 Organization of the paper

This report consists of the following sections: Background Information, Problem Formulation, Methodology, Experiments & Results, and Conclusion & Future Work. We described the simplification & constraints of our dataset in Section 2. Then we formulated our problem, primal, dual, and KKT conditions in section 3. We described our data, collection process, and processing in section 4 and demonstrated our results and observation in section 5. Last, we concluded our report and provided suggestion of future work in section 6.

2 Problem Specification

We are trying to solve a constrained optimization problem under the settings of wildfire intervention. The primary question is that when multiple fires start simultaneously, which subset of locations should the central fire department devote their manpower to under certain budget constraints.

We aim to model this problem with a few simplifications.

- The search domain is restricted to counties in California only. Each location is also generalized to individual county, instead of cities or other geographical units.
- The impact of a neglected wildfire location's impact is calculated with considerations of social, economical and other metrics.
- The budget and cost of sending manpower to contain the wildfire will be approximated using publicly available historical data.
- The number of regions that receive help will be determined by a fixed parameter, which will be experimented on.

More details about the data collection will be discussed in the later section.

3 Problem Formulation

We define the graph G as the undirected graph mapping the counties of a state. Since there are 58 counties in California, we will have a diagonal matrix $M \in \mathbb{R}^{58\times 58}$ that denotes the financial, social and geological impact of each area if it was damaged by wildfire. Given the difficulty of the terrain of each node in the graph, we will also have a diagonal matrix D denoting the relevant cost required for the firefighters to reach and thus contain the wildfire should there be one. Additionally, there would be another matrix C denoting the connectivity of each pairwise counties, in other words, how likely would the fire spread between them. Let Matrix S denote the sum of the surrounding impact at each location if fire was spread to each connected region. This matrix can be computed given M and C.

Our assumption is that we will only save a subset of the regions, and thus needing to minimize the aggregate impact of the wildfire under budget constraints.

We also have the following definitions related to the constraints:

- Let $e \in \{0, 1\}^{58}$ denote the regions that currently experiencing a wildfire. Where $e_i = 0$ implies that county i does not have a wildfire, and vice versa.

- Let f denote the total number of regions neglected by the fire department. This is a parameter that we will tune in the experiment stage. The constraint associated with f is $1^T x = f$, this is simply stating that we have to save f regions, otherwise the problem has a solution of $x = 0$.
- Let t denote the total budget that the central department has for deployment.
- Let x denote the set of regions to send the firefighters. $x_i = 0$ will imply firefighters were sent to region i and thus negating its possible impact.

3.1 Primal Formulation

The problem in primal form is:

$$\begin{array}{ll} \text{minimize} & e^T Mx + e^T Sx \\ \text{subject to} & 1^T x = f \\ & 1^T Dx \leq t \\ & 0 \leq x \leq 1 \end{array}$$

In standard form:

$$\begin{array}{ll} \text{minimize} & e^T Mx + e^T Sx \\ \text{subject to} & 1^T x - f = 0 \\ & 1^T Dx - t \leq 0 \\ & x - 1 \leq 0 \\ & -x \leq 0 \end{array}$$

3.2 Lagrangian Dual

The Lagrangian dual can be defined as:

$$\begin{aligned} L(x, \lambda, v) &= e^T Mx + e^T Sx - \lambda_{1i}^T x_i + \lambda_{2i}^T (x_i - 1) + \lambda_3^T (1^T Dx - t) + v(1^T x - f) \\ &= (e^T M + e^T S)x - \lambda_1^T x + \lambda_2^T (x - 1) + \lambda_3^T Dx - \lambda_3^T t + v^T x - vf \\ &= (e^T M + e^T S - \lambda_1^T + \lambda_2^T + \lambda_3^T D + v^T)x - (\lambda_2^T + \lambda_3^T t + vf) \end{aligned}$$

The dual function $g(\lambda, v)$ is:

$$\begin{aligned} g(\lambda, v) &= \inf_x L(x, \lambda, v) \\ &= -(\lambda_2^T + \lambda_3^T t + vf) + \inf_x (e^T M + e^T S - \lambda_1^T + \lambda_2^T + \lambda_3^T D + v^T)x \end{aligned}$$

Thus, analytically we can see that

$$g(\lambda, v) = \begin{cases} \text{-}\ (\lambda_2^T + \lambda_3^T t + vf) & (\mathrm{e}^T M + e^T S - \lambda_1^T + \lambda_2^T + \lambda_3^T D + v^T) = 0 \\ \text{-}\infty & \text{otherwise} \end{cases}$$

The dual optimization problem is :

$$\begin{array}{ll} \text{maximize} & -(\lambda_2^T + \lambda_3^T t + vf) \\ \text{subject to} & \lambda \geq 0 \\ & v \geq 0 \\ & (e^T M + e^T S - \lambda_1^T + \lambda_2^T + \lambda_3^T D + v^T) = 0 \end{array}$$

3.3 KKT conditions

The first-order KKT condition that lead to zero duality gap are the following:

- Primal constraints: $1^T x - f = 0$, $1^T Dx - t \leq 0$ and $0 \leq x \leq 1$.
- Dual constraints: $\lambda \geq 0$.
- Complementary slackness: $\lambda_1^T x = 0$, $\lambda_2^T (x-1) = 0$ and $\lambda_3^T (1^T Dx - t) = 0$.

Since the objective is affine, there is no duality gap between primal and dual problem.

4 Methodology

4.1 Data Collection & Processing

The data on the adjacency of counties in California was obtained from the United States Census Bureau's county adjacency file[6]. This file provides a list of all counties in the United States and identifies which counties are adjacent to each other. The data was parsed using `Python` by extracting all counties in other states except for California and all disturbing information. All counties in California are stored in a list, and all adjacency are stored in a dictionary, with the key being the county and the value being the list of counties that are adjacent to that county in the key. The $C \in \{0,1\}^{58\times 58}$ adjacency matrix can be easily computed. This adjacency matrix will be used as input for our wildfire analysis, allowing us to examine the potential spread of fires from one county to another.

In order to guarantee the authenticity of connectivity matrix C, we integrated counties' humidity and fire probabilities on the basis of adjacent matrix. For the humidity data, we referred California Average Humidity County Rank[7]. Since the lower humidity levels, the higher wildfire spread rate, we use one minus the normalized humidity to represent part of connectivity. Moreover, each counties' inherit probability of having a wildfire on its own should also be Incorporated into the connectivity metric. This data was obtained from CAL FIRE's fire probability model, which approximated the fire probabilities of all CA counties from 2021 to 2050[8].

A diagonal matrix $M \in \mathbb{R}^{58\times 58}$ denotes each area's financial, social, and geological impact if it was damaged by wildfire. We extracted Dollar Damage data from the 2021 Wildfire Activity Statistics Redbook CAL FIRE(California Department of Forestry and Fire Protection) provided [9]. It is important to note that these statistics only covered the Direct Protection Area by CAL FIRE; however, there are other agencies, such as the United States Forest Service, United States Forest Service also have direct responsibilities over 17 counties within California. Since we have limited access to the dollar damage data among those

17 counties, we took their adjacent counties' average dollar damage to represent each counties' dollar damage. The diagonal matrix $M \in \mathbb{R}^{58\times 58}$ will be used as an input for further analysis.

The matrix $S \in \mathbb{R}^{58\times 58}$ represents the aggregate impact of a county's surrounding counties. We only consider CA counties in this study, where some counties border counties that are outside of California, we simply ignored them. We can represent S in the following way:

$$S = \begin{bmatrix} s_1 & & \\ & \ddots & \\ & & s_{58} \end{bmatrix}$$

where s_i can be calculated as the following:

$$s_i = 1^T(Mc_i)$$

Before we state the data about cost and damage, it is worth to mention that CAL FIRE is responsible for fire protection in State Responsibility Areas of California totaling 31 million acres, which is more than 75% of the state of California. There are several other partnering agencies including USFS(United States Forest Service), BLM(Bureau of Land Management), NPS(National Park Service), CC(Contract County) and other federal fire protection agencies. We only consider the areas that is under CAL FIRE's jurisdiction.

Data of the D matrix should illustrate the cost of each county spend on suppressing the wildfire. Therefore, $D \in \mathbb{R}^{58\times 58}$ diagonal matrix with each diagonal element representing each county's suppression cost. Since there is no direct data for counties' cost, we made use of each county's state responsibility area of CAL FIRE(California Department of Forestry and Fire Protection) and the cost of wildfire per acre to build the data for each county's suppression cost. The state responsibility area data are provided by the Redbook of wildfire which includes related Wildfire Activity Statistics[9]. The cost of wildfire per acre was fixed to 100 dollars per acre according to Stanford research publication[10]. There are several counties that are not under the jurisdiction of CAL FIRE, and the responsibility area is zero. Thus, we replace zero with their adjacent counties' area average for calculation.

The total budget t is a scalar given by recent four years' average emergency fire suppression costs. And the data can be retrieved from Department of Forestry and Fire Protection budget files.

4.2 Connectivity Analysis

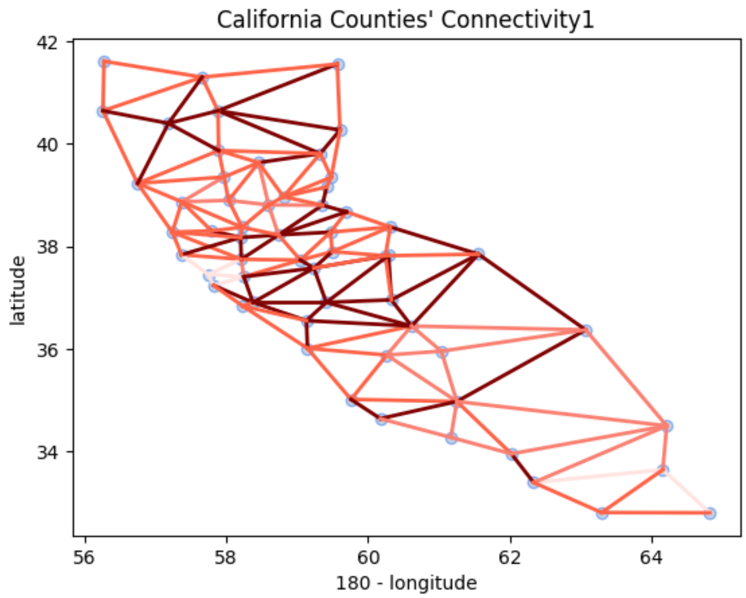


Figure 1: California County Connectivity1

Connectivity is of great importance to the spread wildfire among counties, and the biggest wildfires in California history are all related to high extension rate of wildfire. Therefore, we gathered the latitude and longitude of 58 counties of California and drew their connectivity according to the connectivity matrix C. There are two connectivity graphs since the connectivity from county A to county B and the connectivity from county B to county A are not the same value. In these two graphs, we used darker color to represent higher connectivity and lighter color for lower connectivity.

Most of the dark red connections are concentrate on the county of Trinity, Shasta, Sacramento, El Dorado and San Joaquin. And most of the dark green connections are about the county of Kern, Placer and a great amount of counties in the mid-California. Hence, we were supposed to observe the optimization models' results about these counties. However, the final conclusion can have minor differences from what we expected since different counties might have greats gaps in wildfire damages.

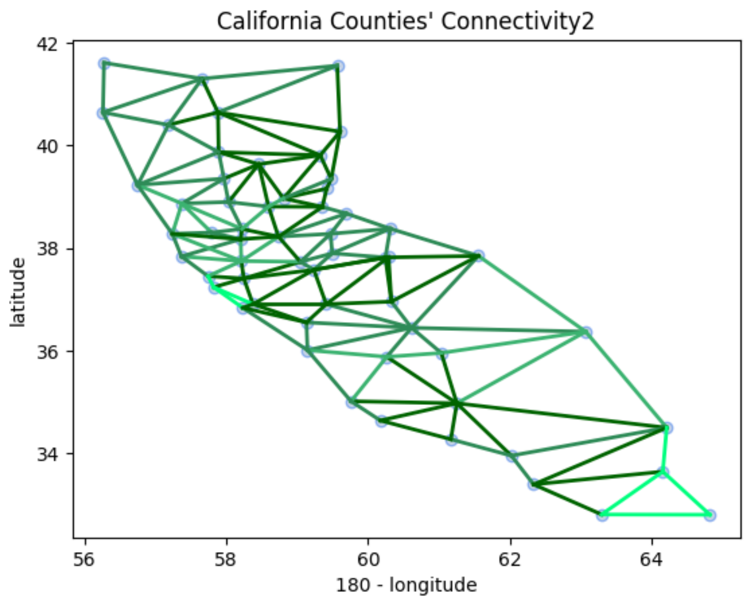


Figure 2: California County Connectivity2

4.3 Algorithm & Code

In this project, we used `Numpy`, `Pandas`, `Tabulate` for data processing. For calculating the solutions, we used `Cvxpy`. All code and gathered data sets are available on Github. [11]

5 Experiments & Results

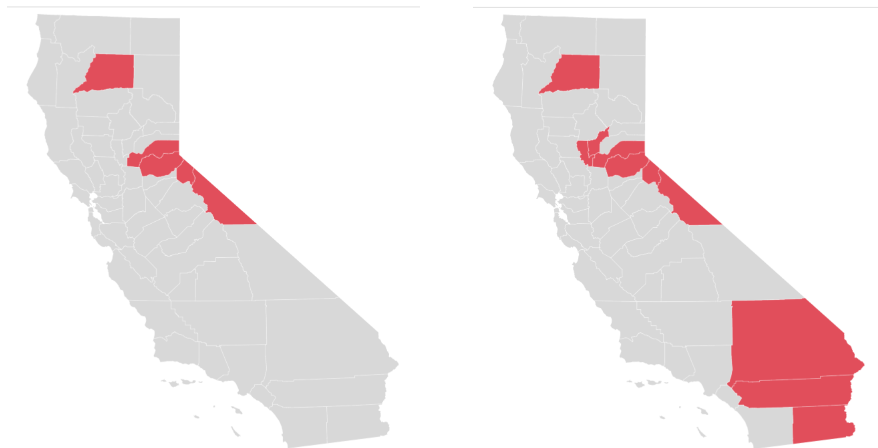
We intend to experiment our constraint optimization model under the following two scenarios:

- All 58 counties in CA experiencing wildfire simultaneously.
- All counties that experienced a wildfire during 2021 calendar year.

In each case, we wish to compute the 5 or 10 top counties that the institutions should priorities their manpower to. Our conjecture about the results is that the top counties should be ones that have the highest impact of wildfire if not controlled, as well as where major national forests are concentrated in California.

5.1 Even Distributed Wildfire

Given 5 and 10 regions that would be chosen by fire department, we conducted two convex optimization experiments. Since f is a scalar stands for the total number of regions neglected, the corresponding values should be 53 and 48.



(a) Top 5 counties to save

(b) Top 10 counties to save

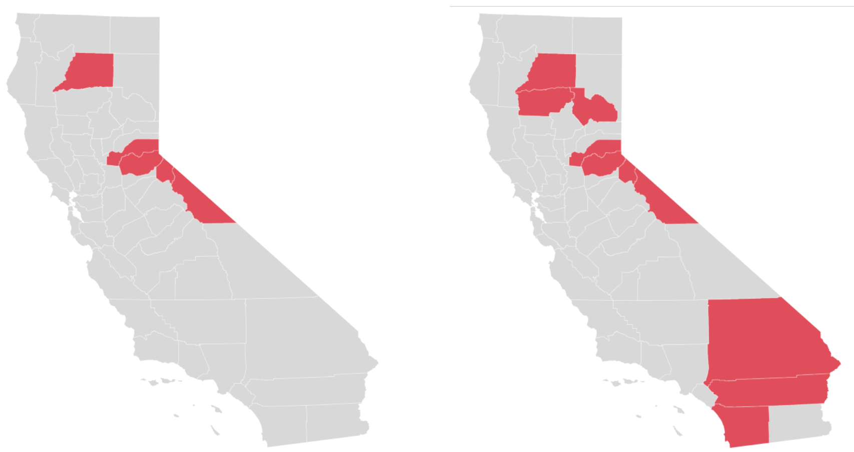
Figure 3: All counties experiencing wildfire

According to the results, it is obvious that all counties that would be saved have x_i extremely close to 0. On the contrary, the counties that are not chosen by the optimization algorithm have x_i close to 1. This phenomenon is matched with our prior assumption about x that $x_i = 0$ will imply firefighters were sent to the region i and 1 would imply the converse condition. Although we applied convex relaxation to the problem that does not require the solution to be strictly integer, the results were close enough to $0, 1$ which is clear to separate from.

The figure shows that El Dorado is the county that should be saved with the first priority. The Caldor Fire in 2021 burned 221,835 acres in the Eldorado National Forest and other areas of the Sierra Nevada in El Dorado, Amador, and Alpine County, California, which cost more than 1.2 billion dollars. Similarly, the largest wildfire in 2021, Dixie Fire caused great damages to Butte, Plumas, Lassen, Shasta, and Tehama Counties, but since CAL FIRE does not take big state responsibility area for these counties, the only one county been chosen is Shasta.

5.2 2021 Wildfire

In order to put the optimization model into practice, we bring the biggest 23 real wildfire counties in 2021 into the experiment. From the following figure, we can infer that most of the counties are almost the same as the results in the Even Distributed Wildfire. Several counties like Tehema, San Diego, Plumas replaced the county of Yuba, Sutter and Imperial. The reason behind this change is CAL FIRE has a great amount of state responsibility area in the latter counties, which means these counties cost much more than suppressing wildfire in other counties.



(a) Top 5 counties to save

(b) Top 10 counties to save

Figure 4: 2021 Wildfire incidents

6 Conclusion & Future Work

The application of this project is for CAL FIRE to reasonably allocate wildfire related resources when there are several wildfire in different counties. In retrospect of connectivity analysis, Trinity, Shasta, Sacramento, El Dorado, Kern and Placer should be the central counties we are supposed to observe. These radial-sized couties and mid-California counties become the counties that the optimization model would choose, which is rational considering about connectivity analysis. The biggest 5 fire happened in 2021 which includes Dixie Fire(Plumas, Butte, Lassen, Shasta, Tehama), Monument Fire(Trinity), Caldor Fire(El Dorado), River Complex and Antelope Fire(Siskiyou). Among these counties, Trinity and Siskiyou are not in any result of our experiments. This bias may be caused by the fact that we only considered CAL FIRE's state responsibility area and their corresponding damages, which would constitute one of the future work we could do.

As mentioned before, suppressing wildfires are joint responsibilities for agencies including CAL FIRE, USFS, BLM, NPS, CC and other federal fire protection agencies. So if these agencies are interested, they could also apply their data into our models to acquire corresponding results.

Our future work may also include adding other factors other than humidity and fire probability to construct the connectivity matrix like solar radiation, streamflow, temperature, average wind speed and precipitation, etc. Furthermore, the cost matrix can also be modified with more parameters including firefighting equipment, personnel, and aircraft allocation costs. In addition to that, our optimization model gives us definitive results on which subset of counties to save. In the future, we would like to modify the definition of x to have a result showing how many firefighting resources should be distributed to different counties. If it is possible, we can add more parameters to build a directed connectivity model to better capture the causal relationships between the wildfires and geographical characteristics of the counties.

7 Task Assignment

- Literature Survey - Yiqing Li
- Data processing - Yiqing Li, Hanxiang Jiang, Yuxin Liu, Yuan Chang
- Algorithm - Yuan Chang
- Experiment - Yuxin Liu
- Report - Yiqing Li, Hanxiang Jiang, Yuxin Liu, Yuan Chang

References

[1] CAL FIRE. 2021 Incident Archive. https://www.fire.ca.gov/incidents/2021/

[2] Malik A, Rao MR, Puppala N, Koouri P, Thota VAK, Liu Q, Chiao S, Gao J. Data-Driven Wildfire Risk Prediction in Northern California. Atmosphere. 2021; 12(1):109. https://doi.org/10.3390/atmos12010109

[3] G. E. Sakr, I. H. Elhajj, G. Mitri and U. C. Wejinya, "Artificial intelligence for forest fire prediction," 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Montreal, QC, Canada, 2010, pp. 1311-1316, doi: 10.1109/AIM.2010.5695809.

[4] Ghisu T, Arca B, Pellizzaro G, Duce P. An Improved Cellular Automata for Wildfire Spread. Science Direct. 2015. https://www.sciencedirect.com/science/article/pii/S1877050915011965

[5] Cheng, T., Wang, J. (2007). Application of a Dynamic Recurrent Neural Network in Spatio-Temporal Forecasting. In: Popovich, V.V., Schrenk, M., Korolenko, K.V. (eds) Information Fusion and Geographic Information Systems. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-37629-3_12

[6] United States Census Bureau. County Adjacency File. https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html

[7] World Media Group. California Average Humidity County Rank 2022. http://www.usa.com/rank/california-state–average-humidity–county-rank.htm

[8] CAL FIRE FRAP. Annual Probability of Wildfire Occurance 2021-2050 of California. https://egis.fire.ca.gov/FireProbability/

[9] Joe Tyler, Wade Crowfoot and Gavin Newsom. 2021 Wildfire Activity Statistics from California Department of Froestry and Fire Protection. https://34c031f8-c9fd-4018-8c5a-4159cdff6b0d-cdn-endpoint.azureedge.net/-/media/calfire-website/our-impact/fire-statistics/

[10] Marshall Burke, Sam Heft-Neal and Michael Wara. Managing the growing cost of wildfire, 2020. https://siepr.stanford.edu/publications/policy-brief/managing-growing-cost-wildfire

[11] Li Y, Jiang H, Liu Y, Chang Y (2023). Code and data sets for CSE203B Final Project. https://github.com/itis2010me/CSE203B