Estimating Wildfire Smoke Impact on Fresno, California

Informed by Chronic Lower Respiratory-Related Mortality

Sarah Kilpatrick - Data Science

University of Washington

kilpas@uw.edu

CENTER FOR STUDIES IN DEMOGRAPHY & ECOLOGY

Introduction

Fresno is a mid-sized city in the California Central Valley, an economic hub predominantly contributing to the agricultural industry. As of 2020, the city boasts over half a million inhabitants; it is geographically centered between population centers in Northern California, Southern California, and Nevada. There is evidence to suggest that wildfires are increasing in not only size but also intensity. Airborne pollutants from wildfires in forests hundreds of miles away from Fresno still pose the risk of exacerbating chronic heart and lung conditions (Environmental Protection Agency).

Data: A Legacy of Wildfires

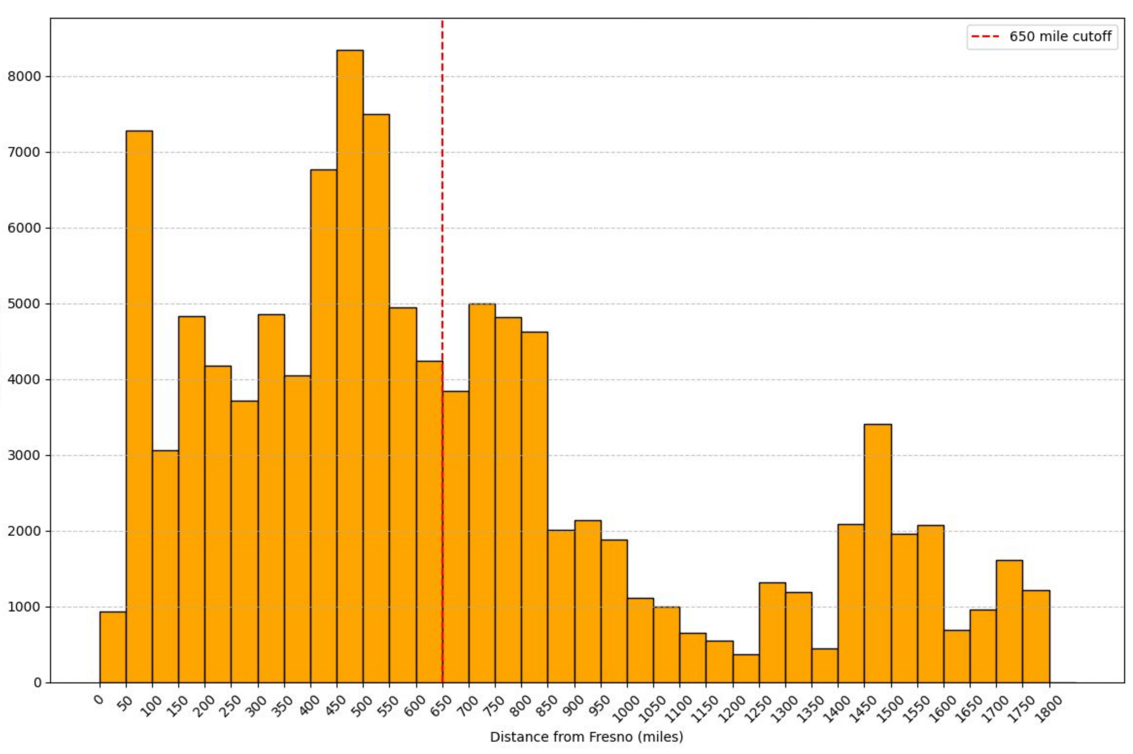


Fig. 1 Wildfire data from the United States Geological Survey (USGS) includes 135,026 documented wildfires over the course of 128 years (Welty et. al). 47.92% of these fires occurred within 650 miles of Fresno.

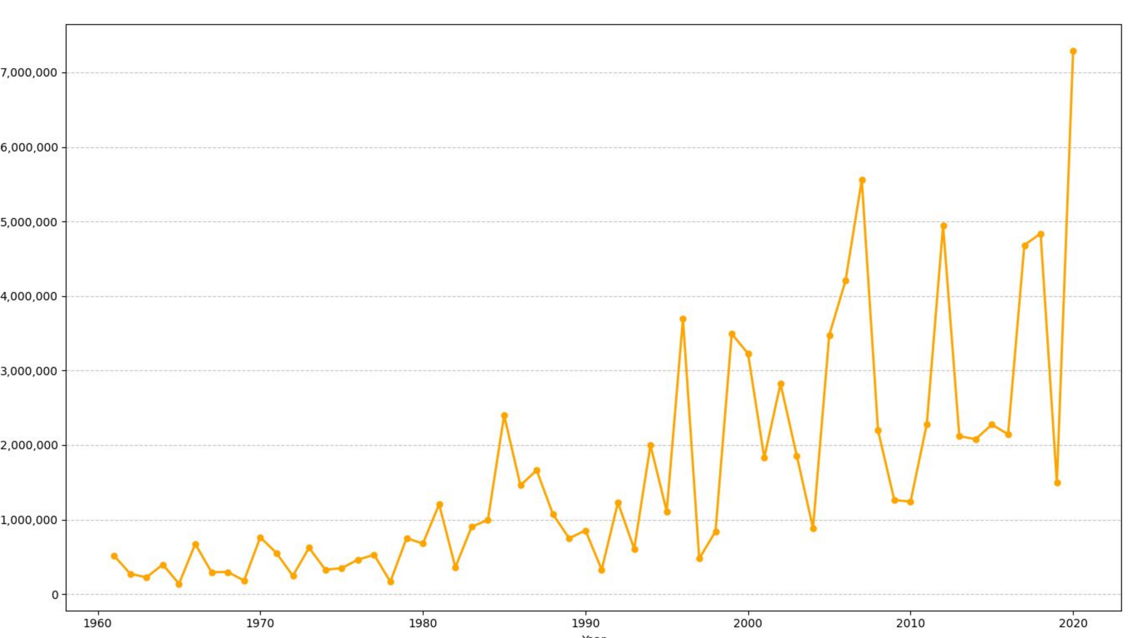
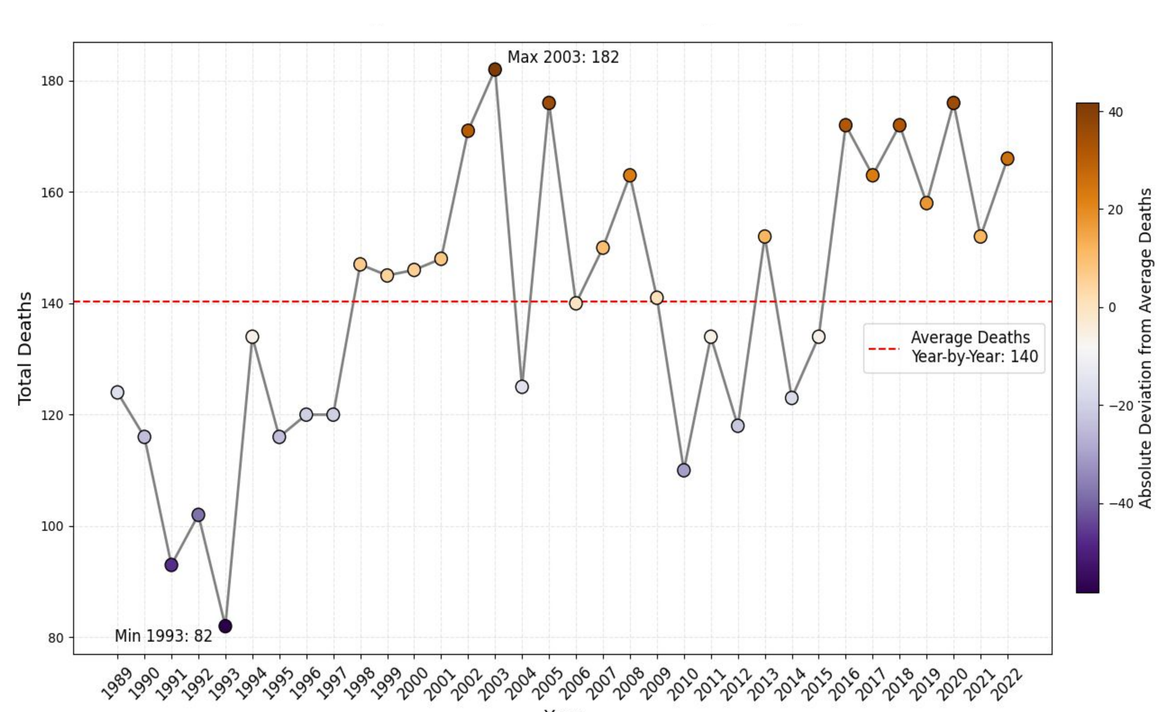


Fig. 2 Total Acres Burned Per Year (1961-2020). Total acres burned in North America have become increasingly variable over the years. While large fire years are not inherently harmful to fire-dependent ecosystems, the rise in particularly large, destructive fires is likely to significantly impact respiratory health (Donato et al.).



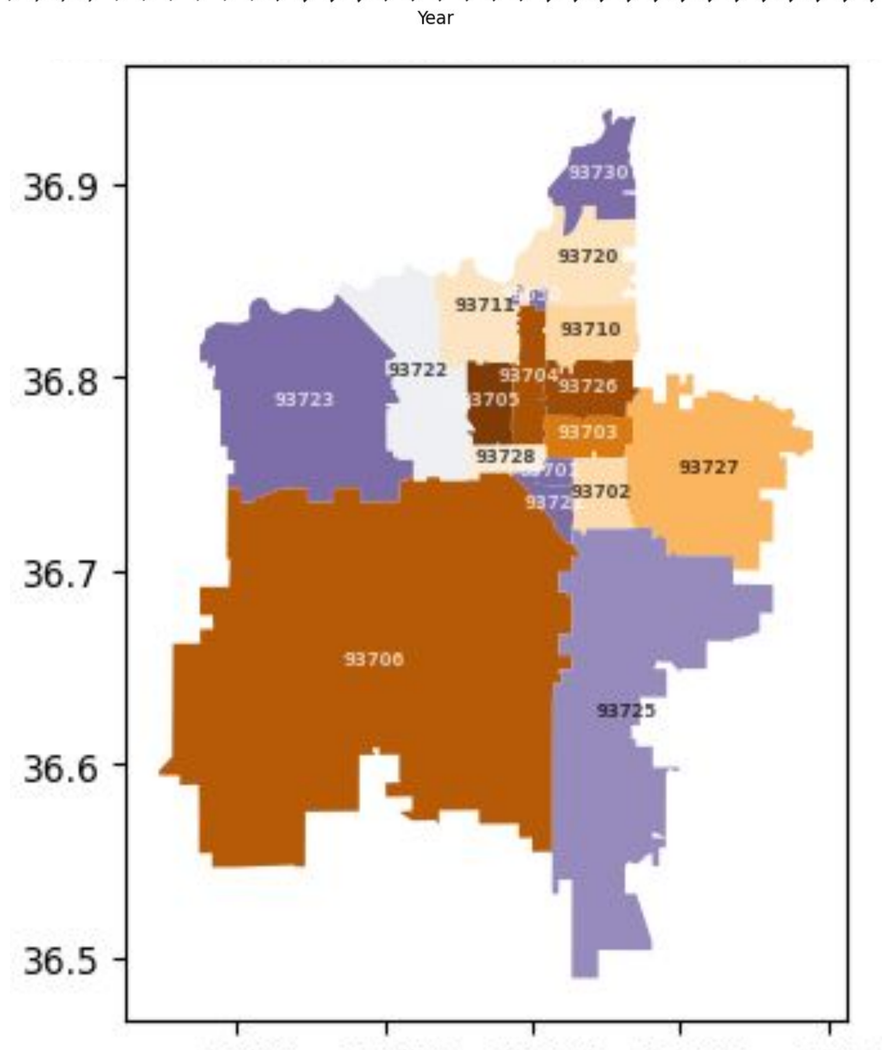


Fig. 3 Fresno, CA (1989-2019) Deviation from Average Deaths Due to CLRD Controlling per 100,000 people.
Fig. 4 Total deaths due to CLRD by year (1989-2019) provided by California HHS.

The "Death Profiles by Zip Code (1989-2022)" dataset, hosted on the California Health and Human Services (CalHHS) Open Data Portal and provided by the California Department of Public Health, contains comprehensive mortality data for California residents. Derived from death certificates, this dataset offers counts stratified by ZIP Code of residence, age, gender, and cause of death.

Methodology

The links between increased wildfire-related pollution and increased mortality in California are well-documented (Chen et al., Gwon et. al). This analysis creates a predictive estimate for the impact of smoke caused by wildfires within 650 miles of Fresno. Then, data on deaths caused by chronic lower respiratory diseases inform the predictive estimate to quantify future impacts of smoke from a public health perspective. This project presents a refinement of the preliminary smoke estimate discusses the use of non-periodic time series forecasting techniques. This project generates correlations between the refined smoke estimate Fresno's deaths caused by respiratory conditions. In turn, calling for a study detecting causal connections between the two phenomena.

S : *Smoke Impact Score*
A : *Median Acres Burned Annually*
W : *Median Cardinal Weight Annually*
C_l : *Distance Correlation between first-round smoke estimate and $l = [0-5]$ years of lag in deaths due to CLRD*
l : *Number of years of lag between Wildfires and deaths due to CLRD*
D : *Median Distance to Fresno Annually (in miles)*
F : *Scaling Factor*

$$S = \frac{A \cdot W \cdot C_l}{(D+1)^2/F}$$

Fig. 5 Correlation between two smoke estimates and associated variables with deaths due to CLRD 0-5 years following the wildfire years.

During the fire season (May 1–October 31), winds frequently originate from the northwest, amplifying the transport of wildfire smoke into the region. The scaling factor ensures that the resulting score is intuitive and comparable to established air quality indices, making it a practical tool for understanding wildfire smoke impacts.

Distance correlations detect both linear and nonlinear associations. It is multiplied by 10 to represent that a higher statistically-significant correlation means a higher, and therefore more intense, final smoke estimate. The lag in years (l) can be set between 0-5 to measure the association between a wildfire in year x and deaths in year $x+l$.

The estimate also incorporates the inverse square law to model smoke dispersion, accounting for diminishing intensity over distance and ensuring mathematical stability in rare cases where a wildfire's perimeter intersects Fresno's city boundaries. The use of medians (A, W, and D) improves the robustness of the estimate by reducing the influence of outliers, such as unusually large fires or extreme distances; this final smoke estimate builds upon the preliminary smoke estimate, which used averages.

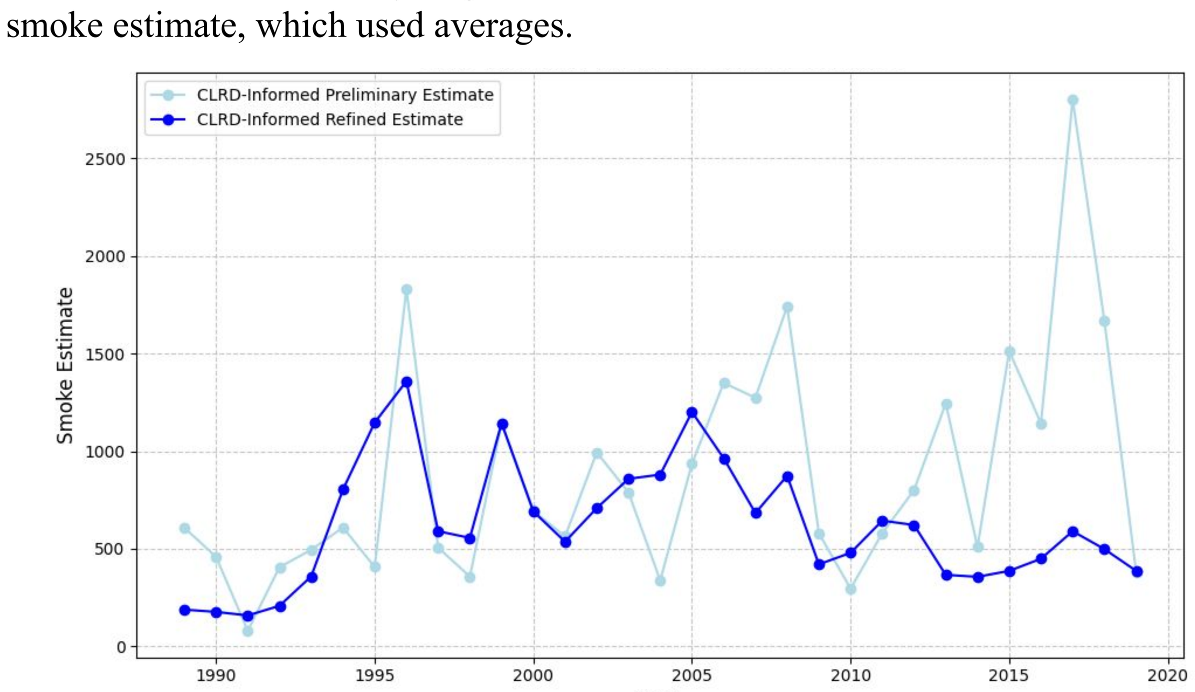


Fig. 6 Comparison of first and final smoke estimates. The preliminary smoke estimate measures smoke 'prevalence', which indicates that smoke from larger, more intense fires is coming into Fresno as years go on. However, in the same timeframe, the refined smoke estimate shows that the impact on CLRD-related deaths is not necessarily increasing at the same rate.

Forecasting Methods

Holt-Winters Exponential Smoothing Model, which is relatively simple, fast, known to work well with trend and seasonality and has low parameter tuning requirements. However, the lack of flexibility for irregular data limited this model's capability to capture the historical data's variability.

Convolutional Neural Network (CNN) was the most promising, adapted for time series data, as found on the Tensorflow website. This method is promising but fails to generalize well to unseen data. It can handle non-linear patterns, flexible input, but it is most useful for large datasets, rather than univariate data over only 30 years.

Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX) Model is the most flexible out of the three for handling seasonality and non-stationary data; however, the forecasted results were not very statistically significant.

Opportunities for Continuation

- Account for annual income and occupation, key predictors of health outcomes.
- Incorporate evidence suggesting that wildfire-induced air pollution in the stratosphere may significantly influence pollutant dispersion alongside tropospheric winds (J.M. Katich et al.).
- Incorporate a more refined dispersive mathematical model for wildfire smoke.
- Include or restrict different time periods for a more specific analysis.
- Investigate the confounding effects of air pollutants originating from the northwestern San Francisco Bay Area, which may impact Fresno independently of wildfire activity.