Wildfire Smoke and its Impact on Respiratory Health - Stockton, California

Sushma Vankayala

Introduction **1**
Related Work **2**
Methodology **3**
- Historic Wildfires 3
- Smoke Estimates 4
- Health Data 7
  - Data Integration and Preparation 8
  - Visual Analysis 9
  - Correlation Analysis 9
  - Time Series Analysis 9

Findings **10**
- Visual Analysis 10
- Correlation Analysis 11
- Time Series Analysis 12

Discussions **13**
Limitations **14**
Conclusions **15**
References **15**
Data Sources **15**

Introduction

Wildfires have become an increasing threat worldwide, driven by factors such as climate change, prolonged droughts, and changing land use. In California, wildfires are not just an environmental concern but also a significant public health issue. The smoke generated from wildfires contains fine particulate matter (PM2.5) and harmful chemicals, which can penetrate deep into the lungs and even enter the bloodstream, causing widespread health problems.

Wildfire smoke has been linked to various short- and long-term health effects. Short-term exposure can lead to breathing difficulties, irritation of the eyes and throat, and worsening of pre-existing conditions such as asthma and chronic obstructive pulmonary disease (COPD). Long-term exposure increases the risk of chronic diseases, including cardiovascular conditions, lung disease, and even premature death.

Stockton, a city in California's Central Valley with over 320,000 residents, is frequently affected by wildfire smoke due to its proximity to fire-prone regions and the tendency of smoke to settle in the valley's low-lying areas. As wildfires become more frequent and intense, understanding their health impacts is critical for protecting communities. This study focuses on analyzing the effects of wildfire smoke on Stockton's residents, particularly related to respiratory diseases, to provide actionable insights that can guide health policies and interventions.

Related Work

Existing research provides valuable insights into the health impacts of wildfire smoke, emphasizing the critical need to investigate its effects on urban communities like Stockton, CA. A UCLA study estimated that wildfire smoke exposure in California caused over 50,000 deaths within a 11-year period, largely due to respiratory and cardiovascular illnesses exacerbated by PM2.5 exposure[1]. This research highlights how even short-term exposure to wildfire smoke can significantly increase mortality rates.

Long-term studies also underscore the chronic health risks associated with wildfire smoke. Researchers at UC Davis found that individuals exposed to prolonged wildfire smoke were at a higher risk of developing chronic obstructive pulmonary disease (COPD) and lung cancer[2]. These effects were particularly pronounced in populations with pre-existing conditions, such as asthma and heart disease. The findings also suggested systemic inflammation caused by PM2.5 and its potential to harm other organ systems beyond the lungs.

Many investigations have revealed the correlation between wildfire events and hospitalizations. A study focusing on Northern California wildfires reported sharp increases in emergency room visits during and immediately after wildfire seasons, especially for respiratory conditions such as asthma[3]. This aligns with findings from multiple studies that link wildfire smoke exposure to spikes in hospital admissions for respiratory distress, cardiovascular complications, and even mental health conditions like anxiety and depression.

Guided by these findings, this study poses two primary research questions:

1. How does wildfire smoke exposure impact mortality rates in Stockton, CA, particularly concerning respiratory illnesses and cancers?
2. Is there a correlation between hospitalization data and wildfires, considering both short-term and long-term effects of smoke exposure?

Methodology

Historic Wildfires

To address my research questions about the health impacts of wildfire smoke in Stockton, CA, I began by analyzing the occurrence and nature of smoke events in the area. Using USGS data, I assessed wildfire frequency, scale, and proximity to Stockton, considering variables such as distance from the city and acres burned.

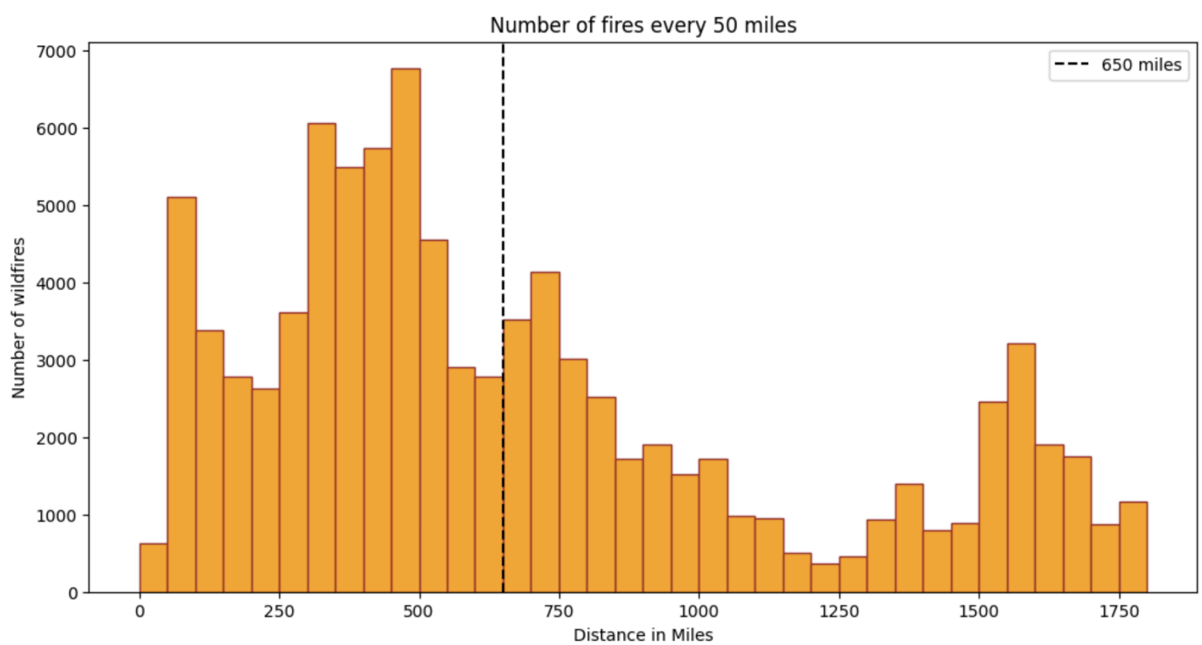


Figure 1: Histogram showing the number of fires occurring every 50 miles from Stockton, CA.

The distribution of wildfires around Stockton shows a distinct pattern. The histogram in Fig.1 illustrates wildfire occurrences within incremental 50-mile distances from Stockton. A notable peak occurs around the 400-500 mile range, with approximately 6,000-7,000 fires recorded. The analysis reveals significant fire activity within a 650-mile radius (marked by the dashed line), suggesting that Stockton is potentially affected by smoke from fires across a large geographical area.

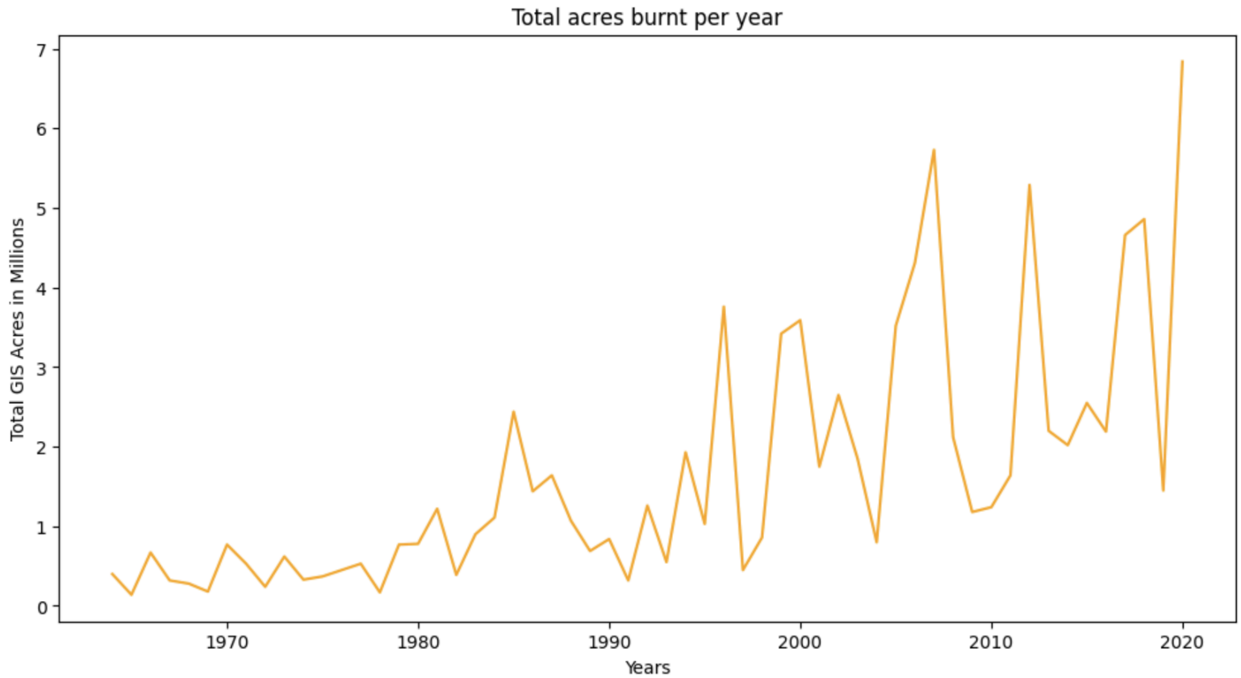


Figure 2: Time series graph of total acres burned from the fires within 650 miles of Stockton, CA

The temporal analysis of total acres burned near Stockton, shown in the second graph, reveals a concerning upward trend from 1963 to 2020. The line graph in Fig.2 shows dramatic increases in burned acreage starting from the late 1980s, with particularly severe spikes occurring in the 2000s and 2010s. The most recent years show unprecedented levels of burning, with peaks reaching nearly 7 million acres in a single year. This trend suggests a significant intensification of wildfire activity in regions that could affect Stockton's air quality.

These patterns indicate several key points:

- The frequency of fires varies significantly with distance from Stockton, with certain distance ranges showing higher fire activity
- The total area burned has increased substantially over time, particularly in recent decades
- The data suggests a potential correlation between increasing wildfire activity and climate change impacts

This analysis forms the foundation for understanding how wildfire patterns may influence smoke exposure and subsequent health impacts in Stockton. The increasing trend in both fire frequency and intensity suggests a growing potential for smoke-related health issues in the community.

Smoke Estimates

To quantify wildfire smoke impact in Stockton, California, I developed a Smoke Estimate that incorporates multiple key factors. The formula considers:

$$\text{Smoke estimate} \sim \frac{\text{Acres} * \text{Circleness} * \text{Type}}{\text{Distance}}$$

This formula captures several crucial aspects of smoke dispersion:

- The size of the fire (Acres) directly influences smoke production. Larger fires produce higher quantities of smoke.
- The shape of the fire (Circleness) affects smoke distribution patterns. Smoke from rounder fires tends to be more contained.
- The type of fire impacts smoke intensity and characteristics. A prescribed fire is generally under control, and hence potentially have lower smoke emissions.
- The distance factor accounts for smoke dissipation over space.

The formula for estimating smoke exposure reflects key factors that influence smoke impact. It accounts for how the effect of smoke typically diminishes with distance from Stockton and increases with the size of the wildfire. Parameters such as the "circleness" (a measure of how circular the fire's spread is) and fire type further refine the calculation, capturing characteristics that affect smoke production and dispersal. These refinements make the formula more representative of real-world smoke behavior.

To evaluate whether the smoke estimates meaningfully represent conditions in Stockton, I turned to Air Quality Index (AQI) data. AQI is a widely used measure that quantifies air pollution levels, particularly pollutants linked to wildfire smoke, such as particulate matter (PM2.5 and PM10) and gases like CO, SO2, NO2, and O3. These pollutants are critical for understanding the impact of wildfire smoke on air quality and public health.

Daily AQI records were gathered for wildfire-prone months (May to October) spanning 1964 to 2023. Since air quality can vary significantly across a region, data from multiple monitoring stations in San Joaquin County, where Stockton is located, was aggregated to provide a comprehensive view of local air quality trends. This dataset allowed me to establish connections between wildfire activity and its potential effects on Stockton's air quality.

The graph in Fig.3 shows both the metrics (smoke estimates for 1963 to 2020, and the AQI estimates for 1963 to 2023) revealing several interesting patterns.

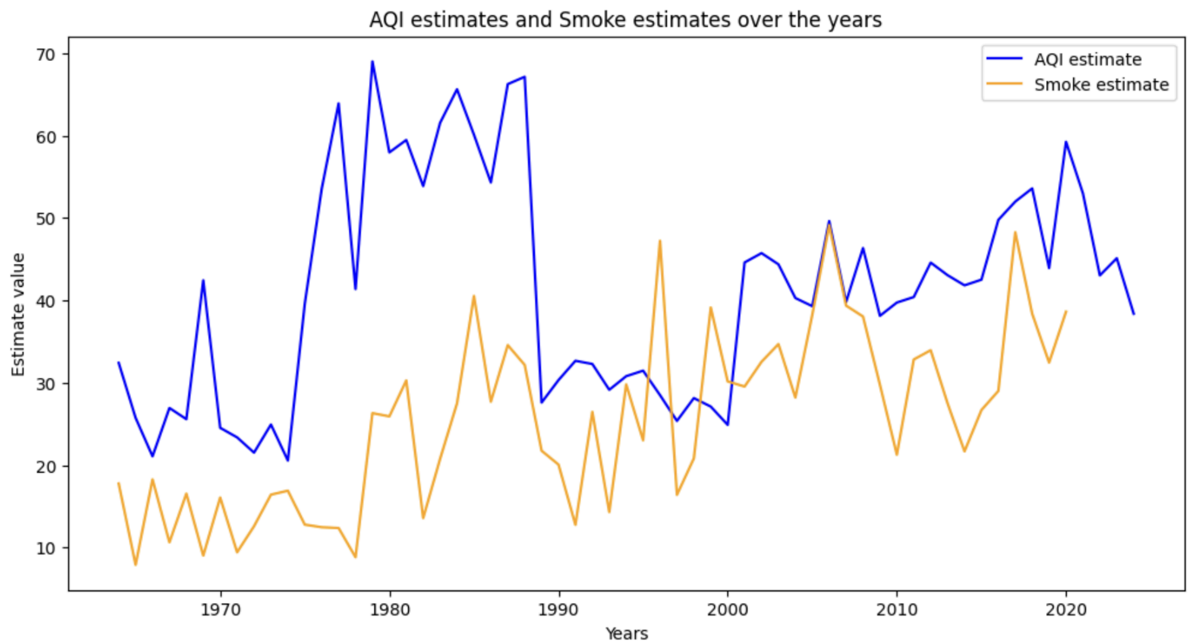


Figure 3: Comparison graph of AQI estimates and Smoke estimates

- A notable shift in AQI values occurs between 1975 and 1990, with values increasing from around 20-30 to 60-70. This dramatic increase likely reflects a combination of factors:
  - Implementation of more sophisticated air quality monitoring systems
  - Changes in industrial activity and urbanization in the Stockton area
  - Improved accuracy in measurement techniques
  - Possible changes in AQI calculation methodology
- After 1990, the relationship between smoke estimates and AQI becomes more apparent, with both metrics showing similar patterns of variation

To validate the smoke estimates derived from the formula, I performed a correlation analysis with the aggregated AQI data from the EPA's Air Quality System. The Pearson correlation coefficient of 0.364 (p-value=0.005) between smoke estimates and AQI values validates the effectiveness of quantifying wildfire smoke's impact on Stockton's air quality. Note that the AQI data for this analysis is restricted to the fire season each year, i.e., May to October.

To understand future wildfire risks around Stockton, I decided to predict smoke levels for upcoming years. I chose the SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors) model for this prediction because it works well with data that shows both seasonal patterns and long-term trends, like wildfire smoke.

The SARIMAX model makes sense for our smoke estimates because:

- It can capture yearly patterns in wildfire activity (seasonal component)
- It accounts for the overall increasing trend in wildfires over decades
- It considers how past years' smoke levels influence future ones

- It can handle the irregular spikes we see in particularly bad fire years

The values predicted for the years 2020-2050 were plotted on a graph, along with the 95% confidence interval for those predictions. Looking at our prediction graph in Fig.4 below, we can see several important patterns:

- The model predicts a continuing upward trend in smoke estimates through 2050
- The predicted values show seasonal variations, similar to historical patterns
- The confidence interval (shown by the shaded area) widens as we look further into the future, reflecting increased uncertainty in long-term predictions
- The model suggests particularly high smoke level fluctuations could occur in the years closer to 2020.

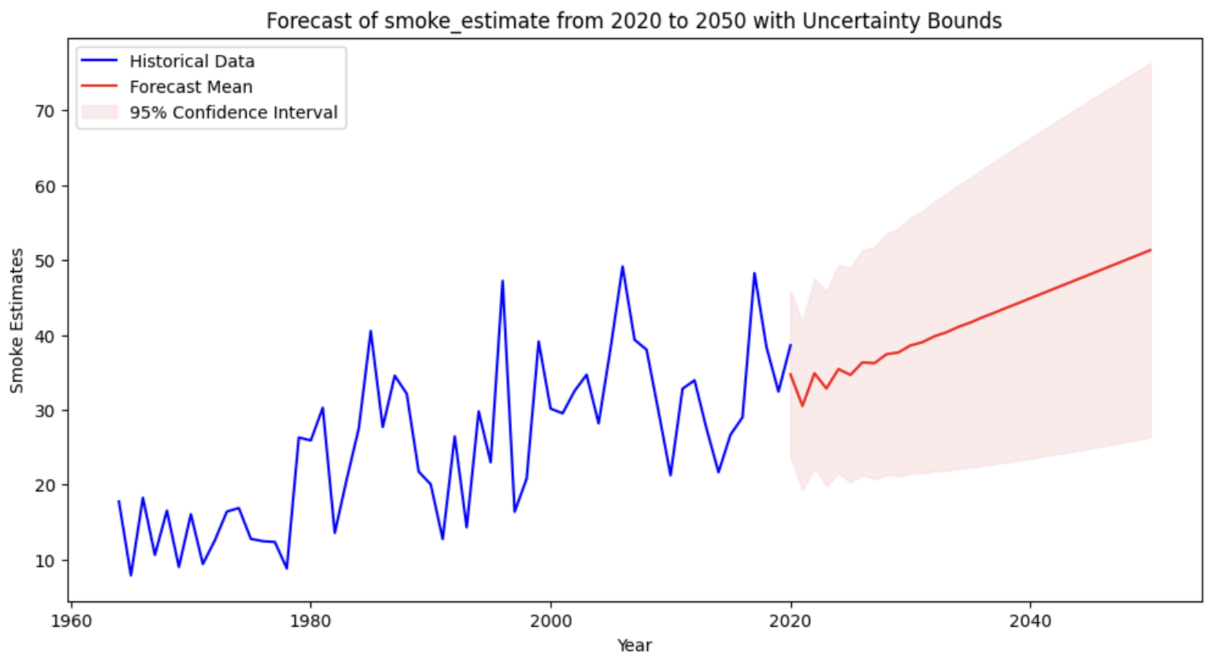


Figure 4: Forecast of Annual Smoke Estimates

The widening confidence interval in our predictions is realistic - it's harder to predict exact smoke levels for years further in the future. It's worth noting that these predictions assume current conditions and trends will continue. Major changes in climate patterns, fire management practices, or land use could affect the actual future smoke levels.

Health Data

To understand how smoke exposure affects people's health, we need to analyze key health data. This study focuses on mortality rates and hospitalization data as these indicators give a clear picture of the community's health challenges. These measures were chosen because they are closely linked to the health problems caused by smoke exposure, helping us explore its impacts on public health in Stockton. Below is the health-related data that I used for my analysis:

1. **Mortality Data**:
    a. Source: Multiple Causes of Death - CDC: The Multiple Cause of Death database provides mortality and population statistics for all U.S. counties, based on death certificates of U.S. residents. Each certificate includes a primary cause of death, up to twenty additional contributing causes, and demographic information.
    b. Link: https://wonder.cdc.gov/mcd.html
    c. Key Columns: Year, Multiple Cause of death, Deaths
    d. Available Data: 1999-2020
2. **Hospitalization Data**:
    a. Source: California Health and Human Services: The California Health and Human Services Agency (CalHHS) has launched its Open Data Portal initiative in order to increase public access to one of the State's most valuable assets – non-confidential health and human services data. The portal offers access to standardized data that can be easily retrieved, searched, analyzed, and re-used by individuals.
    b. Link to dataset: https://data.chhs.ca.gov/dataset/hospital-inpatient-characteristics-by-patient-county-of-residence
    c. Key columns: dsch_yr (Discharge Year), mdc_desc(Major Diagnostic Categories), Discharges
    d. Available Data: 2012-2023

Note: Per my initial extension plan, I intended to use data from the Healthcare Cost and Utilization Project (HCUP) by the Agency for Healthcare Research and Quality (AHRQ) for my analysis. I recently found the data from CalHHS directly provides aggregated data that includes the information from the AHRQ dataset. Hence, I will use the CalHHS data in place of AHRQ data for this analysis.

I decided to use both exploratory analysis (visual analysis) and statistical analysis(correlation analysis) to examine the relationships between smoke exposure and health metrics.

Data Integration and Preparation

For this analysis, I integrated three key datasets to examine the relationship between wildfire smoke and health outcomes:

1. **Mortality Data**: Sourced from the CDC, this dataset provided information about deaths from diseases related to PM2.5 exposure, including respiratory and certain cancer-related conditions.
2. **Hospital Discharge Records**: Obtained from California Health and Human Services (CalHHS), this data contains hospitalization rates for respiratory conditions across the region.

3. **Smoke Estimates**: Calculated in Part 1 of this project, these estimates quantified wildfire smoke exposure within 650 miles of Stockton, offering a localized metric for smoke-related impacts.

Note: To account for the temporal dependency between smoke events and mortality rates, the smoke estimates were shifted by 1 year. This was the duration with best correlation.

The datasets were cleaned and standardized to enable consistent comparisons over the analysis period (1999–2020). Tools like Python's **pandas** and **polars** packages were utilized to handle missing data, normalize formats, and ensure seamless integration. The health metrics of interest included:

- Deaths from specific conditions: asthma, pneumonia, COPD, lung cancer, and colon cancer.
- Total respiratory disease deaths.
- Hospital discharge rates for respiratory conditions.

Visual Analysis

To explore trends in health metrics over time and their potential correlation with wildfire smoke estimates, I created visual representations of the data using Python’s **matplotlib** library. Normalizing the data with the **MinMaxScaler** from **scikit-learn** ensured that metrics with different ranges could be plotted on comparable scales. This visual analysis provided an intuitive way to observe temporal changes and identify patterns that warranted further investigation.

Correlation Analysis

To quantify relationships between smoke exposure and health outcomes, I calculated **Pearson correlation coefficients**. This statistical measure helped assess the strength and direction of associations between smoke estimates and health metrics. The analysis was performed using Python’s **scipy** library, offering insights into how wildfire smoke might influence mortality and hospitalization trends.

Time Series Analysis

To predict future health impacts, I implemented the **SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors)** model. This choice was guided by the model's ability to:

- Account for seasonal variations in both smoke exposure and health outcomes.
- Integrate the relationship between smoke estimates and health metrics as external influencing factors.
- Handle the time-dependent nature of health impacts, capturing both short-term fluctuations and long-term trends.

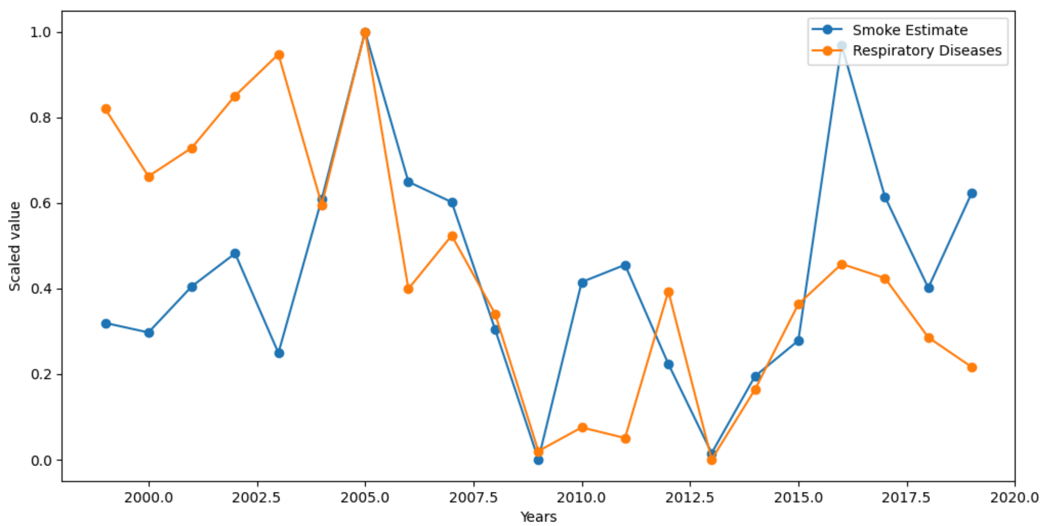
Using Python's **statsmodels** package, the **SARIMAX** model allowed us to project health outcomes under various smoke exposure scenarios, offering valuable predictions for public health planning and interventions.

Findings

Visual Analysis

The graphical analysis highlighted several significant patterns in the relationship between wildfire smoke exposure and health outcomes:

1. **Alignment Between Smoke Estimates and Mortality Rates**:



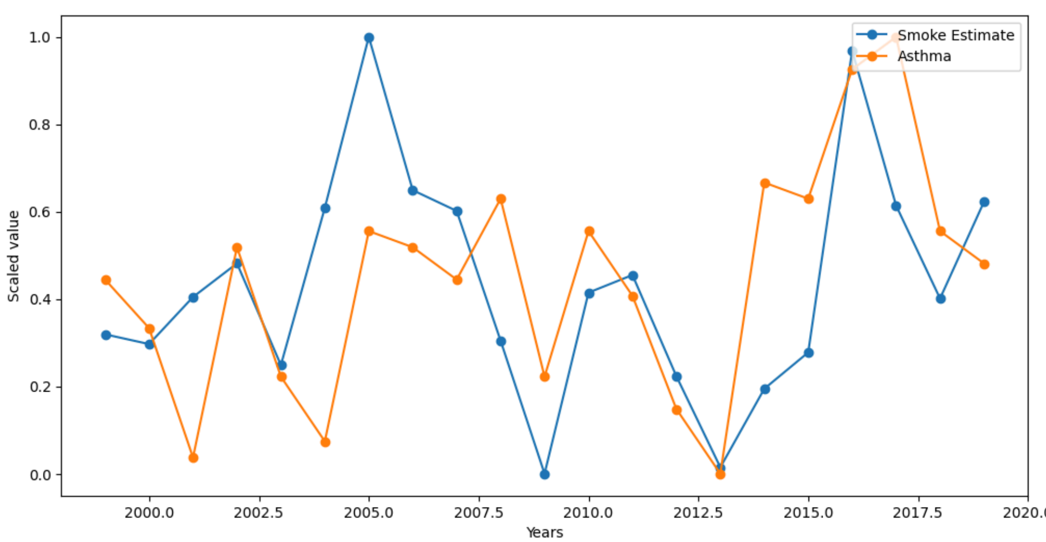


Figure 5: Comparison of Smoke Estimates and Deaths due to Respiratory Diseases(above) and Asthma(below)

Visual overlays of smoke estimates and mortality rates in Fig.5 revealed striking alignments during peak years. Notable spikes were observed in both metrics around the years 2005 and 2016.

2. **Delayed Hospital Discharge Peaks**:

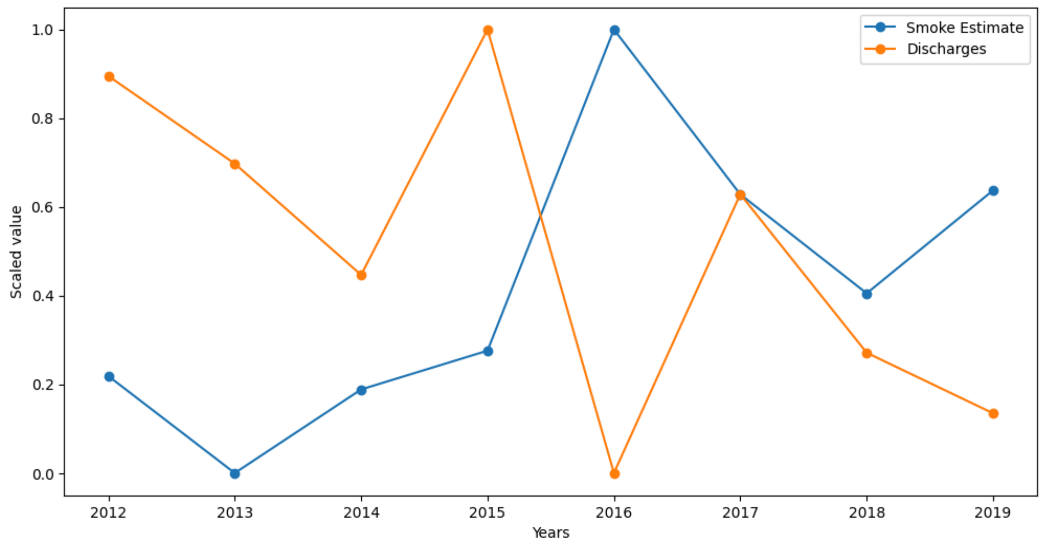


Figure 6: Comparison of Smoke Estimates and Hospital discharges related to Respiratory Diseases

As shown in Fig.6, the hospital discharge rates showed a delayed response to smoke exposure, typically peaking after major wildfire events. This delay suggests that wildfire smoke potentially impacts the hospitalization durations of individuals already admitted.

Correlation Analysis

Pearson correlation coefficients were calculated to assess the strength and direction of the relationships between smoke estimates and various health outcomes. The analysis provided the following insights:

- **Positive Correlation with Respiratory Disease Deaths**: A strong positive correlation (0.39) was observed between smoke estimates and deaths caused by respiratory diseases. This finding aligns with the visual patterns observed over the years, suggesting that as smoke exposure increases, the number of respiratory-related deaths also rises.
- **Negative Correlation with Cancer Deaths**: An unexpected negative correlation (-0.25) was found between smoke estimates and cancer-related deaths. This might warrant further investigation, as the relationship between wildfire smoke and cancer mortality may involve other indirect factors.

In addition, the positive correlation between smoke estimates and various individual respiratory diseases was also noteworthy:

- **Asthma**: 0.49
- **Chronic Obstructive Pulmonary Disease (COPD)**: 0.27
- **Pneumonia**: 0.21

One of the more intriguing findings was the **negative correlation (-0.69)** between smoke estimates and hospital discharges for respiratory diseases. This suggests that during periods of intense wildfire smoke exposure, the number of discharges tends to decrease. Several possible explanations for this include:

1. **Longer Hospital Stays**: Individuals experiencing more severe respiratory symptoms due to smoke exposure may require extended hospitalization, which would result in fewer discharges within a given period.
2. **Healthcare System Strain**: During wildfire events, hospitals may become overwhelmed, prioritizing critical care and delaying routine discharges or admissions for less severe cases.

Further analysis of hospitalization trends, patient demographics, and healthcare resource availability during wildfire seasons is needed to understand the full scope of this complex relationship.

Time Series Analysis

To gain deeper insights into future health impacts, I performed a time series analysis on health metrics closely correlated with smoke estimates, specifically total deaths from respiratory diseases and asthma-related deaths. The analysis used the SARIMAX model to predict future trends:

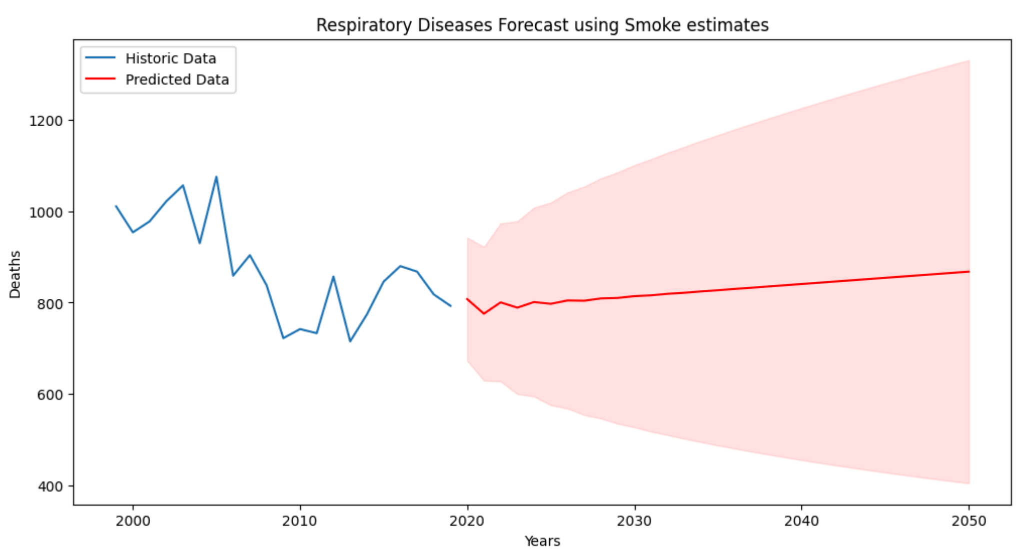


Figure 7: Forecast of deaths due to Respiratory Diseases

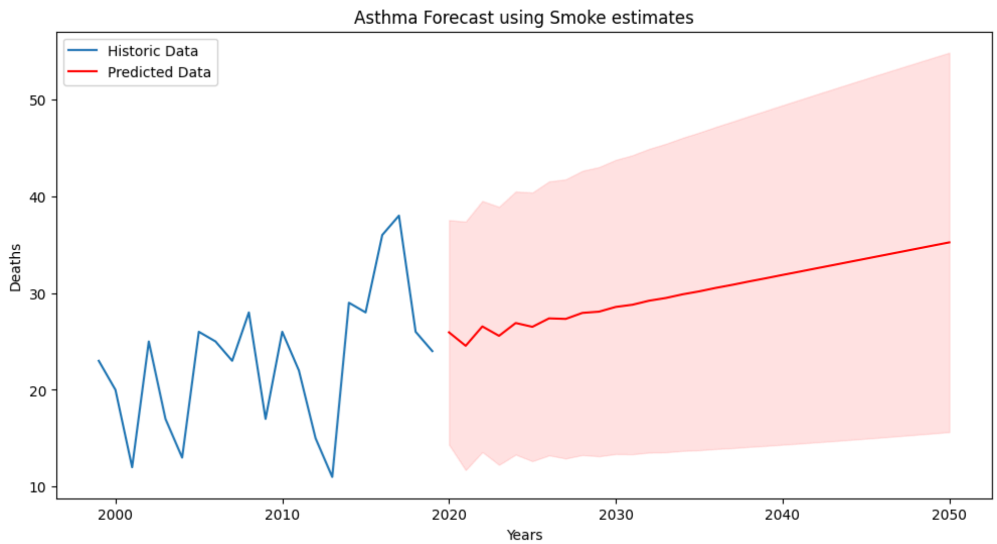


Figure 8: Forecast of deaths due to Asthma

The SARIMAX model offered valuable predictions of future health outcomes under increasing wildfire smoke exposure:

- **Projected Increase in Respiratory Health Impacts**: Predictions from 2020 to 2050 suggest an upward trend in the number of deaths due to respiratory diseases. A steeper curve is observed in predictions of deaths due to asthma.
- **Uncertainty in Long-Term Predictions**: While the widening confidence intervals in future projections reflect uncertainty, the overall trend remains concerning, underscoring the need for targeted public health interventions.

This comprehensive analysis sheds light on the complex relationship between wildfire smoke and health outcomes, offering actionable insights for mitigating future impacts in Stockton and similar regions.

Discussions

A significant portion of Stockton's population earns below the state's average, amplifying the urgency of addressing health concerns linked to wildfire smoke. Many lower-income residents may face increased vulnerability due to limited healthcare access, higher rates of pre-existing respiratory conditions, and financial barriers to relocation or preventive measures during high smoke levels. Additionally, the region's poverty rate—around 15%—underscores the need for targeted interventions that support economically disadvantaged communities, ensuring access to resources and knowledge to protect from smoke-related health risks.

Given these socioeconomic challenges, Stockton’s council must act swiftly. With respiratory health issues, particularly asthma, projected to rise, the city must focus on enhancing healthcare infrastructure and expanding resources for underserved populations in the next 1-2 years.

Immediate actions should include increasing access to respiratory care, telemedicine, and emergency response capabilities. Long-term efforts, such as improving air quality monitoring and investing in smoke-resilient infrastructure, must begin within 5-10 years to reduce the health burden and improve resilience to future wildfire events.

Human-centered data science principles guided this analysis, ensuring that vulnerable residents' needs were prioritized. By integrating health and environmental data, the analysis provided actionable insights for policy interventions, focusing on equity in public health. Prioritizing groups like the elderly, children, and those with pre-existing conditions reflects this understanding while promoting real-time data sharing and community education empowers individuals and healthcare providers to minimize the health impacts of wildfire smoke.

Limitations

Several limitations impacted the analysis, primarily related to missing data and technical challenges. Missing particulate matter (PM2.5) data for certain years and gaps in about 20% of daily AQI records potentially skewed our AQI estimates. To address these gaps, methods such as data imputation could be explored in future work. The wildfire dataset's coverage was another limitation, extending only up to 2020, while the analysis sought data through 2024. Incorporating updated datasets or supplementary sources is necessary to bridge this gap. Additionally, errors in calculating the shortest distance for some wildfires led to the exclusion of a small number of records, which could be mitigated by refining the calculation process.

The methodology also had limitations, particularly in the smoke estimate calculations. The simplified formula used did not account for variables such as wind patterns, fire shape, or atmospheric dispersion, which could affect smoke dispersion and health outcomes. A more advanced model incorporating these factors would improve accuracy. The health data used in this analysis was available only at the county level (San Joaquin County), whereas the focus was on Stockton city, potentially introducing inaccuracies when generalizing county data to the city level. Furthermore, the COVID-19 pandemic likely confounded the analysis in 2020, as the spike in deaths and hospitalizations due to COVID-19 may have obscured or exaggerated the effects of wildfire smoke exposure.

Lastly, while strong correlations were observed between smoke estimates and health outcomes, causality was not established. Further studies with more advanced statistical techniques, such as randomized controlled trials or longitudinal data analysis, would be needed to confirm causal links. The modeling assumptions relied on past trends, which may not account for changing factors such as healthcare access, wildfire prevention, or public health interventions. Data aggregation by year could also mask short-term trends or extreme events, so a more granular analysis, such as seasonal or monthly data, would provide additional insights.

Conclusion

In conclusion, this analysis highlights the significant and growing connection between wildfire smoke exposure and adverse public health outcomes, particularly in Stockton, California. The findings point to a concerning relationship between increased smoke exposure and higher mortality rates from respiratory diseases, with delayed peaks in hospital discharge rates indicating prolonged health effects. Given projections of rising respiratory issues, especially asthma, through 2050, immediate action is needed to address these health risks.

Stockton must prioritize strengthening its public health infrastructure, improving air quality monitoring systems, and enhancing healthcare preparedness. Early warning systems, expanded emergency response capabilities, and increased access to respiratory care and telemedicine will be crucial for mitigating future health impacts. Public health campaigns educating residents about wildfire smoke risks and preventive measures will also play a key role in protecting vulnerable populations. The timeline for action is urgent, with immediate steps required in the next 1-2 years, followed by medium-term investments in infrastructure and resilience.

Despite these promising insights, the analysis does have limitations, including incomplete AQI data, potential biases from the use of county-level mortality data instead of city-specific data, and the confounding effect of the COVID-19 pandemic on health outcomes. These factors should be considered when interpreting the results and formulating public health strategies. Nonetheless, the findings offer a solid foundation for data-driven interventions to mitigate the health impacts of wildfire smoke on Stockton's residents.

References

1. UCLA Study on Wildfire Smoke's Long-Term Health Effects: https://innovation.luskin.ucla.edu/2024/06/18/ucla-study-of-wildfire-smokes-long-term-health-effects-finds-upwards-of-50000-deaths-in-11-years/
2. UC Davis Environmental Health Impacts of Wildfires: https://www.environmentalhealth.ucdavis.edu/wildfires/environmental-health-impacts
3. UC Davis on Wildfire Smoke and Lung Disease Risk: https://health.ucdavis.edu/news/headlines/wildfire-smoke-increases-inflammation-lung-disease-risk/2022/02
4. Health Effects of Fine Particulate Matter (PM2.5): https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm
5. Stockton, CA: Data on Demographics and Economy: https://datausa.io/profile/geo/stockton-ca/
6. U.S. Centers for Disease Control and Prevention: https://www.cdc.gov/
7. SARIMAX official documentation: https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

8. Blog on SARIMAX: https://datascientest.com/en/sarimax-model-what-is-it-how-can-it-be-applied-to-time-series

Data Sources

1. **Wildland Fire Data**
   - **Source**: Combined Wildland Fire Datasets for the United States and Certain Territories, 1800s-Present (USGS)
   - **Link:** https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81
   - **Key Variables**: Fire Name, Fire Start/End Date, Fire Type, Burned Area, Location (Latitude/Longitude)
2. **Air Quality Index (AQI) Data**
   - **Source**: EPA Air Quality Service (AQS) API
   - **Link:** https://aqs.epa.gov/aqsweb/documents/data_api.html
   - **Key Variables**: Date, Station ID, AQI Value, Pollutant, Latitude/Longitude, FIPS Codes (State/County/City)
3. **Multiple Causes of Death - CDC**
   - **Source**: CDC's Multiple Cause of Death Database
   - **Link:** https://wonder.cdc.gov/mcd.html
   - **Key Variables**: Year, County, Cause of Death (ICD-10 Codes)
4. **California Health and Human Services (CalHHS)**
   - **Source**: California Health and Human Services Open Data Portal
   - **Link:**https://data.chhs.ca.gov/dataset/hospital-inpatient-characteristics-by-patient-county-of-residence
   - **Key Variables**: Year, County, Diagnosis/Procedure Codes, Discharge Status, Hospital Type, Race/Ethnicity