Part 4: Final Report
Wildfire Smoke Impact Analysis on Murfreesboro, TN

By: Elizabeth Holden

Introduction

Wildfires are becoming increasingly frequent and intense all across the world. While many think of the fires as being destructive and posing a large threat to communities, fire smoke is just as, or maybe even more, harmful. Due to smoke released from wildfires, the impact of them travels so much further than just the ground they touch. I grew up in Washington state and have had many wildfire seasons filled with smokey skies and bad air quality, most of the time due to fires not even in my city. In fact, almost every year of my high school soccer tryouts, which take place in peak wildfire season, had to be held inside due to harmful air quality from smoke outside. Having experienced the effects of wildfire firsthand, I wanted to dive into the smoke impact of wildfires on the city of Murfreesboro, Tennessee.

Murfreesboro is a city about 40 minutes outside of Nashville, in Rutherford County, and according to its Wikipedia page, it is known as the "geographical center of Tennessee" (*Murfreesboro, Tennessee*). This analysis seeks to explore the health impacts of wildfire smoke, particularly to see if a relationship with chronic respiratory illness deaths exists. By examining these impacts and predicting them for the future, this study aims to provide valuable insights that can guide city officials, healthcare providers, and residents in their efforts to mitigate risks and prepare for the future. To perform this analysis, I used data containing information about wildfires occurring in fire season (May - October) and within 650 miles of Murfreesboro from 1961 to 2021. Then along with chronic respiratory illness data from the Institute for Health Metrics and Evaluation, I predicted the impact of smoke on Murfreesboro till 2050.

Background/Related Work

There's growing concern about the health impacts of wildfires, and more and more studies are coming out that show how wildfire smoke can worsen existing health problems. According to Dr. Carrie Redlich, a doctor at Yale Medicine who specializes in occupational and environmental health and asthma, wildfire smoke particles are small enough to get deep into the lungs and even enter the bloodstream, which can make certain respiratory and heart conditions worse (Yale Medicine, 2023). With this knowledge I was able to conduct analysis to analyze the impact of smoke, which includes chronic respiratory illness.

Data

The data used in this project came from multiple sources. Wildfire data was obtained from the USGS Wildland Fire Combined Dataset, which includes information on the year, size, duration, and location in exact coordinates of wildfires' perimeter. The data was filtered through for fires that occurred in the years 1961 to 2021. The fires were also filtered for ones that were within 650 miles of Murfreesboro, TN. The distance calculations were calculated as the average of the distances between Murfreesboro and every point in the fire ring which is a ring of geospatial coordinates. For all of these calculations, the coordinates have to be converted from the ESRI:102008 coordinate system to the EPSG:4326. Additionally, meters have to be converted to miles.

AQI data was collected from the US EPA's Air Quality System (AQS) API from a monitoring station in Rutherford County. The monitoring station had daily data available for gaseous parameters, SO_2, NO_2, and O_2, from 1988-2012.

Health data on chronic respiratory illness deaths in Tennessee was sourced from IHME's Global Burden of Disease study, covering the years 1990 to 2020. This dataset provides annual mortality rates from diseases such as COPD and asthma, broken out by age and sex. The data is only available at a state level, but was used to examine the trend of deaths as years go on. The dataset includes columns for location, year, age group, gender, cause of death or injury, type of measure (such as deaths), number of deaths (value), and the lower and upper bounds of the confidence interval for the value. For this analysis, I focused solely on the year and value columns, which represent the total number of deaths for each year. This approach was chosen because the analysis did not involve breaking down the data further by age, gender, or specific type of chronic respiratory illness.

Model

Selecting the ARIMA (Auto-Regressive Integrated Moving Average) model for this analysis was an essential decision that required comparing it to other methods. ARIMA is specifically designed for time series data and can make long-term forecasts based on historical patterns (Autoregressive integrated moving average, n.d.). Another model I looked at was a Gradient Boosted Regressor (GBR) but its application to time series forecasting requires a lot of preprocessing of the data. According to Snowflake's engineering blog, GBR requires extensive preprocessing, including the transformation of time series into supervised learning tasks by creating lagged features, trend indicators, and seasonality components. This preprocessing can significantly increase the model's complexity and introduce a higher risk of overfitting, particularly with smaller datasets (Snowflake, 2024). ARIMA seemed like the better option as my data was already set up to be used without many more preprocessing steps. Additionally, later in my analysis, I added chronic respiratory deaths to my model and, after research I learned, there is a variant of ARIMA, named ARIMAX, that has the ability to include external regressors.

Methodology

The first step in this analysis, after obtaining the data as described above, was to analyze how many fires there were within 650 miles of Murfreesboro and with these fires, see how many acres were burned per year. The acres of wildfires during fire season, May 1 to October 31, within a 650-mile radius from Murfreesboro, were summed up by year to get the total acres burned for each year.

The next step of the analysis was to define an estimate of the impact that the smoke has on Murfreesboro each year using this data. In order to define this, I wanted an equation that accounted for a larger estimate when more acres are burned, but also to account for how smoke disperses with distance. After researching, I found the Inverse Square Law which, according to Wikipedia, is any scientific law stating that the observed "intensity" of a specified physical quantity is inversely proportional to the square of the distance from the source of that physical quantity” (Inverse-square law). Using this I defined my smoke impact estimate as follows:

$$\text{Smoke Impact} = \text{intensity} \times \left(\frac{1}{(\text{distance} + 1)^2} \right)$$

The intensity is the acres burned, and distance is the distance of the fire from Murfreesboro. I added 1 to the distance to account for the edge case where distance equals zero. For each fire meeting criteria (occurring between 1961-2021 and within 650 miles of Murfreesboro), the smoke impact was calculated, then summed by year to produce an annual smoke impact estimate.

In order to assess how good of an estimate this smoke impact is, I compared it to the AQI data collected from the EPA monitoring station in Rutherford county. The AQI estimate represents an annual average for Murfreesboro, TN, based on daily data from a single monitoring station capturing SO_2, NO_2, and O_2 levels from 1988-2012. To calculate the AQI for each day I used the following calculation which is based on the “Technical assistance document for the reporting of daily air quality: The air quality index (AQI)" document from the EPA (U.S. Environmental Protection Agency, 2024). For each day, the maximum AQI across all available pollutants was selected, and an annual AQI average was calculated.

After computing and analyzing the smoke impact estimates for 1961 to 2021, the next step was to predict them till 2050. In order to do this, I used the time series model ARIMA. ARIMA has three main parameters, p, d, q, so I used cross validation to find the best parameters, which were the ones where the model had the lowest root mean square error (Statsmodels Documentation, 2024)

I wanted to take this analysis further by including health data to see how the smoke may be affecting the people in Murfreesboro, TN over time. Specifically, I added the IHME Global Burden of disease data on deaths from chronic respiratory illnesses, which could be related to smoke exposure. The goal of this extension is to model how smoke exposure may connect to death rates from respiratory illnesses and to use this model to predict how health impacts from smoke could look in the future, up to 2050.

In order to do this, I plotted the available and cleaned IHME chronic respiratory death data to see the trend for Tennessee. I also computed the Pearson correlation coefficient to see if the number of deaths from chronic respiratory illness and smoke impact values were linked and followed the same trend.

In order to add the number of deaths from chronic respiratory illness to the ARIMA model as an external regressor, and use ARIMAX, I needed the values for chronic respiratory illness deaths for every year that I wanted to predict my smoke impact estimate for. Since the deaths follow a fairly linear pattern trend, I used linear regression to predict the values for the deaths till 2050.

With values for chronic respiratory illness deaths from 1990 to 2050, I then was able to use the ARIMAX model. The chronic respiratory illness deaths were used as an external regressor and the model predicted the smoke impact through 2050, with a 95% confidence interval. As with the first ARIMA model, I used cross validation to find the best parameters.

Ethical considerations

When conducting this analysis, I kept reproducibility at the forefront. The code is well documented, as well as decisions that were made. Reproducibility was a large reason why I chose to use the ARIMA model as it is a standard model for time series analysis, making it accessible for people to find resources on and reproduce. Additionally, since I was working with public health data, I used a reliable source, IHME, that takes many steps to ensure their research is trusted and keeps those whose data is involved safe.

Findings

The following graph is a histogram of the number of wildfires within 1800 miles of Murfreesboro, TN from 1961- 2021. The data in blue are the fires that are within 650 miles of the city and are the fires that were used for my predictive model.

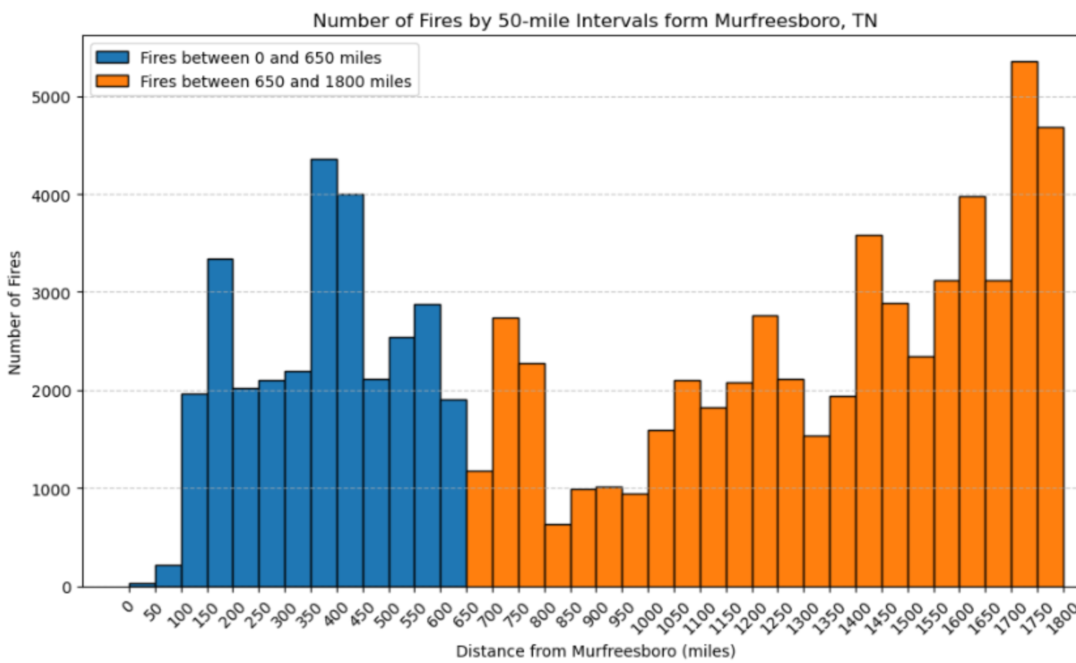

Figure 1: A histogram of the number of wildfires within 1800 miles of Murfreesboro, TN from 1961- 2021. The fires are binned in intervals of 50 miles for the histogram. The y-axis represents the number of fires, and the x-axis represents the distance of the fire in miles from Murfreesboro. The data in blue are the fires that are within 650 miles of the city and are the fires that were used for my predictive model. The data in orange are fires that are between 650 and 1800 miles from the city.

Due to Murfreesboro's central location in the state and Tennessee being in the Southeastern region of the United States, there are many fires beyond the 650-mile radius this analysis used.

The following time series graph shows the total number of acres burned annually during fire season, May 1 to October 31, within a 650-mile radius from Murfreesboro, TN.

For each fire meeting these criterion, the acres of all fires for each year were summed up to get the total acres burned for each year.

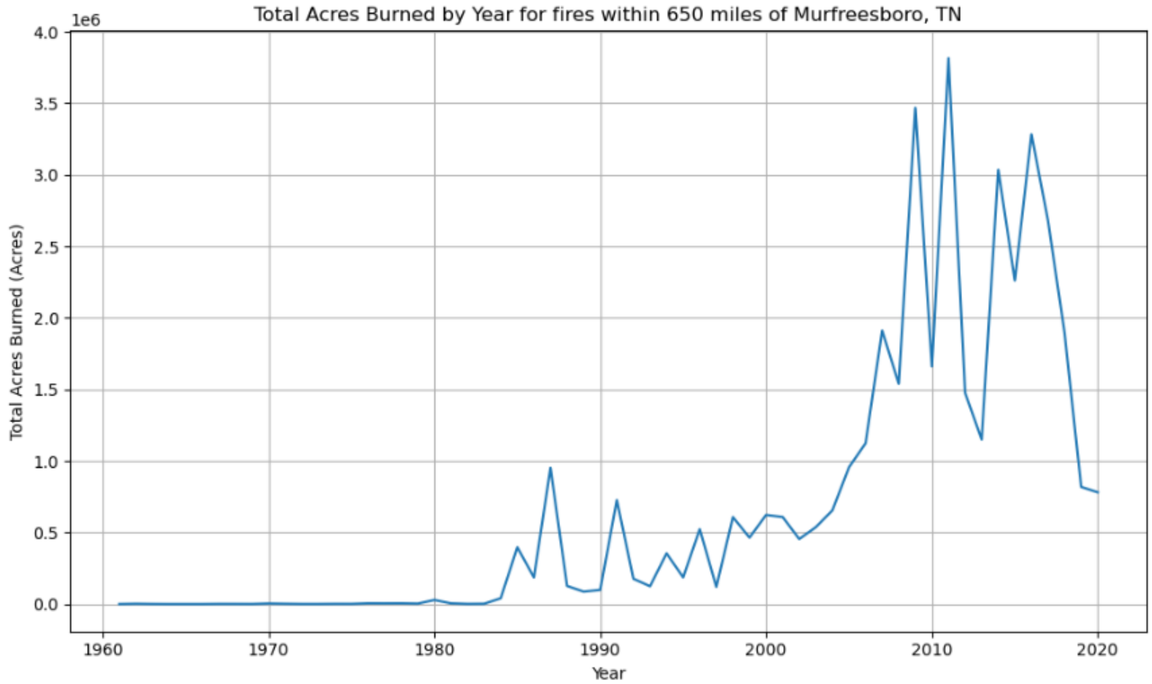

Figure 2 A time series graph shows the total number of acres burned annually during fire season, May 1 to October 31, within a 650-mile radius from Murfreesboro, TN. The x-axis represents the year from 1961 to 2021, while the y-axis shows the total acres burned for each year. Each data point on the line represents the total acres burned across all fires for that year. The line shows the overall trend of the total acres burned by wildfires from 1961 to 2021.

We can see that as the years go on, the total number of acres burned generally increases, peaking around 2012. These large peaks can be attributed to large well-known fires. Additionally, the increase in acres burned may be slightly attributed to the better tracking and record keeping of fires in recent years.

The following graph shows the smoke impact estimates from 1961 to 2021:

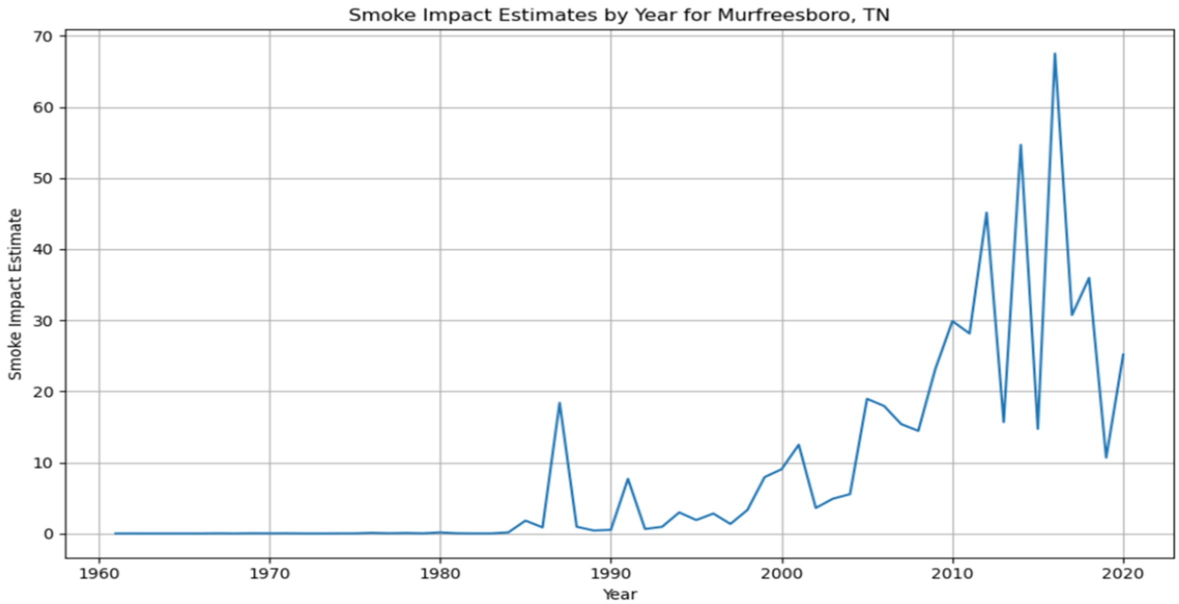

Figure 3 The x-axis represents time, where each point is a year between 1961 and 2021, while the y-axis shows the smoke impact estimate indicating levels of smoke intensity. The graph's trend line allows users to observe changes over time.

From analyzing this graph, we can see that the estimates generally are increasing as years progress, except for a slight recent decrease. This graph is very similar to the graph showing the acres burned per year, due to the acres burned (intensity) being a large part of the smoke impact calculation.

The following time series graph presents wildfire smoke impact estimates and AQI estimates for Murfreesboro, TN, from 1961 to 2021.

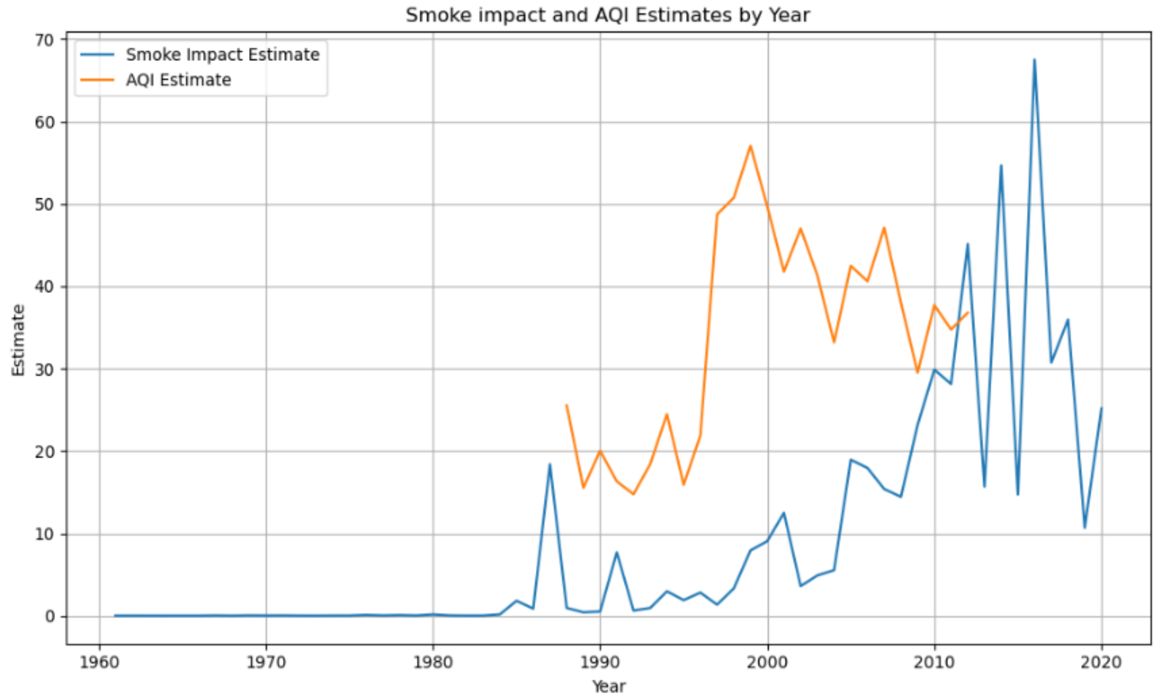

Figure 4 The x-axis represents time, where each point is a year between 1961 and 2021, while the y-axis shows the smoke impact estimate and AQI values, indicating levels of smoke intensity and air quality, respectively. The graph's two trend lines allow users to observe changes over time, as well as to compare the two and see the relationship between smoke impact estimates and AQI. The blue line is the smoke impact estimate, and the orange line is the AQI estimate.

While both trends show a slight upward movement over the years, they differ in their peaks and dips, suggesting that other factors also contribute to the estimates. The computed correlation of 0.28 suggests a weak positive relationship. While smoke may influence AQI, the metrics differ in how they were computed and the components of them. Oxygen is not really impacted by wildfire smoke, but sulfur levels can be if the fire is burning vegetation with a lot of sulfur. NO2 however, is a part of what makes up wildfire smoke, so these levels should be directly correlated with wildfire smoke levels. If I had a monitoring station that had particulate values like PM2.5, my AQI estimate would be a better detection of wildfire smoke.

The following graph is the smoke calculated smoke estimates from 1961 to 2021 and then predicted estimates through 2050, as well as a 95% confidence interval for the predictions.

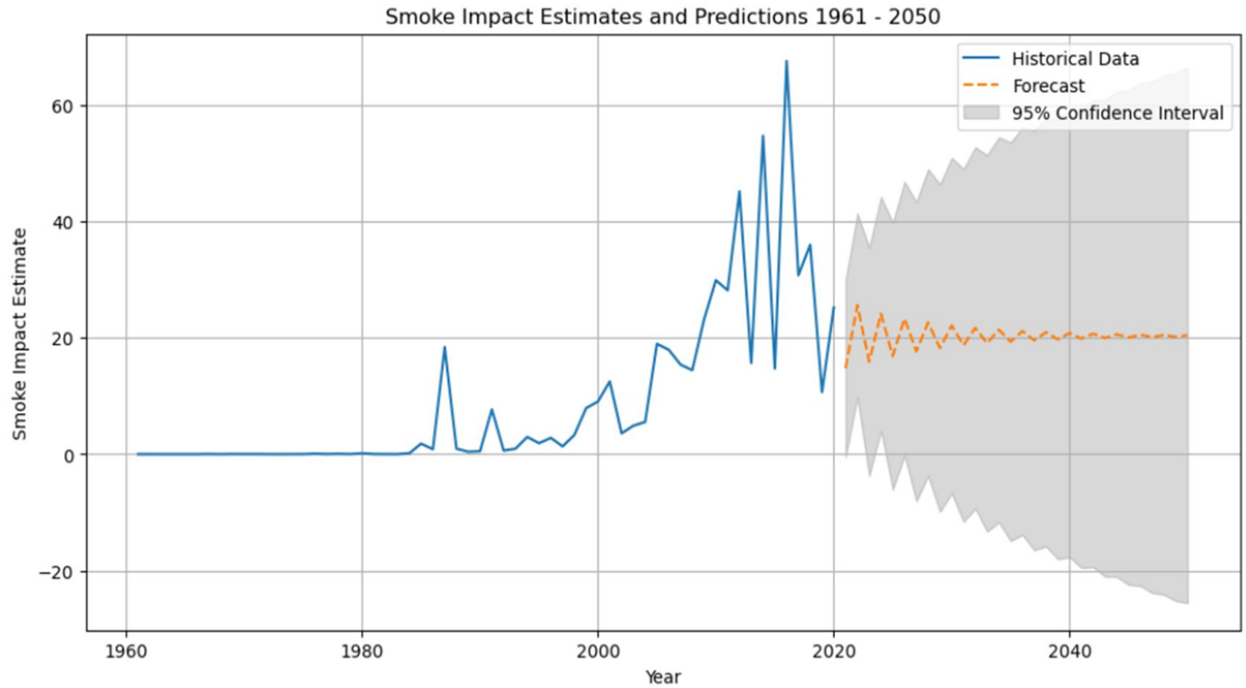

Figure 5 The x-axis represents time, where each point is a year between 1961 and 2050, while the y-axis shows the smoke impact estimate, indicating levels of smoke intensity. The blue line is the calculated smoke impact estimate, and the orange line is the forecasted values from the model. The grey shaded area represents the 95% confidence interval for the model predictions.

From this graph, you can see that the predictions are pretty constant around 20. The 95% confidence interval is very large and goes well below zero, which is not a possible value, for the smoke impact estimate. From this we can see that the model is not very good and will not give us much information about the impact of smoke into the future.

To extend the analysis, IHME number of deaths by chronic respiratory illness data was explored and added.

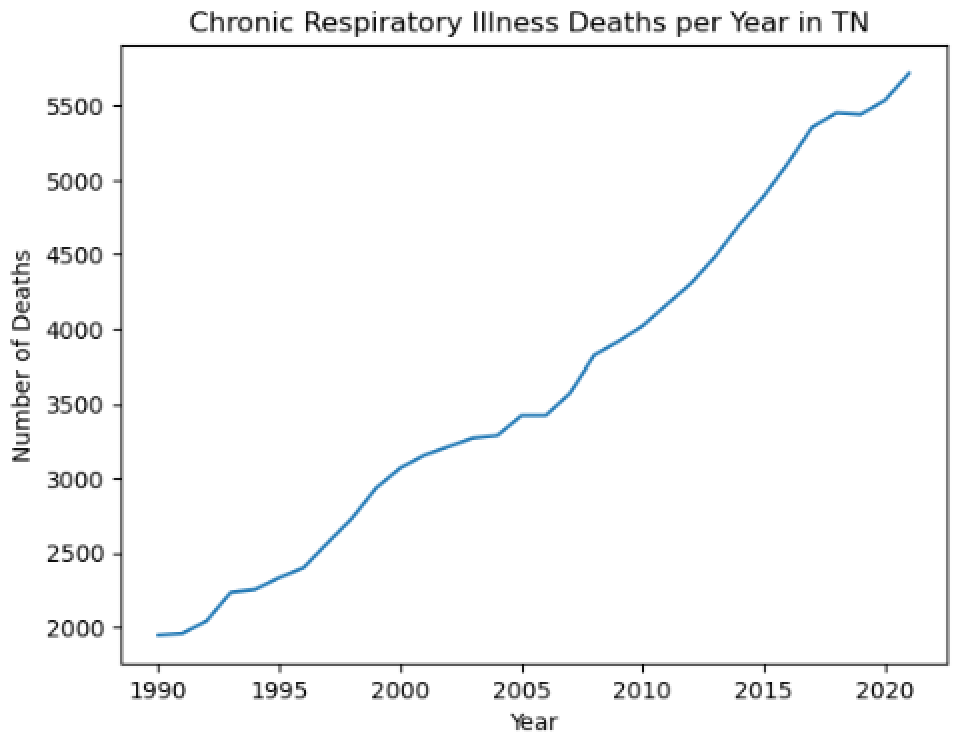

Figure 6 The x-axis represents time, where each point is a year between 1990 and 2020, while the y-axis shows the number of deaths from chronic respiratory illness. The graph's trend line allows users to observe changes over time.

From this graph we can see that the number of deaths is increasing as years go on, in a fairly linear way. To analyze if there was a correlation between the number of chronic respiratory illness deaths and smoke impact estimates, I computed a Peason correlation coefficient and found that they are positively correlated due to the coefficient being about 0.7. This indicates that as smoke impact scores increase, there is a corresponding rise in respiratory deaths. This finding is important as it provides evidence of the link between environmental factors and health outcomes, emphasizing the need for targeted interventions to mitigate these risks.

IHME number of deaths by chronic respiratory illness data was predicted till 2050 using a linear regression model. The following graph shows the results of this model:

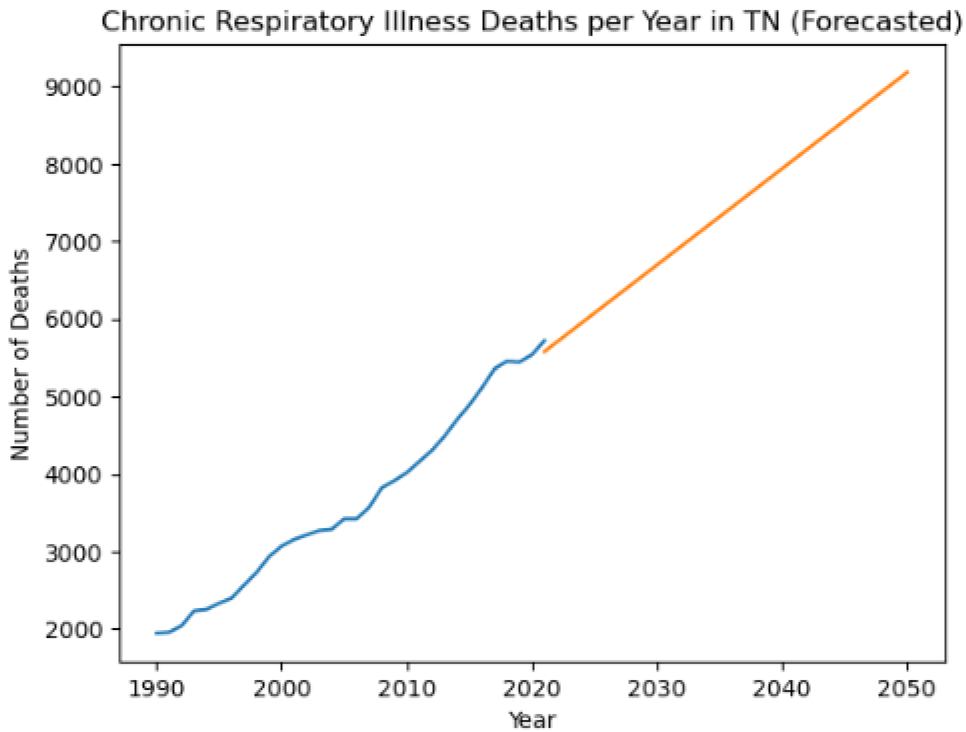

Figure 7 The x-axis represents time, where each point is a year between 1990 and 2050, while the y-axis shows the number of deaths from chronic respiratory illness. The graph's trend line allows users to observe changes over time. The blue line is the historical values obtained from the dataset and the orange line represents the predicted values.

The IHME number of deaths by chronic respiratory illness data and predictions were added to the smoke impact model and the following graph shows the model predictions:

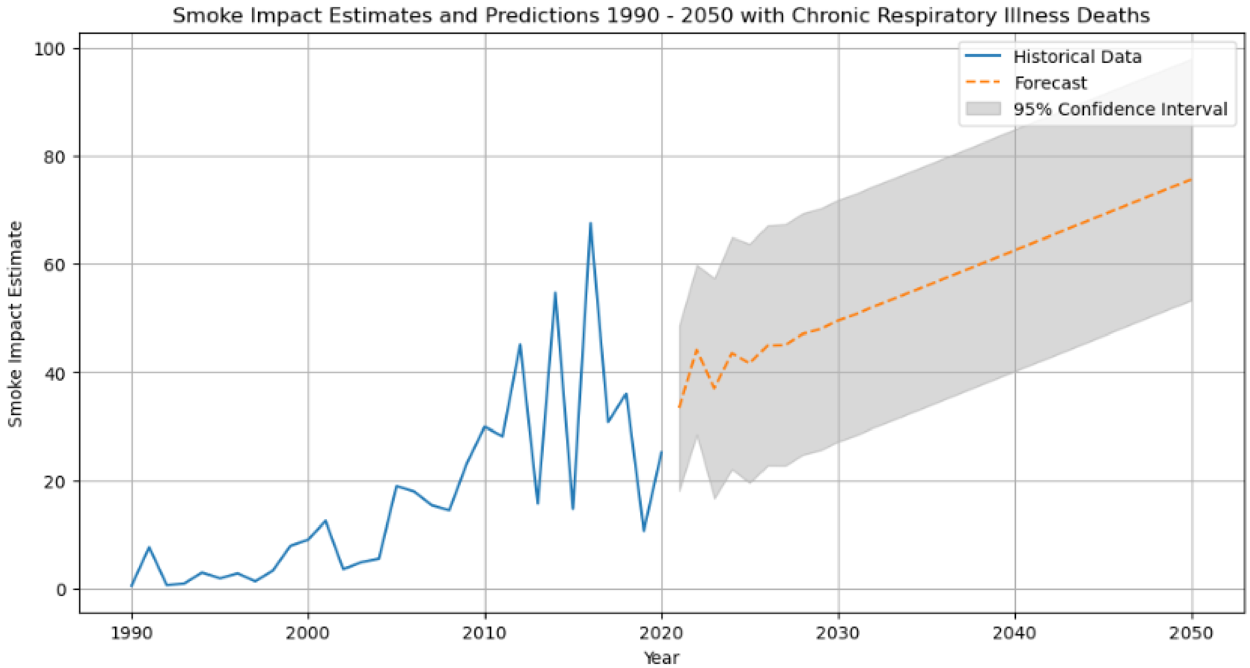

Figure 8 The x-axis represents time, where each point is a year between 1990 and 2050, while the y-axis shows the smoke impact estimate. The graph's trend line allows users to observe changes over time. The blue line is the historical values obtained from the dataset and the orange line represents the forecasted values. The grey shaded region represents the 95% confidence interval for the model's predictions.

From these results we can see that the smoke impact with chronic respiratory illness deaths as a factor, is increasing in the future. These predictions have a much smaller and consistent confidence interval as time progresses. We can see that the lower bound of the confidence interval begins around where the smoke estimate is for the last year of historical data. These results show that the impact of smoke will rise and continue to rise, so Murfreesboro must be proactive and implement policies and educational programs to combat this.

Discussion / Implications

The findings of this analysis show the urgent need for action to mitigate the health impacts of wildfire smoke. The link shown between chronic respiratory illness and wildfire smoke by the correlation shows the threat to public health as wildfires continue to become more frequent and intense. This is confirmed by Figure 8 showing the smoke impact to increase at a steep rate beginning immediately. Not only will this cause a lot of lost lives but it will also put the healthcare system at risk. The more people with chronic respiratory illnesses who are towards the end of their lives, it only increases the number of resources and space that health care facilities need.

By focusing on actionable insights, the analysis aims to support informed decision-making in this time of increasing wildfire risks. City officials should prioritize public health initiatives, such as distributing air filters and issuing emergency smoke alerts during wildfire seasons. The more the city officials can educate their population about the risk and how to avoid them, the more people can take serious action. During high smoke days, people can take actions such as using air purifiers and limiting outdoor activities on smoky days. Healthcare providers must prepare for an increase in respiratory-related hospital admissions, particularly from those who have chronic respiratory illnesses.

Limitations

Despite its findings, this study had several limitations. There were several unknowns and dependencies that impacted the reliability of this analysis. One major challenge was that time series forecasting, especially over a long period like 25 years, was difficult to get right. Predicting this far into the future assumed that trends from the past would continue, but in reality, things could change due to climate change, new wildfire management strategies, or health advancements. This could lead to more error over time, especially since wildfire activity and smoke impact can vary so much from year to year. Additionally, the smoke impact scores used were estimates based on fire acreage and distance, so they were not exact measurements of the smoke Murfreesboro experienced. AQI is heavily determined by particle matter and wind, both of which I did not have access to.

Another unknown was the chronic respiratory illness death data itself. Since an external regressor was used to add the death rates to the existing model, historical and future values were required to predict the impact estimates. Assumptions had to be made for future death rates, which could introduce more uncertainty, especially if healthcare or policies changed significantly over the next few decades. The assumptions were that rates were based on historical trends and modeled predictions. The death data covered only the years from 1990 to 2020, while the smoke impact data spanned back to 1961. This mismatch meant that only overlapping years were used, resulting in a smaller subset of historical trends to base the future predictions on.

There was also a limitation in the way the mortality data was recorded. The dataset tracked all deaths from chronic respiratory illnesses, not just those that may have been related to wildfire smoke exposure. Chronic respiratory diseases could be caused by many factors, such as genetics, smoking, or other environmental pollutants, making it challenging to pinpoint how much wildfire smoke specifically contributed to these deaths. This means that any conclusions drawn from this analysis should be interpreted with caution, as other causes could be influencing the results. Additionally, the mortality data was for the entire state of Tennessee, but this analysis focused on Murfreesboro. Statewide data might not fully capture what was happening in Murfreesboro specifically, as there could be differences in population health, healthcare access, or air quality. However, given the wide range of wildfires included in the data, it was assumed that statewide mortality data was an acceptable trend.

Conclusion

This study highlights the growing threat of wildfire smoke to public health in Murfreesboro, Tennessee, by combining wildfire, AQI, and health data to provide a comprehensive analysis. The findings show a concerning trend which is that wildfires have worsened over the past six decades, with predictions suggesting it will continue to increase alongside the associated health impacts. By incorporating chronic respiratory illness data, this project goes beyond air quality metrics to focus on human centered health outcomes.

One of the most significant findings was the strong correlation of 0.7 between smoke impact estimates and respiratory deaths, illustrating the potential health burden of wildfire smoke. This highlights the need for Murfreesboro's city officials to prioritize public health measures, such as issuing smoke alerts and providing resources like air purifiers to vulnerable populations. Healthcare providers should also prepare for an increase in respiratory-related conditions during wildfire seasons, as trends suggest these issues will intensify. This analysis is just the beginning of what hopefully will be a much larger effort to keep the residents of Murfreesboro safe and healthy.

References

Environmental Protection Agency. (2024, May). Technical assistance document for the reporting of daily air quality: The air quality index (AQI). Retrieved December 4, 2024, from https://document.airnow.gov/technical-assistance-document-for-the-reporting-of-daily-air-quailty.pdf

Institute for Health Metrics and Evaluation. (n.d.). Global burden of disease (GBD). IHME. Retrieved November 4, 2024, from https://www.healthdata.org/research-analysis/gbd

Snowflake Engineering. (2024). Time Series Forecasting: Comparing Transform Techniques and Tree-Based Models. Retrieved from https://www.snowflake.com/engineering-blog/time-series-forecasting-comparing-transform-techniques-tree-based-models/

Statsmodels Documentation. (2024, Dec). statsmodels.tsa.arima.model.ARIMA. Retrieved December 4, 2024, from https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html

VizHub- GBD Compare, n.d. GBD compare | Viz hub. IHME. Retrieved November 4, 2024, from https://vizhub.healthdata.org/gbd-compare/

Wikipedia contributors. (n.d.). Murfreesboro, Tennessee. In Wikipedia. Retrieved December 4, 2024, from https://en.wikipedia.org/wiki/Murfreesboro,_Tennessee

Wikipedia contributors. (n.d.). Inverse-square law. In Wikipedia. Retrieved December 4, 2024, from https://en.wikipedia.org/wiki/Inverse-square_law

Wikipedia contributors. (n.d.). Autoregressive integrated moving average. In Wikipedia. Retrieved December 4, 2024, from https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

Yale Medicine. (2023). How bad is wildfire smoke for your health? Yale Medicine. Retrieved from https://www.yalemedicine.org/news/how-bad-is-wildfire-smoke-for-your-health

Data Sources

Wildfire data: The wildfire data comes from: https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81 . The file is too large to upload to the repository, but the file used in analysis was the USGS_Wildland_Fire_Combined_Dataset.json file from the GeoJSON download.

AQI data: The AQI data came from the US EPA Air Quality System (AQS) API.

Chronic Respiratory Illness Death data: The chronic respiratory illness death data comes from IHME's Global Burden of Disease Study. The link to access this data is https://vizhub.healthdata.org/gbd-compare/ . You must use the visualization tool to filter for the state of Tennessee, all age groups, both genders, the death metric, and for the disease.