Marques Chacon
Professor McDonald
DATA 512
11 December 2023

Final Report

Introduction

The western US has experienced a dramatic increase in the number of wildfires in recent decades. These fires have been linked to chronic health issues, especially when it comes to breathing problems due to smoke. However, wildfires also have an overlooked impact – local communities are suffering from the effects of wildfire smoke in unforeseen ways, such as loss in sales for a city's primary industries, school cancellations due to wildfire smoke, less tourism, and disproportionately affecting unhoused people. Nowhere are these effects more apparent than in the city of Idaho Falls, ID, which sits at the confluence of many different wildfire events each year. This report will detail my analysis of the effects of wildfire smoke on the city over the past 60 years, and then project the future impacts, specifically on flight traffic, along with explaining the ramifications, and urge city officials to take action and prepare for these looming effects. While most research focuses on climatic and health impacts due to wildfires, very little has focused on the effects of wildfire smoke and flight patterns. This analysis will illustrate the importance of addressing tourism-related impacts (on Idaho Falls) caused by wildfire smoke, so that this city can sustain its local tourism industry and minimize its labor displacement.

Background

One of the most obvious impacts of wildfire smoke includes its effects on health, particularly when it comes to breathing issues. When I first ventured into this project, I thought about analyzing how wildfire smoke can lead to worse health outcomes and the development of COPD, asthma, and other lung conditions. After all, there has been much research on how smoke from fires can irreversibly damage the lungs, so I was considering if there had been any research done on the long term effects of wildfire smoke when it comes to the residents of cities that are annually impacted. To my surprise, this evidence has been sparse, as most studies focus on effects 1-3 years after the wildfire event, and most of them are focused on mental health impacts. Longer term effects are harder to identify, since much of the health effects of wildfires are immediate. I attempted to look into surveys from the CDC on the prevalence of COPD and other breathing-related issues in areas like Idaho Falls, which receives an ever-increasing amount of smoke from wildfires each year. However, the data was sparse and tenuous at best, so I shifted my analysis elsewhere.

As mentioned in the introduction, very little research has been devoted to analyzing the effects of wildfire smoke on flight traffic. With that being said, people have looked into the impacts that wildfires have on tourism in general. According to an article from Jane Marsh titled "The Impacts of Wildfires on Tourism", wildfires have reduced annual tourism income by 11% in California, and have been responsible for destroying many tourist attractions which generate revenue for local communities. Furthermore, they cause hotel rooms to become more expensive, which deters many tourists from visiting. Using this knowledge as the foundation for my analysis, I looked into flight data, as the Department of Transportation Statistics captures a ton of metrics for all flights in the United States. This includes metrics which I believe are relevant to wildfire smoke, including total number of passengers flying into an affected city, and the average arrival delay for flights into the city. Specifically, I think that the increase in wildfire smoke in Idaho Falls will reduce the annual number of passengers flying into the city, as well as cause longer arrival delays, since smoke can reduce visibility, making flights more dangerous, and deter more people from visiting the city. Furthermore, since wildfire smoke causes various breathing problems, tourists will view the city as more hazardous, leading to a negative stigma on the city's image and less people willing to risk their health to visit the city.

Methodology

To conduct this analysis, I used different methods of time-series forecasting to predict various measures. Specifically, I used an Autoregressive Integrated Moving Average, or ARIMA, to estimate the annual smoke impact on Idaho Falls in future years, because we have already witnessed an increase in the amount of smoke from wildfires. ARIMA tends to have stable predictions since it uses moving averages based on a certain number of lags (or past data values) which inform a particular prediction. If the effect of smoke is projected to grow, then we should expect any health or economic impact to worsen as well. Therefore, it's important to project how much smoke the city is expected to receive in the coming years.

In addition to ARIMA, I used a Vector Autoregressive (VAR) Model to predict the effects that wildfire smoke has on the number of passengers and the average flight delays into a city. I decided not to use ARIMA on these measures because I believe that there are confounding factors when it comes to flight traffic into a city. ARIMA tends to assume that there are no known predictor variables at play, and while that may be more true when it comes to the amount of wildfires in general, I believe it to be less true when it comes to flight traffic, since I believe those measures to be directly impacted by smoke. Furthermore, VAR not only takes these interactions into account, but it can also predict future values much like ARIMA. Thus, I am not predicting these values independently of one another – I am considering how trends for each of these variables will impact the trajectories of one another as well. In other words, using this model, we can attribute any patterns in future flight trends to trends in wildfire smoke as well, as opposed to mere coincidence. This will help us identify what steps to take in light of the findings we uncover from the flight data.

Findings

Using data from the USGS, which houses information on individual wildfires since the 1800s, I first calculated an estimate that represents the impact of smoke on Idaho Falls, by taking into consideration the size of the fire, along with the average distance to each of the fire's perimeter coordinates. Then, after calculating the total value by year, I plotted these values onto a graph to see the trends that have been occurring since 1963.

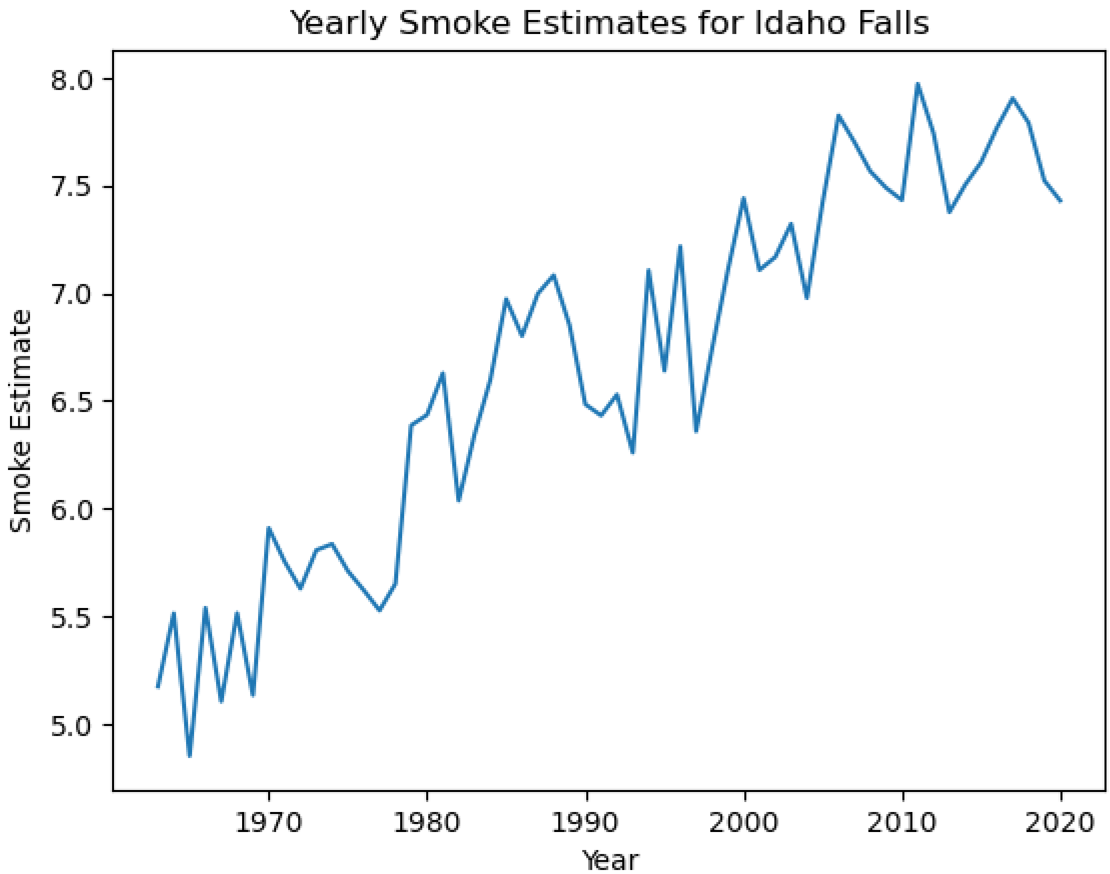

Figure 1: Estimated Smoke Impacts on Idaho Falls since 1963

As we can see from the graph above, we see a very clear trend. The annual smoke impact from wildfires around Idaho Falls has steadily increased since 1963, increasing from a value of 5.0 in 1963 to a value of 8.0 by the 2010s. This suggests that there are worsening effects on the city from wildfires over the years. However, interestingly, we also notice that the trend appears to be plateauing since 2010. Does this indicate that wildfire smoke could potentially be reaching a maximum? It's possible, but nevertheless there is quite a bit of variability from year to year, so it's probable that Idaho Falls could experience unforeseen effects in future years. How likely this will occur depends on our ARIMA forecast, which can be seen in Figure 2 below:

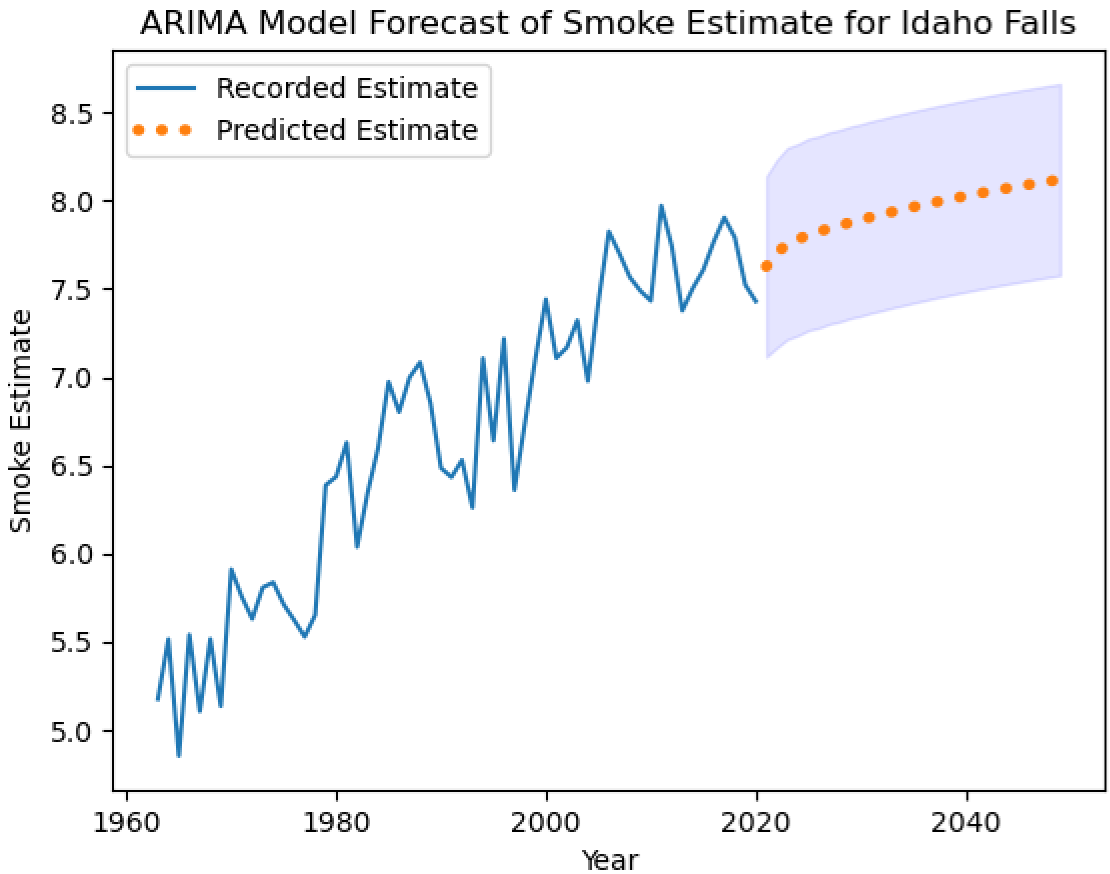

Figure 2: Predicted Smoke Impact on Idaho Falls in Future Years up to 2050

From this forecast, we see that the smoke impact on Idaho Falls is projected to increase for the foreseeable future. Furthermore, the shading around our predicted estimate indicates the level of uncertainty we have on our prediction. Thus, by 2050, we can expect a minimum smoke estimate of around 7.5, which is the same levels experienced in the 2010s. However, for that same year, more than half of the possible range in our prediction interval is above 8.0, which means that if things continue on this trend, then we will almost certainly experience uncharted territory when it comes to the devastation of wildfires and their associated smoke impact.

After analyzing the smoke data from the USGS, I decided to incorporate some external data to supplement this analysis. As mentioned in the background section, I gathered data from the Department of Transportation Statistics, specifically looking at arrival delays and number of passengers, in an attempt to find any correlations between my smoke estimate and those measures. Before considering each of the variables together, I first looked at any trends that might be occurring regardless of any external factors. To calculate the average flight delay, I considered only flights that occurred during the wildfire season, as flights outside of it are more likely to not be as affected by wildfires, and that could skew the results. These findings are captured below:

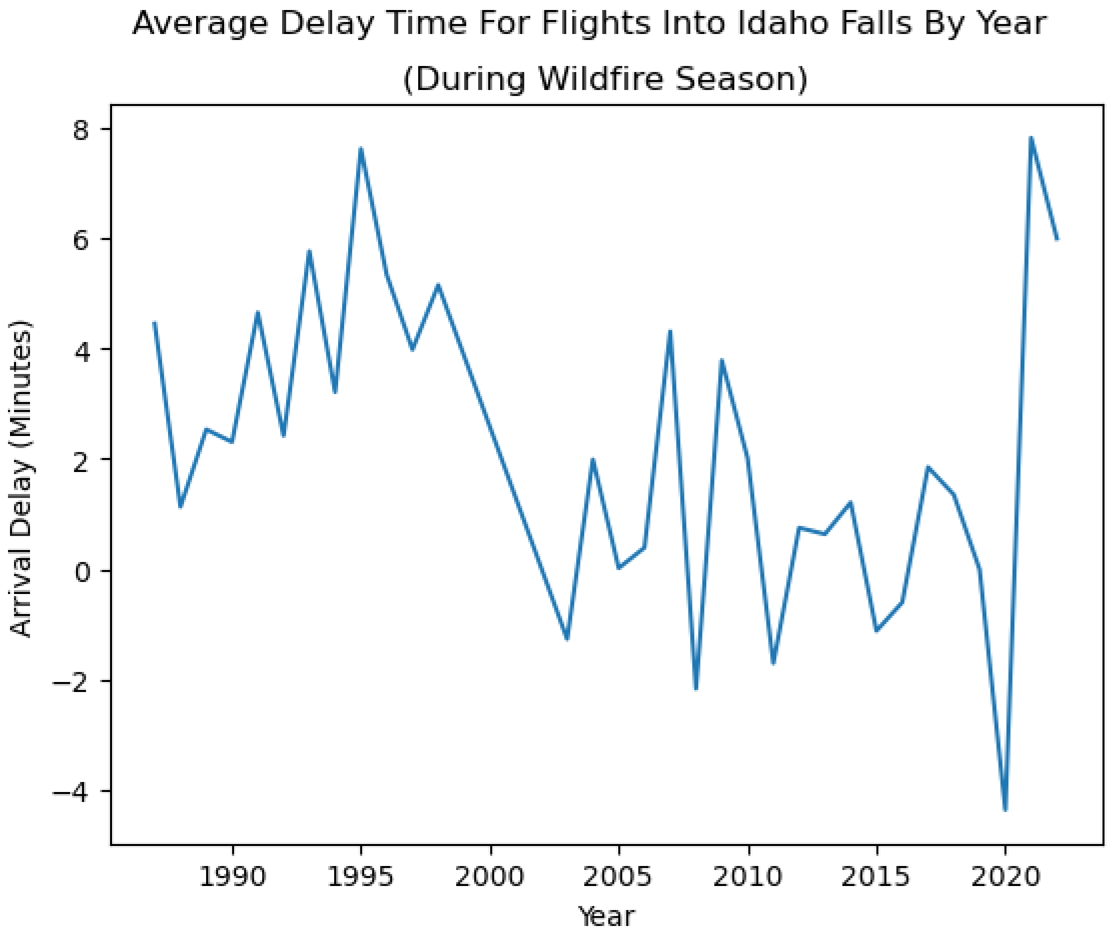

Figure 3: Average Flight Delay Time for All Flights into Idaho Falls By Year

As we can see from Figure 3, there appears to be two distinct periods which may have measured flight delays differently. The earlier period, spanning from 1987-1998, shows a gradual increase in the average arrival delay each year, peaking with a maximum of almost 8 minutes per flight in 1995. However, there is a 4 year gap in this data, as the years 1999-2002 are not tracked. The period from 2003-2020 shows a different trend, with the average flight delay hovering between -2 and 2 minutes. This means that on average, many flights were actually early as opposed to late. In 2020, the average flight was actually 4 minutes early. This could be due to reduced flight traffic from COVID-19. The most intriguing thing about this graph is the dramatic increase in the arrival delay from 2020-2021, with the average annual delay now around 8 minutes, which is a stark contrast to the levels from 2003-2020, reaching levels only seen once before, around 1995. I am not sure why this is the case, but it could be due to a rebound effect from COVID, as people may have wanted to travel more after being isolated for a whole year. As we will see later on, this trend is reflected in the number of passengers as well.

I calculated the number of passengers in a similar way that I calculated the average arrival delay time. I considered only flights during the wildfire season so as to not bias my data too much on external factors, and then I took the sum for each year to represent the total number of passengers that flew into Idaho Falls each year (during wildfire season). We can see these results in Figure 4 below:

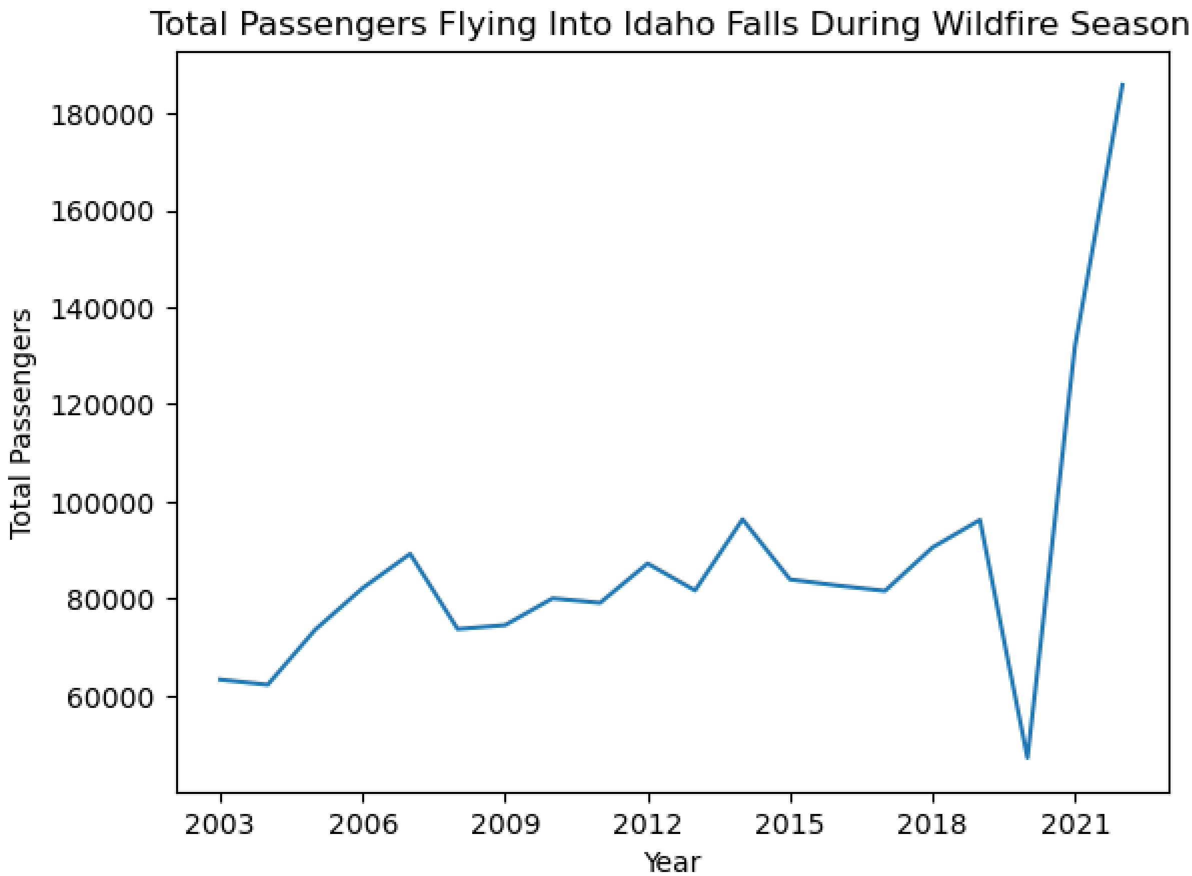

Figure 4: Total Number of Passengers Flying into Idaho Falls Each Year During Wildfire Season

With this graph, we see a similar trend that we saw in Figure 3. From 2003-2019, the number of passengers has been staying relatively consistent, hovering around 80000 passengers per year, with perhaps a very slight increase overall. In 2020, the total number of passengers dips below 60000, likely due to COVID-19, which restricted air travel across the country. After 2020, we see a dramatic increase in the number of passengers to unforeseen levels, reaching 130000 in 2021 and 185000 in 2022. This is more than double the number of passengers who had been flying into the city on average in the past 20 years. It’s unclear if these trends will continue, or if they will revert back to levels pre-COVID. The rationale behind this may be the same as the rationale behind the dramatic increase in average flight delay during the same period – people may be tired of being cooped up inside of their homes and wanted to travel more once restrictions had lifted.

While the previous graphs show trends based on independent perspectives of the data, I wanted to use VAR to forecast trends in flight delays and number of passengers, using the smoke estimate I calculated from earlier. I combined all of this information into one dataset, imputed any missing values, then created a VAR model to forecast my results. More details can be found in the accompanying notebook for this repository, but the following figures summarize my results:

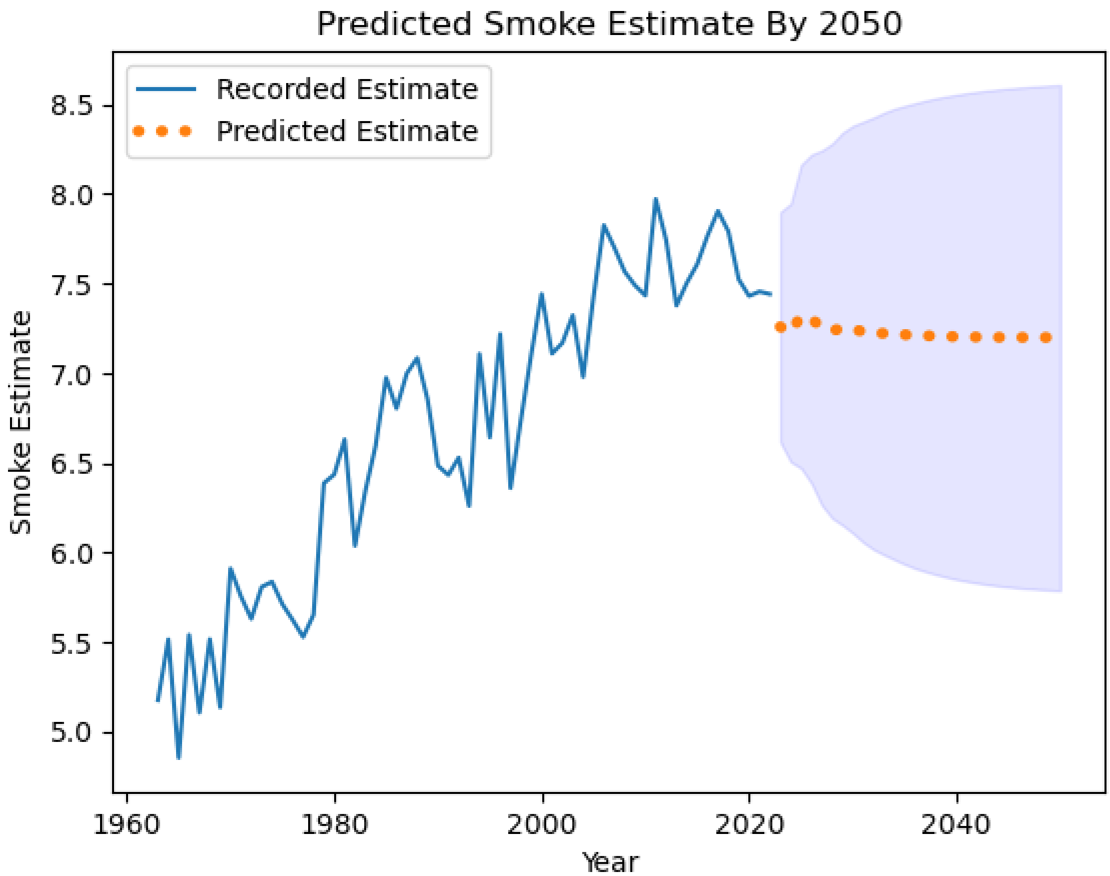

Figure 5: VAR Forecast of Smoke in Idaho Falls Up To 2050

From Figure 5, we can see that the smoke estimate is actually projected to be lower than levels seen in the 2010s. This could be due to the fact that the VAR model is factoring in the number of passengers into the city, as well as the delay time. As I mentioned in the methodology section, the VAR model forecast on the smoke estimate may not be as informative since VAR assumes that the predictors influence each other, while in reality, the influence may go in only one direction. Let's take a look at the flight metrics to see if there are more meaningful patterns:

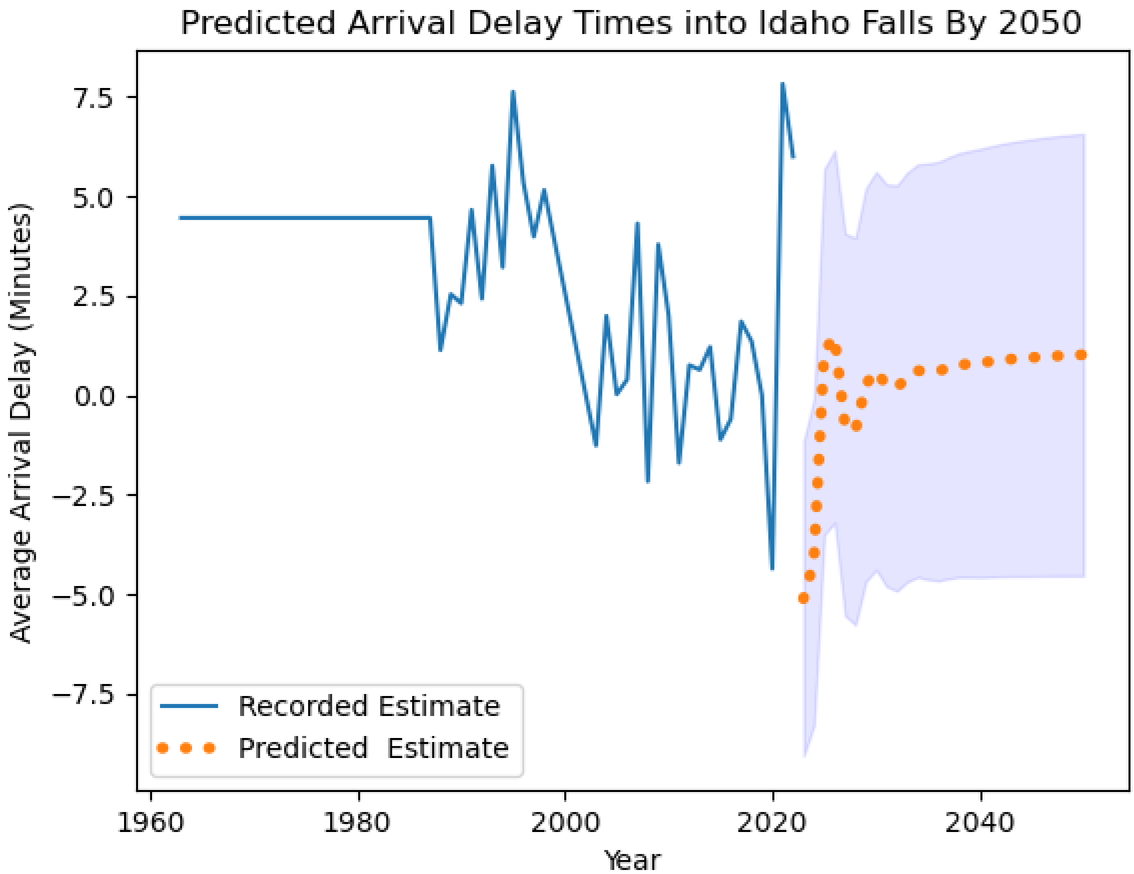

Figure 6: VAR Forecast of Average Arrival Delay Time for Flights into Idaho Falls, Each Year

From this plot, we can see that there is some variability in our prediction estimate, and that may be due to the variability inherent from year to year in the average arrival delay. However, we ultimately see an increase in the average flight delay as the years go on. Since we used VAR to forecast this, then this increase can be attributed to the number of passengers that are expected to fly into the city in future years, as well as the increasing smoke impact. Finally, let’s take a look at what our VAR model predicted in terms of the total passengers flying into the city in future years:

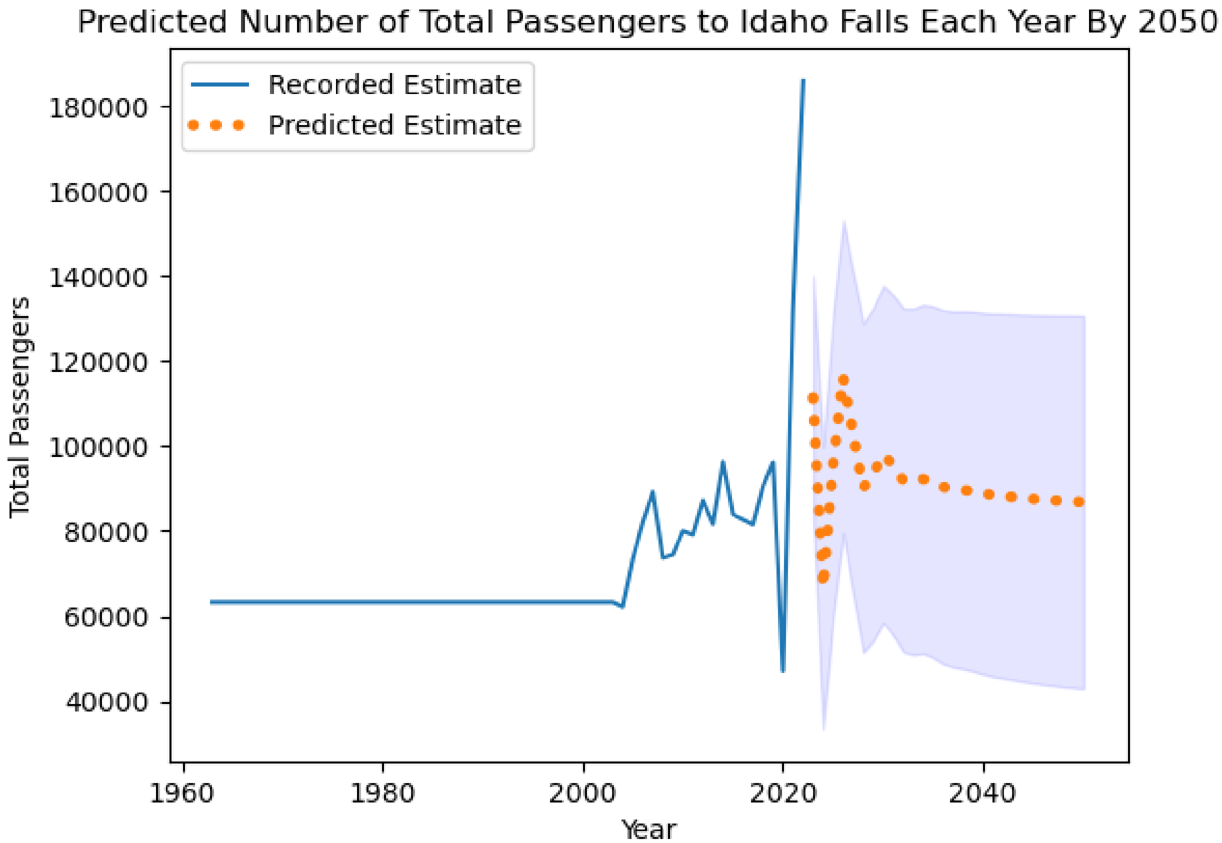

Figure 7: VAR Forecast of the Total Annual Passengers into Idaho Falls, Up to 2050

As we can see from Figure 7, our prediction shows the most oscillation when compared to the previous two plots. This could be biased due to the unforeseen increase in the number of passengers from 2020-2022, which broke a nearly 20-year long trend of steady increases. Our VAR model attempts to take this into account by predicting a number of passengers close to the 2021 value for most of the years in the 2020s, but by the 2030s, it is predicting a gradual decline in the number of passengers flying into the city.

My analysis shows that although the number of passengers has been increasing since being tracked, when taken in the context of rising smoke (from the ARIMA model) and increased flight delays, then this trend does not hold up in future years. As a matter of fact, we see that this is projected to decrease in the coming years, with a not so unsubstantial probability that it can dip to levels that are lower than when it was first being tracked. While intuitively, we may think that this would lead to less flight delays, our VAR model actually shows the opposite. Flight arrival delays are expected to increase in the future, even though we project a decrease in the number of passengers. Thus, at least with the data that we have right now, our smoke estimate must be the culprit behind these trends. While there may be more confounding factors at play, these results do support my initial hypothesis that wildfire smoke negatively affects flight traffic.

Discussion

These results can be summarized to three key findings:

1. Smoke in the city is projected to increase in the future.
2. Flights to Idaho Falls are expected to arrive later than scheduled in the coming years.
3. The number of annual passengers into the city is expected to decrease in the coming years.

This is important because we have shown a link between increased wildfire smoke and worsening travel outcomes, specifically in the number of passengers and increased flight delays. As I alluded to earlier, these metrics act as proxies for tourism, with the number of passengers acting as a more direct indicator. More people flying into a city is generally a strong indicator of a bustling tourism sector. Since we expect less people flying into the city in future years, then we can expect less money generated from tourism. That's a worrying sign, since Idaho Falls depends on tourism to sustain its economy, especially during the summer months. My analysis shows that wildfire smoke is potentially the root cause of these projected trends, so this supports the idea that tourism is negatively affected by wildfire smoke in a city, and thus, the city should not only directly address the tourism concern, but also mitigate any effects from wildfire smoke.

The city can take tangible steps right now to handle these future concerns. For example, they could provide incentives for local businesses to cater to tourists. Giving tax breaks or promotional support to businesses who attract tourists (e.g. hotels, souvenir shops, etc.) could bolster the city's image and make the city more attractive for people to visit. The government can also partner with travel agencies to market the city, distributing brochures and advertisements that promote different landmarks and tourist attractions. For longer term results, the city should focus on sustainability and prevention. For example, we could develop green spaces that help reduce the amount of wildfires that could potentially spread to the city. Furthermore, the city could allocate more resources to medical supplies such as respirators, and ensure that there are enough medications and doctors to handle the increased number of people who are expected to be hospitalized in future years due to wildfire smoke. Finally, the city should enforce responsible land use. Urban green spaces should be well hydrated, and residents should not be burning anything in fire-prone areas. If there is a hot and dry day forecasted, then enforce burn bans on those days to reduce the chance that a fire could ignite. Since my analysis predicts negative tourism growth by the 2030s, then the city should create a plan that incorporates each of these suggestions by then, investing in green infrastructure as soon as possible so that the smoke trends could potentially reverse themselves by then.

Limitations

While this analysis has brought forth some insightful conclusions, there are quite a bit of caveats to discuss when it comes to the data and models that I employed. I mentioned this a few times in the methodology and results sections, but I am using a VAR model to forecast the expected number of passengers and the average arrival delay for flights into Idaho Falls. VAR models assume that each time series variable is correlated with one another. My model uses three predictors (smoke estimate, number of passengers, arrival delays) which I believe are correlated with each other. However, I did not actually test this assumption. For instance, I did not check if there was a true correlation between the number of passengers flying into a city, and the amount of smoke that the city got from wildfires. If it ends up that these predictors are uncorrelated with each other, then the results from my VAR model are not explainable, and my conclusions would be invalid. VAR models also assume that the predictors influence each other, but in reality, the direction of influence seems to go only one direction. Amount of smoke affects the number of passengers flying into a city, which affects the average arrival delay, although one could argue that the average arrival delay into a city could affect the number of passengers, since longer delays could deter more people from visiting. When it comes to my ARIMA model for predicting future smoke estimates, I am assuming that there are no known predictor variables, but in reality, the rise in smoke is coming from the rise in destructive wildfires, which in turn is a product of climate change. While the results of my model aren't necessarily invalid, they could be much more predictive if I took into account these external factors.

Outside of the statistical assumptions for my models, I had incomplete data for each of my predictors. The average arrival delays were compiled from a dataset that only goes back to 1987, and the total number of passengers for each full wildfire season only went as far back as 2003. This meant that my VAR model would either suffer from the lack of training data, or from false/misleading data. I decided to use a simple linear function that entered estimated values for each of these predictors so that they went back to 1963. Unfortunately, that meant taking the earliest known value for each of these variables and copying them for each year back to 1963. We could have tried using some backwards ARIMA forecasting, but I did not want to make the data more inaccurate than it is, as there is the assumption that these patterns we are seeing currently have been trending in the same direction since that time, which I am not sure is something we can safely assume.

Finally, my analysis assumes that all flights landing in Idaho Falls are not layovers. This is actually a huge limitation, since flights that use Idaho Falls as a layover would be counted in my data, but the actual destination for those passengers would be some other city. This would confound the results of my data, and it may falsely explain the rise in the number of passengers flying into the city for a given year. For instance, a flight that may end in Seattle may have a layover in Idaho Falls, and if there are increases in flights to Seattle that include Idaho Falls as a layover, then my data would show an increase in the number of passengers flying into the city, even though in actuality, there aren't any new tourists at all. This would render my findings invalid, and it is something that should be investigated in future analyses.

Conclusion

This study aims to emphasize the ramifications of wildfire smoke on the city of Idaho Falls by measuring its impact on flight delays and passengers visiting the city. I hypothesized that the effects from wildfires would have negative impacts on the tourism industry, since smoke would deter people from wanting to visit. In turn, we should see less passengers flying into the city in future years. My results indicate exactly that, in addition to projecting longer delays amidst more smoke in future years. This project has made me realize just how far reaching seemingly benign events like wildfires can actually pose onto communities, and it reinforces the need for data scientists to approach their analyses through a sociological lens. Without any consideration into socioeconomic factors like tourism, our analytical conclusions would fall on deaf ears, and city governments would have no reason to make changes. As data scientists, it is important to place societal issues at the foundation of our analysis if we truly want to inspire change.

References

Marsh, Jane. "The Impact of Wildfires on Tourism." SeaGoingGreen, 19 May 2022, https://www.seagoinggreen.org/blog/the-impact-of-wildfires-on-tourism. Accessed 11 Dec. 2023.

Miller, Naseem S. "Do Wildfires Have Long-Term Health Effects? Study Shows Current Evidence." The Journalist's Resource, 11 Oct. 2023, journalistsresource.org/health/wildfires-longterm-impact-on-health/.

NCHS. "Prevalence of Selected Measures Among Adults Aged 20 and Over: United States, 1999-2000 through 2017-2018." 21 Oct. 2020.

Richards, Bella. "How Aviation and Tourism Are Intrinsically Linked." Simple Flying, 2 Oct. 2022, simpleflying.com/how-aviation-and-tourism-are-intrinsically-linked/.

Data Sources

https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

https://www.transtats.bts.gov/ONTIME/Arrivals.aspx

https://www.transtats.bts.gov/Data_Elements.aspx?Qn6n=F