Modeling and Forecasting Wildfire Trends in Colorado, Montana, Utah, and Wyoming: A Spatio-Temporal Approach with SAS

Update

Prepared by:

Princess Tagoe

May 14, 2025

1 Introduction

Wildfires are among the most destructive and unpredictable natural disasters, affecting ecosystems, human lives, and economies worldwide [6]. In the United States, particularly in the western region, wildfires have become more frequent and intense due to climate change, land-use changes, and human activities [2]. States such as Colorado, Montana, Utah, and Wyoming have experienced a significant rise in wildfire incidents, resulting in loss of biodiversity, property damage, and air quality deterioration. The increased frequency of wildfires is linked to prolonged droughts, rising temperatures, and altered precipitation patterns [10]. The interaction of these climate variables creates highly flammable conditions, leading to catastrophic fires.

To illustrate, a closer look at historical wildfire data from Florida's St. Johns River Water Management District (SJRWMD) provides valuable insights into how fires start, spread, and impact communities. Between 1981 and 2001, SJRWMD, which spans 31,681 km^2, experienced an average of 1,509 wildfires per year, burning nearly 50,857 acres annually [4]. The main culprits behind these fires were lightning (340 fires per year), arson (445 fires per year), and human-related accidents (724 fires per year). Many of these accidental fires were sparked by campfires, discarded cigarettes, burning debris, and faulty equipment [4].

Some areas were hit harder than others. Volusia County, home to Daytona Beach, saw the highest number of fires, with 4,250 ignitions and 249,838 acres burned [7]. Jacksonville's Duval County followed closely with 2,905 fires, while Brevard County lost 173,789 acres to flames. A deep dive into the data shows that while most fires were small, a handful of massive fires accounted for the majority of the burned land [5]. The picture isn't much different today—if anything, it's worse. Climate change, urban expansion, and the accumulation of dry vegetation have made wildfire even more dangerous. In 2020 alone, California lost over 4 million acres to wildfires, while Colorado and Montana have seen fire activity increasing due to drier conditions and rising temperatures [5].

As wildfire becomes more unpredictable, better forecasting and prevention strategies are crucial. Using advanced data analysis and machine learning, we can identify high-risk areas and improve fire management efforts [3]. The goal is simple: to protect lives, homes, and natural landscapes from the devastating effects of wildfires. Traditional wildfire management strategies rely on historical fire records and expert judgment [3]. However, with the growing complexity of wildfire behavior, there is a pressing need for advanced analytical approaches that integrate spatial and temporal data. The integration of Geographic Information Systems (GIS), statistical modeling, and machine learning offers a robust framework for understanding past wildfire trends and forecasting future fire events. By leveraging historical data and environmental factors, predictive models can identify high-risk areas and optimize resource allocation for fire prevention and response efforts [1].

This study focuses on conducting a comprehensive spatio-temporal analysis of wildfires in Colorado, Montana, Utah, and Wyoming. This study will utilize wildfire records from the U.S. Forest Service's Fire Occurrence Database (FPA-FOD), which contains comprehensive data on wildfires from 1992 to 2020. The findings of this study will have significant implications for policymakers, emergency responders, and environmental agencies. The ability to predict wildfire-prone areas with high accuracy will enhance decision-making in resource allocation, fire mitigation strategies, and land-use planning [9, 8]. This research aligns with global efforts to develop data-driven solutions for wildfire risk assessment and management, ultimately contributing to sustainable environmental and community resilience.

2 Data and Methodology

The dataset used for this study was obtained from the U.S. Forest Service Fire Occurrence Database, containing wildfire records from 1992 to 2020. The analysis was conducted using SAS, utilizing procedures such as PROC SGMAP for spatial visualization, PROC FASTCLUS for clustering, and PROC ARIMA for forecasting wildfire occurrences.

This comprehensive dataset comprises 2,303,566 recorded wildfires, with 39 variables describing each incident. Key variables include temporal information (fire year, discovery and containment dates/times), spatial data (latitude, longitude, state, county), fire characteristics (size, classification), cause information (human vs. natural causes), and administrative details (reporting agencies, land ownership).

The dataset supports multiple analytical approaches, including temporal trend analysis, spatial pattern detection, causal factor assessment, and predictive modeling. While some entries contain missing values, the overall dataset provides a robust foundation for understanding wildfire patterns across the United States over nearly three decades.

2.1 Data Preprocessing

The wildfire dataset was cleaned and processed by:

- Converting date formats for consistency.
- Calculating fire burn duration (days between discovery and containment).
- Categorizing fire sizes into standard classes (0-0.25 acres to 5000+ acres).

3 Results and Analysis

3.1 Temporal Trends in Wildfires

The temporal analysis reveals patterns in wildfire occurrences over time. Figure 1 shows the daily number of wildfires, highlighting seasonal peaks. Figure 2 demonstrates long-term trends in monthly fire frequency.

3.2 Wildfire Causes and Burn Time

The wildfire causes were analyzed to determine their impact on fire duration and size. Figures 5 and 6 indicate that fires classified under "Other" causes have the longest burn time and largest average size.

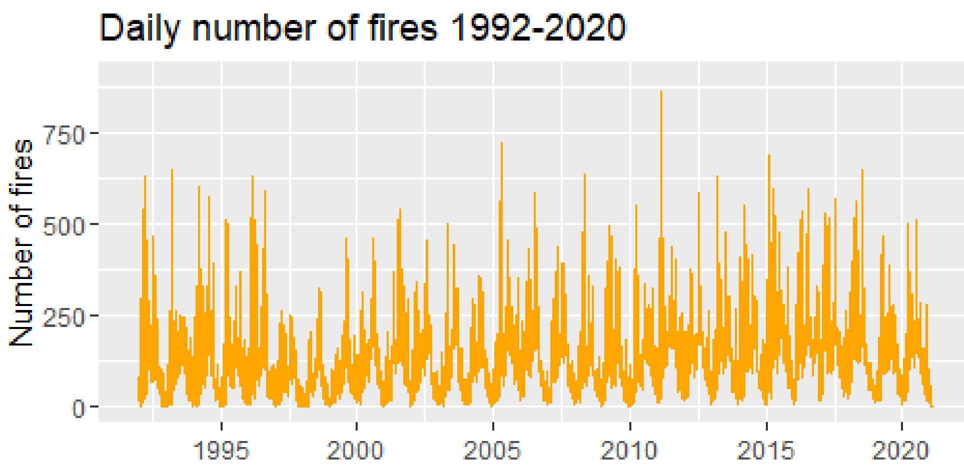

Figure 1: Daily number of wildfires (1992-2020)

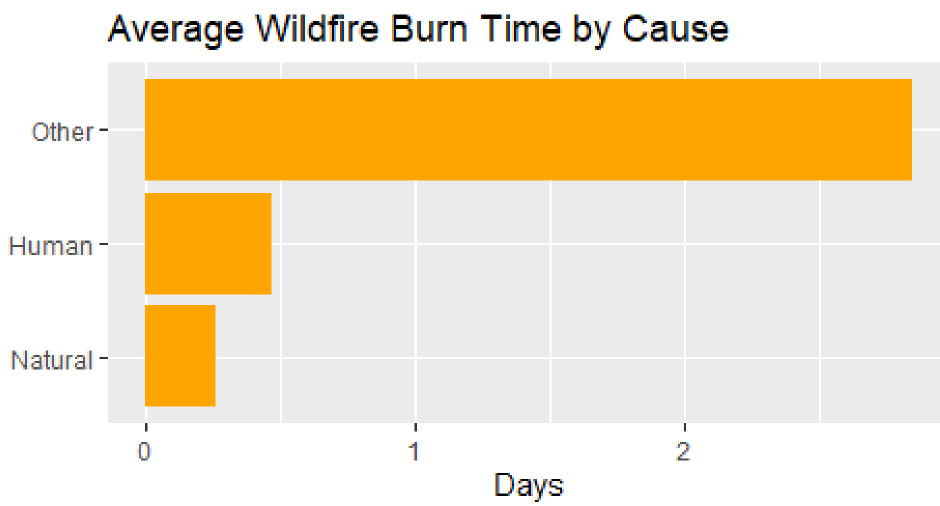

Figure 3: Average wildfire burn time by cause

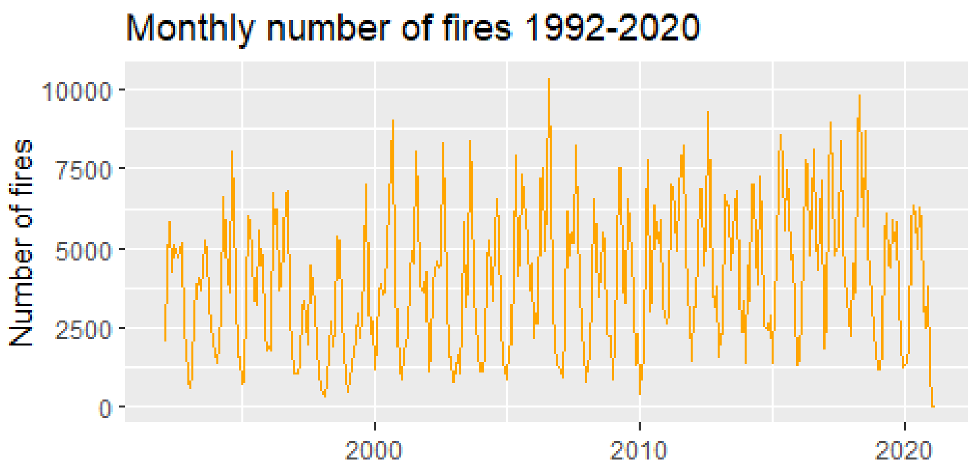

Figure 2: Monthly number of wildfires (1992-2020)

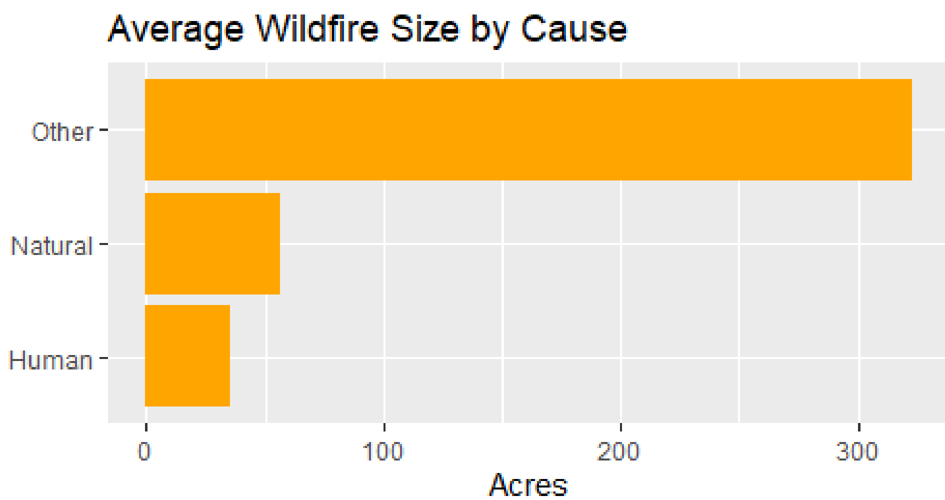

Figure 4: Average wildfire size by cause

3.3 Spatial Distribution of Wildfires

Spatial analysis reveals that California, Texas, and Florida experience the highest number of wildfires, as shown in Figure 5. When filtering for natural causes, California remains the most affected state (Figure 6).

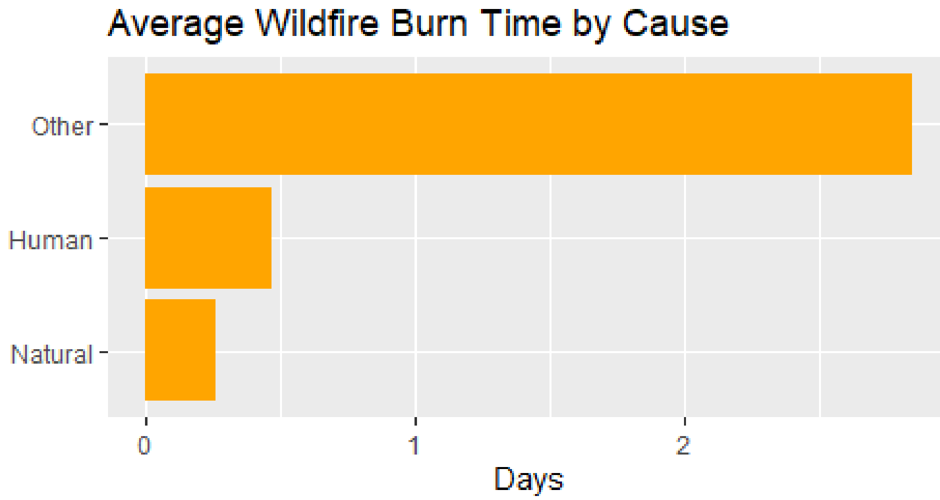

Figure 5: Average wildfire burn time by cause

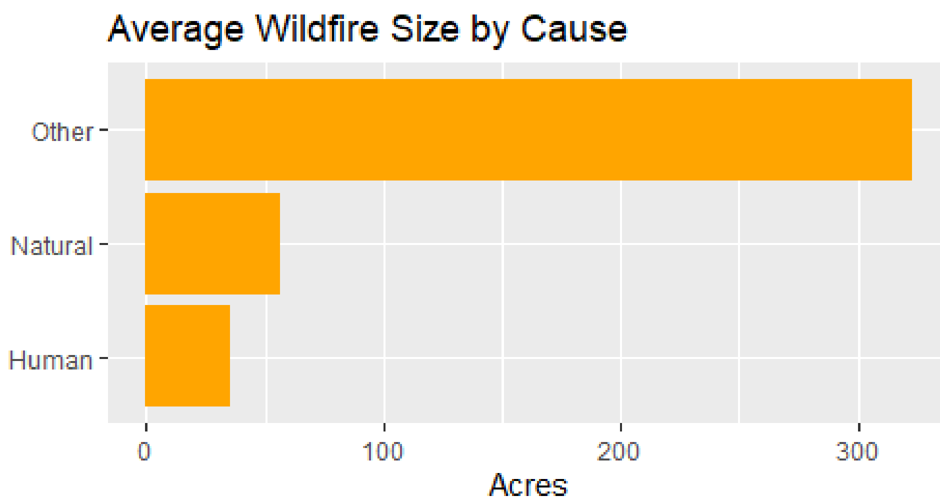

Figure 6: Average Wildfire size by cause

Figures 7 to 8 provide a national overview of wildfire distribution across the United States from 1992 to 2020. However, our primary focus is on the states of Colorado, Montana, Utah, and Wyoming, which lie in the western U.S. and represent ecologically sensitive regions that frequently experience large-scale wildfires.

In Figure 8, which shows the total number of wildfires per state, all four states of interest exhibit moderate to high wildfire activity relative to the rest of the country. Colorado and Montana, in particular, show higher fire frequencies, likely due to a combination of dry summers, forest density, and topographic complexity. Utah and Wyoming, while slightly lower in raw counts, still reflect significant wildfire presence, particularly in rural and mountainous areas.

Figure 7, depicting wildfires caused by natural factors such as lightning, further highlights that these four states are particularly susceptible to naturally ignited fires. This is consistent with their climate characteristics—long dry spells and frequent thunderstorms—especially in summer. Montana and Wyoming stand out with relatively high natural ignition counts compared to many eastern states.

The map of human-caused wildfires indicates that anthropogenic activity also plays a substantial role in fire ignition across these western states. This includes recreational land use, urban-wildland interface expansion, and transportation corridors. The human-caused fire pattern complements the natural causes, resulting in a dual-pressure fire regime that poses challenges for fire prediction and management in this region.

Although it is known and can be seen that California is consistently the most wildfire-prone state, regardless of cause, this regional analysis sets the foundation for a more detailed spatio-temporal investigation of wildfire patterns in Colorado, Montana, Utah, and Wyoming, including seasonal trends, cause-specific frequency shifts, and the impact of climate variability. Understanding the historical patterns will also support more accurate forecasting models tailored specifically to these four states. Understanding these spatial and causal patterns is essential for developing effective wildfire prevention strategies and allocating firefighting resources appropriately.

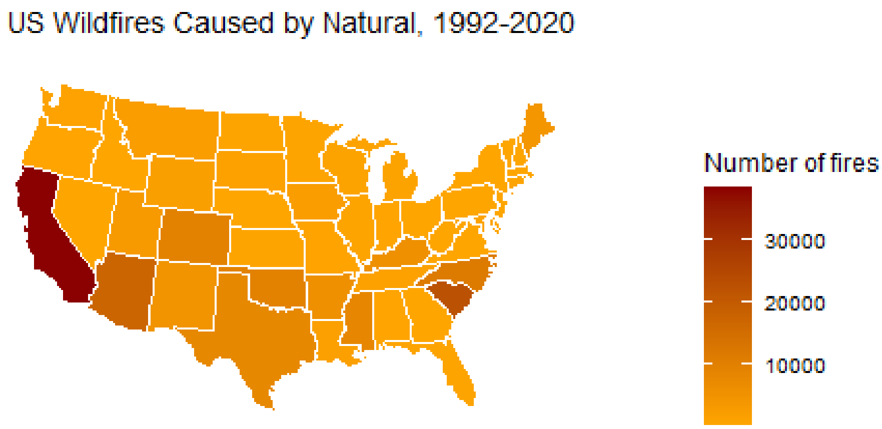

Figure 7: Natural cause by state

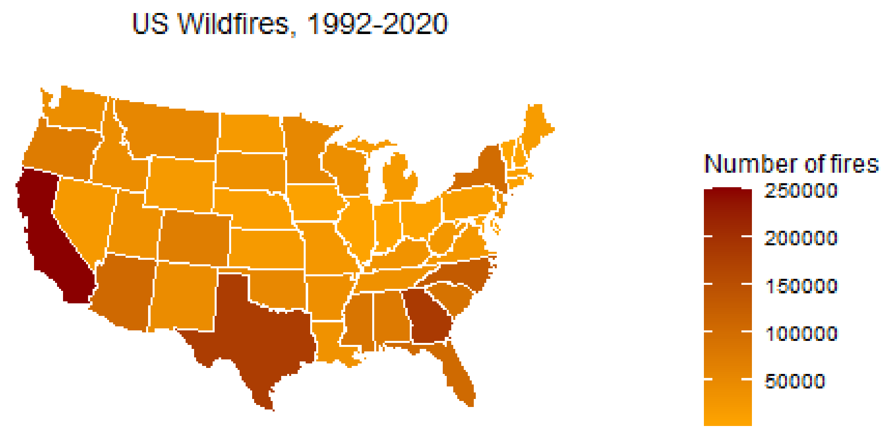

Figure 8: Number of wildfires in all states

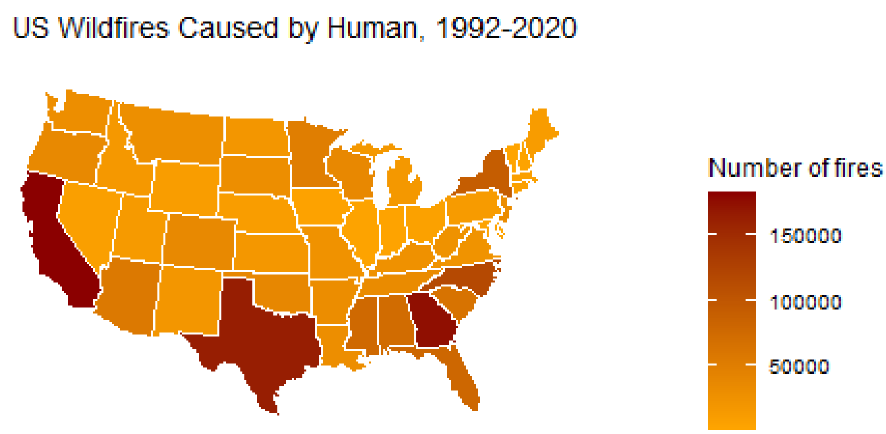

Figure 9: Wildfire caused by human state map

4 Mathematical and Statistical Framework

4.1 1. Spatial Clustering of Wildfires

To analyze wildfire clustering, we use Local Moran's I, which measures spatial autocorrelation. It is used to check if wildfires are clustered in certain areas (hotspots). The formula is:

$$I_i = \frac{z_i}{S^2} \sum_{j=1}^{n} w_{ij} z_j \tag{1}$$

where:

- $z_i = x_i - \bar{x}$ is the deviation from the mean.
- $S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ is the variance.
- w_{ij} is the spatial weight between locations i and j.

If $I_i > 0$, it indicates clustering of high values (hotspots), while $I_i < 0$ suggests clustering of low values (cold spots).

4.2 2. Time Series Forecasting

We use an ARIMA (AutoRegressive Integrated Moving Average) model to predict future wildfire occurrences.

The general ARIMA model is:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

where:

- Y_t is the observed wildfire count at time t.
- ϕ_p are autoregressive coefficients (how much past values influence the current value).
- θ_q are moving average coefficients (how much past errors affect the current value).
- $\epsilon_t \sim N(0, \sigma^2)$ is white noise.

To determine the best ARIMA model, we select parameters (p, d, q) using Akaike Information Criterion (AIC):

$$AIC = -2 \log L + 2k$$

where L is the likelihood function and k is the number of parameters.

5 Spatio-Temporal Analysis

5.1 Spatial Clustering of Wildfires

Using Moran's I statistic, we conducted spatial clustering to identify wildfire hotspots. The results indicate strong clustering in regions with frequent wildfires.

5.2 Time Series Forecasting

We applied an ARIMA model to forecast future wildfire occurrences. The model suggests an increasing trend in fire frequency over the next decade.
To analyze long-term wildfire trends and predict future fire activity, we employed a time series approach using STL decomposition and ARIMA forecasting.

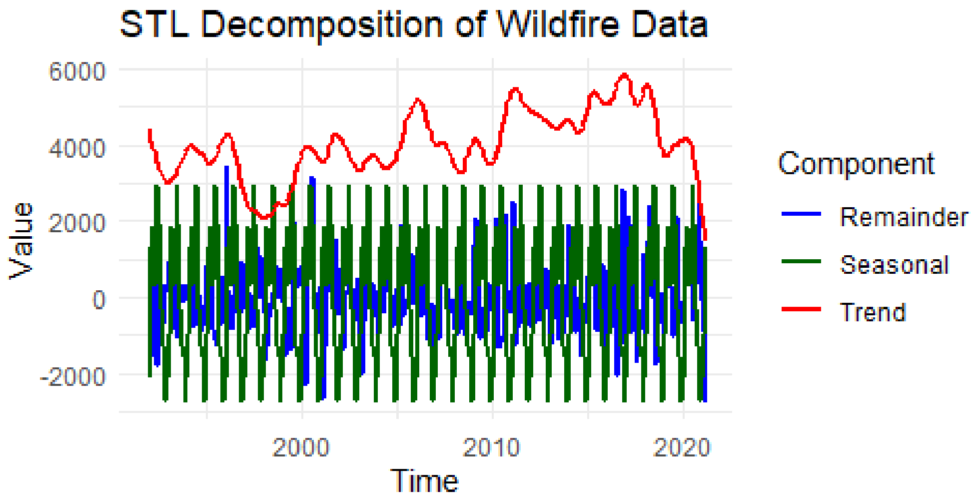

Figure 10: STL Decomposition of Wildfire Data

An ARIMA model was applied to predict wildfire activity for the next 12 months. The forecast results, shown in Figure 11, suggest that wildfire occurrences will continue to follow seasonal patterns, with expected fluctuations. The model's prediction intervals indicate increasing uncertainty over time. However, the overall trend suggests that wildfire occurrences may remain at elevated levels. These results align with the long-term increase observed in the STL trend component.

These findings indicate that wildfire occurrences are both seasonal and increasing over time. The forecasted wildfire activity emphasizes the importance of preparedness and resource allocation during peak wildfire months. Future work could incorporate external climate variables, such as temperature and precipitation, to improve the forecasting accuracy.

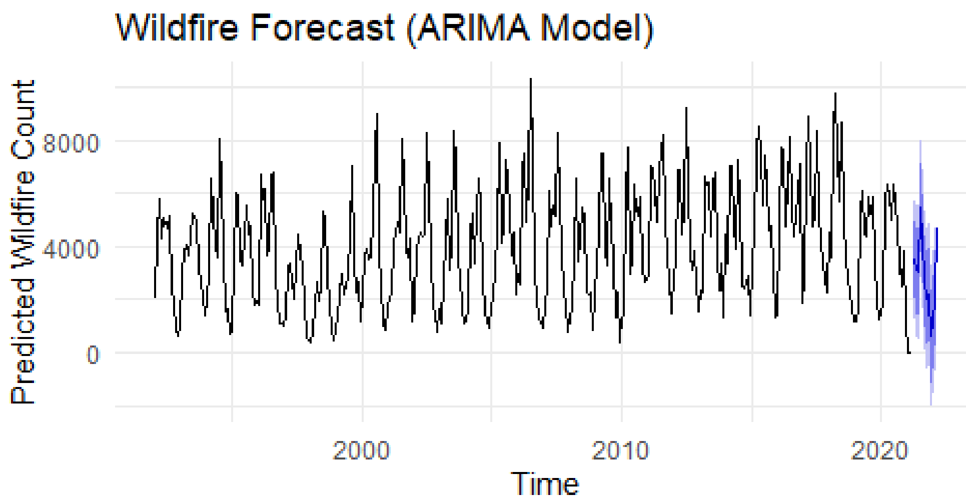

Figure 11: Wildfire Forecast Using ARIMA Model

6 Conclusion

This study highlights critical trends in wildfire occurrences and predicts future risks using spatio temporal techniques.

References

[1] Bhowmik, R. T., Jung, Y. S., Aguilera, J. A., Prunicki, M., and Nadeau, K. A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework. *Journal of Environmental Management 341*, 23423 (Sep 2023), 117908.

[2] Cerda, A. *Fire Effects on Soils and Restoration Strategies.* CRC Press, 2009.

[3] Chu, S., and Majumdar, A. Opportunities and challenges for a sustainable energy future. *Nature 488*, 7411 (Aug 2022), 294–303.

[4] Genton, M. G., Butry, D. T., Gumpertz, M. L., and Prestemon, J. P. Spatio-temporal analysis of wildfire ignitions in the st johns river water management district, florida. *International Journal of Wildland Fire 15*, 1 (Mar 2006), 87–97.

[5] Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., and Flannigan, M. D. A review of machine learning applications in wildfire science and management. *Environmental Reviews 28*, 4 (Dec 2020), 478–505.

[6] Kramer, H. A., Mockrin, M. H., Alexandre, P. M., Stewart, S. I., and Radeloff, V. C. Where wildfires destroy buildings in the us relative to the wildland–urban interface and national fire outreach programs. *International Journal of Wildland Fire 27*, 5 (2018), 329.

[7] Podur, J., Martell, D. L., and Csillag, F. Spatial patterns of lightning-caused forest fires in ontario, 1976–1998. *Ecological Modelling 164*, 1 (Jun 2003), 1–20.

[8] Short, K. C. Spatial wildfire occurrence data for the united states, 1992-2012 [fpa_fod_20140428] (2nd edition). Forest Service Research Data Archive, Apr 2014.

[9] Short, K. C. Spatial wildfire occurrence data for the united states, 1992-2020 [fpa_fod_20221014]. Forest Service Research Data Archive, Nov 2022.

[10] Sippel, S., et al. Drought, heat, and the carbon cycle: a review. *Current Climate Change Reports 4*, 3 (Jun 2018), 266–286.