

Wildfire Smoke and Tourism: Analyzing Impacts on National Parks Visitor Counts Near Palmdale, California

Course: DATA512 - Introduction to Human-Centered Data Science

Name: Baisakhi Sarkar

Date: 12/03/2024

Abstract

Wildfires have become an increasingly frequent occurrence in the western United States, with wide-ranging implications for public health, tourism, and local economies. This project investigates the relationship between wildfire smoke and visitor counts to national parks near Palmdale, California, using historical data from 1964 to 2020. By applying statistical and time-series analysis, the study aims to uncover patterns and correlations that inform regional planning and decision-making. The findings provide actionable insights for city councils, policymakers, and park management to mitigate potential economic impacts and ensure sustainable tourism practices in the face of growing environmental challenges.

Table of Contents

ABSTRACT	1
1. INTRODUCTION	4
2. BACKGROUND/RELATED WORK	4
2.1 HYPOTHESES/RESEARCH QUESTIONS	5
2.2 METHODOLOGY INSIGHTS FROM RELATED WORK	5
2.3 DATASETS	6
3. METHODOLOGY	6
3.1 ANALYTICAL METHODS	6
3.2 HUMAN-CENTERED CONSIDERATIONS	7
3.3 ETHICAL DESIGN	8
4. FINDINGS: RESILIENCE AND VARIABILITY AMONG NATIONAL PARKS TO WILDFIRE SMOKE	8
4.1 PARKS SHOWING RESILIENCE TO SMOKE	8
4.2 PARKS SHOWING SENSITIVITY TO SMOKE	9
4.3 UNEXPECTED TRENDS AND POSSIBLE EXPLANATIONS	10
4.4 ADDITIONAL FINDINGS ON WILDFIRE COUNTS AND SMOKE ESTIMATES	10
4.5 KEY OBSERVATIONS	12
5. DISCUSSION/IMPLICATIONS	12
5.1 IMPORTANCE OF THE FINDINGS	12
5.2 RECOMMENDATIONS FOR STAKEHOLDERS	12
5.3 TIMELINE FOR ACTION	13
5.4 ROLE OF HUMAN-CENTERED DATA SCIENCE	13
5.5 BROADER IMPLICATIONS	13
6. LIMITATIONS	14
6.1 DATA LIMITATIONS	14
6.2 ASSUMPTIONS IN SMOKE ESTIMATE CALCULATION	14
6.3 MODELING LIMITATIONS	14
6.4 GEOGRAPHICAL AND CONTEXTUAL VARIABILITY	15
6.5 ETHICAL AND HUMAN-CENTERED CONSIDERATIONS	15
6.6 EXTRAPOLATION RISKS	15
6.7 GENERALIZABILITY	15
7. CONCLUSION	15

7.1 SUMMARY OF FINDINGS.....	16
7.2 IMPLICATIONS FOR HUMAN-CENTERED DATA SCIENCE.....	16
7.3 FINAL TAKEAWAY.....	16
8. REFERENCES.....	17
9. DATA SOURCES	18
10. APPENDIX.....	19

1. Introduction

Wildfires in the western United States have become a recurring challenge, exacerbated by climate change, population growth, and land-use changes. California, a region known for its iconic national parks and natural beauty, is particularly vulnerable to wildfire smoke. These fires not only pose environmental and health hazards but also threaten the state's thriving tourism industry, a critical component of local economies. As national parks attract millions of visitors annually, any disruption to tourism has cascading effects on regional employment, revenue generation, and community well-being.

This analysis is motivated by the pressing need to understand how wildfire smoke impacts tourism, particularly in California's national parks. While the adverse health effects of wildfire smoke have been widely studied, its implications for tourism - a sector that heavily depends on favorable environmental conditions - remain less explored. Addressing this gap provides an opportunity to assess the broader socio-economic consequences of wildfires and equip policymakers with actionable insights.

Understanding the relationship between wildfire smoke and visitor patterns is vital for multiple stakeholders. For local governments and park authorities, it enables better planning and resource allocation during wildfire seasons. For businesses reliant on park tourism, it informs strategies to mitigate revenue loss and maintain operations. Finally, for residents and community leaders, it underscores the importance of resilience planning to preserve livelihoods and regional economic stability.

This analysis matters because it goes beyond examining environmental impacts—it connects them to human-centered concerns like economic stability and community well-being. By answering unresolved questions about the interplay between wildfire smoke and tourism, this study contributes to the growing body of knowledge needed to tackle the challenges of living in an era marked by increasingly frequent and severe wildfires.

2. Background/Related Work

The intersection of wildfire impacts, and tourism has garnered significant attention in academic and applied research due to the increasing frequency and intensity of wildfires globally. While much of the existing literature focuses on the environmental, health, and property damage caused by wildfires, a growing body of work explores the broader socio-economic impacts, including those on tourism. Research by Yau-Huo Shr, Xibo Wan, and Wendong Zhang (2024) provides national-scale evidence on how wildfire smoke affects outdoor recreation and quantifies the impact in terms of reduced visits and economic costs. Similarly, studies like those by Jacob Gellman, Margaret Walls, and Matthew J.

Wibbenmeyer provides detailed research on the impact of wildfires and smoke on outdoor recreational activities in the western United States, analyzing campground use, visitor behavior, and welfare implications. These works have informed the hypothesis that tourism in wildfire-prone regions, particularly in California's national parks, is significantly influenced by smoke events.

Research conducted by the National Park Service (NPS) Visitor Use Statistics division has been pivotal in understanding visitor trends in U.S. national parks. These datasets reveal annual and seasonal visitation patterns, offering a foundation for assessing how environmental events like wildfires disrupt tourism. Additionally, reports from the U.S. Geological Survey (USGS) have provided essential data on the spatial and temporal distribution of wildfires, enabling the creation of smoke estimates that correlate with park proximities and visitor behavior.

2.1 Hypotheses/Research Questions

This study is guided by the following hypotheses:

1. **Tourism Correlation Hypothesis:** Wildfire smoke negatively correlates with visitor numbers in California's national parks.
2. **Seasonality Hypothesis:** The impact of smoke on visitor patterns varies by season, with summer months—traditionally the peak tourism period—showing resilience or adaptation to wildfire conditions.
3. **Regional Vulnerability Hypothesis:** Parks closer to urban areas or those with unique characteristics may experience varying levels of disruption due to wildfire smoke.

2.2 Methodology Insights from Related Work

As part of **Course Project - Part 2**, the extension plan focused on exploring the impact of wildfire smoke on tourism in California's national parks, particularly its correlation with visitor counts over time. This extension was chosen because tourism is a significant socio-economic driver in California, and understanding its vulnerability to environmental disruptions like wildfires is crucial for policymakers, park authorities, and local communities.

Motivation for the Extension Plan

The decision to focus on tourism was guided by:

- **Human-Centered Relevance:** Tourism-dependent regions face economic challenges during smoke events, impacting livelihoods, local businesses, and public services.
- **Unique Research Opportunity:** While the effects of wildfire smoke on health and property damage have been extensively studied, its impact on tourism, particularly in natural spaces like national parks, remains less explored.
- **Availability of Data:** The rich historical visitor datasets provided by the National Park Service (NPS) offered an opportunity to perform time-series analysis and assess trends over several decades.

During the development of Course Project - Part 2, I chose the **SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors)** model to analyze the relationship between

wildfire smoke and tourism trends. This model was selected after considering alternatives such as ARIMA and LSTM, which are commonly used for time-series forecasting. While ARIMA is effective for capturing trends and seasonality, it does not account for external influencing variables. SARIMAX, on the other hand, allows the inclusion of exogenous variables, making it well-suited for this study where wildfire smoke serves as a critical external factor influencing park visitation. Its ability to handle both seasonal patterns and external variables provided a robust framework for understanding how wildfire smoke impacts visitor counts over time. This choice was further validated by previous studies that effectively used SARIMAX in disaster impact forecasting, demonstrating its relevance in scenarios requiring complex temporal and external factor analysis. Additionally, SARIMAX's interpretability and ability to model seasonality aligned well with the dataset's monthly and yearly patterns, making it the most appropriate choice for this analysis.

2.3 Datasets

This study integrates data from multiple reliable sources to extend the common analysis in Part 1. The key datasets include:

1. **Combined Wildland Fire Datasets (USGS)**: Offers historical wildfire data, including year, area burned, and proximity to Palmdale, California.
2. **National Park Service Visitor Use Statistics**: Provides monthly and yearly visitor counts for California's national parks from 1979 to 2020, enabling a detailed analysis of tourism trends.
3. **Smoke Estimate Calculations**: Derived from USGS wildfire data to estimate smoke exposure for each park based on proximity, fire size, and seasonal effects.

By synthesizing insights from prior research and leveraging advanced time series models like SARIMAX, this study aims to provide a comprehensive understanding of how wildfire smoke impacts California's tourism sector. The findings will help inform data-driven strategies for enhancing the resilience of tourism-dependent regions to future wildfire events.

3. Methodology

The methodology for this study was carefully designed to answer the research questions while incorporating human-centered considerations and ethical principles. The approach was divided into three major components: data preprocessing, modeling, and analysis, with a strong emphasis on ensuring transparency, fairness, and reproducibility throughout the process.

3.1 Analytical Methods

1. **Data Preprocessing:**
 - **Cleaning and Filtering:** The wildfire dataset from the USGS and the visitor dataset from the National Park Service were cleaned to remove inconsistencies, irrelevant fields, and

duplicates. Overlapping wildfires and circular fires larger than a predefined threshold were filtered to ensure data reliability.

- **Standardization:** Yearly visitor counts for each park were standardized to enable direct comparison across parks with varying visitor scales. This ensured that the analysis was not skewed by differences in baseline visitor numbers.
- **Smoke Estimate Calculation:** A custom smoke estimate was derived by considering factors like fire size, proximity, and seasonal effects. This metric served as an exogenous variable in the modeling process.

2. Modeling:

- **Selection of SARIMAX Model:**
 - The **SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous Regressors)** model was chosen for its ability to handle seasonal patterns and incorporate external variables like smoke estimates. This was critical for understanding the potential causal relationship between wildfire smoke and visitor counts.
 - SARIMAX also accounted for the inherent seasonality in tourism trends, making it a robust choice for this study.
- **Rationale for Exclusion of Other Models:**
 - While simpler models like ARIMA could model trends, they lacked the capacity to include external influencing factors.
 - Advanced machine learning models, such as LSTMs or XGBoost, were deemed less interpretable for stakeholders like policymakers, who require actionable insights rather than opaque predictions.

3. Analysis:

- **Correlation Analysis:** The relationship between smoke estimates and visitor counts was evaluated using Pearson correlation coefficients for each park, revealing nuanced patterns in the data.
- **Trend Forecasting:** SARIMAX was used to predict visitor counts for the years 2021–2050, providing future scenarios for park managers and policymakers.
- **Visualization:** Time series plots and scatter plots were generated to communicate findings effectively. These visualizations emphasized clarity and accessibility for non-technical audiences.

3.2 Human-Centered Considerations

1. Stakeholder Relevance:

- The study was designed to address the concerns of local governments, national park managers, and residents. By focusing on the impact of wildfire smoke on tourism, the analysis provided actionable insights for managing park operations and mitigating economic losses.

2. Ethical Data Usage:

- **Transparency:** Data sources were clearly documented, and intermediate data files were stored for reproducibility.

- **Accuracy:** The analysis avoided overgeneralization by highlighting limitations, such as the uncertainty in wildfire reporting for earlier decades.
 - **Fairness:** Potential biases in data collection, such as unequal reporting across regions or demographic groups, were acknowledged and discussed in the limitations section.
3. **Societal Impact:**
- Understanding how wildfire smoke affects tourism supports equitable policy decisions. For instance, identifying parks disproportionately affected by smoke can help allocate resources effectively.
 - The study also emphasized long-term planning by forecasting future visitor trends, enabling stakeholders to prepare for evolving wildfire patterns.
4. **Reproducibility:**
- The entire workflow, including data preprocessing, modeling, and visualization, was implemented in Jupyter Notebooks with detailed documentation to ensure that others can replicate the study or extend it to other regions.

3.3 Ethical Design

The study adhered to the principles of **beneficence** and **justice** by aiming to benefit communities dependent on tourism while ensuring that all regions and stakeholders are represented fairly. By choosing interpretable models and highlighting assumptions, the study prioritized ethical transparency over computational complexity.

4. Findings: Resilience and Variability Among National Parks to Wildfire Smoke

The analysis revealed varying levels of resilience and sensitivity among national parks to wildfire smoke. These findings are based on observed visitor trends, historical smoke estimates, and predictions made using SARIMAX models. Below is a summary of the key insights for each park, including explanations for observed trends and possible reasons for unexpected patterns.

4.1 Parks Showing Resilience to Smoke

Note - Based on Fig 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 in the Appendix.

1. **Joshua Tree National Park**
 - **Findings:** Visitor counts in Joshua Tree show consistent growth over the years despite increasing smoke estimates. Predictions also indicate a stable or slightly increasing trend in visitation.
 - **Possible Explanations:**

- Proximity to urban areas like Los Angeles and the park's unique desert landscape may make it a resilient destination, as visitors are drawn to its distinctive features regardless of smoke events.
 - The relatively lower tree density in the desert ecosystem may also limit direct wildfire impacts, reducing the psychological impact of smoke.
2. **Death Valley National Park**
 - **Findings:** Death Valley exhibits a steady increase in visitor counts over the decades, with minimal observable correlation to smoke estimates. Predicted trends show continued growth.
 - **Possible Explanations:**
 - The park's extreme desert environment may attract visitors seeking unique, rugged experiences.
 - The perception of Death Valley as a "smoke-tolerant" destination due to its arid climate and sparse vegetation could reduce deterrence.
 3. **Santa Monica Mountains National Recreation Area**
 - **Findings:** Santa Monica has maintained stable visitor numbers over the years, with little impact from smoke estimates. Predictions suggest continued resilience.
 - **Possible Explanations:**
 - Proximity to major metropolitan areas like Los Angeles provides consistent tourist demand.
 - Visitors may view this park as a short-term recreational area, making it less vulnerable to changes in air quality compared to parks requiring extended stays.

4.2 Parks Showing Sensitivity to Smoke

Note - Based on Fig 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 in the Appendix.

1. **Kings Canyon National Park**
 - **Findings:** Kings Canyon shows a noticeable decline in visitor counts during years with high smoke estimates. Predicted visitor numbers reflect vulnerability to increasing smoke events.
 - **Possible Explanations:**
 - The park's dense forests and wilderness appeal may be less attractive during heavy smoke events, which obscure views and pose health risks.
 - Accessibility issues caused by wildfires or road closures may exacerbate visitor deterrence.
2. **Sequoia National Park**
 - **Findings:** Visitor trends for Sequoia exhibit fluctuations aligning with smoke estimates. Predictions suggest increasing variability in visitation due to smoke.
 - **Possible Explanations:**
 - Iconic giant sequoia trees may become less of a draw during smoke events due to limited visibility and health concerns.
 - The park's location in regions prone to large wildfires amplifies direct impacts, including evacuation and road closures.
3. **Lassen Volcanic National Park**

- **Findings:** Lassen Volcanic has shown inconsistent visitor trends that correlate with smoke estimates. Predictions indicate potential sensitivity to future smoke events.
- **Possible Explanations:**
 - Remote location and reliance on clear weather for outdoor activities may deter visitors during smoke events.
 - Limited infrastructure to handle emergency situations or mitigate smoke impacts may exacerbate the issue.

4.3 Unexpected Trends and Possible Explanations

Note - Based on Fig 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 in the Appendix.

1. Channel Islands National Park

- **Findings:** Visitor counts show limited correlation with smoke estimates despite its proximity to areas prone to wildfires.
- **Possible Explanations:**
 - The park's island location provides natural isolation from mainland smoke, reducing direct impact.
 - Visitors may view the park as a unique experience, less affected by mainland conditions.

2. Yosemite National Park

- **Findings:** Visitor counts remain relatively stable despite increasing smoke estimates, though some dips are visible during extreme wildfire years.
- **Possible Explanations:**
 - Yosemite's iconic status as a must-visit destination may override concerns about smoke exposure.
 - Active management efforts to keep major areas accessible during wildfires may reduce visitor deterrence.

4.4 Additional Findings on Wildfire Counts and Smoke Estimates

Note - Based on Fig 4.41, 4.42, 4.43, 4.44 in the Appendix.

The findings reveal significant patterns in wildfire frequency, total acres burned, and smoke estimates, shedding light on the broader environmental and socio-economic context of the study. Below are the key observations:

1. Increasing Wildfire Frequency and Severity

- **Findings:** Analysis of historical wildfire data highlights an increase in both the frequency and intensity of wildfires over the last several decades. This trend is particularly evident in the total acres burned, which shows sharp peaks during certain years (e.g., 2020) and a noticeable upward trajectory over time.
- **Possible Explanations:**
 - Climate change has led to drier conditions and prolonged fire seasons, increasing wildfire susceptibility.

- Human activities, such as land use changes and fire suppression policies, have contributed to fuel accumulation, intensifying fires.

2. **Smoke Estimates and Air Quality Correlation**

- **Findings:** The cumulative smoke estimate strongly correlates with maximum AQI levels, particularly in recent decades, emphasizing the worsening air quality caused by wildfire events. This correlation indicates that wildfire smoke directly impacts air quality in regions near the affected parks.
- **Possible Explanations:**
 - Larger wildfires release significant amounts of particulate matter, which disperses over large areas, degrading air quality.
 - Increased wildfire activity near densely populated regions exacerbates health risks, with noticeable AQI spikes during active fire years.

3. **Proximity and Wildfire Risk**

- **Findings:** Parks located closer to high wildfire occurrence zones (e.g., Sequoia and Kings Canyon) experience higher smoke exposure and associated visitor declines, compared to parks farther from wildfire-prone areas (e.g., Channel Islands).
- **Possible Explanations:**
 - Proximity to active fire zones increases the likelihood of direct and indirect impacts, including evacuation orders, road closures, and visibility issues.
 - Parks like Channel Islands, which are geographically isolated, benefit from natural barriers that limit smoke intrusion.

4. **Temporal Trends in Wildfire and Smoke Impacts**

- **Findings:** Peaks in wildfire activity align closely with summer months, which are also the peak tourism season for many parks. This overlap indicates that visitor patterns are likely adapting to seasonal smoke exposure, potentially driven by the inevitability of wildfire activity during peak vacation months.
- **Possible Explanations:**
 - Visitors may perceive smoke exposure as a manageable risk during limited vacation windows, particularly for iconic destinations like Yosemite and Joshua Tree.
 - Seasonal infrastructure enhancements, such as improved air filtration in visitor centers, may also reduce the perceived severity of smoke impacts.

5. **Regional Variability in Smoke Estimates**

- **Findings:** Parks near urban centers, such as Santa Monica, exhibit lower sensitivity to smoke exposure despite being located in wildfire-prone regions. Conversely, remote parks like Lassen Volcanic show higher sensitivity due to limited access and infrastructure.
- **Possible Explanations:**
 - Urban proximity ensures consistent tourist inflow due to convenience and short-term visits.
 - Remote parks lack the resources to mitigate smoke impacts effectively, leading to greater visitor deterrence during active wildfire periods.

4.5 Key Observations

- **Seasonal Trends:** Summer remains the peak season for most parks, which aligns with wildfire activity. This overlap suggests that visitors may be adapting to the inevitability of smoke during peak tourism months.
- **Regional Variability:** Parks closer to urban centers (e.g., Santa Monica) show greater resilience, whereas remote parks (e.g., Lassen Volcanic) are more vulnerable to visitor declines.

5. Discussion/Implications

The findings from this study shed light on the interplay between wildfire smoke and tourism in California's national parks, with significant implications for city officials, park authorities, and residents. This section highlights the importance of these findings, actionable recommendations for stakeholders, and the role of human-centered data science principles in shaping the study.

5.1 Importance of the Findings

The resilience and vulnerability observed across different parks provide valuable insights into how wildfire smoke impacts regional tourism. While some parks like Death Valley and Joshua Tree demonstrated strong resilience, others, including Kings Canyon and Lassen Volcanic, showed notable sensitivity to smoke events. These results emphasize that smoke does not uniformly affect all parks, underscoring the need for location-specific strategies to mitigate its impacts.

The study's findings are particularly significant in light of the increasing frequency and severity of wildfires driven by climate change. Tourism is a key economic driver for many regions in California, and disruptions in visitor numbers could have cascading effects on local economies, including job losses and reduced revenue for businesses reliant on park tourism.

5.2 Recommendations for Stakeholders

To address the implications of these findings, the following recommendations are proposed for stakeholders:

1. **City Councils and Mayors:**
 - Develop **regional wildfire response plans** that prioritize tourism-dependent areas, ensuring that parks and their surrounding communities are equipped to handle the economic fallout of smoke events.
 - Advocate for increased funding to support **smoke management programs**, including controlled burns and reforestation efforts that reduce wildfire severity.
2. **Park Authorities:**
 - Implement **real-time air quality monitoring systems** at park entrances to provide visitors with transparent information about smoke conditions.
 - Enhance **infrastructure for visitor safety**, such as indoor recreational facilities with filtered air, to sustain visitor engagement during smoke-affected periods.

- Launch targeted **marketing campaigns** to reassure potential visitors about safety measures and promote off-peak or less smoke-prone seasons.
3. **Residents and Local Businesses:**
- Engage in community-driven initiatives to **diversify local economies**, reducing reliance on tourism during wildfire seasons.
 - Encourage businesses to adopt flexible operating models, such as expanding online services or offering incentives to attract tourists during periods of low visitation.

5.3 Timeline for Action

Given the urgency of increasing wildfire severity, stakeholders should aim to initiate actionable steps within the next 2-3 years. Immediate priorities include securing funding for smoke mitigation programs, deploying air quality monitoring systems, and developing targeted marketing strategies. Long-term planning should focus on strengthening park infrastructure and fostering economic diversification to build resilience against future disruptions.

5.4 Role of Human-Centered Data Science

Human-centered data science principles were integral to this project's design and execution, ensuring that the study remained focused on the well-being of affected communities and stakeholders:

1. **Ethical Considerations:**
The analysis prioritized ethical use of data, ensuring that datasets were sourced responsibly and anonymized to protect privacy. Transparency in data preprocessing and modeling was maintained to uphold accountability.
2. **Stakeholder-Centric Design:**
The research questions were formulated with an emphasis on tangible outcomes for city officials, park authorities, and residents. The insights generated were tailored to inform actionable strategies rather than academic exploration alone.
3. **Inclusivity and Equity:**
Special attention was given to parks and communities that showed greater vulnerability to smoke events. By identifying disparities, the study aimed to empower stakeholders to implement equitable solutions that prioritize the most affected regions.
4. **Transparency in Communication:**
Visualizations and predictions were presented in an accessible manner, ensuring that non-technical stakeholders, such as city council members and local residents, could easily interpret the findings.

5.5 Broader Implications

The findings of this study extend beyond California, serving as a framework for understanding the socio-economic impacts of wildfires on tourism globally. By integrating human-centered approaches with advanced data science techniques, this project highlights the importance of bridging research and policy to create sustainable, resilient communities in the face of environmental challenges.

6. Limitations

Every study is constrained by its limitations, and this analysis is no exception. While it provides valuable insights into the relationship between wildfire smoke and tourism, several factors may affect the accuracy, reliability, and generalizability of the results. These limitations are discussed below to provide context for interpreting the findings.

6.1 Data Limitations

- **Data Completeness:** The wildfire dataset, while comprehensive, may not accurately capture all historical wildfires, particularly those in earlier decades where record-keeping was less systematic. Smaller or undocumented fires might have been excluded, impacting the smoke estimate calculations.
- **Visitor Data Gaps:** The National Park Service Visitor Use Statistics data had some missing or inconsistent entries, particularly in earlier years. These gaps were addressed using interpolation methods, which may not fully capture true visitation patterns.
- **Timeframe Discrepancies:** The overlap of data availability for different variables (e.g., smoke estimates, visitor counts, air quality, etc.) varied significantly, leading to truncated time periods for analysis in some cases.

6.2 Assumptions in Smoke Estimate Calculation

- The smoke estimate relied on distance from fire perimeters, fire size, and other factors, but excluded variables like wind direction, meteorological conditions, and terrain, which can significantly influence smoke dispersion. This simplification may result in inaccuracies in the smoke exposure levels used in the analysis.
- Prescribed fires were treated with the same weight as wildfires in terms of smoke contribution, which may not accurately reflect their differing impacts on air quality and tourism.

6.3 Modeling Limitations

- **Choice of SARIMAX:** While SARIMAX was selected for its ability to incorporate exogenous variables, its performance is heavily dependent on parameter tuning and data stationarity. The predictions for some parks (e.g., Channel Islands and Death Valley) showed anomalies like negative values, highlighting potential model misfits.
- **Seasonality Assumptions:** SARIMAX assumes regular seasonal patterns, which may not fully capture the irregularities in visitor behavior caused by external factors like policy changes, extreme weather events, or other disruptions.
- **Limited Feature Inclusion:** The models focused on smoke estimates as the primary exogenous variable but did not account for other factors like economic downturns, park closures, or promotional activities that might influence visitor trends.

6.4 Geographical and Contextual Variability

- The analysis assumed a uniform impact of smoke estimates on visitor behavior across all parks, which may overlook regional differences such as accessibility, park amenities, or visitor demographics.
- Parks with unique characteristics (e.g., urban proximity for Santa Monica or desert landscapes for Death Valley) may exhibit trends that are not generalizable to other parks.

6.5 Ethical and Human-Centered Considerations

- The study relied on aggregated data and did not account for individual experiences, such as visitor feedback or perceptions of smoke risk. This limits the ability to fully understand the human-centered impacts of wildfire smoke on decision-making.
- While steps were taken to ensure data transparency and ethical use, potential biases in the source datasets (e.g., reporting bias in wildfire records or visitor count estimation methods) could influence the results.

6.6 Extrapolation Risks

- The prediction for visitor counts and smoke estimates extend to 2050, but such long-term forecasts inherently carry significant uncertainty. Unforeseen factors like future wildfire policies, climate change impacts, or technological advancements in air quality management could drastically alter trends.

6.7 Generalizability

- This analysis focuses on California's national parks and may not be directly applicable to other regions with different climate conditions, park characteristics, or socio-economic contexts. The findings should be interpreted with caution when generalizing to other locations.

While these limitations do not invalidate the study, they highlight areas where further research and improved methodologies could enhance the robustness of the findings. Acknowledging these limitations ensures transparency and helps stakeholders interpret the results within an appropriate context.

7. Conclusion

This study set out to examine the impact of wildfire smoke on tourism in California's national parks, focusing on three key hypotheses:

1. **Tourism Correlation Hypothesis:** Wildfire smoke negatively correlates with visitor numbers in California's national parks.

2. **Seasonality Hypothesis:** The impact of smoke on visitor patterns varies by season, with summer months—traditionally the peak tourism period—showing resilience or adaptation to wildfire conditions.
3. **Regional Vulnerability Hypothesis:** Parks closer to urban areas or those with unique characteristics may experience varying levels of disruption due to wildfire smoke.

7.1 Summary of Findings

Through detailed analysis and predictive modeling using SARIMAX, the study revealed nuanced relationships between wildfire smoke and park visitation:

- **Correlation with Smoke Estimates:** While some parks, like Kings Canyon and Death Valley, showed a significant correlation between smoke estimates and visitor counts, others, such as Yosemite and Santa Monica, demonstrated a resilience to smoke, with no strong correlation observed.
- **Seasonal Patterns:** Visitor patterns during summer months showed a relative robustness to wildfire smoke, likely due to the fixed vacation schedules of tourists and the high appeal of outdoor activities during this period.
- **Regional Differences:** Parks closer to urban areas, such as Santa Monica, exhibited different trends compared to remote parks like Lassen Volcanic, emphasizing the role of accessibility and visitor demographics in shaping tourism resilience.

7.2 Implications for Human-Centered Data Science

This study demonstrates the value of human-centered data science in addressing real-world challenges. By combining smoke estimates with visitor data, the analysis provides actionable insights for park authorities and policymakers to improve resource allocation, disaster preparedness, and communication strategies for park visitors.

The use of human-centered principles ensured that the analysis focused on the needs and well-being of park visitors and local communities. For instance, ethical data practices, such as transparency in data processing and the acknowledgment of limitations, were emphasized throughout the study. Additionally, the focus on practical implications—such as informing park management strategies and visitor safety protocols—underscores the human-centered nature of this research.

7.3 Final Takeaway

This study highlights the complex interplay between environmental factors and socio-economic activities, such as tourism. While the findings confirm that wildfire smoke can disrupt visitation patterns in some parks, they also reveal the resilience of tourism in others. By integrating robust analytical methods and human-centered insights, this research underscores the importance of tailoring strategies to local contexts and addressing the broader societal impacts of environmental challenges.

Moving forward, continued monitoring, improved data collection, and collaboration between policymakers, researchers, and park management can enhance our ability to predict and mitigate the impacts of wildfires on tourism and beyond.

8. References

- [1] U.S. Geological Survey (USGS). Combined Wildland Fire Datasets for the United States and Certain Territories, 1800s-Present. Accessed from [USGS Wildland Fire Data](#).
- [2] National Park Service (NPS). Visitor Use Statistics for U.S. National Parks. Accessed from [NPS Visitor Use Statistics](#).
- [3] U.S. Environmental Protection Agency (EPA). Air Quality System (AQS) API Documentation. Accessed from [AQS API](#)
- [4] Seasonal AutoRegressive Integrated Moving-Average with eXogenous Factors (SARIMAX) Model Documentation. [Statsmodels Library](#)
- [5] Shr, Y.-H. J., Wan, X., & Zhang, W. (2024). *When Nature Turns Hazy: How Wildfire Smoke Affects Outdoor Recreation*. Retrieved from [When Nature Turns Hazy](#)
- [6] Gellman, J., Walls, M., & Wibbenmeyer, M. J. (2021). *Wildfire, Smoke, and Outdoor Recreation in the Western United States*. Resources for the Future Working Paper 21-22. Available at: Wildfire, Smoke, and Outdoor Recreation in the Western United States
- [7] McDonald D., *wildfire.zip* - Google Drive. Retrieved October 23rd, 2024, from [Google Drive Link](#)
- [8] McDonald D., *wildfire_geo_proximity_example.ipynb* - Google Drive. Retrieved October 23rd, 2024, from [Google Drive Link](#)
- [9] McDonald D., *epa_air_quality_history_example.ipynb* - Google Drive. Retrieved October 23rd, 2024, from [Google Drive Link](#)
- [10] GeoJSON Python Package. (2024). GeoJSON: A Python library for working with GeoJSON. Available at [GeoJSON](#).
- [11] PyProj Library Documentation. (2024). PyProj: Python interface to PROJ (cartographic projections and geodetic transformations library). Available at [PyProj](#).

9. Data Sources

- **Combined Wildland Fire Datasets (US Geological Survey)**
This dataset provides comprehensive historical wildfire data, including attributes such as fire year, area burned, geographic coordinates, and fire perimeter. It was instrumental in identifying wildfire patterns and estimating smoke impact based on proximity to national parks in California. Access the dataset here: [USGS Wildland Fire Data](#)
- **National Park Service Visitor Use Statistics**
This dataset contains monthly and yearly visitor counts for national parks in the United States, including those in California. It allowed for an analysis of tourism trends in the context of wildfire smoke exposure. Access the dataset here: [NPS Visitor Use Statistics](#).
- **US Environmental Protection Agency Air Quality Service (AQS) API**
This API provided historical air quality data (AQI) used to validate and compare smoke estimate models. It played a key role in correlating smoke estimates with air quality trends over time. Access the documentation here: [AQS API](#)

10. Appendix

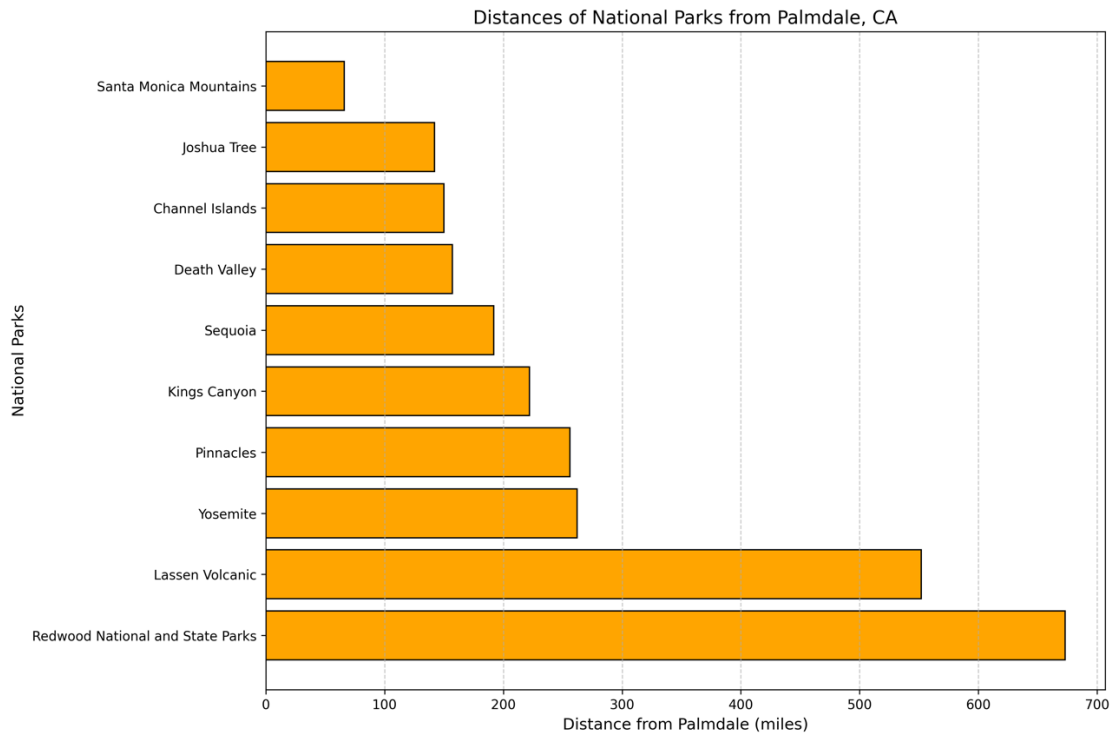


Fig 4.11

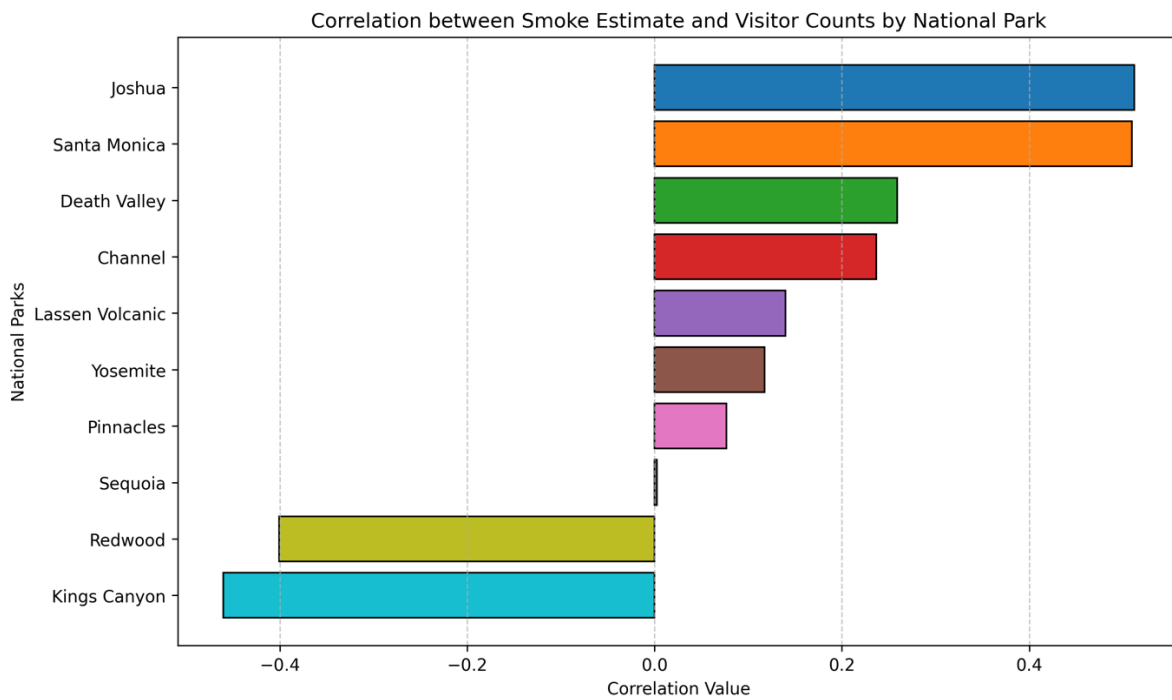


Fig 4.12

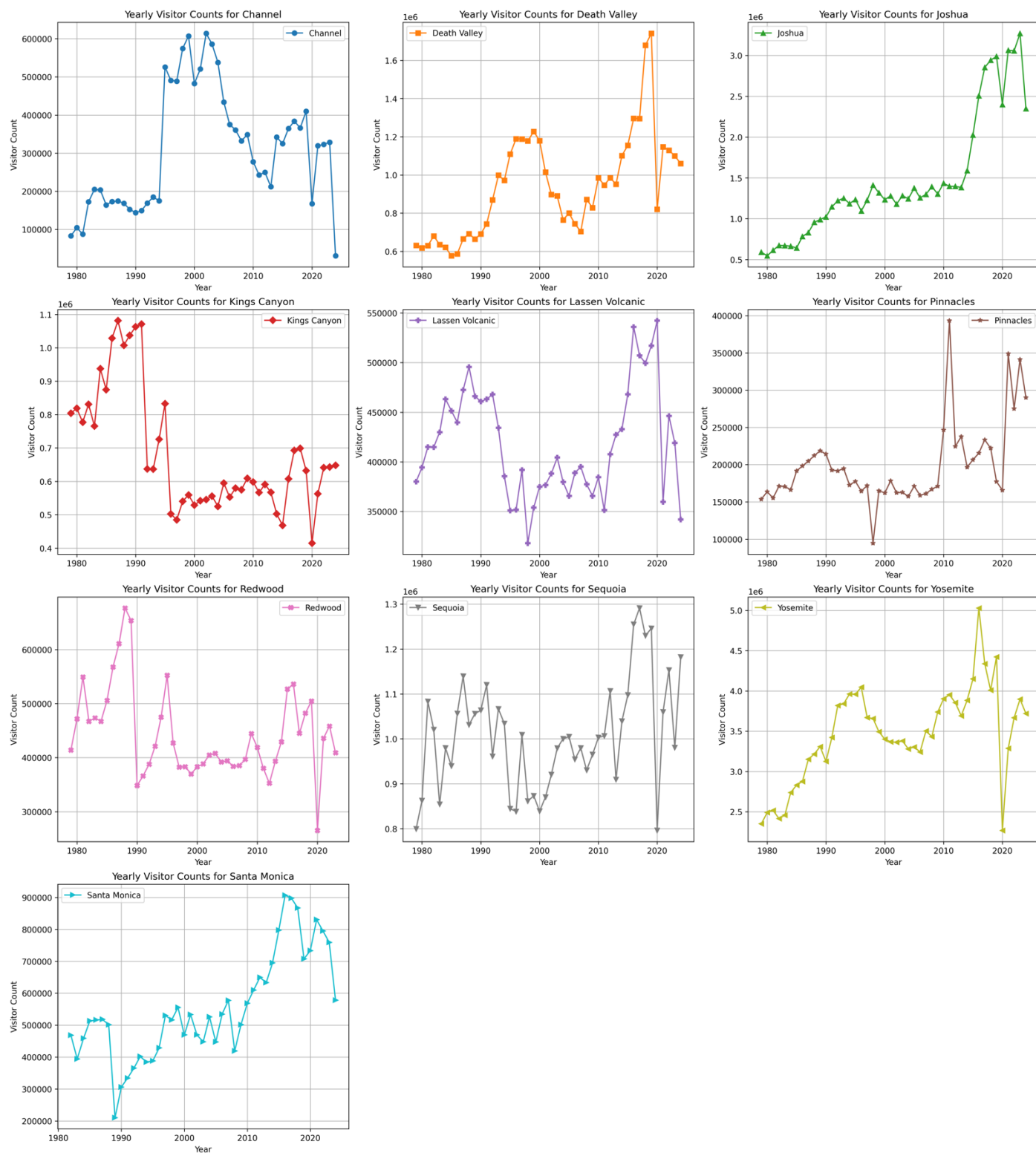


Fig 4.13

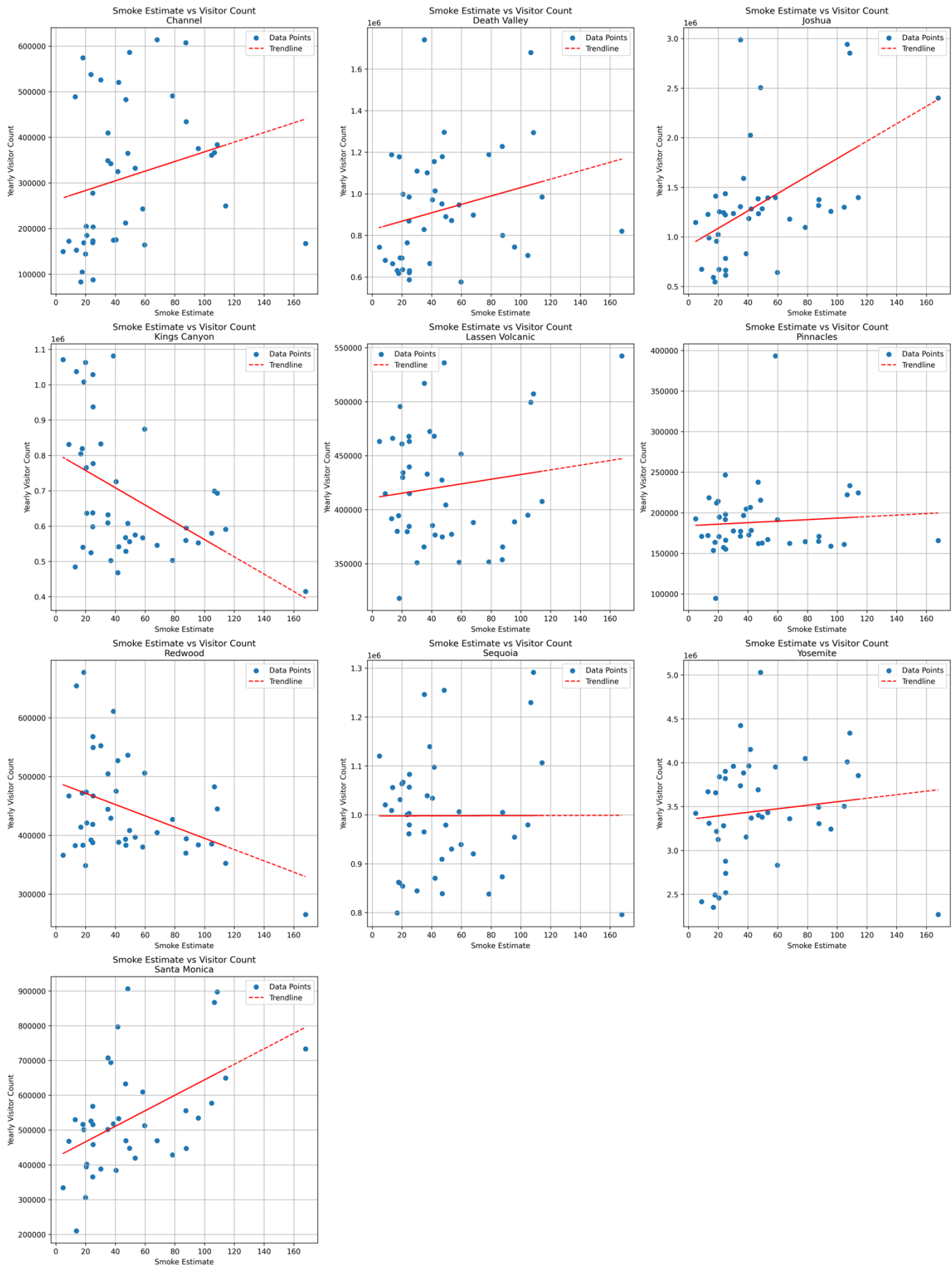


Fig 4.14

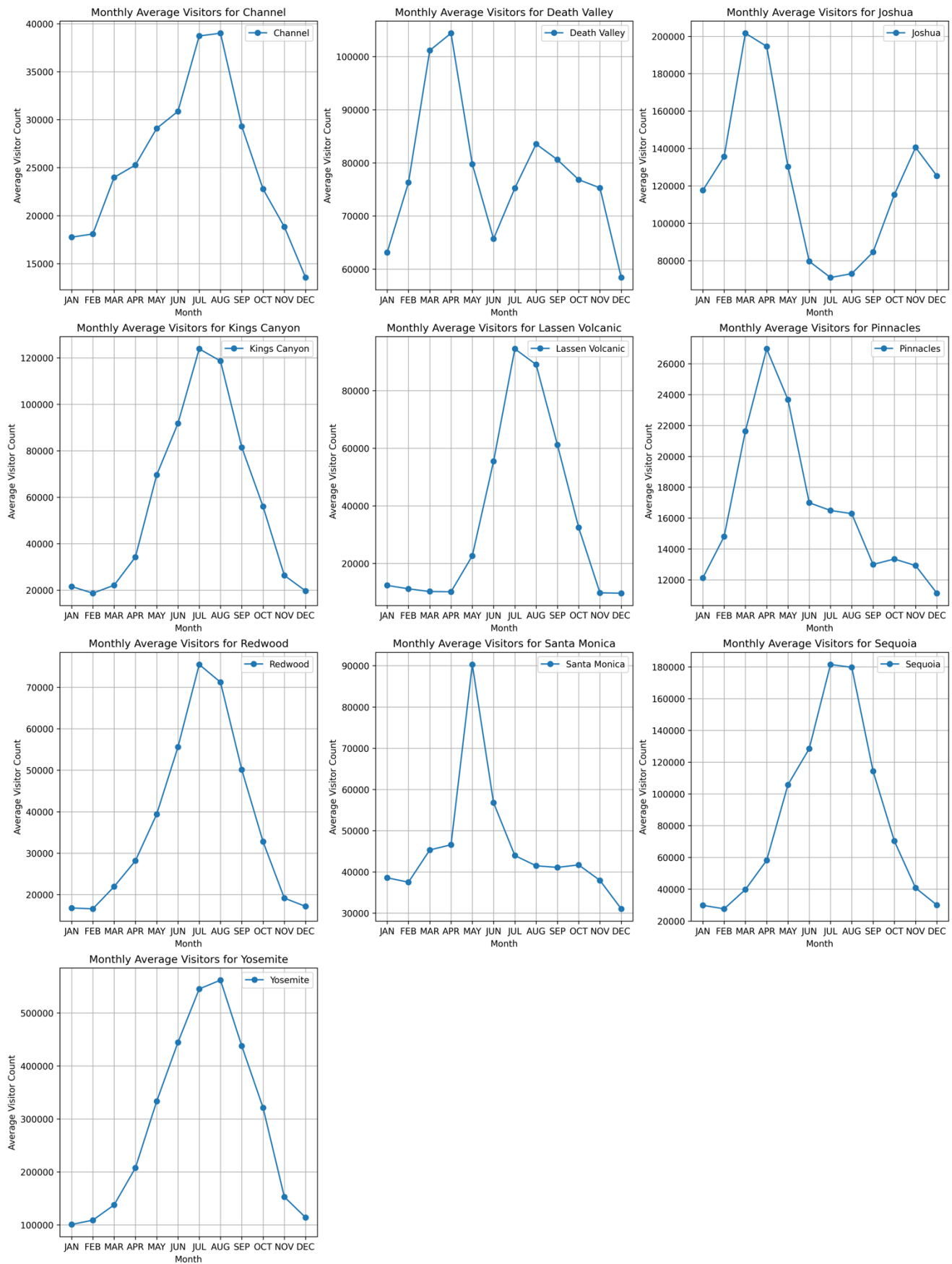


Fig 4.15

Visitor Counts and Smoke Estimates Prediction for National Parks (2021-2050)

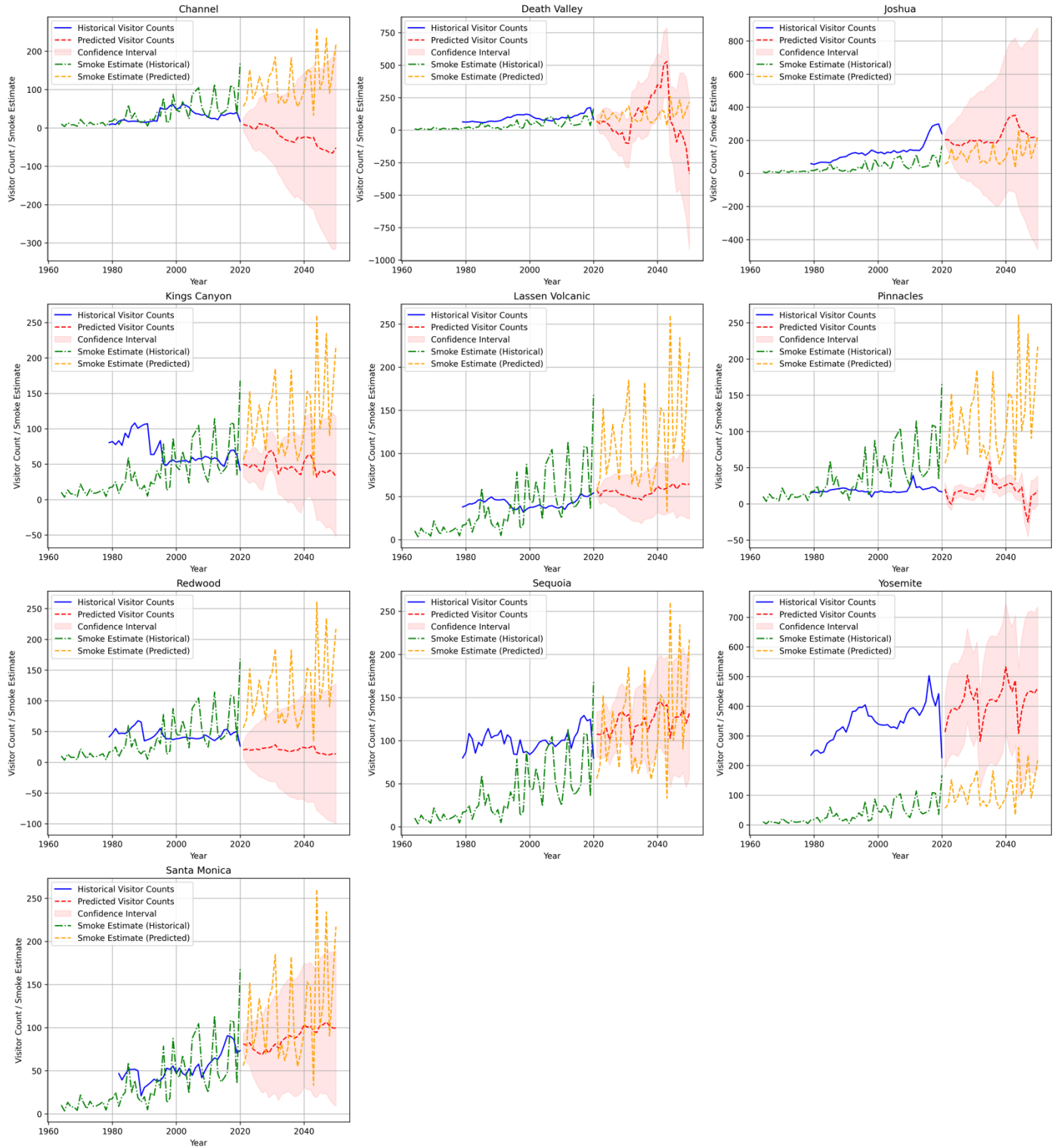


Fig 4.16

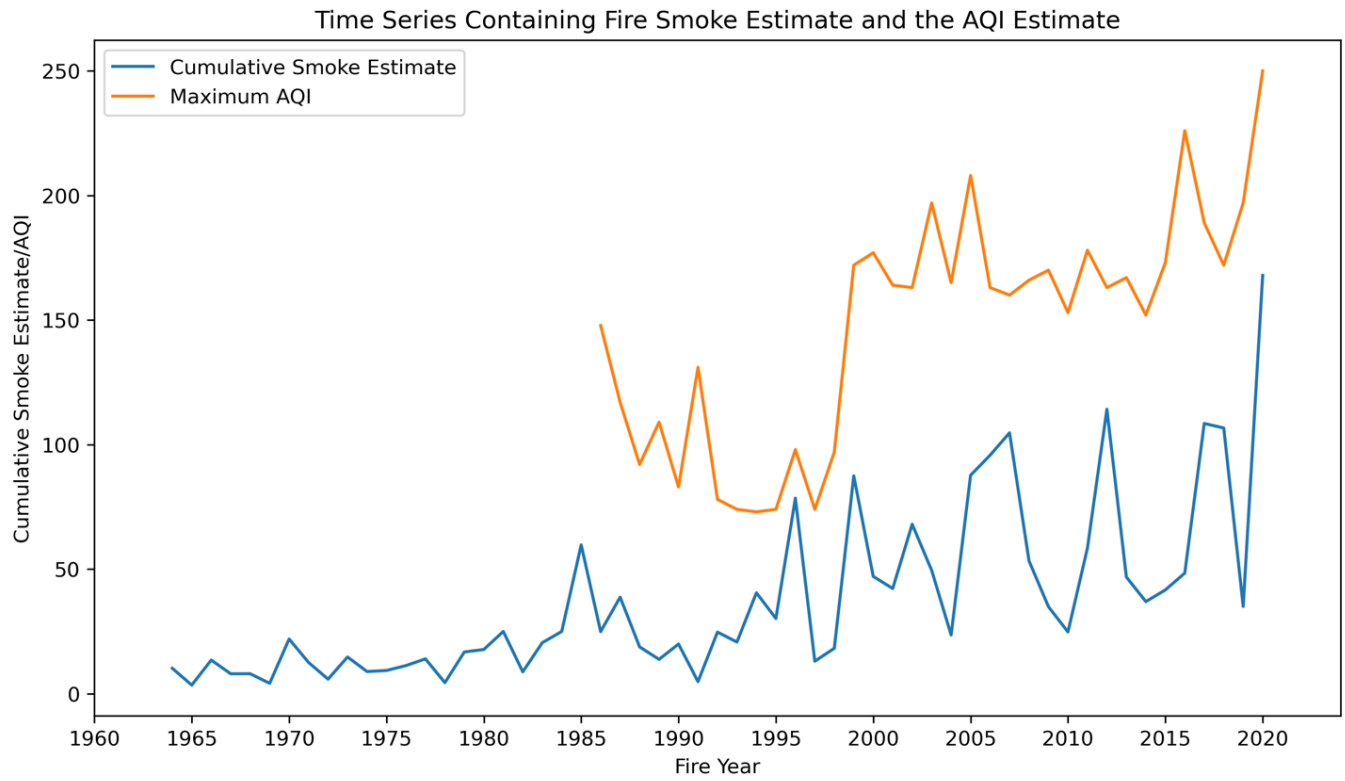


Fig 4.41

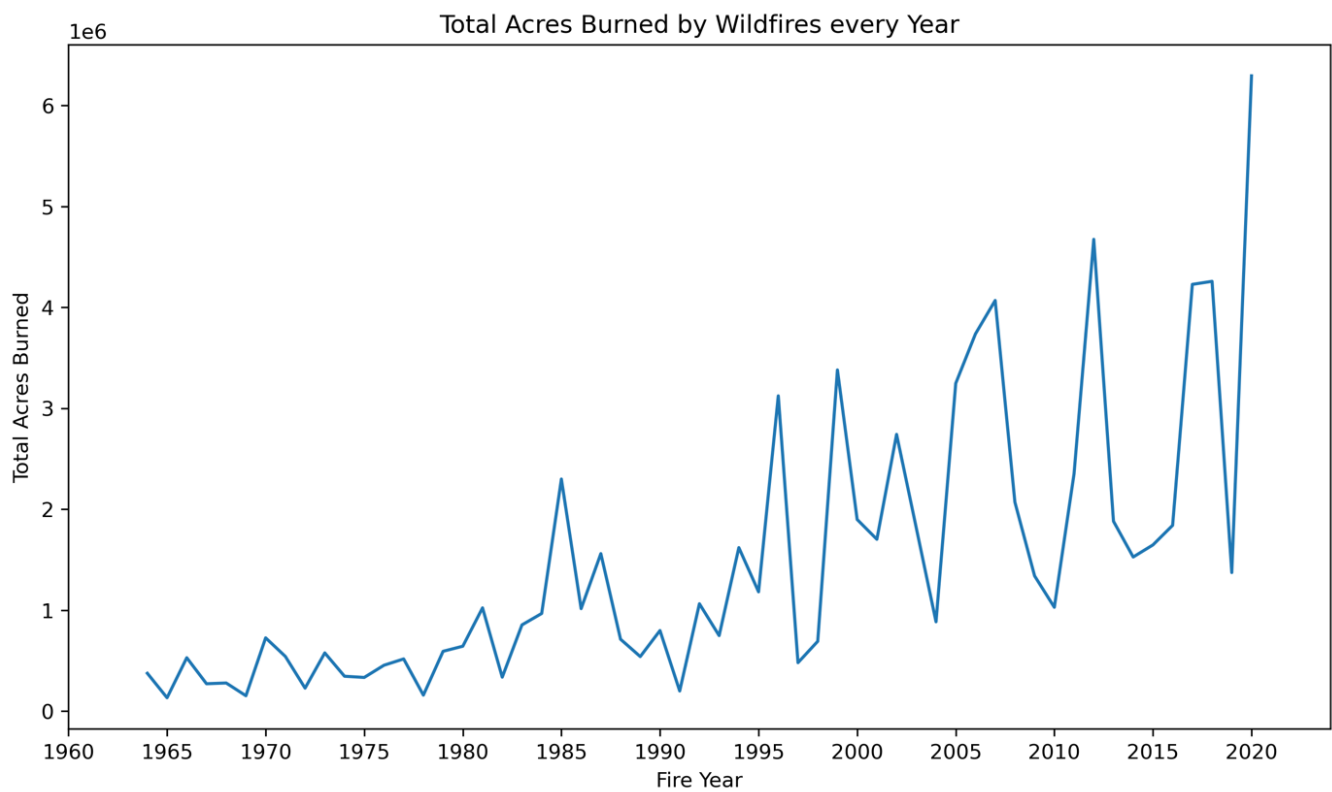


Fig 4.42

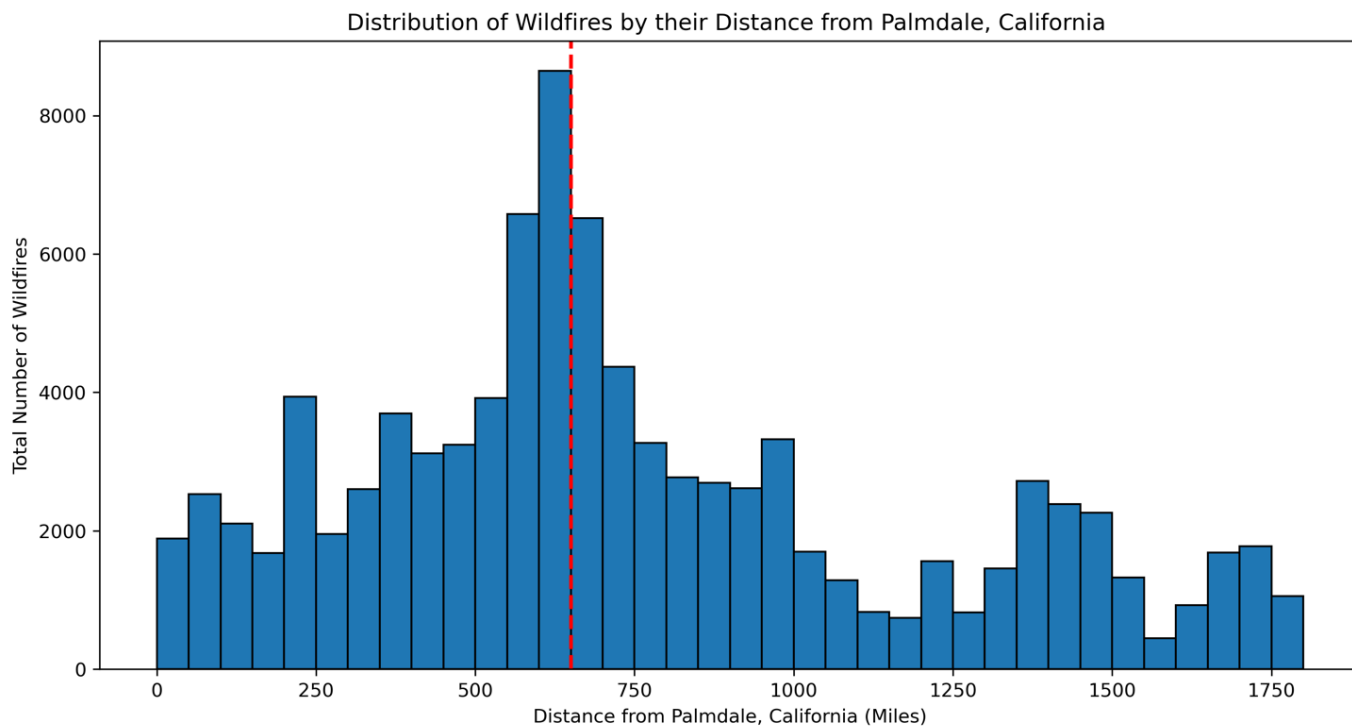


Fig 4.43

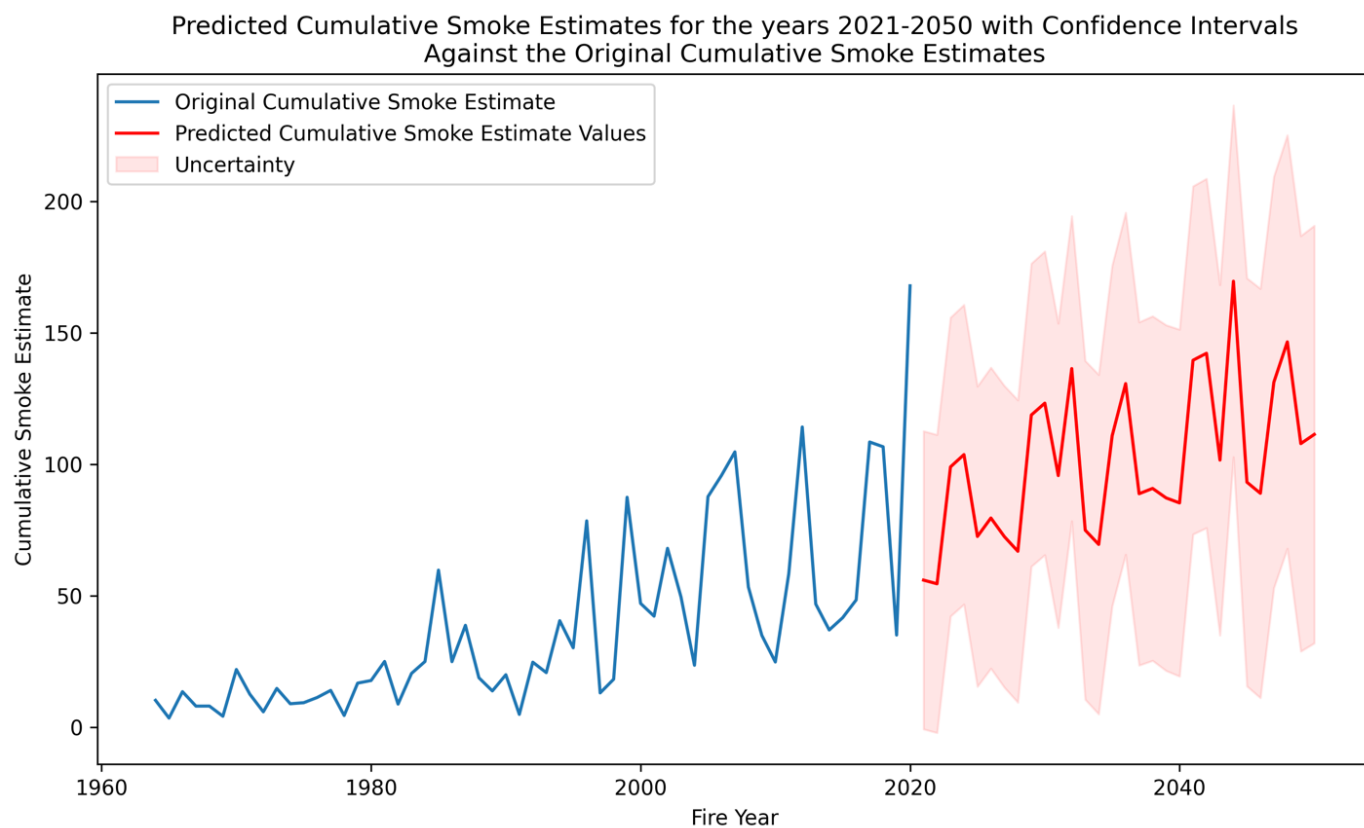


Fig 4.44