

Introduction

This analysis seeks to inform the citizens and council of Charleston about the significant impact of forest fires on the city, specifically examining the relationship between smoke from nearby fires and the Air Quality Index (AQI). The motivation behind this study stems from the urgency of addressing a tangible yet often underestimated problem. Forest fires, whether directly or indirectly, pose severe risks to public health and environmental stability. By understanding these repercussions, particularly the strain on healthcare systems and the well-being of Charleston's residents, this analysis contributes to addressing an unresolved research question in environmental health. It is crucial because the harmful effects of poor air quality, exacerbated by forest fires, demand informed policy-making and community preparedness.

Background/Related Work

Research in this area highlights that Charleston has historically experienced significant impacts from forest fires, underscoring its vulnerability to future events. Previous studies have examined both the frequency of forest fires and their devastating effects on South Carolina, where Charleston is located [1][2]. These findings emphasize the critical importance of analyzing the city's susceptibility to forest fire-related consequences.

The primary wildfire data used in this study was sourced from the [US Geological Survey's Combined Wildland Fire Polygons Dataset \(1800s-Present\)](#), available in both ArcGIS and GeoJSON formats. This dataset provides fire polygons, years, and acreage, offering a comprehensive view of wildfire activity across time. However, it lacks specific start and end dates for fires, which could limit the precision of certain analyses. The Air Quality Index (AQI) data was obtained via the EPA Air Quality System (AQS) API. This dataset includes detailed air quality measurements for Charleston, with a specific focus on PM2.5, the primary pollutant associated with wildfire smoke. These datasets were crucial in quantifying the environmental and health impacts of forest fires on Charleston.

A historical review of devastating fires in South Carolina informs the urgency of this analysis. Additionally, an ARIMA model [3] previously used to predict forest fires was adapted for this study. However, due to time constraints, a simplified version of the model was implemented for the project.

In Part 2 of the course project, the focus shifted to healthcare metrics to evaluate the impact of forest fires. Life expectancy and mortality rates were selected as key indicators to demonstrate the healthcare-related repercussions of fire events. Data for this analysis was sourced from the Institute for Health Metrics and Evaluation ([IHME](#)), covering the period from 2000 to 2019.

These prior studies, datasets, and models provided a robust foundation for developing the hypotheses and framework for this analysis, reinforcing its relevance to Charleston's residents and policymakers.

Methodology

The analytical methods were carefully chosen to ensure accuracy, relevance, and a focus on human-centered outcomes. For wildfire data, the Reader library was utilized for efficient extraction, and shortest distance calculations were employed to assess proximity to the fire periphery. These methods provided a precise understanding of smoke impact, a crucial element for evaluating how wildfires affect Charleston. Ethical considerations played a key role in this choice, as ensuring data accuracy and transparency was critical to maintaining trust in the findings and avoiding biases that could lead to misinterpretations or misinformed public decisions.

Before diving into analytical methods, extensive data cleaning processes were applied to ensure the reliability of the data used. This involved handling missing values through imputation or removal, correcting inconsistencies in data entries, and filtering out outliers that could skew the results. The data was also standardized to ensure uniformity across various metrics, making it easier to integrate and analyze different data sources. These preliminary steps were crucial in establishing a solid foundation for the subsequent analysis, ensuring the data was accurate and relevant to the study's objectives.

PM2.5 data was selected for air quality analysis due to its well-documented health effects and its primary role in air quality degradation during wildfires. By focusing on this pollutant, the analysis directly addressed human health concerns. To maintain relevance, filters were applied to the dataset, such as limiting the study to wildfire season (May to October) and a 650-mile proximity from Charleston. These filters ensured the study captured data directly impacting the local population while avoiding unnecessary complexity. This approach was ethically grounded in providing actionable insights tailored to Charleston's residents, without overgeneralizing findings to unrelated contexts or populations.

The smoke estimate calculation utilized a logarithmic scaling formula: $\text{fire size in acres} / \log(1 + \text{fire distance})$. This method emphasized the significant impacts of nearby fires while appropriately reducing the influence of more distant events. By adopting a realistic scaling approach, the analysis avoided overstating risks and focused on accurately representing threats to air quality, ensuring that the results remained both practical and credible.

Fire estimate is used as $\text{fire size in acres} / \log(1 + \text{fire distance})$

The ARIMA (AutoRegressive Integrated Moving Average) model was selected for its robustness in time-series forecasting and its adaptability to datasets with trends and seasonality. ARIMA models are particularly suited for data that show patterns over time and can be made stationary through differencing, which is critical for handling the non-stationary nature of air quality data influenced by wildfires. The model incorporates three main components: the autoregressive (AR) terms that express the variable's dependence on its previous values, the integrated (I) part involving differencing to make the data stationary, and the moving average (MA) terms that model the error terms influencing the observations. While a simplified version of the ARIMA model was used due to project constraints, it provided reliable predictions that could guide policy and preparedness efforts. The decision to use ARIMA was also informed by its widespread application in similar studies, ensuring the method was credible and its results comparable to related research. Furthermore, ARIMA's transparency and interpretability allowed findings to be communicated effectively to policymakers and the public, aligning with ethical principles of clarity and accessibility in research.

Finally, healthcare metrics such as life expectancy and mortality rates were analyzed to provide a tangible understanding of the long-term consequences of wildfires on human lives. These metrics were chosen because they are universally recognized indicators of public health and directly reflect quality of life and the effectiveness of healthcare systems. By analyzing trends in life expectancy and mortality, the study uncovered the potential effects of wildfire-induced air quality degradation on overall health, with a particular focus on vulnerable populations, such as children, the elderly, and individuals with preexisting conditions. This human-centered approach ensured that the analysis had real-world relevance, emphasizing its importance for policymakers and the residents of Charleston. The findings aimed to encourage proactive community preparedness, equitable healthcare planning, and interventions designed to mitigate the long-term risks associated with wildfires and air quality degradation.

Additionally, to understand the relationships between different variables, such as fire size, proximity, and PM2.5 levels, the Pearson correlation coefficient was employed. This statistic measures the linear correlation between two variables, providing a value between -1 and 1. A value of 1 indicates a perfect positive correlation, where as one variable increases, the other also increases. A value of -1 indicates a perfect negative correlation, where as one variable increases, the other decreases. A value of 0 suggests no linear correlation. This metric was crucial in identifying significant predictors of air quality degradation and helped quantify the strength and direction of these relationships, thereby contributing to more informed, data-driven decisions in the study.

Findings

Geographical Impact of Forest Fires:

Analysis reveals that a significant number of forest fires have occurred within a 200–300-mile radius of Charleston, as shown in **Figure 1**. This spatial concentration of fires highlights the city's vulnerability to wildfire smoke, reinforcing the importance of addressing its long-term impact on air quality and public health.

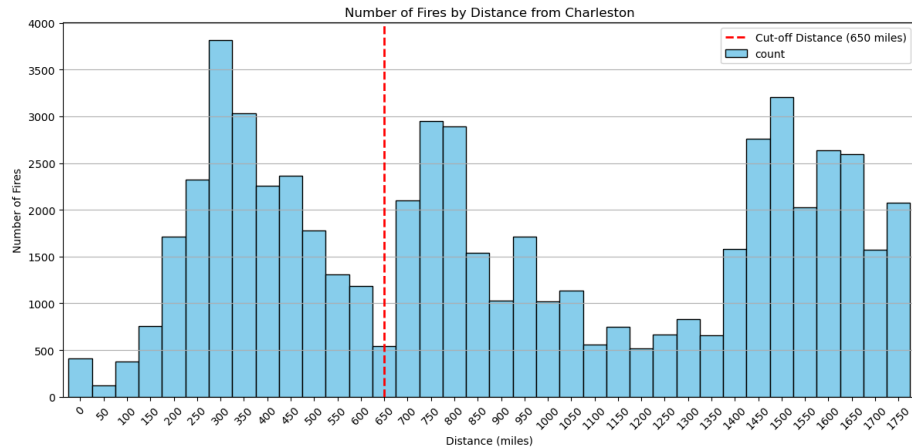


Figure 1

Air Quality and Smoke Estimate Correlation:

A positive correlation (correlation coefficient = 0.35) was observed between the Air Quality Index (AQI) and smoke estimates derived from wildfire data (refer to **Figure 2**). While the correlation is moderate, it indicates that smoke from nearby forest fires has contributed to the deterioration of Charleston's air quality, underscoring the environmental risks posed by such events.

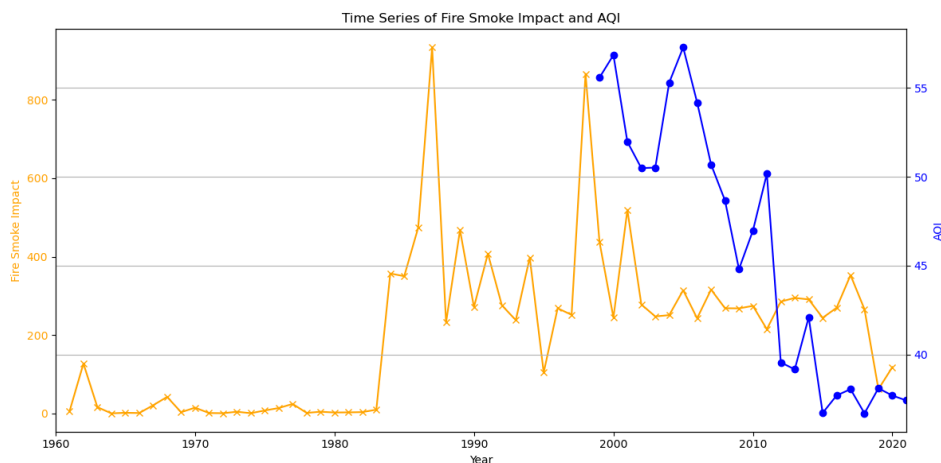


Figure 2

Smoke Estimate Predictions:

Forecasts generated by the ARIMA model project an upward trend in smoke estimates for the next two decades (**Figure 3**). This suggests that Charleston may face an increase in the adverse effects of wildfires, warranting proactive measures to mitigate the

potential risks. These findings serve as a warning for citizens and policymakers to prepare for worsening air quality in the coming years.

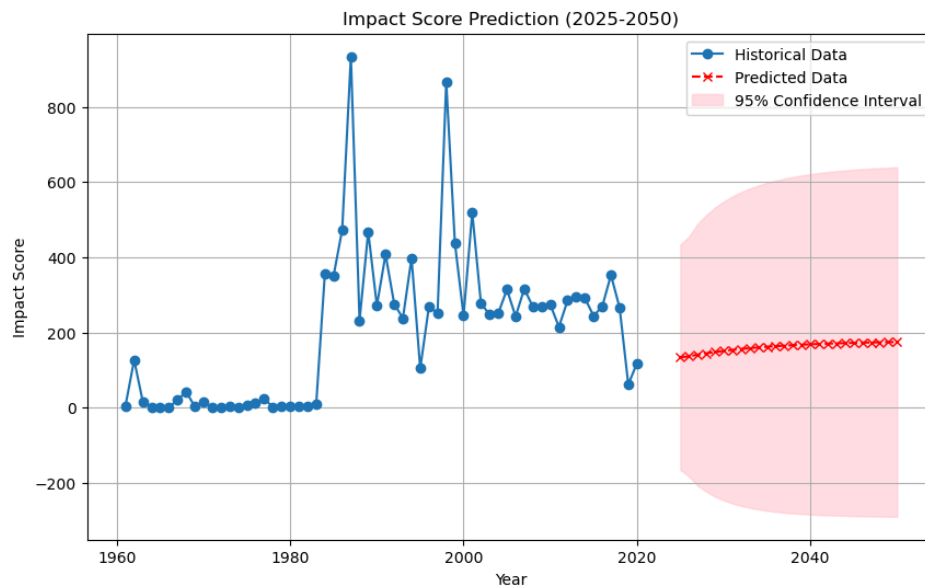


Figure 3

Healthcare Impact:

The analysis of normalized smoke estimates against healthcare metrics revealed significant correlations:

- **Negative correlation with life expectancy:** As smoke estimates increased, life expectancy showed a decline, indicating the detrimental health effects of prolonged exposure to wildfire smoke.
- **Positive correlation with mortality:** An increase in smoke estimates corresponded to a rise in mortality rates (**Figure 4**). This further demonstrates the critical health challenges posed by forest fires, particularly for vulnerable populations.

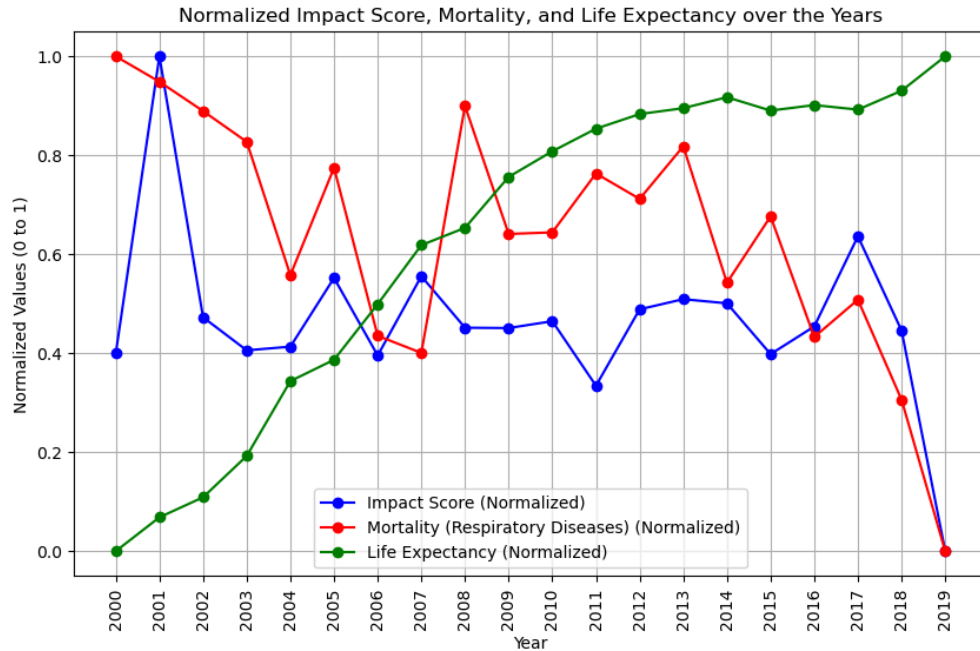


Figure 4

Mortality and Life Expectancy Predictions:

Using the ARIMA model, predictions for future mortality and life expectancy trends were attempted. However, these results were inconclusive due to the simplicity of the model and the presence of confounding factors, such as socioeconomic variables, healthcare access, and other environmental determinants (**Figure 5**). These findings highlight the need for more sophisticated modeling and additional data to fully understand the long-term impacts of forest fires on public health.

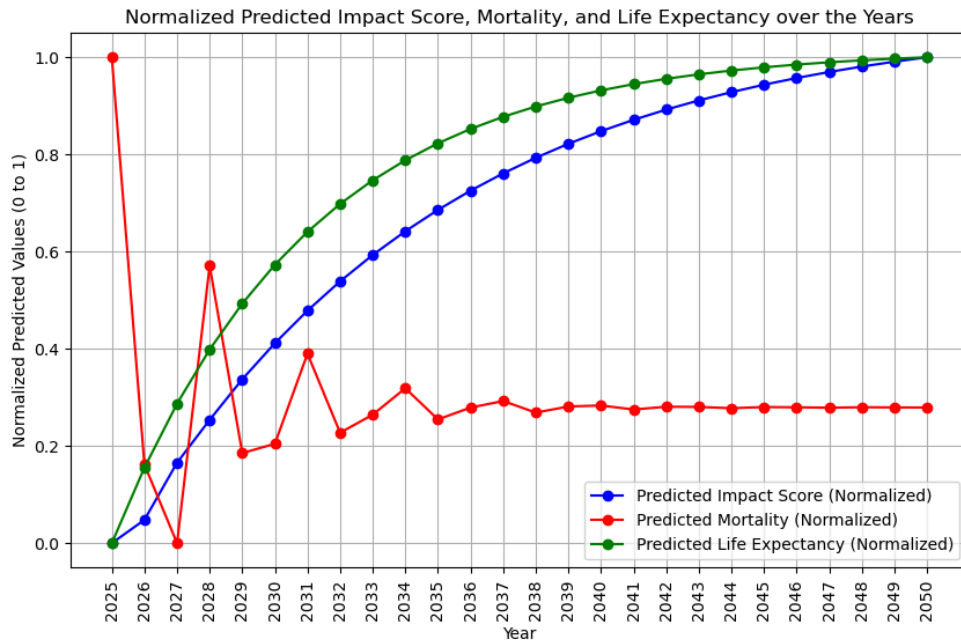


Figure 5

Discussion/Implications

The findings of this comprehensive project are vitally important for a thorough understanding and proactive addressing of the city of Charleston’s vulnerability to forest fires and their extensive, long-term impacts on air quality and public health. The analysis has established a definitive correlation between the presence of wildfire smoke and a concerning decline in life expectancy, coupled with increasing mortality rates. These crucial insights highlight the urgent need for immediate action to safeguard the residents of Charleston. The projections indicating a significant increase in smoke levels over the coming two decades present a dire warning: without timely and effective intervention, the detrimental effects on public health and overall quality of life will only intensify.

These findings were meticulously developed through the application of human-centered data science principles, ensuring that the entire project was deliberately designed to prioritize the well-being and safety of Charleston’s residents. Key metrics such as life expectancy, mortality rates, and air quality indices were strategically chosen for their direct relevance to human health and their potential to resonate meaningfully with both policymakers and the broader public. The choice to

focus on these specific indicators reflects a profound commitment to addressing tangible, day-to-day concerns faced by the community. Moreover, the study employed transparent and ethical methodologies, utilizing publicly available datasets to guarantee accessibility and uphold privacy standards, and communicating the results in a manner that stakeholders can readily comprehend and utilize.

In response to these critical findings, it is imperative that the Charleston city council takes decisive action. This should include the implementation of stricter air quality monitoring systems and the enforcement of robust wildfire-related emergency response protocols. Securing state and federal funding to support public health campaigns and resilience initiatives will be essential. Furthermore, local policymakers must engage collaboratively with environmental and health experts to develop and implement long-term strategies aimed at mitigating the health and environmental impacts of forest fires.

The leadership roles of the city manager and mayor are crucial in spearheading efforts to enhance Charleston's resilience to these environmental threats. This includes prioritizing investments in vital infrastructure upgrades, such as the enhancement of air filtration systems in schools, hospitals, and public buildings. Additionally, exploring urban greening projects, which serve to reduce urban heat islands and bolster environmental resilience, should be considered a priority. Alongside these infrastructural and environmental initiatives, launching public education campaigns will empower residents to adopt practices that reduce their exposure to wildfire smoke and advocate for stronger local and regional protective measures.

Furthermore, residents themselves play a crucial role in their own safety. They can take personal protective measures, including the use of air purifiers, remaining indoors on days when smoke levels are particularly high, and wearing protective masks when necessary. Community engagement is equally vital, as widespread participation in preparedness programs can significantly strengthen the city's overall response to wildfire risks and enhance community resilience.

The timeline for implementing these measures is critically short—with Charleston facing an estimated window of 2–5 years to enact concrete measures before the projected increase in smoke levels exacerbates the existing public health challenges. This urgency is deeply informed by human-centered design principles, which emphasize proactive

and equitable solutions designed to minimize harm to the most vulnerable populations, such as children, the elderly, and those with preexisting health conditions.

By integrating human-centered principles throughout the project—from the careful selection of relevant metrics to ensuring ethical data use and clear, effective communication—the study delivers actionable insights that prioritize equity, relevance, and the long-term welfare of Charleston’s residents. This focused approach ensures that the findings and recommendations will support decision-making that not only benefits the community but also fosters a sustainable and healthier future for all involved.

Limitations

Data Coverage and Completeness:

- For AQI estimates, only PM2.5 concentration was considered. While PM2.5 is a primary pollutant associated with wildfire smoke, other particles (e.g., PM10, NOx, and Ozone) were not included, which may have led to an incomplete assessment of air quality degradation.
- Charleston’s AQI data was available only from 2000 onward, limiting the analysis to approximately 20 years of data. This restricted timeframe may not fully capture long-term trends or earlier correlations between smoke estimates and air quality.
- Data for life expectancy and mortality was similarly limited, spanning only 2000 to 2019. This constrained dataset might overlook long-term effects or historical variations, potentially impacting the robustness of conclusions.

Data Handling and Assumptions:

- For AQI calculations, a simple average of the available data was used, which does not account for seasonal variations or potential outliers. More sophisticated data processing techniques could improve result accuracy.
- Missing values in some datasets may have been handled with imputation or exclusion, potentially introducing bias into the analysis.

Correlation and Statistical Limitations:

- The analysis relied on Pearson's correlation to evaluate relationships between variables, which assumes linearity. While the observed correlation coefficients provided insights, they do not prove causation or capture potential nonlinear relationships.

ARIMA Model Limitations:

- The ARIMA model used for smoke impact predictions assumes stationarity and linear trends. Although differencing was applied to address stationarity, the model may not fully account for complex, nonlinear trends in wildfire activity or smoke dispersion. Additionally, external factors like changing climate patterns or urban development, which could influence smoke levels, were not included in the model.

Influence of External Factors:

- Socioeconomic conditions, access to healthcare, and demographic variables (e.g., age, race, and income) were not explicitly included in the analysis of mortality and life expectancy. These factors likely contribute significantly to health outcomes and could affect the observed correlations with smoke estimates.

Model Accuracy and Generalization:

- Predicting respiratory mortality and life expectancy using smoke exposure data introduces complexities, as these health outcomes are influenced by a range of confounding variables. Further model tuning, validation, and inclusion of additional factors would be necessary to improve predictive accuracy and generalizability.

Data Accessibility and Licensing:

- Some datasets, particularly those related to local health records or detailed wildfire activity, may have licensing restrictions or gaps in availability. Access to more comprehensive datasets could significantly enhance the depth and reliability of the analysis.

This comprehensive study was meticulously designed to address the critical questions concerning the impact of forest fires on Charleston, specifically investigating the intricate relationship between wildfire smoke and air quality, and how these factors collectively influence broader health outcomes such as life expectancy and mortality rates. The hypotheses put forth posited that wildfire smoke significantly contributes to the degradation of air quality, and that these deleterious effects are closely correlated with a decrease in life expectancy and an increase in mortality rates.

The empirical findings from the study substantiated these hypotheses to a notable extent. A moderate positive correlation (0.35) was observed between smoke estimates and the Air Quality Index (AQI), unequivocally demonstrating that wildfire smoke plays a significant role in the deterioration of air quality. Furthermore, the analysis revealed that smoke estimates negatively correlated with life expectancy and positively correlated with mortality rates when normalized, thereby reinforcing the severe health risks posed by prolonged exposure to wildfire-related pollutants. Predictive analyses utilizing an ARIMA model projected a likely increase in smoke exposure over the forthcoming two decades, highlighting the critical urgency for proactive intervention. However, certain limitations were encountered, including gaps in data coverage, the constraints of statistical methods employed, and the influence of external variables, all of which underscore the necessity for continued research to fortify these conclusions.

The study also profoundly contributes to the understanding of human-centered data science by showcasing how data-driven methodologies can be adeptly crafted to prioritize human well-being and equity. By selecting life expectancy, mortality, and AQI as the core metrics, the study ensured that the findings directly addressed significant human impacts. Ethical considerations were meticulously adhered to, featuring transparency in the methodological approaches, reliance on publicly accessible datasets, and the clear communication of results, all integral to the study's framework. Moreover, the analysis underscored the importance of tailoring insights to yield actionable outcomes for Charleston's residents, policymakers, and community leaders, demonstrating the potential of human-centered data science to guide informed decision-making and cultivate sustainable, equitable solutions.

Reflecting on my personal learning journey through the execution of this project, I have gained invaluable insights into the practical application of data science in addressing real-world issues. The challenge of translating complex data into comprehensible, actionable information has enhanced my analytical and communication skills, underscoring the importance of clarity and precision in data interpretation. The ethical dimensions explored, such as data privacy and the potential consequences of misinformation, have deepened my understanding of the responsibilities that accompany the role of a data scientist. This project has not only advanced my technical proficiency but also reinforced the critical need for ethical considerations and human-centric approaches in data science practices. These learnings will undoubtedly shape my future endeavors in the field, driving me to pursue projects that foster positive social impact and advocate for the ethical use of data.

References

1. https://firststreet.org/city/charleston-sc/4513330_fsid/fire
2. <https://www.scfc.gov/protection/fire-burning/fire-resources/wildfire-history/>
3. <https://www.sciencedirect.com/science/article/abs/pii/S2352938522000696>

Data Sources

1. <https://ghdx.healthdata.org/record/ihme-data/united-states-causes-death-life-expectancy-by-county-race-ethnicity-2000-2019>
2. <https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81>