Modeling the Spread of Wildfires and Assessing Urban Energy Infrastructure Vulnerability

Nicholas Cheung, Aziz Shaik
2 October 2025
CS 5834: Intro to Urban Computing

Group Members

Nicholas Cheung	MEng Student in Computer Science & Applications. DA&AI Concentration
Aziz Shaik	MEng Student in Computer Science & Applications. DA&AI Concentration

Introduction

Wildfires have become increasingly frequent across the United States due to increasing global temperatures and droughts, presenting a growing threat to natural ecosystems and urban communities. Since the 1980s, the average number of acres burned in wildfires has almost tripled [1], prompting the need for greater research into how these fires can be controlled and how their potentially disastrous effects can be mitigated.

In this project, we aim to be able to model the spread of wildfires to predict the scale and potential effect of a particular fire. Furthermore, we want to be able to determine the impact that the fire could have on nearby urban populations, by assessing the vulnerability of critical energy infrastructure that large urban populations rely on.

Planned Work

To begin, we will be using the spatial wildfire occurrence data to create a graph that creates edges on the basis of spatio-temporal proximity between distinct ignition points to observe the spread of wildfires (e.g. within 5 miles and 24 hours). This will form the basis of our project and give insights into how quickly a fire might spread and where it might spread to. We can use metrics such as centrality or clustering coefficients to identify key ignition locations to give a sense of which locations tend to be under the greatest threat of wildfires.

In order to assess the vulnerability of energy infrastructure, we will create a separate network of crucial energy infrastructure, namely power plants and substations, with transmission lines forming edges between substations. With this, we can see which substations would be impacted and the downstream effects through the transmission lines. These two graphs would be overlaid on top of each other to see the proximity of the wildfires to key infrastructure and to relay the risk of the wildfires to each of the identified components of the power grid.

We can then use several unsupervised clustering algorithms such as k-Means or DBSCAN to find clusters of substations or power plants for later interpretation. Analysis of these clusters could reveal trends with regards to the types of grid assets that are most at risk, and which ones are least at risk. With this data and analysis, it could allow for better and safer planning of future energy asset construction, as well as inform resilience planning decisions for existing infrastructure.

Preliminary Research

Wildfire behavior has been recognized as the outcome of a complex interplay between fuel availability, weather conditions, and climate. Traditional methods provide important insights into these systems, but fall short of capturing the complex non-linear processes that drive the spread of wildfires [2]. Recent studies have shown using machine learning techniques to estimate the importance of these environmental factors proves to be useful in predicting where future fires might start and how climate change and wildfires are connected.

One article explains how support vector machines and ensemble methods have shown promise with environmental datasets, but seem to struggle with the dynamic nature of wildfires [3]. Deep learning approaches, such as convolutional neural networks and convolutional recurrent networks, demonstrated the ability to capture spatio-temporal complexities of wildfire data that the previous models had missed.

At the same time, we see published articles bringing attention to the importance of finding vulnerabilities in urban energy infrastructure. One study explains the impact power supply downtime has on health and mortality rate [4]. Another paper describes the vulnerabilities in urban energy infrastructure through three stages in the novel COVID analogy [5]. Stage 1 (virus entry) describes an extreme event disrupting a component of the energy ecosystem. Stage 2 (propagation through respiratory system) describes an electricity sector failing in a whole area. Stage 3 (heart, kidneys, digestive system, and other organs) describes penetration beyond the energy ecosystem and to other interconnected systems such as transportation, water supply, and communication. This explains why building a resilient energy infrastructure is pivotal for the success of an urban system. Through this research we hope to find weaknesses in energy infrastructure.

Datasets

- **Dataset**: Spatial wildfire occurrence data for the United States, 1992-2020
  Description & Use: This dataset is an SQLite file that contains information about wildfires from 1992 to 2020, with specific location, time, and cause data for each fire, determined based on separate ignition events. Our main interest with this dataset will be the time and location of each fire, which will then be compared to other fires that are close in time and space to form edges.
  `https://www.kaggle.com/datasets/braddarrow/23-million-wildfires`

- **Dataset**: US Electric Power Plants
  Description & Use: This data is pulled from the Homeland Infrastructure Foundation Level Database (HIFLD) maintained by the Department of Homeland Security (DHS). Structured as a CSV file, it provides valuable longitude and latitude data for electric power plants in the US that can allow us to assess the threat of wildfires based on the wildfire occurrence data.
  `https://www.kaggle.com/datasets/behroozsohrabi/us-electric-power-plants`

- **Dataset**: US Electric Power Transmission Substations
  Description & Use: This data is also from the HIFLD, with a CSV file structured similarly to the electric power plant data. Substations form another crucial component of the

energy infra structure system so this allows us to observe the impact of wildfires on this as well. https://www.kaggle.com/datasets/behroozsohrabi/us-electric-power-transmission-substations

- **Dataset**: US Electric Power Transmission Lines
  Description & Use: This data is also from the HIFLD, with a CSV file structured similarly to the electric power plant data. Although this does not have specific location data, it contains the substation connections for these lines. We would have to make our own connections between substations with this dataset.
  https://www.kaggle.com/datasets/behroozsohrabi/us-electric-power-transmission-lines

References

[1] Wildfire data in the US (NIFC) – processed by Our World in Data, “Acres burned - comparable data (NIFC)” [dataset]. Wildfire data in the US (NIFC) [original data], National Interagency Coordination Center; National Interagency Fire Center. Retrieved: Sep. 20, 2018. [Online]. Available: https://www.nifc.gov/fireInfo/fireInfo_stats_totalFires.html

[2] Wang, S. S.-C. and Wang, Y.: Quantifying the effects of environmental factors on wildfire burned area in the south central US using integrated machine learning techniques, Atmos. Chem. Phys., 20, 11065–11087, https://doi.org/10.5194/acp-20-11065-2020, 2020. https://acp.copernicus.org/articles/20/11065/2020/

[3] Andrianarivony, H.S.; Akhloufi, M.A. Machine Learning and Deep Learning for Wildfire Spread Prediction: A Review. Fire 2024, 7, 482. https://doi.org/10.3390/fire7120482 https://www.mdpi.com/2571-6255/7/12/482

[4] J. A. Sathaye et al., ”Rising Temps, Tides, and Wildfires: Assessing the Risk to California’s Energy Infrastructure from Projected Climate Change,” in IEEE Power and Energy Magazine, vol. 11, no. 3, pp. 32-45, May-June 2013, doi: 10.1109/MPE.2013.2245582. keywords: Global warming;Globalization;Energy management;Environmental factors;Terrestrial atmosphere;Temperature measurement;Environmental management;Supply and demand, https://ieeexplore.ieee.org/document/6506884

[5] A.T.D. Perera, Tianzhen Hong, Vulnerability and resilience of urban energy ecosystems to extreme climate events: A systematic review and perspectives, Renewable and Sustainable Energy Reviews https://doi.org/10.1016/j.rser.2022.113038