Urbanization and Fire Risk

Patrick Chang
Henry M. Gunn High School, 780 Arastradero Rd, Palo Alto, CA 94306

Abstract
Fires, whether wild or human-caused, are destructive and fairly common. A main method of damage mitigation comes from halting the spread of fires, affected by factors such as fire fuel availability, human intervention, and the land condition. People are capable of altering all of those factors, most effectively with urbanization and land development. By understanding the relationship between urbanization and fire risk, city planning can be altered to optimally prevent fire spread. Since fire frequency is strongly attributed to erratic factors including human activity and lightning, fire risk is defined as the probability of the spread in the area of a fire in this study. Urbanization includes many factors which influence fire risk, such as building concentration and infrastructure including roads and fire stations. By using historical land usage maps—generally from censuses—to separate regions to specific levels of urbanization and finding the areas of fires within those regions, regression methods including Neural Networks are used to search for correlation between urbanization and fire risk.

Key Words: Fire Risk, Urbanization, Regression, Correlation, Neural Networks

1. Introduction

The inspiration for this project comes from the August Lightning Complex in 2020 that rolled over most of the Bay Area of California, being the cause of multiple large wildfires near my hometown. Even when the fires shrouded the sky orange from smoke, the fires never seemed to reach any more densely populated suburban regions, causing me to be intrigued about the effects of human development on fire spread. I witnessed this firsthand later in the year when hiking near Napa Valley and seeing the fire scar on the mountain slopes next to the valley, yet the valley itself appeared to be untouched—either by firefighter intervention or quick repair.

Fire vulnerability is increased by many factors: lack of forest management, human development, and the climate or specific weather conditions. The first three can create or allow more growth of possible fuel for fires whilst the latter increases possible risk of causing an accidental fire. Human development is the simplest factor to control for the purpose of decreasing fire risk, thus making it the main point of my research. Fire risk is assessed with the total area of fires, as the causation of fires relies on too many unpredictable factors, although humans are reputed to cause many more accidental fires. For example, campfires can easily cause a forest fire without human supervision. Also, human development creates significantly more factors to account for in fire risk: the increased human accidents and the primary goal of firefighters to stop destruction of property, most found in urban areas.

Firstly, fire stations are more commonly found within more developed areas, so there would be more resources available to halt fire spread or prevent fires with good forest management. Also, firefighting prioritizes human lives and property over wilderness, focused on mitigating fire spread towards urban areas. Roads serve as evidence of infrastructure that can be used to effectively distribute firefighting resources, also occasionally serving as firebreaks. Firefighters can far more easily deploy resources from the solid roads. Without roads, they may need to traverse often rugged terrain as many fires start within mountainous regions. Roads may also be evidence of superior forest management.

However, smaller amounts of development without sufficient fire protection can instigate or further spread fire through creating more potential fire fuel.

To somewhat isolate human development as a factor, other factors such as climate must be kept similar. Data for fires and urbanization is focused on California due to the frequency of large fires along with consistent statewide building codes and climate. To further keep similar climates, data from certain counties is removed for extreme climates usually concerning the amount of precipitation. The more southern and inland counties have far more arid climates while the two north coast counties receive too much precipitation. Even with consistent climate, other factors are also difficult to control, for example, the types of vegetation and how mountainous the region is. Mountainous terrain creates more unpredictability for wildfires.

2. Methods

2.1 Data Sources

Data on fires and urbanization is all from California to create consistency for building codes and climate.

Figure 1 shows the counties that were selected for this study. The similarities of the weather conditions within these counties will attempt to limit factors other than urbanization to affect fire risk.

Fire data is retrieved from CAL fire's website [1] for the years from 1970 to 2020. CAL fire's data consists of all fires of over 10 acres and prescribed burns used for forest management, the latter of which are omitted since they are intentional fires. Example of fire perimeters data is shown in Figure 2. The red outlines all the fires of 2020, whereas other grey perimeters are of fires in years before the year of 2020. Prescribed burns are not shown due to them being intentional, for the purpose of limiting potent fire fuel. Due to the limits of urbanization data, fire data is only used from 1970 to 2020.

Urbanization data including road density data and housing density data were extracted from two sources: US census and a recent study on wildland urban interface (WUI) [2], where US decennial census data is compiled as housing density, wildland urban interface, and land usage data. WUI represents the transitioning of wilderness to urban areas, while the rest are split into non-WUI vegetated (e.g., farmland) and non-vegetated, such as in large cities. These are split into their respective housing densities. The provided data includes vegetation and agriculture land; however only housing density percentages and housing unit data are used in this study.

Road density (total road length normalized by the land area) can also be a way of classifying urbanization, where more roads equal more urbanization along with possible land uses. As stated previously, roads themselves can also affect fire spread, so road density is used to predict fire risk in this study.

Figure 1: California Counties within the red boundary, where there is the most similar weather, are used for this Study.

Based on these raw datasets (Fire and Urbanization), two datasets were constructed for this study: county-based dataset and fire incidence-based dataset. The county-based dataset aggregates fire and urbanization data county by county. The fire incidence-based dataset aggregates fire and urbanization data based on each individual fire incidence.

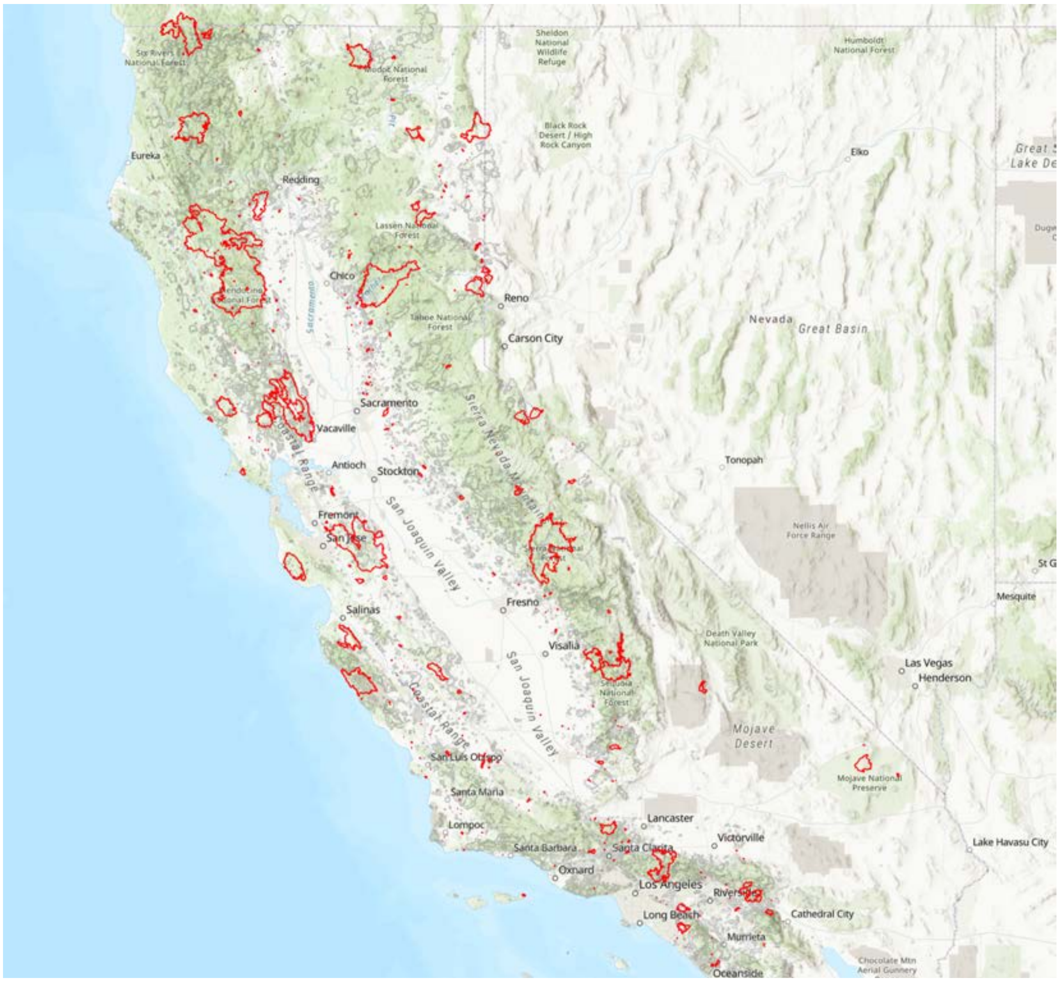

Figure 2: Example of fire perimeters data (2020)

2.2 Data Sets Construction

The dataset construction leverages on the use of open-source GIS software [3], which provides the capability of analysing spatial and geographic data. For example, to construct the county-based dataset, a fire perimeter map (See Figure 4) overlays a county map of California. For each county, one can clip the overlaid map to extract fire data for each clipped layer. Figure 4 shows fire spread in Napa County in 2020. Additional maps such as the US census road maps can overlay on top of the overlaid map. Figure 5 shows the clipped map resulting from overlapping fire map, county map, and US census road map for Napa County.

For the first dataset, data collection consisted of clipping each year's fire data with the specified counties then finding the total fire area in the county. WUI Urbanization data is already split into counties with percentages and areas of certain housing densities, which can be combined into a single value of housing density.

Roads that signify certain land uses such as agriculture is another way to understand fire risk. For road data, a road map layer undergoes the same process as the previous fire data, except all features are summed as total road length instead of area. Lastly, all data is divided by a calculated county area to find the housing, road, and fire densities so smaller counties can be compared with larger counties.

For the second dataset, a similar process is used to extract the dataset with county-based clipped layers replaced by fire incidence-based clipped layers. In the first dataset, fire density is used as the fire risk measurement. Fire risk in the second dataset is replaced by the total fire area of each incident.

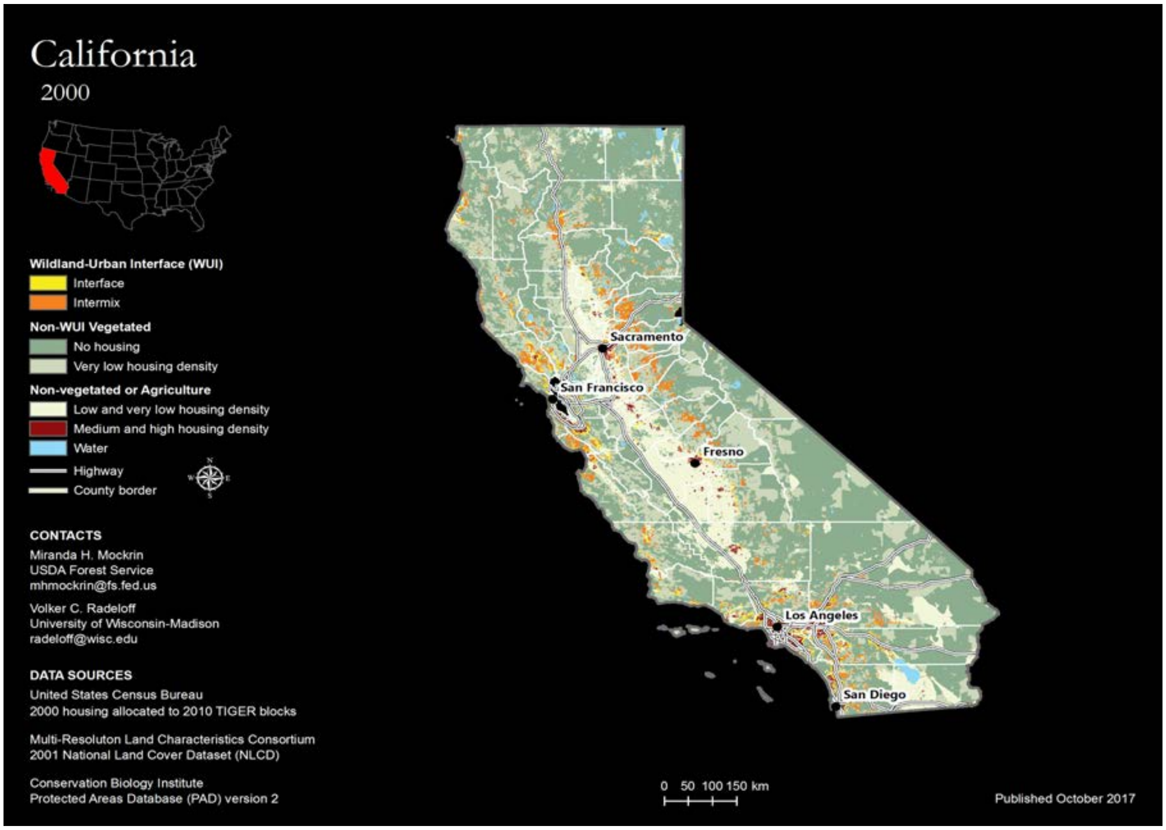

Figure 3: Urbanization Dataset (California Example)

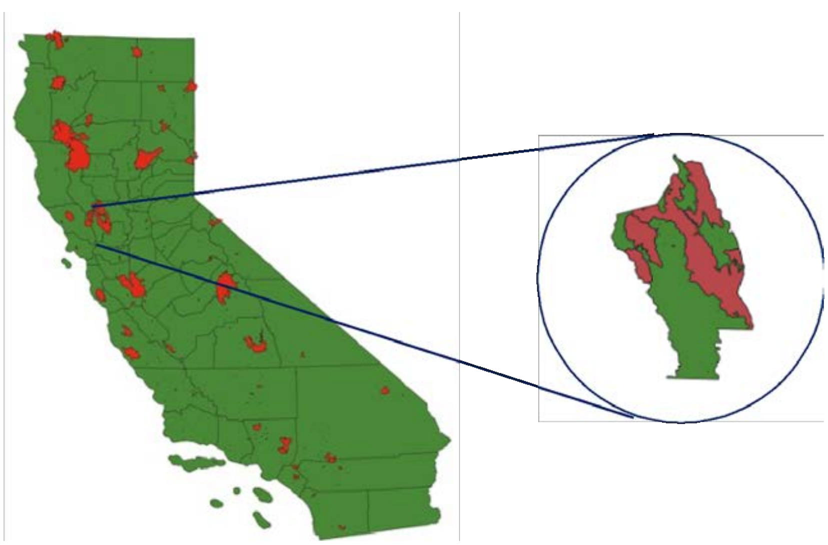

Figure 4: Fire Spread in Napa County, 2020

Detailed housing data is provided by the US census every 10 years, which is compiled to housing densities and specific land usage types. The data only includes years 1990, 2000, and 2010, thus linear interpolation and extrapolation are used to match the fire data in the county-based first dataset. Figure 6 shows the housing density data from 1970-2020 after interpolation and extrapolation. US census Road Density data is available from 2010-2020. Similarly, Figure 7 shows the interpolation/extrapolation results of the total road length for years 1970 to 2020. The second dataset is fire incidence-based; therefore, no interpolation/extrapolation was used to expand the dataset.

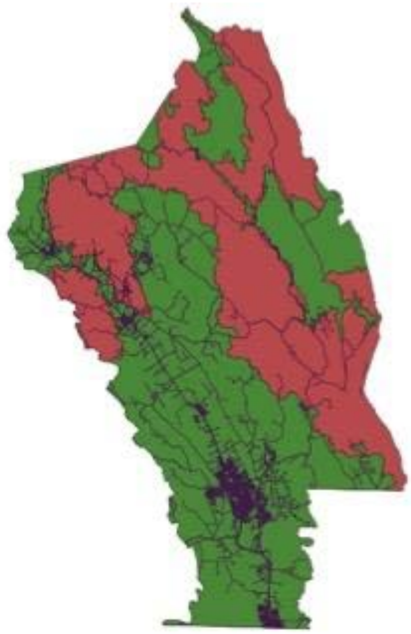

Figure 5: Fire Spread and Roads from Napa County in 2020. Road data extracted from the 2020 US Census.

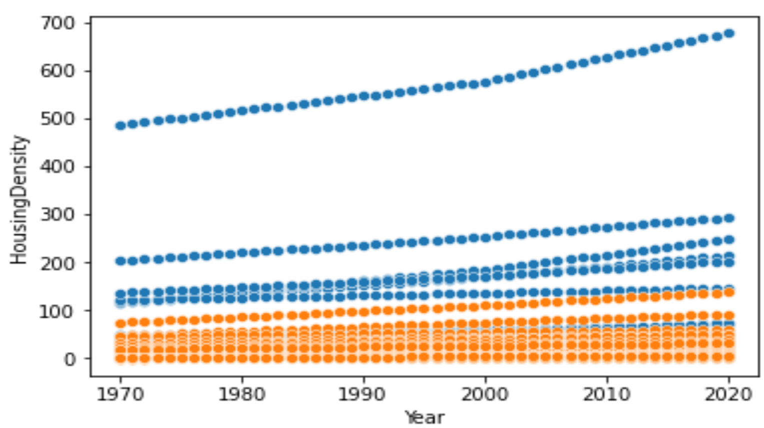

Figure 6: Housing Density Data after Interpolation and Extrapolation

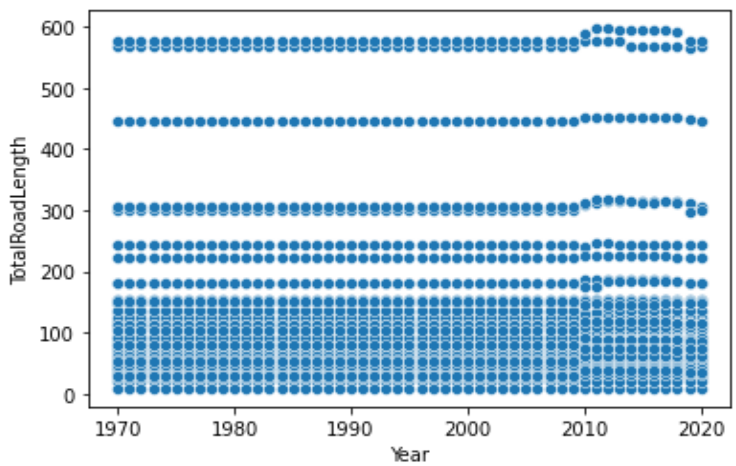

Figure 7: Total Road Length Data Extrapolated for 1970-2020

2.3 Linear Regression

Linear regression was used initially to find how correlated urbanization factors such as road density, housing density, and WUI housing density is to fire density using the first dataset, to provide insight on whether urbanization was an effective factor for fire risk. The road density is defined as the total road length normalized by the land area. The housing density is the number of housing units normalized by the land area. The WUI housing density is defined as the number of housing units in the WUI region normalized by the WUI land area. The fire risk is defined as the fire spread normalized by the land area. Linear regression assumes that each urbanization factor is linearly correlated to the fire risk. The fire risk can be predicted with: $fire\_risk = w \cdot urbanization\_factor$, where w is a scalar number.

2.4 Multiple Linear Regression

Since fire risk may not just be impacted by one urbanization factor alone, multiple linear regression was used with the assumption that the fire risk is linearly correlated to multiple independent factors. Both road density and housing density are assumed to be linearly correlated to fire risk. Based on the multiple linear regression, the fire risk is predicted by the following equation:

$$fire\_risk = w_1 \cdot road\_density + w_2 \cdot housing\_density,$$

where w_1 and w_2 are both scalar numbers.

Multiple linear regression was also used to compare the correlation between the fire risk and road density and the correlation between the fire risk and housing density.

In the Results Section, we will learn that fire risk and urbanization factors are mostly linearly correlated in the logarithmic scale.

2.5 Neural Networks

Since the correlation between fire risk and urbanization factors may not be linear, neural networks were used to apply non-linear fits to predict fire risk more accurately. Figure 8 models the fire risk using road density and housing density. Here the multilayer perceptron neural network is used to model the fire risk using inputs of road density and housing density. Due to limited data, the network only consists of one hidden layer with 2 nodes. The output of the neural network is the predicted fire risk. Before feeding into the neural network, the input data was normalized to resemble a distribution centered around 0 with standard deviation of 1. The RELU activation function is used in this study.

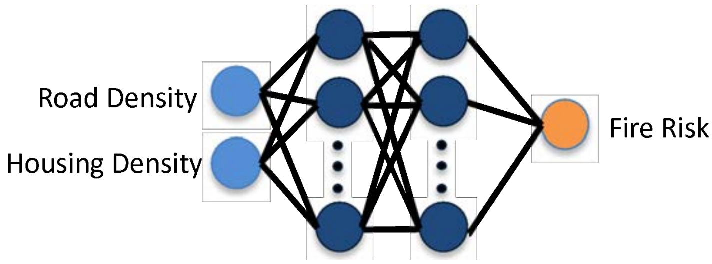

Figure 8: Neural Network Design

3. Results

3.1 Correlation Analysis & Linear Regression

Scatter plots and heat maps are used in the correlation study for both datasets. Figure 9 through Figure 13 show the analysis results of the first dataset. Low correlations between the fire density and the urbanization factors are observed. Since most data clustered near low fire density, correlation analysis was also performed in the logarithmic scale for better visualization of the data. However, the data still does not appear to show any useful trends. Very small R^2 values ($R^2 < 0.01$) were obtained after linear regression. The heat map also displays that there is little correlation between fire density and the other variables. This may be due to the excessive normalization of the data, with all variables being averaged over counties as well as with interpolation/extrapolation of urbanization data.

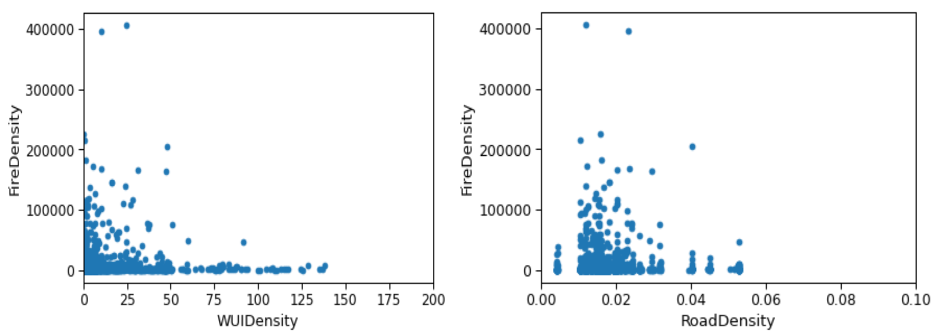

Figure 9. Scatter Plot of Fire Density vs. WUI Density (Left), Road Density (Right)

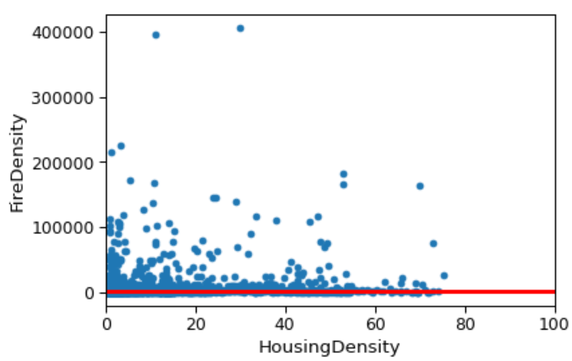

Figure 10. Scatter Plot of Fire Density vs. Housing Density. Linear regression result (red line) is shown.

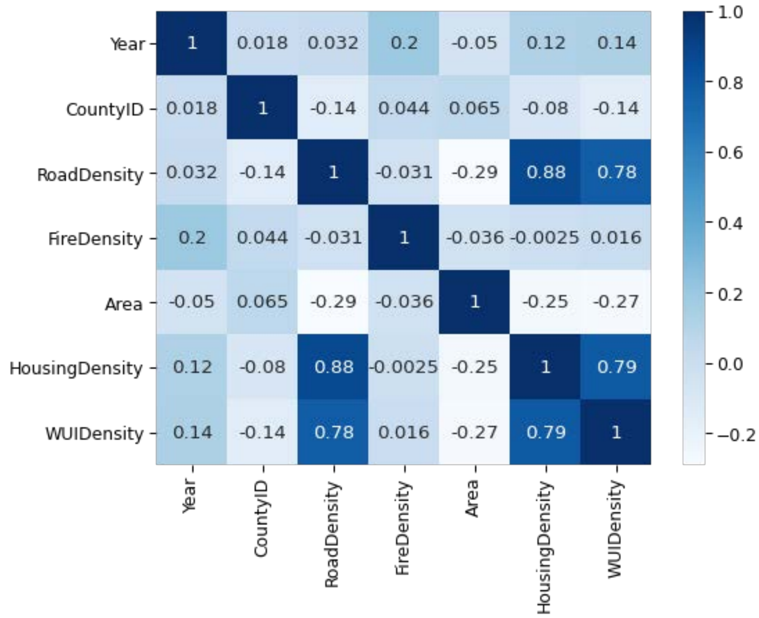

Figure 11. Heat Map of Various Variables

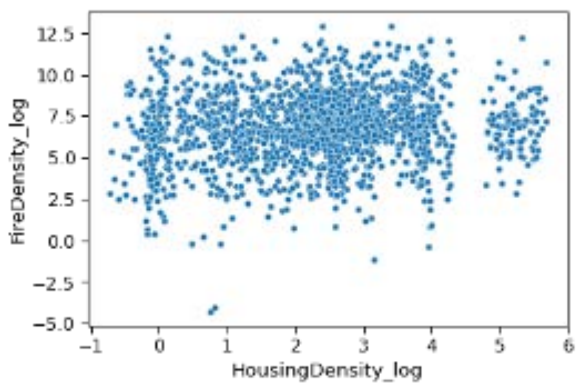

Figure 12. Scatter Plot of Fire Density vs. Housing Density (Logarithmic Scale)

Similar to the first dataset, in the second dataset, there is a strong imbalance in data towards smaller fires, resulting in poor regression models (See Figure 14). After scaling the dataset to the logarithmic scale, there are many data points found with very small housing and road density values. Data points with very small values are removed, resulting in data points that can be significantly better fitted with regression models. The housing and road densities with very small values may represent rural areas with little to no population.

The right-hand side plot in Figure 15 shows the scatter plot of total fire area vs. road density after removing the very small road density data points. The heat map depicted in Figure 16 shows that road density is negatively correlated to the total fire area. The polynomial fitting result is also shown in Figure 15 with R^2 value of 0.3067. Contrarily, roads can be a sign of non-residential development (i.e., logging, farming, transmission lines), which are easier to burn through with the lack of more fire-resistant vegetation such as most trees. However, once road density exceeds a certain point, it usually signifies residential property.

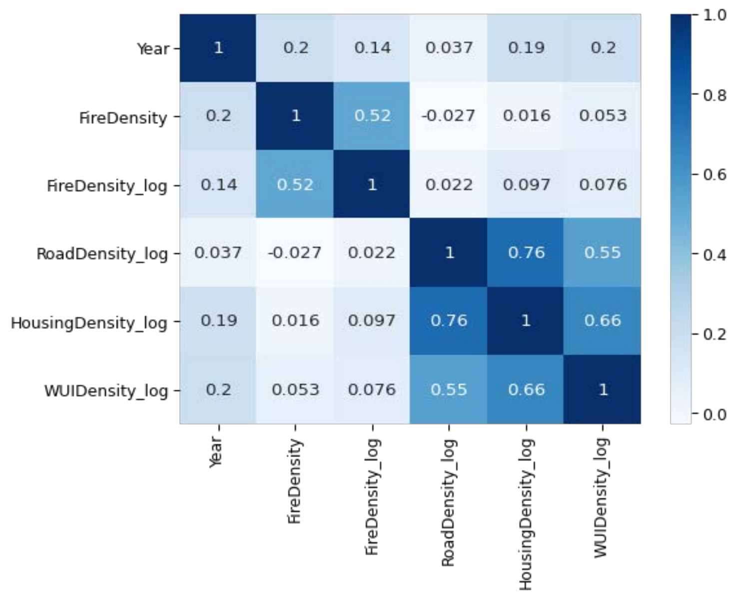

Figure 13. Heat Map of Various Variables in Logarithmic Scale

The right plot in Figure 17 shows the scatter plot of total fire area vs. housing density after removing the very small housing density data points. The heat map depicted in Figure 18 shows that housing density is negatively correlated to the total fire area. The linear fitting result is also shown in Figure 17 with R^2 value of 0.5988. There is an inverse linear relationship in logarithmic scale between housing density and fire area. This may be due to firefighters prioritizing the protection of residential properties, then dealing with the containment of fires.

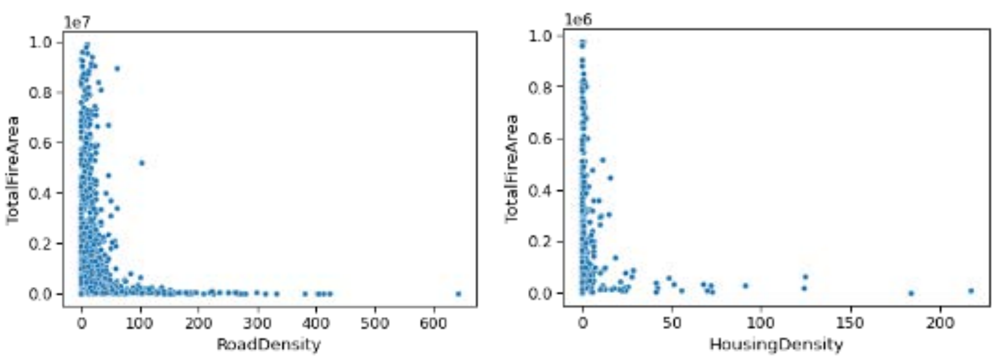

Figure 14. Total Fire Area vs. Road Density (Left), Housing Density (Right)

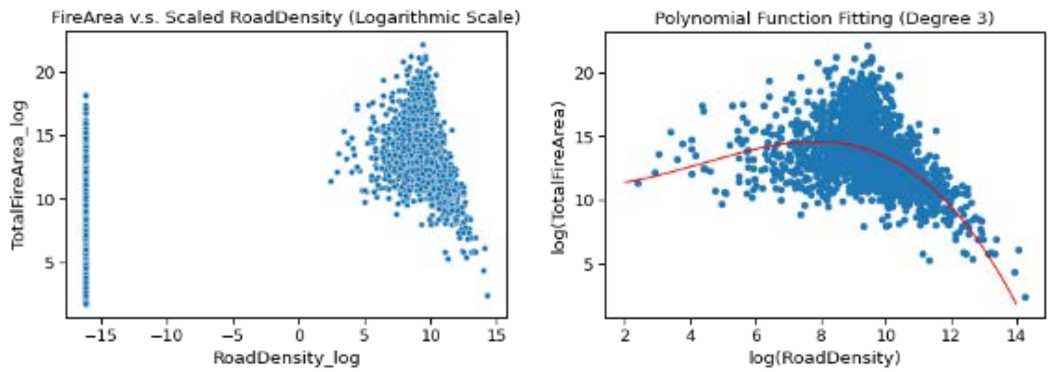

Figure 15. Total Fire Area vs. Road Density in Logarithmic Scale (Left), Polynomial Fitting of Total Fire Area via Road Density in Logarithmic Scale (Right)

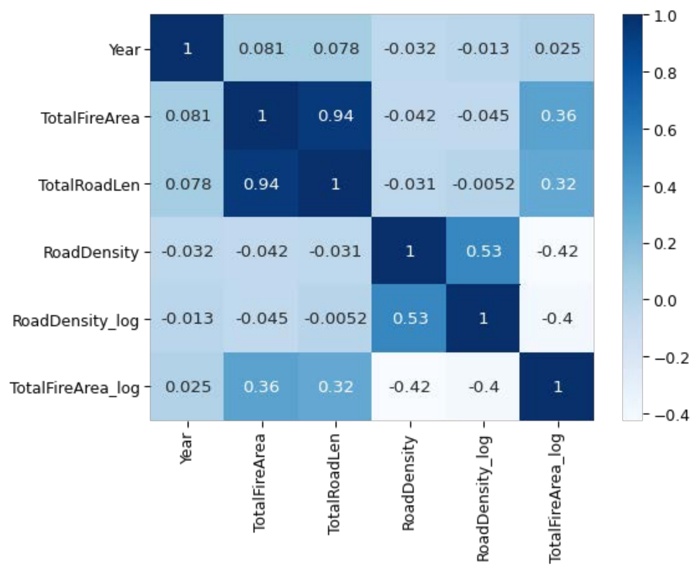

Figure 16. Heat Map (Road Density vs. Total Fire Area)

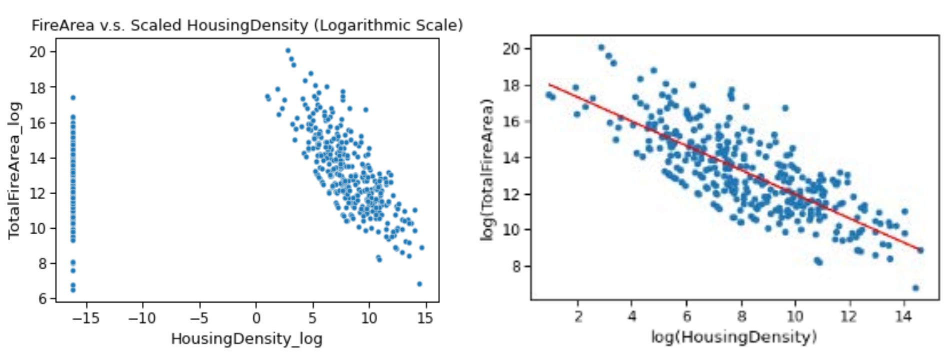

Figure 17. Total Fire Area vs. Housing Density in Logarithmic Scale (Left), Linear Fitting of Total Fire Area via Housing Density in Logarithmic Scale (Right)

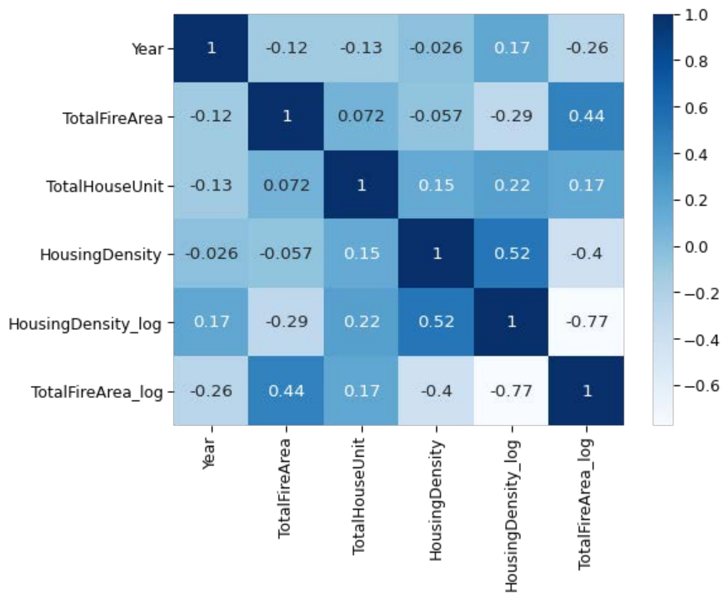

Figure 18. Heat Map (Housing Density vs. Total Fire Area)

3.2 Multiple Linear Regression & Neural Networks

Using both road and housing density logarithmically scaled, multiple linear regression and neural network regression models are constructed based on the second dataset. For neural networks, the two inputs are first scaled to resemble a distribution of zero mean and standard deviation of 1, then go through one hidden layer to return fire risk, or predicted fire area. Figure 19 shows the converged loss history for both training data and validation data. Neural network regression appears to have improved on multiple linear regression in test errors (See Figure 20). The mean absolute test error for neural networks is 0.71 and the mean absolute test error for multiple linear regression is 0.76. However, there is too little data that can be used for considering both road density and housing density factors, resulting in only 90 data points for training and 10 for testing. Hence, no significant conclusion can be made through these models. Future study is required using a larger set of data.

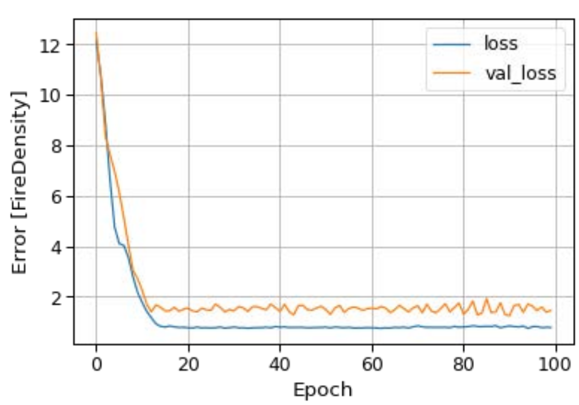

Figure 19. Loss History (Training & Validation)

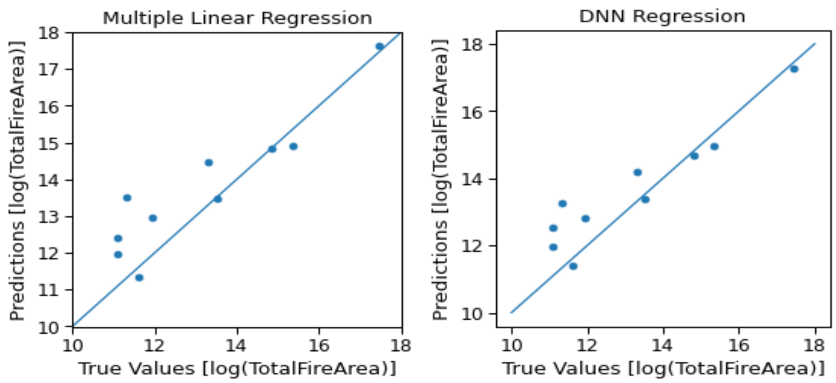

Figure 20. Test Errors (Left: Multiple Linear Regression, right: Neural Network)

4. Conclusion

In the end, firefighting is about preventing fires from being able to cause damage to properties. Forest fires should be a normal occurrence in order to keep more burnable vegetation away and revitalize forests. However, the presence of firefighting along with other factors such as climate change and human development adds or keeps more potential fuel. This has made fires more aggressive and larger than before, being able to severely harm ecosystems and affect air quality. This is not to mention the strong possibility of the fire passing through populated regions.

Thus, human development should minimize ‘leapfrog’ sprawl (where human development is sporadic and discontinuous) to reduce the amount of other infrastructure that may increase fire risk with increased fire damage costs. This type of development can be partially represented by the area denoted as wildland-urban interface. An example of a leapfrog sprawl that was devastated by fires in 2018 were the outskirts of Redding, California. Reducing development near wilderness or high fire risk regions leads to less demand for immediate firefighter response, in turn creating normal and healthier patterns of forest fires with frequent yet smaller fires.

5. Acknowledgements

This study leverages several open-source components including QGIS Geographic Information System, Cal Fire database, WUI dataset, and US Census data.

6. References

[1] Cal Fire website, https://www.fire.ca.gov (Last Access September 24, 2021)

[2] Radeloff, V. C., D. P. Helmers, H. A. Kramer, M. H. Mockrin, P. M. Alexandre, A. Bar-Massada, V. Butsic, T. J. Hawbaker, S. Martinuzzi, A. D. Syphard, and S. I. Stewart, “Rapid growth of the U.S. Wildland Urban Interface raises wildfire risk,” *Proceedings of the National Academy of Sciences*, 115(13): 3314-3319, 2018.

[3] QGIS Geographic Information System. Open-Source Geospatial Foundation Project. https://qgis.org/en/site/ (Last Access September 24, 2021)