Predicting Burned Area of Wildfires Using Environmental and Fire Characteristics

Info Innovators: Arnav Meduri, Kevin Mao, Ricardo Urena, and Ben Trokenheim

Background & Motivation

- Wildfires are among the most **destructive** and **costly** natural disasters in the United States.
    - Since 1983, the U.S. has averaged approximately 70,000 wildfires per year (EPA).
    - Estimated annual economic losses from wildfires in the U.S. range from $394 billion to $893 billion (Senate Joint Economic Committee).
    - Wildfires contribute 5–8 billion metric tons of CO_2 to the atmosphere globally each year (International Fund for Animal Welfare Coordination Center).

- There is a need to better understand the factors that drive fire behavior and **burned area** to improve **resource allocation** and **mitigation** efforts.

- **Research Questions:**
    - What environmental and fire-specific factors influence wildfire size?
    - Can we develop models to estimate burned area based on geographic factors, weather conditions, land cover, and fire-specific information?

Data Description

- **Primary Source:** Forest Service Research Data Archive (U.S. Forest Service)
  - 1.88 million wildfire records collected from 1992 to 2015
- **Additional Sources:**
  - Integrated Surface Hourly Dataset (NOAA)
  - World Cities Database (SimpleMaps)
  - Land-Cover Changes Global Dataset (Meiyappan and Jain)
- **Dataset Overview:**
  - 14,229 observations (after filtering for middle 50% of fire sizes)
  - 41 variables per fire event (including fire attributes, weather conditions, vegetation characteristics)

Category	Variables
Wildfire Characteristics	**fire_name:** Name of fire **fire_size:** Area burned (acres) **fire_size_class:** Size class (A–G) **fire_mag:** Scaled fire magnitude **fire_cause:** Descriptive cause of fire
Location & Time	**latitude, longitude:** Fire coordinates **coordinates_state:** U.S. state **discovery_month:** Month discovered **putout_time:** Duration to extinguish
Weather (7, 15, 30 days prior and on containment day)	**Temp_*:** Avg. temperature (°C) **Wind_*:** Avg. wind speed (m/s) **Hum_*:** Avg. humidity (%) **Prec_*:** Total precipitation (mm)
Other	**vegetation:** Dominant land type **remoteness:** Scaled distance from nearest city (unitless)

Table 1. Key attributes in compiled dataset (**categorical**, **numerical**, **date/string**, response)

Exploratory Data Analysis

Univariate EDA

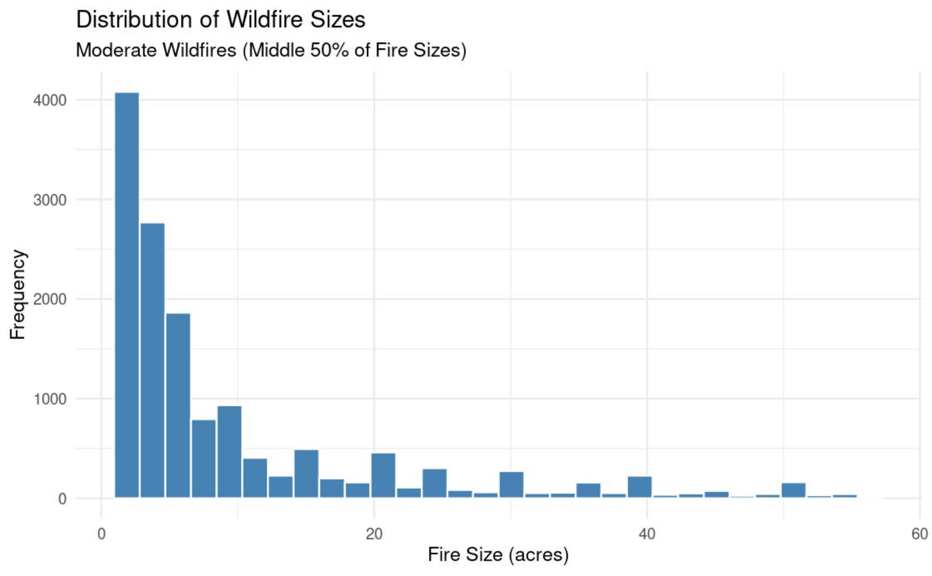

Figure 1. Histogram showing the distribution of wildfire sizes (in acres) for the middle 50% of wildfires in the dataset. Extremely small and large fires were excluded to focus on fires that are more likely to be informative for modeling.

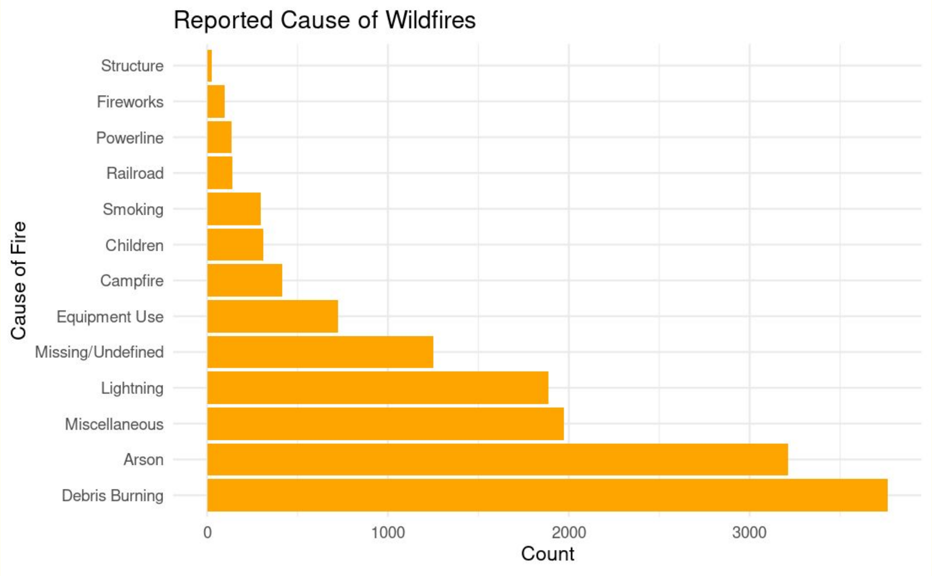

Figure 2. Bar plot showing the distribution of reported wildfire causes in the dataset. The most common causes include debris burning, arson, miscellaneous, and lightning.

Exploratory Data Analysis

Bivariate EDA

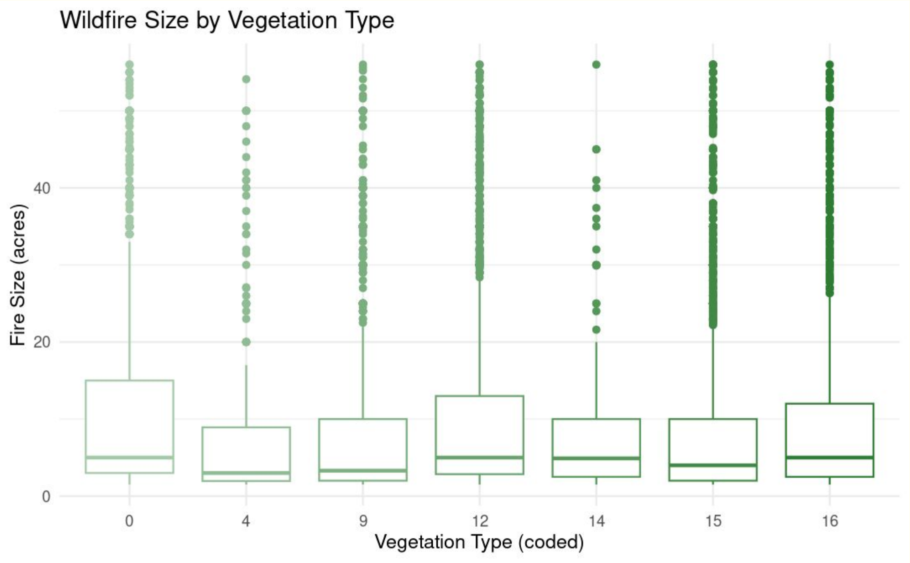

Figure 3. Boxplot of wildfire sizes (in acres) by vegetation type (coded as: 0 = Other, 4 = Temperate Evergreen Needleleaf Forest, 9 = C3 Grassland/Steppe, 12 = Open Shrubland, 15 = Polar Desert/Rock/Ice, 16 = Secondary Tropical Evergreen Broadleaf Forest).

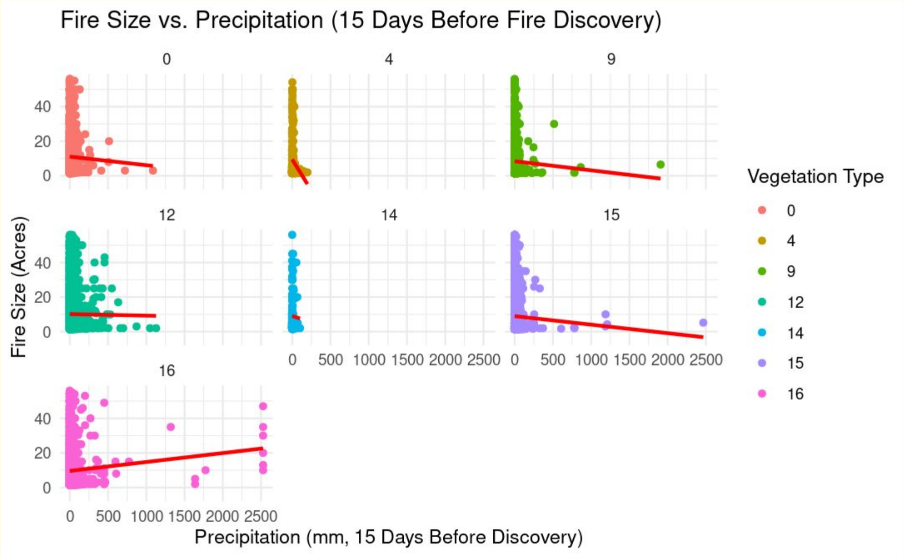

Figure 4. Scatterplots showing the relationship between precipitation 15 days prior to fire discovery and wildfire size (in acres), faceted by vegetation type.

Initial Modeling & Results

- So far, we have fit two linear regression models to estimate fire size:
  - **Main effects model** (remoteness, Prec_pre_15, Temp_pre_15, Vegetation, stat_cause_descr, Wind_cont)
  - **Interaction effects model** (same predictors as above, plus Prec_pre_15 * Vegetation interaction)

- Both models have limited explanatory power (low adjusted R-squared values):
  - **Main effects: 0.0217**
  - **Interaction Effects: 0.0226**

```
fire_main_fit <- lm(fire_size ~ remoteness + Prec_pre_15 + Temp_pre_15 + Vegetation + stat_cause_descr + Wind_cont,
data = wildfire_clean_50)

fire_int_fit <- lm(fire_size ~ remoteness + Temp_pre_15 + Prec_pre_15 * Vegetation + stat_cause_descr + Wind_cont, data
= wildfire_clean_50)

glance(fire_main_fit)$adj.r.squared
glance(fire_int_fit)$adj.r.squared
```

```
[1] 0.02168202
[1] 0.02259321
```

Figure 5. Model fit results for two linear regression models (main effects model and interaction effects model) predicting fire size.

Next Steps/Questions

- Explore additional predictors in the dataset which may provide useful information for modeling fire size
    - Lagged weather variables (**Wind_pre_15**, **Hum_pre_7**, **Temp_pre_30**)
    - Temporal variables (**discovery_month**, **disc_pre_year**)
    - Geographic variables (**state**)
- Consider alternative modeling approaches for analyzing fire size
    - Fit a logistic regression model to classify whether a fire falls in the upper or lower half of "moderate" fires (middle 50% of fire sizes)
    - Would have to create binary outcome variable based on the **fire_size** within the already-filtered dataset
- Improve/refine our current linear regression model
    - Apply a log transformation to reduce skewness in fire size
    - Evaluate potential interaction terms (e.g., between precipitation and vegetation)
    - Simplify categorical variables to reduce complexity and improve interpretability
    - Combine vegetation types (**Vegetation**) into broader groups (e.g., forest, shrubland, grassland)
- **Any feedback on to how to improve our model/modeling approach?**

References

[1] United States Environmental Protection Agency. “Climate Change Indicators: Wildfires | US EPA.” *US EPA*, EPA, July 2016, www.epa.gov/climate-indicators/climate-change-indicators-wildfires.

[2] Committee, United States Joint Economic. “Climate-Exacerbated Wildfires Cost the U.S. Between $394 to $893 Billion Each Year in Economic Costs and Damages.” *www.jec.senate.gov*, 16 Oct. 2023, www.jec.senate.gov/public/index.cfm/democrats/2023/10/climate-exacerbated-wildfires-cost-the-u-s-between-394-to-893-billion-each-year-in-economic-costs-and-damages.

[3] International Fund for Animal Welfare. “Does Climate Change Cause Wildfires?” *ifaw.org*, 24 Oct. 2024, www.ifaw.org/journal/climate-change-wildfires.

[4] Short, Karen C. “Spatial Wildfire Occurrence Data for the United States, 1992-2015 [FPA_FOD_20170508] (4th Edition).” *Forest Service Research Data Archive*, https://doi.org/10.2737/rds-2013-0009.4.

[5] “Global Hourly - Integrated Surface Database (ISD).” *National Centers for Environmental Information (NCEI)*, 17 Mar. 2021, www.ncei.noaa.gov/products/land-based-station/integrated-surface-database.

[6] Meiyappan, Prasanth, and Atul K. Jain. “Three Distinct Global Estimates of Historical Land-Cover Change and Land-Use Conversions for over 200 Years.” *Frontiers of Earth Science*, vol. 6, no. 2, June 2012, pp. 122–39, https://doi.org/10.1007/s11707-012-0314-2.

[7] “World Cities Database | Simplemaps.com.” *simplemaps.com*, 2019, simplemaps.com/data/world-cities.


Thank You for Listening!