U.S. Wildfires: Linear Regression Analysis to Predict Relationship between Environmental Factors and Spread of Wildfires Using R Studio

Abstract

There is always a source, a specific climate condition, and a certain temperature to start a wildfire and eventually widespread. The long-term goal is to predict the spread of the fire and take preventive measures before fatal and serious damages to the human life and property happens. In this research project, I considered various environmental factors (independent variables) on the day of the wildfire instance, for example, temperature, wind speed, dew point, and precipitation, to determine a linear model for predicting wildfire spread (dependent variable). By employing statistical analysis on the linear regression model (RStudio), the significance of each factor's effect on the spread of the wildfire was determined. The data for average temperature, high temperature, and dew point is normally distributed and these individual variables seem to show a relationship with the wildfire area, but only high temperature showed a significant relationship (p value of 0.031). The linear regression analysis of combined independent variables refuted my hypothesis that a warm, dry, and windy conditions will have a significant impact on the creation and spread of a wildfire. It appears that a simple linear model will not be an effective form of prediction for spread of wildfires.

Background

(Source: https://www.nasa.gov/topics/earth/features/climate-fire.html)

Damaging Effects of Wildfires:

- Fatalities
- Health issues (Respiratory, Cardiovascular, Allergy, COPD etc.)
- Damage to public and private property
- Water and air quality
- Economic impact due to property loss and costs of containing and managing wildfires
- Carbon emission
- Ecosystem alteration

Multiple Studies have found that climate variations impact the extent, frequency and area coverage of wildfires (1-3). The warmer and drier conditions have always contributed towards wildfires in the western United States (3, 4). In order to minimize the damages, scientists and researchers have tried to predict and understand the behavior (speed, direction, modes of spread) of wildfires. However, there are a lot of factors involved, such as complex fuel configurations, complicated chemical kinetics, and balances between different modes of heat transfer, that can lead to different behaviors and difficult to accurately predict. Current models for numerical description of wildfire propagation are mainly algebraic and based on statistical or physical attributes, while others are based on ideas that incorporate rough approximations to the complex physical and chemical processes.

Results

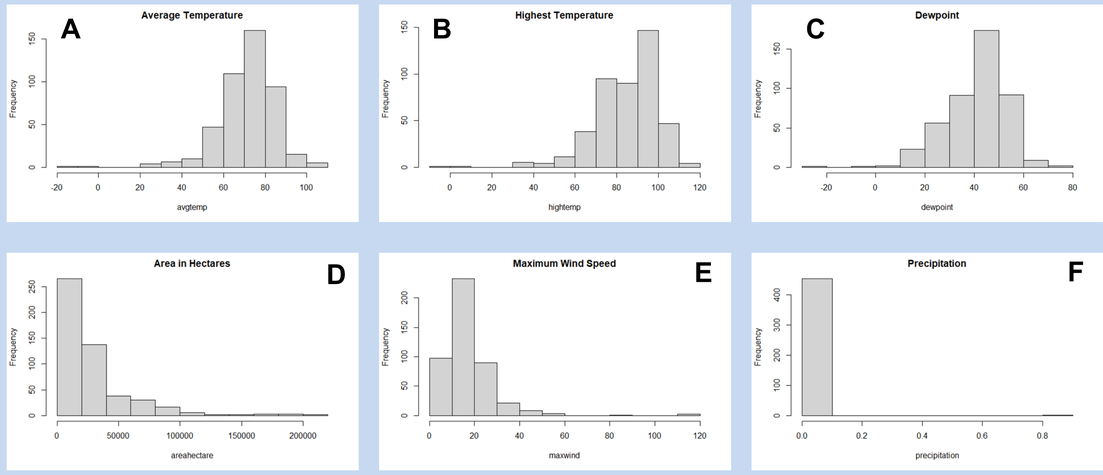

Figure 1: Histograms of normality Test of variables. (A) average temperature, (B) high temperature, and (C) dewpoint distribution in general are normally distributed with few outliers. (D) area hectares and (E) maximum wind speed distribution is skewed to the right with (F) precipitation distribution generally collective at 0.0.

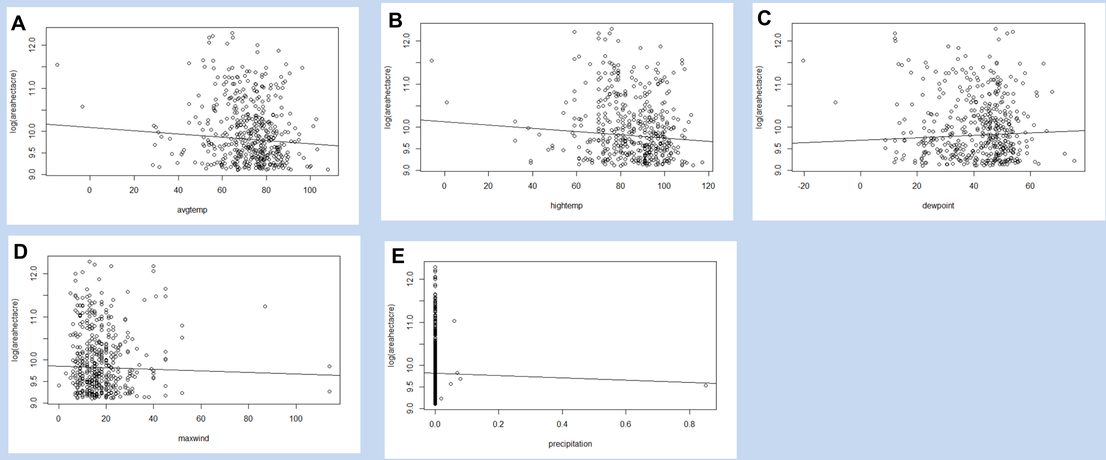

Figure 2: Scatter plots of linear regression analysis showing frequency of independent variables and relationship with dependent variable (area hectare).

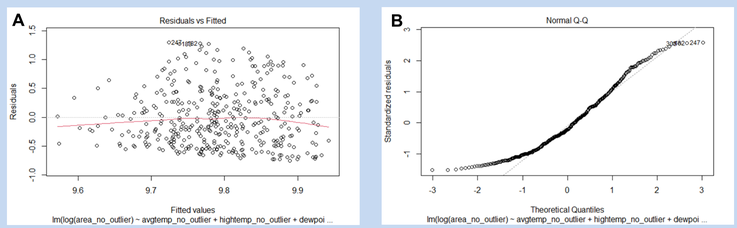

Figure 3: Model Diagnostics. (A) shows that the residuals have constant variance when plotted against fitted values. (B) shows that normality assumptions of my model is not met though some of the data points are completely aligned with the regression line and there are multiple outliers.

Methods

Datasets and Variables

- Data for wildfire instances in the U.S. from 2000 to 2009 in the form of latitude and longitude obtained from Latlong, www.latlong.net). This is the only data I could find in the public domain.
- For each event, information was collected for the city/county (location) from historical weather analyzer (https://www.wunderground.com/history)
- Data for the average temperature, highest temperature, dew point, precipitation, and wind speed in Microsoft Excel for Modeling and Prediction work
- 506 U.S. locations of wildfire events and over 2500 data points including 5 different independent variables and one dependent variable.

Application

- **RStudio:** An open-source application for programming, data analyses, and graphics
- **Independent Variables**
  - Average Temperature
  - Highest Temperature
  - Dew Point
  - Wind Speed
  - Precipitation
- **Prediction model**
- Linear Regression

Conclusions

- Individual variables (avg temp, high temp, dew point) seem to show a relationship with the logarithm of the area but only high temperature showed a significant relationship with p value less than 0.05
- Multivariate model resulted with constant variance but violated linearity conditions
- Hypotheses refuted and there are a few possible causes
  - Outliers
  - Linear Model may be too simple for complex set of data

Future Research

- Refine the prediction analysis. Clean up the current data by removing the outliers and re-run the linear model.
- Collect additional data for more recent U.S. wildfires and increase the sample size.
- Analyze the data using complex regression analysis assuming that climatic conditions could be cross talking to each other and impacting spread of wildfire.
- Improve analyses programming skills. I started learning R very recently and from this project I can see room for improvement.
- Explore potential collaborations and get access to data from reliable sources such as EPA and NASA.

References

1. https://www.epa.gov/climate-indicators/climate-change-indicators-wildfires
2. Stevens LE; et al. Climate change in the human environment: Indicators and impacts from the Fourth National Climate Assessment. J Air Waste Manag Assoc. 2021 Oct;71(10):1210-1233..
3. Kitzberger T; et al. Contingent Pacific-Atlantic Ocean influence on multicentury wildfire synchrony over western North America. Proc Natl Acad Sci U S A. 2007 Jan 9;104(2):543-8.
4. Westerling AL. Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. Philos Trans R Soc Lond B Biol Sci. 2016 Jun 5;371(1696):20150178.