

Wildfire Vulnerability Across Crop Types

Andrea Zantek

3/13/2024

Abstract

Climate change has led to an increase in both the frequency and severity of wildfires. This raises the question: 'How does wildfire vulnerability differ among crops?' Through implementing a variety of modeling techniques, we will aim to analyze how wildfire vulnerability varies among crops. Findings indicate that Grapes are the most susceptible to wildfire events.

Background and Significance

Climate change has led to an increase in both the frequency and severity of wildfires. Since the 1970's, California wildfire occurrence has increased by 140%, with average wildfire size nearly doubling (Environmental Market Solutions Lab, Appendix Figures 9 and 10). The trend is attributed to climate change-related weather disturbances, most notably, the increase in surface temperature and the drying of organic matter and other fire fuels (NOAA). Furthermore, wildfires have multifaceted impacts on agriculture. The majority of crops suffer following wildfire events due to a combination of air pollutants, direct disruptions, and secondary impacts of wildfires such as flooding. Thus, climate-change induced shifts in wildfire patterns puts our most important resource, food, at stake.

This research focuses on understanding wildfire vulnerability in two of California's most important agricultural domains: Grapes and Pastureland, which is used for agricultural grazing. Pastureland and Grapes were chosen due to their known susceptibility to wildfire events and significant contribution to California's agricultural sector. Through my analysis, I hope to provide meaningful contributions that can be used by agricultural producers to address varying wildfire vulnerabilities, thus mitigating future wildfire-induced crop losses.

Intuitively, I expect to find that Grapes and Pastureland will be more vulnerable to wildfires when compared to alternate crops. My intuition is based on the observation that Grape farms and Pastureland tends to be situated in different areas than alternate crops. While agricultural production is concentrated in California's Central Valley, Grapes tend to be grown on the Western side of the state. Coincidentally, the Central Valley experiences few wildfire events, while the perimeter of the state frequently experiences wildfire events and thus, is associated with higher wildfire vulnerability. Similar justification follows for Pastureland, although a significant amount of Pastureland does lie within Central Valley.

Data

This project required the use of spatial data, or data with a geometry column which contains information about the location and/or size of each object in the dataset. There are two types of spatial data utilized—raster and polygon. Raster data involves splitting up an area of land into equally sized squares, also known as pixels, such that each pixel has values for each attribute. Polygon data differs in that it allows for organic boundaries determined by a collection of points, with attribute information stored within individual polygons. Figure 1 illustrates the differences between the raster and polygon formats.

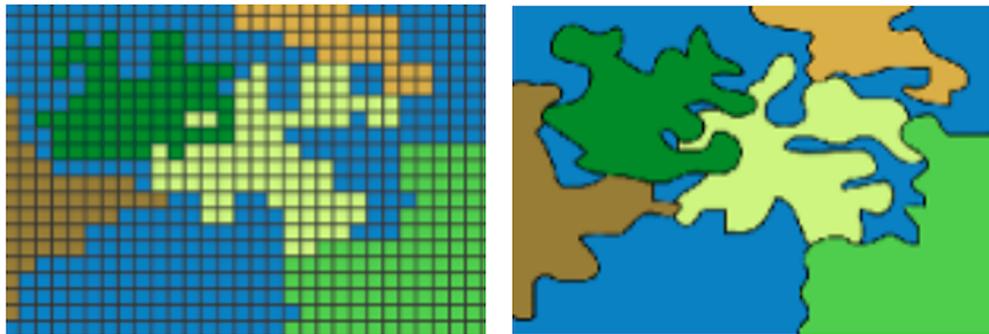
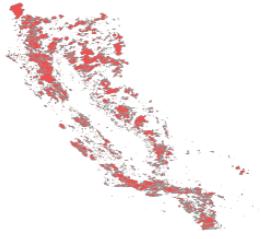


Figure 1: Raster Versus Polygon Data

The spatial data used throughout this project comes from several sources. The California Fire Perimeters Data was sourced from the California Open Data Portal. The dataset is jointly maintained by several governmental organizations including CalFire and the Bureau of Land Management. Depicted in Figure 2, this dataset has a polygon format, where each row pertains to a specific wildfire event, spanning from 1950 to 2023. One limitation arising from this dataset is the presence of measurement error, especially in older fire records. While a variety of modern tools such as satellite imagery are now able to accurately capture wildfire boundaries, the earlier wildfire boundaries were determined from ground surveys, which are notorious for inaccuracies.



ObjectID	Year	Name	Acres	ShapeLength	ShapeArea	Geometry
21440	2020	NELSON	109.60228	0.0357330	0.0000461	MULTIPOLYGON (((-121.3484 3...
21441	2020	AMORUSO	685.58502	0.1011780	0.0002878	MULTIPOLYGON (((-121.3528 3...
21442	2020	ATHENS	27.30048	0.0174496	0.0000115	MULTIPOLYGON (((-121.3333 3...
21443	2020	FLEMING	12.93154	0.0165571	0.0000054	MULTIPOLYGON (((-121.2732 3...
21444	2020	MELANESE	10.31596	0.0109196	0.0000044	MULTIPOLYGON (((-121.3007 3...
21445	2020	PFE	36.70193	0.0242667	0.0000154	MULTIPOLYGON (((-121.3824 3...

Figure 2: Wildfire Perimeters

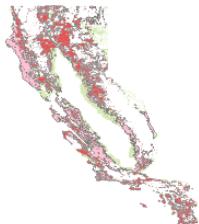
The Crop Data was sourced from the Cropland Data Layer and has the raster format. In this dataset, each row pertains to a specific 30-meter plot of land, with the crop classification information stored under the *Crop Type* variable. Figure 3 illustrates the distribution of agricultural production in California. From this map, we see that agricultural production is concentrated in the Central Valley region. The perimeters of the state are dominated by non-agricultural pixels like shrubland and barren land.



ObjectID	CropType	Geometry
10817	176	POINT (-2286900 2452680)
11067	176	POINT (-2286870 2452650)
11318	176	POINT (-2286900 2452620)
11321	176	POINT (-2286810 2452620)
11573	176	POINT (-2286900 2452590)
11832	176	POINT (-2286870 2452560)

Figure 3: Agricultural Map

The Hazard Class Data, provided by CalFire, utilizes the polygon format to describes an area's likelihood of experiencing a wildfire event. The *Hazard Class* variable is a categorical variable with three levels- Moderate, High, and Very High, represented in Figure 4 by the colors green, pink, and red, respectively. Due to California's extreme susceptibility to wildfires, the Moderate Class is the least severe *Hazard Class* observed in the state. One limitation in this variable is the significant amount of area with NA values. We see a particularly notable lack of data in the Central Valley, where the majority of Californian agricultural land is located.



ObjectID	HazardClassCode	HazardClass	ShapeLength	ShapeArea	Geometry
1	1	Moderate	2013.464	108898.67	POLYGON ((-1938430 1268163,...
2	1	Moderate	2758.326	148499.93	POLYGON ((-1938532 1271467,...
3	1	Moderate	2506.648	149217.93	POLYGON ((-1944405 1275171,...
4	1	Moderate	4256.729	157700.55	POLYGON ((-1887530 1261956,...
5	1	Moderate	9088.942	1084231.78	POLYGON ((-1908571 1268742,...
6	1	Moderate	1523.170	65261.64	POLYGON ((-1873286 1259942,...

Figure 4: Hazard Classes (Green = Moderate, Pink = High, Red = Very High)

The Soil Data is sourced from the gSSURGO Database and was accessed through the ArcGIS Soil Development Toolkit. This dataset uses the polygon format, where lands with similar soils are grouped together, as depicted on the left-hand side in Figure 5. One concern in this dataset is the presence of relatively large polygons located in the Southeast corner of the state, representing public lands, visible on the right-hand side of Figure 5. Alternative soil sampling methods are used on public lands, thus resulting in the larger polygons. Due to the limited amount of agricultural activity in these areas, we will consider this acceptable for the purpose of our research.

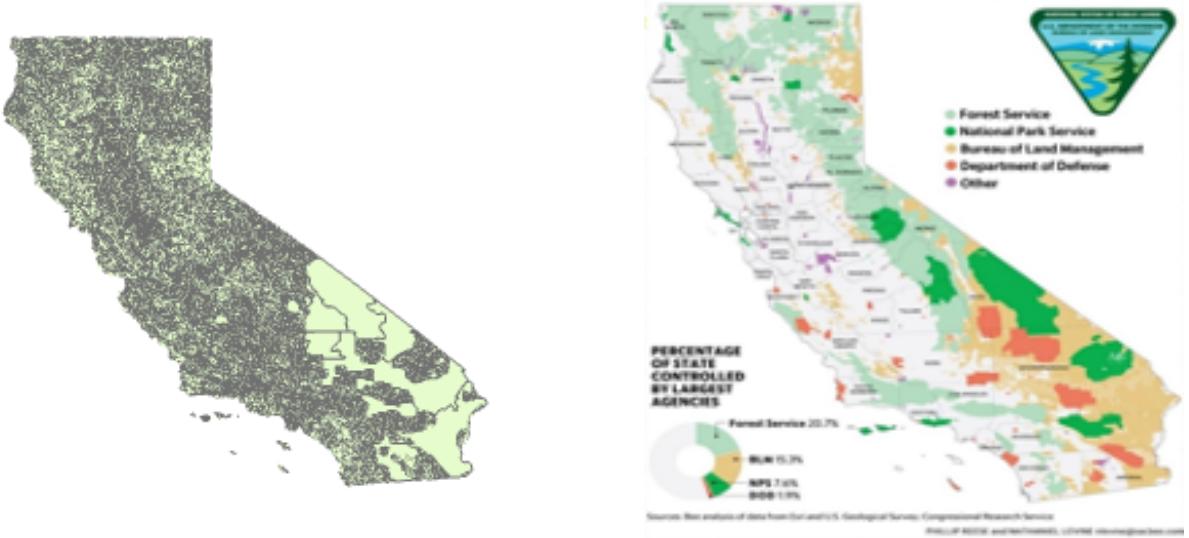


Figure 5: Soil Polygons and Public Lands

ObjectID	ShapeLength	ShapeArea	SoilCompactionPercentage	NumberOfFrostFreeDays	HydrologicGroup	Geometry
1	6736.832	929159.1	25	225	C	POLYGON ((-2244979 2353731,...
2	8147.547	1052839.9	50	125	B	POLYGON ((-2208144 2353732,...
3	9901.538	1760741.4	5	NA	NA	POLYGON ((-2231336 2354345,...
4	4734.245	717553.4	35	175	A	POLYGON ((-2219626 2353632,...
5	3284.710	457928.2	5	NA	NA	POLYGON ((-2222754 2354433,...
6	6495.876	1382703.9	5	NA	NA	POLYGON ((-2237880 2353413,...

Figure 6: Soil Data

The Soil Data provides the following variables: *Number of Frost-Free Days*, *Soil Compaction Percentage*, and *Hydrologic Group*. The *Number of Frost-Free Days* variable helps account for variation in weather. Furthermore, the *Soil Compaction Percentage* and *Hydrologic Group* variables play an integral role in the modeling phase by accounting for how variation in soil properties may impact wildfire vulnerability. The *Hydrologic Groups* are determined by a combination of the soil texture, runoff potential, infiltration rate, and water transmission rate. Infiltration rate describes the speed at which water first enters the soil, whereas water transmission rate describes the speed of water movement through the soil profile once it has entered the soil.

Hydrologic Group Descriptions:

1. Group A: Soils with high infiltration rates and low runoff potential. Deep well-drained sands with high water transmission rate. Group A soils are typically most suitable for agricultural purposes.
2. Group B: Soil with moderate infiltration rates, runoff potential, and water transmission rates. Deep well-drained soil with moderately fine to moderately coarse texture.
3. Group C: Soils with slow infiltration rates and slow water transmission rates, leading to high runoff potential. Soils with a layer that impedes downward movement or fine textured soils.
4. Group D: Soils with an extremely slow infiltration rate and runoff potential, like soils with a high-water table or clay layer near the surface.
5. Other: If a soil receives a “Group D” rating because of a high-water table, multiple classes are assigned: A/D, B/D, or C/D. Observations with multiple classes were updated to be classified as “other”.

The secondary soil variable, *Soil Compaction Percentage*, describes the extent to which soil particles are pressed together. Soil compaction is associated with an inhibited ability to hold and transmit both water and nutrients, thus making it an essential measure of soil fertility. *Soil Compaction Percentage* is an especially relevant variable in agricultural modeling, as compaction is a known by-product of harmful agricultural practices, like overgrazing. The inclusion of the *Soil Compaction Percentage* variable allows us to control for the heightened susceptibility of compacted soils to wildfire conditions, stemming from their inhibited water retention capabilities.

To aggregate the datasets, several data wrangling steps were required. The preliminary data wrangling was completed in ArcGIS due to its ability to process large datasets and wrangling tools catered for spatial data. These capabilities were especially relevant in this project, as there were an enormous number of observations in the Crop Data! First, it was essential to align the coordinate systems and extents, or dimensions, of all layers. This was essential to ensure that all subsequent data wrangling steps were executed properly.

To join the Crop Data with the other datasets, it was necessary to convert the raster layer into a point layer. This process involved creating a new layer of points, located at the center of the pixels, and pulling all associated attributes to the new point layer. Next, a spatial join was required between the Crop Data and the Soil Data. This join ensures that the observations from the Crop Data receive the attributes from the soil polygon they reside in. In the case of a point falling on a boundary between polygons, the attributes from one of the polygons are randomly chosen. A similar spatial join was completed with the Hazard Class Data to pull the associated *Hazard Class* variable for each observation.

The next step was creating the *Distance* variable. Before this variable could be created, it was mandatory to dissolve the overlapping polygon boundaries in the California Fire Perimeters Data. This simply involves aggregating polygons with overlapping boundaries into a single polygon. This was necessary to calculate the planar distance in kilometers from each point to the edge of the nearest polygon boundary. Completing this step provides the *Distance* variable, which represents the distance from a plot of land to the boundary of the nearest historical wildfire.

The next phase involved exporting the data for use in R Studio using a randomly resampled dataset. While having many observations can be helpful in modeling, the full dataset, while glorious, was simply too large to process efficiently using R Studio, where the modeling phase was implemented. A random subset was created such that 2,500 complete observations, meaning no NA values, were pulled for each crop of interest- Grapes, Pastureland, and “Other”, which encompasses all non-grape and non-pastureland agricultural pixels.

One notable limitation is the sole use of the complete observations, as this likely leads to bias. It is probable that the complete observations tend to be more alike than observations within the full dataset. This is especially relevant given the significant number of NA values observed in the Hazard Class Data. However, the complete dataset was necessary to accommodate the more complex modeling techniques. Table 1 displays the first five rows of the final dataset, followed by a summary of all variables.

Table 1: First 5 rows of the dataset

Distance	HazardClass	CropType	Latitude	Longitude	SoilCompactionPercentage	HydrologicGroup	NumberOfFrostFreeDays
0.068	Very High	Grapes	-13726373	5138033	70	A	175
0.606	High	Grapes	-13713814	5136558	35	A	175
0.209	Very High	Grapes	-13602353	5160107	85	A	125
1.625	Very High	Grapes	-13595986	5161887	85	C	90
1.682	Very High	Grapes	-13595836	5161512	85	C	90

Variables of Interest:

- *Distance*= Represents the distance from a plot of land to the boundary of the nearest historical wildfire.
- *Hazard Class*= Categorical variable with three levels (Moderate, High, and Very High), representing an area's likelihood of experiencing a wildfire event.
- *Crop Type*= Categorical variable with three levels (Grapes, Pastureland, and “Other”), representing the dominant crop grown within a plot of land.
- *Number of Frost-Free Days*= Estimate of the yearly number of frost-free days.
- *Soil Compaction Percentage*= Captures the extent to which soil has been compacted, measured on a scale from 0-100%.
- *Hydrologic Group*= Categorical variable with five levels (A, B, C, D, and “Other”) which captures differences in soil texture. Classes A and B are more suitable for agriculture, whereas C, D, and “Other” soils are less suitable for agriculture.
- *Latitude*= Represents the latitudinal position.
- *Longitude*= Represents the longitudinal position.

Methods

The main purpose of this research is to analyze variations in vulnerability to wildfire events between crop types, with Grapes and Pastureland serving as the crops of interest. To determine the relationships of interest, we will implement several modeling techniques, which rely on *Hazard Class* and *Distance* as the response variables. *Hazard Class* and *Distance* serve as proxies for the likelihood of experiencing a wildfire event. Inspiration for these response variables comes from a study on cannabis wildfire vulnerability (Dillis). The justification for the use of the *Distance* variable is based on the observation that Grape farms and Pastureland, in addition to cannabis farms, tend to be in more unconventional areas than other crops. For example, grape production is concentrated on the Western coast, whereas most agricultural activity is concentrated in the Central Valley region. Coincidentally, the perimeters of the state experience many wildfire events, whereas the central region remains relatively protected from wildfires. Furthermore, *Distance* serves as a proxy for wildfire vulnerability because of the assumption that observations with smaller distances, for example 0.01 kilometers, are more prone to wildfires than observations with larger distances, such as 100 kilometers.

In addition to the inclusion of *Hazard Class* and *Distance* as response variables, the following explanatory variables will be employed: *Crop Type*, *Number of Frost-Free Days*, *Latitude*, and *Longitude*. We will also test two different soil variables with each modeling technique: *Hydrologic Group* and *Soil Compaction Percentage*. By testing how models perform with different soil variables, we are able to determine the optimal variable for modeling. Furthermore, this will serve as a robustness check, and we expect that the results should not change significantly with the inclusion of either soil variable. For consistency, all modeling techniques are denoted by a letter, A-E, and a number 1 or 2, indicating the use of *Soil Compaction Percentage* (1) or *Hydrologic Group* (2). In the first modeling stage, we will implement preliminary pixel-level OLS regressions to get a better understanding of how wildfire proximity varies among crops. The first modeling stage utilizes *Distance* as the response variable alongside the soil variables. Model A1 incorporates the *Soil Compaction Percentage* variable, where Model A2 utilizes the *Hydrologic Group* variable. The equations for Model A1 and A2 are listed below and the outcomes are detailed in Appendix Tables 7 and 8:

Model A1:

$$Y_i = \alpha_0 + \beta_1 \text{SoilCompactionPercentage}_i + \beta_2 \text{NumberOfFrostFreeDays}_i + \beta_3 \text{CropType}_i + \epsilon_i \text{ where, } \epsilon_i \sim N(0, \sigma^2).$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Model A2:

$$Y_i = \alpha_0 + \beta_1 \text{HydrologicGroup}_i + \beta_2 \text{NumberOfFrostFreeDays}_i + \beta_3 \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

While these models offer a promising starting point, evident in Appendix Tables 7 and 8, they exhibit severe violations of the linear framework. In Appendix Figures 11 and 12, we observe violations of the constant variance assumption in both models from the Residuals Versus Predicted Plots. The constant variance assumption requires the normal distribution of *Distance* to remain consistent across all values of the explanatory variables. However, a downward-sloping cone is prevalent in both plots, indicating violations of the constant variance assumption. Furthermore, the parabola shapes in the Residuals Versus Predicted Plot for Model A1 indicate the presence of non-linear relationships. In addition, the Histogram of Residuals and QQ Plot shed light on severe normality violations, suggesting that the distribution of *Distance* is not normal across all values of the explanatory variables.

In addition, due to the nature of our data, the independence assumption is violated. The independence assumption requires that observations are unrelated. Given that the average farm size is 445 acres and the observations were derived from a dataset with 30-meter pixels, there are clearly many farms with multiple observations. Furthermore, we should also expect crops to cluster in landscapes suited for their growth. For example, grape production is centered in the Coachella and San Joaquin Valleys. Weather patterns and soil properties also tend to be similarly clustered, resulting in spatial autocorrelation, which refers to the correlation in variables across geographic locations. This implies that neighboring areas have similar crops, weather, and soil.

To quantify the extent of spatial autocorrelation present in the models, we will deploy Moran's I Statistic. Moran's I Statistic takes on values between -1 and 1, where -1 represents perfect dispersion and 1 represents perfect clustering of similar values. Model A1 and Model A2 yield Moran's I Statistic values of .924 and .925, respectively, indicating severe autocorrelation as expected.

Noting the significant autocorrelation, we will aim to improve upon the preliminary regressions by including *Latitude*, *Longitude*, and their interaction. The inclusion of *Latitude* and *Longitude* helps to account for spatial trends in the data. Their interaction allows for the impact of *Latitude* on *Distance* to depend on *Longitude*, and vice versa. This interaction is necessary given California's shape and diverse geography. In addition, at the same longitudinal coordinate, observations could be inland, on the coast, or in the ocean, thus providing further justification for the interaction term. Adding these terms in addition to the baseline explanatory variables generates Model B1 with *Soil Compaction Percentage* and Model B2 with *Hydrologic Group*. The equations are shown below and the results are displayed in Appendix Tables 9 and 10:

Model B1:

$$Y_i = \alpha_0 + \beta_1 \text{SoilCompactionPercentage}_i + \beta_2 \text{NumberOfFrostFreeDays}_i + \beta_3 \text{CropType}_i + \beta_4 \text{Latitude}_i + \beta_5 \text{Longitude}_i + \beta_6 \text{Latitude}_i \text{Longitude}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Model B2:

$$Y_i = \alpha_0 + \beta_1 \text{HydrologicGroup}_i + \beta_2 \text{NumberOfFrostFreeDays}_i + \beta_3 \text{CropType}_i + \beta_4 \text{Latitude}_i + \beta_5 \text{Longitude}_i + \beta_6 \text{Latitude}_i \text{Longitude}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Models B1 and B2 provide a more comprehensive approach with the inclusion of *Latitude* and *Longitude*, demonstrated in Appendix Tables 7 and 8. However, these models also fail to adhere to the assumptions of the linear framework. In Appendix Figures 13 and 14, we see severe constant variance violations in the Residuals Versus Predicted Plots. This violation is evident in the downward-sloping cones, implying that the variance in the normal distribution of *Distance* varies based on values of the explanatory variables. Furthermore, the significant deviance from the line in the QQ plot, in addition to the skew present in the Histograms of Residuals, provides evidence of a normality violation. This implies that *Distance* does not follow a normal distribution at all levels of the explanatory variables.

To address these severe violations, in addition to autocorrelation, we will leverage Geographically Weighted Regression (GWR). GWR is a modeling technique that serves as an extension of the familiar OLS regression but allows for coefficients to vary spatially. Unlike the traditional OLS regression, where a single global coefficient is calculated using the entire dataset, GWR computes a regression at each point in the dataset. To compute a regression at each point, GWR relies on Tobler's Law of Spatial Dependence, which states that neighboring observations are more likely to exhibit similarly. GWR utilizes Tobler's Law to borrow a number of neighboring observations to determine the aforementioned local regressions. The bandwidth describes the number of neighboring observations used in the regression calculations and was optimized using AIC. After optimizing the bandwidth, Model C1 and Model C2 utilize the GWR framework with *Distance* as the response variable and both soil variables. This produces the following equations below, with the outcomes detailed in Appendix Tables 11 and 12:

Model C1:

$$Y_i = \alpha_0 + \beta_{1i}(u_i, v_i) \text{SoilCompactionPercentage}_i + \beta_{2i}(u_i, v_i) \text{NumFrostFreeDays}_i + \beta_{3i}(u_i, v_i) \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Model C2:

$$Y_i = \alpha_0 + \beta_{1i}(u_i, v_i) \text{HydrologicGroup}_i + \beta_{2i}(u_i, v_i) \text{NumFrostFreeDays}_i + \beta_{3i}(u_i, v_i) \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

In the equations, (u_i, v_i) represents the *Latitude* and *Longitude* of observation i. Thus, the $\beta_{ij}(u_i, v_i)$ coefficients represent the local coefficient calculated for each variable and observation i. The error term represents the error for observation i.

Next, we will aim to improve upon the results from the preliminary pixel-level model with a Generalized Additive Model (GAM). A GAM bears resemblance to the GWR framework in that they both serve as extensions of the linear model infrastructure. Where GWR allows for spatially variant relationships, GAM adds the possibility of having non-linear relationships between the explanatory variables and response variable. The GAM framework allows for non-linear relationships by leveraging a smoothing spline for each quantitative explanatory variable, which captures the complexities of the trends while improving the accuracy of the model.

To illustrate this concept, Figures 7 and 8 capture the relationship between the quantitative explanatory variables, *Number of Frost-Free Days* and *Soil Compaction Percentage*, and *Distance*, respectively. Each graph includes a smoothed curve in addition to a linear line of best fit. Both plots provide evidence of non-linear relationships with the outcome variable, *Distance*. In Figure 7, a significant peak in *Distance* is observed among observations exceeding 300 frost-free days annually. Furthermore, Figure 8 features peaks in *Distance* corresponding to specific *Soil Compaction Percentage* ranges, most notably around 0-5%, 35-40%, and exceeding 80%.

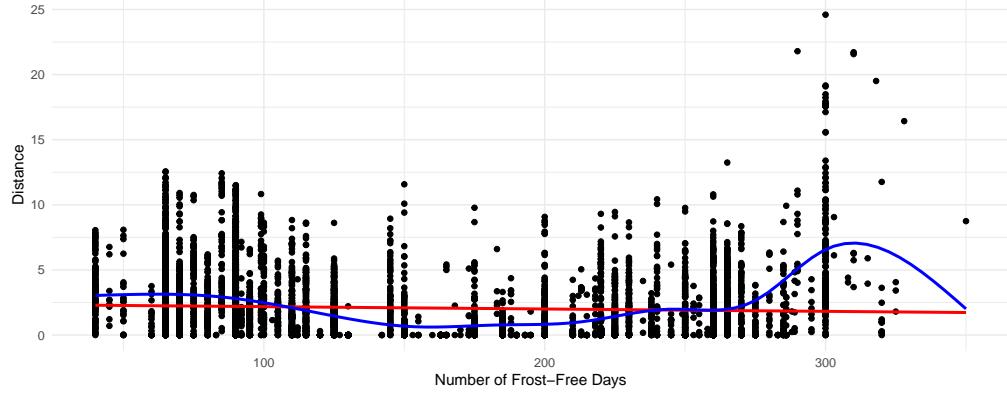


Figure 7: Number of Frost-Free Days Versus Distance

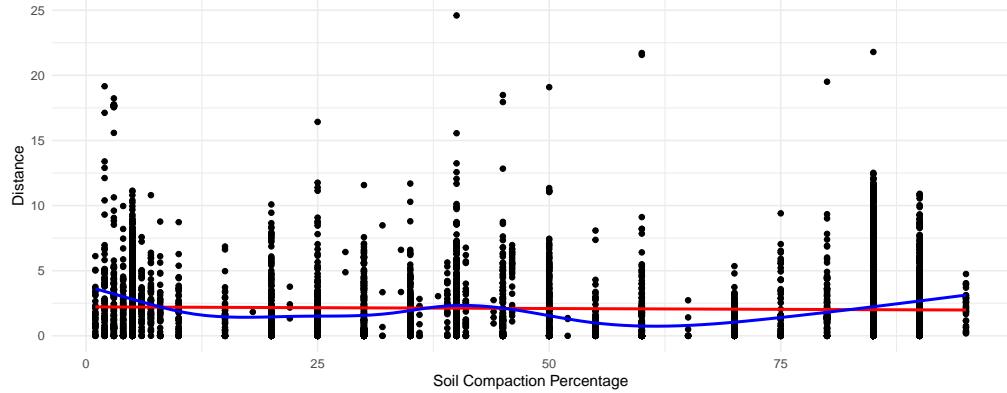


Figure 8: Soil Compaction Percentage Versus Distance

To account for the non-linear relationships evident in Figures 7 and 8, Models D1 and D2 utilize the GAM framework. Model D1 and D2 employ *Distance* as the response variable, alongside the two soil variables. The equations are listed below and the results are detailed in Appendix Tables 13 and 14:

Model D1:

$$Y_i = \alpha_0 + \beta_1 \text{SoilCompactionPercentage}_i + f_1 \text{NumFrostFreeDays}_i + \beta_3 \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Model D2:

$$Y_i = \alpha_0 + f_1 \text{HydrologicGroup}_i + f_2 \text{NumFrostFreeDays}_i + \beta_3 \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

In these equations, we see that for the quantitative explanatory variables, the typical β coefficients are replaced with f_j . The f_j terms represent the smoothing splines these variables are constrained to.

For the final modeling phase, we will transition to using *Hazard Class* as the response variable. Recall that *Hazard Class* is a categorical variable with 3 categories (Moderate, High, and Very High). Since *Hazard Class* has multiple output classes, we will rely on the multinomial framework. A multinomial regression is analogous to a binomial regression, only differing in the number of output classes allowed. As such, the multinomial framework allows us to measure the relationship between the explanatory variables and the odds of the various output classes. Two different iterations of the multinomial technique were implemented using the baseline variables and both soil variables. Model E1 utilizes *Soil Compaction Percentage* and Model E2 incorporates *Hydrologic Group*. The associated equations are listed below and the results are detailed in Appendix Tables 15 and 16:

Model E1:

$$\text{logit}(\text{HazClass})_i = \alpha_0 + \beta_1 \text{SoilCompactionPercentage}_i + \beta_2 \text{NumFrostFreeDays}_i + \beta_3 \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Model E2:

$$\text{logit}(\text{HazClass})_i = \alpha_0 + \beta_1 \text{HydrologicGroup}_i + \beta_2 \text{NumFrostFreeDays}_i + \beta_3 \text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Concluding the modeling phase, we are left with 10 total models. To compare model performance, we will rely solely on AIC and BIC. While conventional practice involves use of multiple comparison metrics, the complex modeling techniques lack simple metrics like R^2 . Nonetheless, AIC and BIC are robust comparison metrics that aim to capture a balance between goodness-of-fit and model complexity, with lower values being preferred. Table 2 provides the comparison metrics for all models.

Table 2: AIC and BIC Values For Model Comparison

Model	AIC	BIC
ModelA1	34912.11	34953.64
ModelA2	34849.41	34911.71
ModelB1	34625.07	34687.37
ModelB2	34599.70	34682.77
ModelC1	29603.25	23405.20
ModelC2	31888.16	25020.39
ModelD1	33914.79	33956.33
ModelD2	33933.46	33995.77
ModelE1	12433.80	12503.03
ModelE2	12660.46	12771.22

Table 2 demonstrates that among the models which rely on *Distance* as the response variable, Model C1 achieves the lowest AIC and BIC values, recording values of 29,603.25 and 23,405.20, respectively. Model C1 employs the GWR framework with *Soil Compaction Percentage*, *Crop Type*, and *Number of Frost-Free Days* as explanatory variables. Model E1 yields the lowest AIC and BIC values between Model E1 and Model E2, the two multinomial models with *Hazard Class* as the response variable. Model E1 achieves AIC and BIC values of 12,433.80 and 12,503.03, respectively. Model E1 includes *Soil Compaction Percentage*, *Crop Type*, and *Number of Frost-Free Days* as explanatory variables. We will proceed with further analysis of Model E1 and Model C1.

Results

Following the model comparison stage, we will proceed with further interpretations of the superior models, Model C1 and Model E1. For the reader's convenience, the equation and coefficient table for Model C1 are provided below:

Model C1:

$$Y_i = \alpha_0 + \beta_{1i}(u_i, v_i)\text{SoilCompactionPercentage}_i + \beta_{2i}(u_i, v_i)\text{NumFrostFreeDays}_i + \beta_{3i}(u_i, v_i)\text{CropType}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Table 3: Model C1 Local Coefficient Summaries and Global Coefficient

	Min.	1st Qu.	Median	Mean	3rd Qu.	Global
Intercept	-40.8	0.2	2.1	1.6	3.5	16.8
NumberOfFrostFreeDays	-0.1	0.0	0.0	0.0	0.0	0.2
CropTypeOther	-15.6	-0.2	0.4	0.5	1.0	12.3
CropTypepasture	-21.6	-0.4	0.1	0.3	1.0	10.6
SoilCompactionPercentage	-0.1	0.0	0.0	0.0	0.0	0.1

Table 3 presents the coefficient table for Model C1, featuring a five-number summary of the local coefficients, alongside the global coefficients in the right-most column. The global estimations are equivalent to Models A1 and A2, as the GWR leverages the same explanatory variables. The local coefficients are reported in the format of a summary including the minimum, first quartile, median, third quartile, and maximum values of the local coefficients. This summary provides insight into the spatial variability of the coefficients across the study area. In addition, we'll focus our interpretations on the local coefficients and their deviance from the global coefficients. In general, we tend to see much larger global coefficients than local coefficients.

This discrepancy implies that the overall impacts of the variables are greater globally than locally, which is expected due to spatial variation in the variables.

The intercept, which represents Grapes with 0 frost-free days and 0% soil compaction, is associated with a median and mean *Distance* value of 2.1 and 1.6 kilometers, respectively. The local intercept coefficients vary from .2 to .5 between the first and third quartile. Furthermore, the median local coefficient exceeds mean local coefficient, demonstrating a slight negative skew in the distribution of local coefficients. This implies the presence of a number of extremely small local coefficients. In addition, the local coefficients differ significantly from the global coefficient of 16.8 kilometers.

The *Soil Compaction Percentage* variable attains a value of 0 for both the median and mean local coefficients. In addition, the local coefficients remain constant at 0 across the first and third quartiles, illustrating a negligible relationship between *Soil Compaction Percentage* and *Distance* at the local level. The global coefficient of .1 indicates a slight positive relationship between compaction and wildfire proximity. This positive relationship can be attributed to spatial variation as it is exclusively observed in the global regression.

Furthermore, the *Number of Frost-Free Days* variable yields a median and mean local coefficient of 0. The local estimates do not vary between the first and third quartile, but remain constant at 0. This implies that after accounting for all other variables, additional frost-free days have a negligible impact on the *Distance* variable at the local level. In addition, the global coefficient is .2, which implies a slight positive relationship between *Number of Frost-Free Days* and proximity to wildfires. However, it is important to note that this positive relationship is only observed in the global regression. Thus, we can attribute this relationship to spatial variation rather than causal effect.

Next, we will delve into the interpretations of the *Crop Type* coefficients. Pastureland achieves median and mean local coefficient values of .1 and .3, respectively. This suggests that pastureland is associated with an increase in *Distance* of about .1 to .3 kilometers when compared to Grapes and all other variables are held constant. The local coefficients range from -.4 to 1 between the first and third quartile, illustrating that the impact of Pastureland when compared to Grapes on *Distance* differs significantly throughout the study area. Furthermore, the mean local coefficient exceeds the median local coefficient, providing evidence of a positive skew in the distribution of local coefficients. This implies the presence of extremely large local coefficient values, which causes the mean to be skewed upwards. Moreover, the global coefficient of 10.6 is much larger than the central values of the local coefficients.

“Other” farms yield median and mean local coefficients of .4 and .5, respectively, suggesting that “Other” farms are situated around .4 to .5 kilometers closer to wildfires than Grape farms on average, when all other variables are held constant. The local coefficients vary between -.2 and 1 between the first and third quartile, demonstrating the spatial variability of the impact of being an “Other” farm compared to Grape farms on *Distance*. Furthermore, the global coefficient of 12.3 kilometers indicates greater distances for “Other” farms when compared to Grape farms. Once again, the global coefficient is much larger than the local coefficients, which relates to the spatial variability of the impact of being an “Other” farm on *Distance*. From these results, it is plausible to conclude that “Other” farms and Pastureland are associated with slightly greater *Distance* values when compared to Grape farms, implying higher wildfire vulnerability for Grape farms. However, it is important to note the magnitude of the coefficients produced. “Other” farms and Pastureland being located only a few kilometers further from historical wildfires is a result that is practically meaningless, since the magnitude of the impact is minuscule. As such, these results do not provide any true insight into a severe disparity in wildfire vulnerability, even if they may help build our intuition.

Next, we will delve into Model E1 interpretations with *Hazard Class* as the response variable. The equation is listed below, followed by Table 4, which contains the exponentiated coefficients. When interpreting a multinomial regression, exponentiating the coefficients provides insight into the relative likelihood of each category compared to the reference category, which is taken to be the High Class.

Model E1:

$$\text{logit}(\text{HazClass})_i = \alpha_0 + \beta_1 \text{SoilCompactionPercentage}_i + \beta_2 \text{NumFrostFreeDays}_i + \beta_3 \text{Crop}_i + \epsilon_i$$

where, $\epsilon_i \sim N(0, \sigma^2)$.

Table 4: Model E1 Coefficients

	Intercept	SoilCompactionPercentage	CropTypeOther	CropTypepasture	NumberOfFrostFreeDays
Moderate	0.812	1.006	10.230	3.155	0.994
Very High	1.514	1.025	0.131	0.749	0.994

The *Soil Compaction Percentage* variable yields an exponentiated coefficient of 1.006 for the Moderate Class and 1.025 for the Very High Class, displayed in Table 4. Thus, holding all other variables constant, if there is a one-percent increase in *Soil Compaction Percentage*, the odds of being in the Moderate Class are expected to be about 1.006 times greater than the odds being in the High Class. Furthermore, holding all other variables constant, if there is a one-percent increase in *Soil Compaction Percentage*, the odds of being in the Very High Class are about 1.025 times greater on average than the odds being in the High Class, holding all other variables constant. In summary, we see that an increase in the *Soil Compaction Percentage* is associated with an increased likelihood of being within the Moderate or Very High Classes, although the magnitude of this effect is practically insignificant.

The *Number of Frost-Free Days* variable has an exponentiated coefficient of .994 for the Moderate Class and .994 for the Very High Class. The Moderate Class coefficient reveals that if there is an additional frost-free day, the odds of being in the Moderate Class versus the High Class are about .994 times smaller on average, holding all other variables constant. Similarly, for each additional frost-free day, the odds of being in the Very High Class versus the High class are about .994 times smaller on average, holding all other variables constant. Overall, we see that an increase in the *Number of Frost-Free Days* is associated with a slight decrease in the likelihood of being in the Moderate or Very High Classes, although the magnitude of this effect is practically insignificant.

Next, we will delve into interpretations of the *Crop Type* coefficients. Grapes are taken as the reference crop for interpretations, similar to how the High Class serves as the reference *Hazard Class*. For the Moderate Class, Pastureland receives a coefficient of 3.155 and “Other” crops receive a coefficient of 10.230. Holding all other variables constant, the odds of Pastureland being in the Moderate Class versus the High Class are 3.155 times greater when compared to Grapes. Similarly, the odds of “Other” farms being in the Moderate Class versus the High Class are about 10.230 times greater when compared to Grapes. These results suggest that Pastureland and “Other” observations are more likely to be in the Moderate Class versus the High Class when compared to Grapes.

For the Very High Class, Pastureland achieves an exponentiated coefficient of .749, indicating that the odds of Pastureland being in the Very High Class versus the High Class are about .749 times smaller when compared to Grapes. Similarly, the odds of “Other” farms being in the Very High Class versus the High Class are about .1301 the odds of Grapes. This indicates that Pastureland and “Other” crops are less likely to be in the Very High Class compared to Grapes. In conjunction with the Moderate Class coefficients, we see evidence of Grapes having an increased likelihood of more severe classes, signaling that Grapes are more vulnerable to wildfires.

Table 5: Model E1 P-Values

	Intercept	SoilCompactionPercentage	CropTypeOther	CropTypepasture	NumberOfFrostFreeDays
Moderate	0.101	0	0	0	0
Very High	0.001	0	0	0	0

Table 6: Model E1 Confidence Intervals

	2.5 %.Moderate	97.5 %.Moderate	2.5 %.Very High	97.5 %.Very High
(Intercept)	0.633	1.041	1.180	1.942
SoilCompactionPercentage	1.004	1.008	1.022	1.027
CropTypeOther	8.458	12.375	0.104	0.164
CropTypepasture	2.648	3.759	0.649	0.863
NumberOfFrostFreeDays	0.993	0.995	0.993	0.995

Finally, we'll have to verify our results. From Table 5, we see that the coefficients are all statistically significant at the 1% level, except for the intercept of the Moderate Class. This provides strong evidence against the null hypothesis, which assumes there is no relationship between the explanatory and response variable. Furthermore, the confidence intervals in Table 6 reinforce our conclusions, as the exponentiated ranges fall exclusively above or below 1. We rely on the value of 1 as the benchmark for the confidence intervals, since this represents no change in the odds, whereas values not equal to 1 indicate an increase or reduction in the odds of a specific class. Once again, the confidence interval for the Moderate intercept spans 0.633 to 1.041, indicating a lack of statistically conclusive results for this coefficient.

Conclusion

In attempting to analyze how wildfire vulnerability varies among different crops, various modeling techniques were performed. Two superior models emerge after comparing the AIC and BIC metrics- a GWR with *Distance* as the response variable and a multinomial model with *Hazard Class* as the response variable. From the GWR, we have some evidence of Grape farms having closer proximity to wildfires than Pastureland and "Other" crops, indicating Grapes having the highest wildfire vulnerability. However, the magnitude of the coefficients renders the conclusions from this model practically meaningless. From the multinomial model, we have statistically significant evidence that Grapes are more likely to grow in areas with more severe classes, indicating that Grapes are more vulnerable to wildfire events when compared to alternate crops. These findings align with our intuition, given that Grapes farms tend to be located in areas more prone to wildfires. This finding is especially concerning given the sensitivity of Grapes under wildfire smoke. Even minimal exposure can tarnish harvests, rendering them unsuitable for wine making (Zakowski). Further research into variations in wildfire vulnerability across crops is essential, especially in the era of increasing wildfire severity and frequency. Suggestions for future research include exploring alternate variables, testing new crops of interest, and implementing temporal analysis.

Appendix

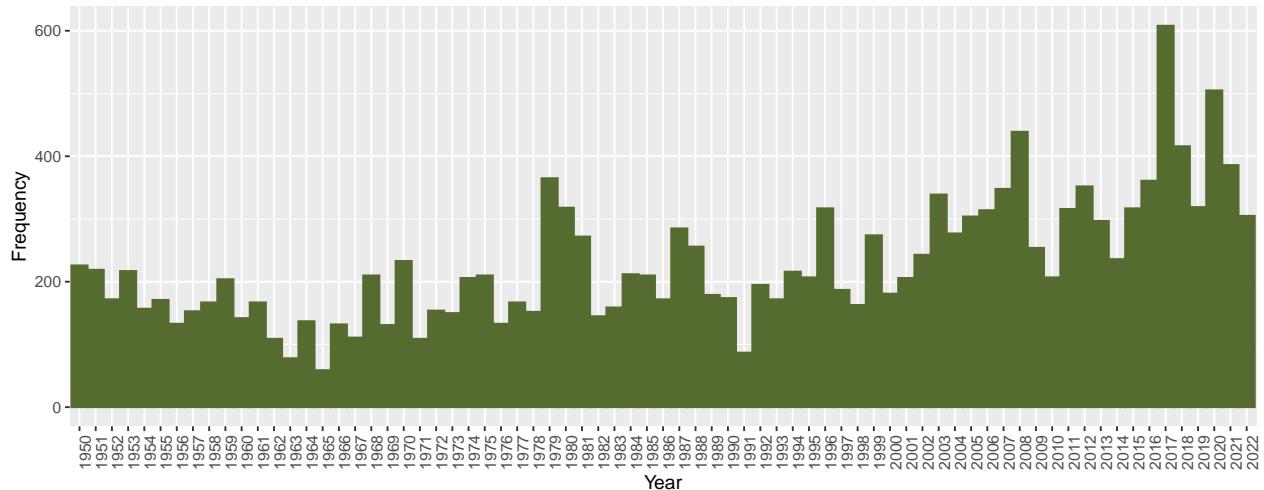


Figure 9: Number of Fires by Year

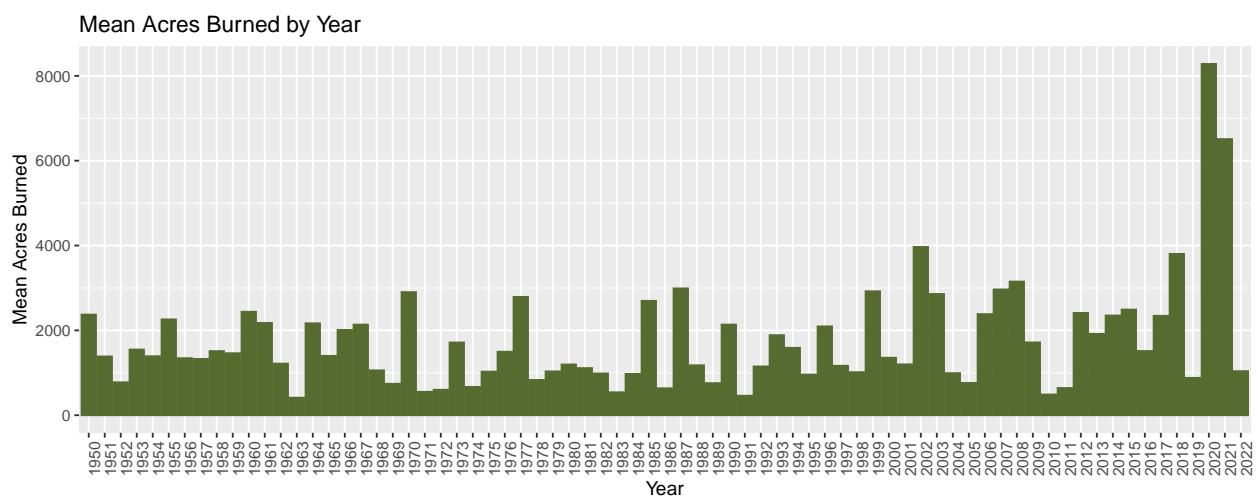


Figure 10: Mean Acres Burned by Year

Table 7: Model A1 Coefficients

	coefficients
(Intercept)	0.706
SoilCompactionPercentage	-0.002
NumberOfFrostFreeDays	0.003
CropTypeOther	2.146
CropTypepasture	0.987

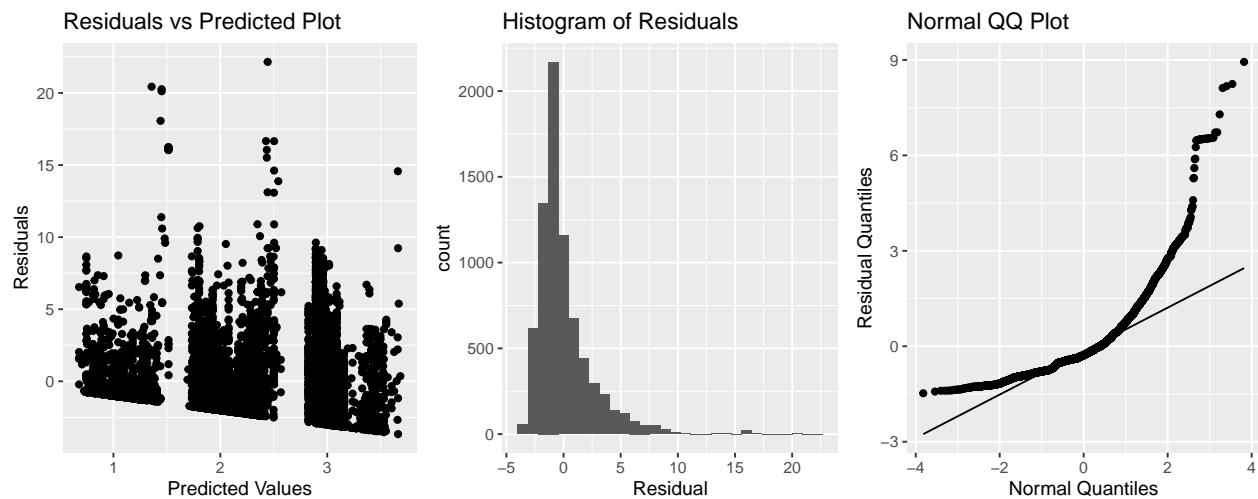


Figure 11: Model A1 Assumption Checks

Table 8: Model A2 Coefficients

	coefficients
(Intercept)	0.594
HydrologicGroupB	-0.198
HydrologicGroupC	0.350
HydrologicGroupD	0.342
HydrologicGroupOther	0.770
NumberOfFrostFreeDays	0.003
CropTypeOther	1.934
CropTypepasture	0.791

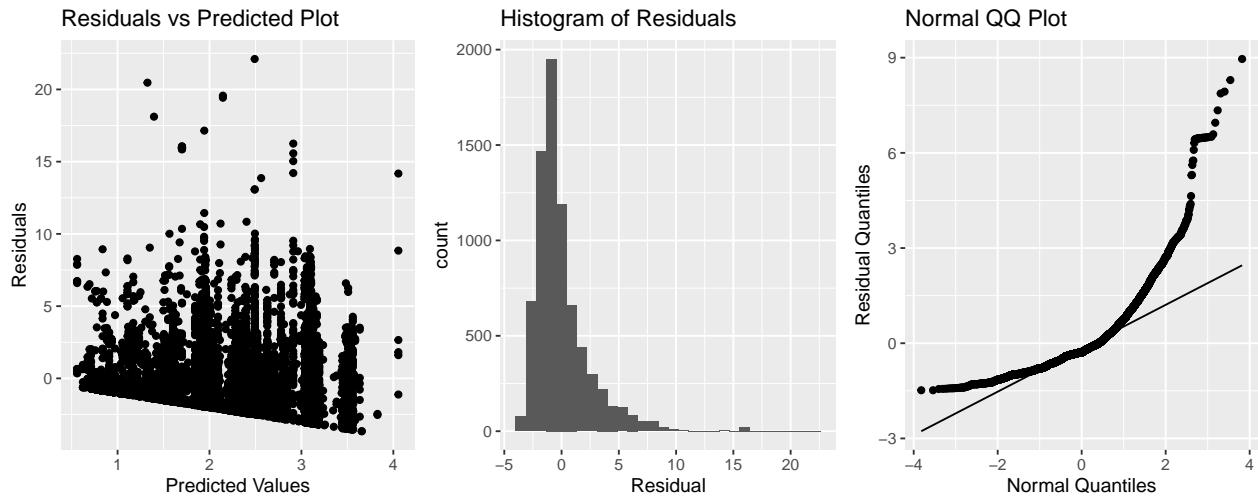


Figure 12: Model A2 Assumption Checks

Table 9: Model B1 Coefficients

	coefficients
(Intercept)	-1260.356
SoilCompactionPercentage	-0.006
NumberOfFrostFreeDays	0.008
CropTypeOther	1.724
CropTypepasture	0.711
Latitude	0.000
Longitude	0.000
Latitude:Longitude	0.000

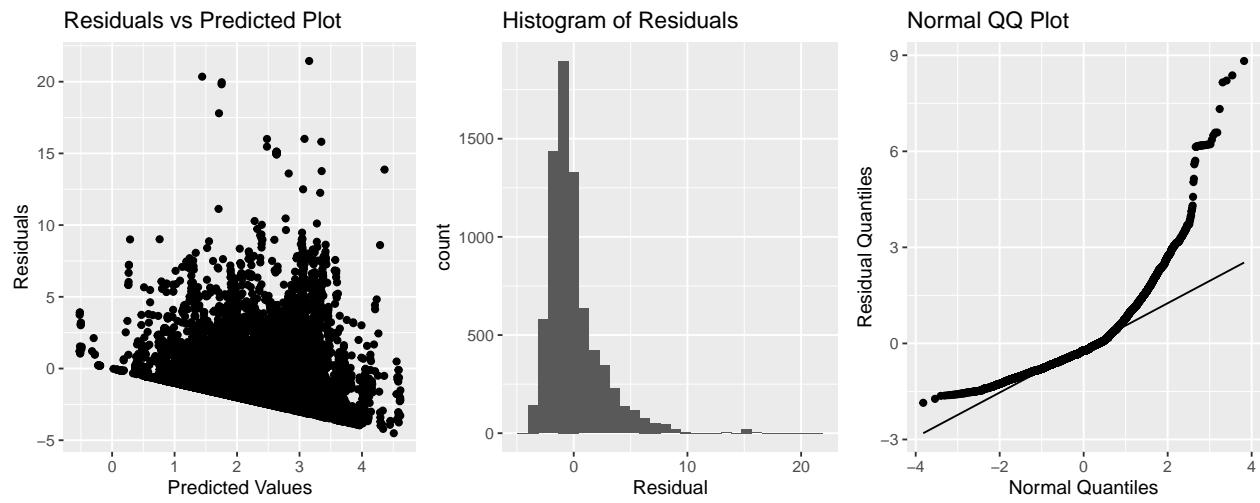


Figure 13: Model B1 Assumption Checks

Table 10: Model B2 Coefficients

	coefficients
(Intercept)	-1191.018
HydrologicGroupB	-0.239
HydrologicGroupC	0.298
HydrologicGroupD	0.311
HydrologicGroupOther	0.660
NumberOfFrostFreeDays	0.007
CropTypeOther	1.636
CropTypepasture	0.639
Latitude	0.000
Longitude	0.000
Latitude:Longitude	0.000

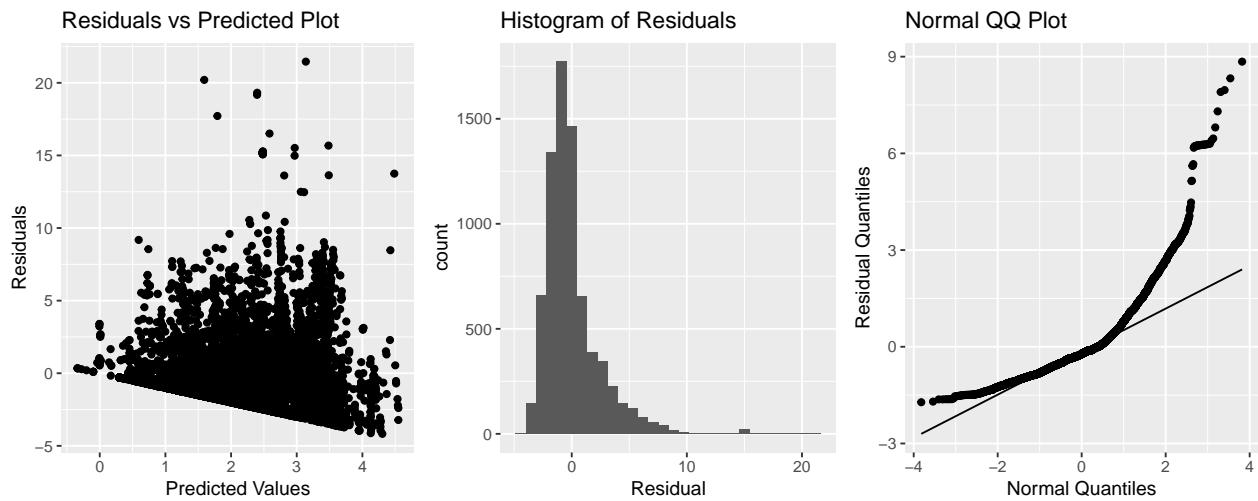


Figure 14: Model B2 Assumption Checks

Table 11: Model C1 Coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Global
Intercept	-40.8	0.2	2.1	1.6	3.5	16.8
NumberOfFrostFreeDays	-0.1	0.0	0.0	0.0	0.0	0.2
CropTypeOther	-15.6	-0.2	0.4	0.5	1.0	12.3
CropTypepasture	-21.6	-0.4	0.1	0.3	1.0	10.6
SoilCompactionPercentage	-0.1	0.0	0.0	0.0	0.0	0.1

Table 12: Model C2 Coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Global
Intercept	-5.9	0.2	1.5	2.3	3.5	14.6
NumberOfFrostFreeDays	0.0	0.0	0.0	0.0	0.0	0.0
CropTypeOther	-1.5	0.3	0.8	0.6	1.1	2.1
CropTypepasture	-2.4	-0.2	0.1	0.2	0.8	1.4
HydrologicGroupB	-17.8	-0.7	-0.1	-1.4	0.5	1.7
HydrologicGroupC	-17.1	-0.3	0.1	-1.0	0.7	2.0
HydrologicGroupD	-17.9	-0.5	0.1	-1.0	0.7	3.6
HydrologicGroupOther	-8.1	-1.3	-0.2	-0.4	1.1	7.1
y	0.0	0.1	1.1	2.1	3.1	24.6

Table 13: Model D1 Coefficients

	coefficients
(Intercept)	1.455
s(NumberOfFrostFreeDays)	0.001
CropTypeOther	1.188
CropTypepasture	0.400
s(SoilCompactionPercentage)	-0.001

Table 14: Model D2 Coefficients

	coefficients
(Intercept)	1.199
s(NumberOfFrostFreeDays)	0.001
CropTypeOther	1.193
CropTypepasture	0.330
HydrologicGroupB	-0.061
HydrologicGroupC	0.418
HydrologicGroupD	0.372
HydrologicGroupOther	0.104

Table 15: Model E1 Coefficients

	(Intercept)	SoilCompactionPercentage	CropTypeOther	CropTypepasture	NumberOfFrostFreeDays
Moderate	-0.208	0.006	2.325	1.149	-0.006
Very High	0.415	0.025	-2.034	-0.289	-0.006

Table 16: Model E2 Coefficients

	(Intercept)	HydrologicGroupB	HydrologicGroupC	HydrologicGroupD	HydrologicGroupOther	CropTypeOther	CropTypepasture	NumberOfFrostFreeDays
Moderate	0.667	-1.050	-0.412	-0.357	-0.722	2.210	0.922	-0.006
Very High	2.669	-0.858	-0.975	-1.465	-2.482	-1.787	-0.297	-0.006

Works Cited

- ArcGIS. (n.d.). ArcGIS Soil Development Toolkit. Retrieved from <https://www.arcgis.com/home/item.html?id=be2124509b064754875b8f0d6176cc4c>
- ArcGIS. (n.d.). *How features are represented in a raster*. Retrieved from <https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/how-features-are-represented-in-a-raster.htm>
- California 4 Wheel Drive Association. (n.d.). What is 30x30 and Why Does It Matter to OHV Recreation? [Image]. Cal4Wheel. <https://cal4wheel.com/component/zoo/item/what-is-30x30-and-why-does-it-matter-to-ohv-recreation?Itemid=101>
- California Department of Forestry and Fire Protection. (n.d.). *Fire hazard severity zones maps*. Retrieved from <https://osfm.fire.ca.gov/what-we-do/community-wildfire-preparedness-and-mitigation/fire-hazard-severity-zones/fire-hazard-severity-zones-maps>
- California State Government. (n.d.). *California fire perimeters (1950-present)*. Retrieved from <https://data.ca.gov/dataset/california-fire-perimeters-1950>
- Comber, L. (2022). *GEOG3915 GeoComputation and Spatial Analysis practicals*. Retrieved from <https://bookdown.org/lexcomber/GEOG3195/spatial-models---geographically-weighted-regression.htm>
- Dillis, C., Butsic, V., Moanga, D., Parker-Shames, P., Wartenberg, A., & Grantham, T. E. (2022). *The threat of wildfire is unique to cannabis among agricultural sectors in California*. *Ecology and Evolution*, 12(17), e4205. <https://doi.org/10.1002/ecs2.4205>
- Dove, N. C., Safford, H. D., Bohlman, G. N., Estes, B. L., & Hart, S. C. (2020). High-severity wildfire leads to multi-decadal impacts on soil biogeochemistry in mixed-conifer forests. *Ecological Applications*, 30(4), e02072. <https://doi.org/10.1002/eap.2072>.
- Environmental Market Solutions Lab. (n.d.). *Long-Term Trends in Wildfire Damage in California*. University of California, Santa Barbara. Retrieved from <https://emlab.ucsb.edu/sites/default/files/documents/wildfire-brief.pdf>
- HydrologyVideos. (2022, May 22). *Creating Curve Number Grid (2/7) - Creating a geodatabase for storing gSSURGO soil data* [Video]. YouTube. https://www.youtube.com/watch?v=ihx6iZeYS_U
- National Oceanic and Atmospheric Administration. (n.d.). *Wildfire-climate connection*. Retrieved from <https://www.noaa.gov/noaa-wildfire/wildfire-climate-connection#:~:text=Research%20shows%20that%20changes%20in,fuels%20during%20the%20fire%20season>.
- QCBS (Quebec Centre for Biodiversity Science). *Generalized Additive Models with Multiple Smooth Terms* [Webpage]. Retrieved from <http://r.qcbs.ca/workshop08/book-en/gam-with-multiple-smooth-terms.html>
- Soil Survey Staff. *Gridded Soil Survey Geographic (gSSURGO) Database for California*. United States Department of Agriculture, Natural Resources Conservation Service. Available online at <https://gdg.sc.egov.usda.gov/>.
- US Department of Agriculture, National Agricultural Statistics Service. (2023). *Crop Reports*. https://www.nass.usda.gov/Statistics_by_State/California/Publications/AgComm/index.php
- Zakowski, E., Wright, C., Smith, J., & Johnson, L. (2023, July 13). *California wine grape growers need support to manage risks from wildfire and smoke*. University of California, Agriculture and Natural Resources. Retrieved from <http://calag.ucanr.edu/archive/?type=pdf&article=ca.2023a0006>