SPATIAL AND TEMPORAL PATTERNS OF WILDFIRES IN CALIFORNIA

A PREPRINT

Shu Li
Department of Civil and Environmental Engineering
University of California, Irvine
Irvine, CA 92697, USA
shul15@uci.edu

Tirtha Banerjee
Department of Civil and Environmental Engineering
University of California, Irvine
Irvine, CA 92697, USA
tirthab@uci.edu

September 25, 2020

ABSTRACT

The environmental pollution, property losses and casualties caused by wildfires in California are getting worse by the year. To minimize the interference of wildfires on economic and social development, and formulate targeted mitigation strategies, it is imperative to understand the scale and extent of previous wildfire occurrences. In this study, we studied the trend of wildfires in different time scales (monthly, seasonal, and yearly), as well as the distribution of wildfires caused by natural and anthropogenic factors across different spatial scales (administrative units, land use) in California from 2000 to 2019. Furthermore, a regression analysis of environmental and human-related variables with the occurrence and frequency of wildfires was carried out, to compare the importance of variables on the risk of wildfire occurrence. The results show that the frequency distribution of the burned area conforms to the characteristics of the Pareto distribution in the twenty years of this research. The spatial distribution of wildfires was closely related to factors such as the causes, population density, and land-use types. In terms of the variables related to the risk of wildfire occurrence, distance to human constructions, the elevation of the terrain, grass cover, and the vapor pressure deficit are crucial. This study reveals the relationship between environmental and social background conditions and the spatial-temporal distribution of wildfires, which can provide a reference for wildfire management, the formulation of future targeted wildfire emergency plans, and the planning of future land use in California.

Keywords wildfire occurrence · burned area · spatial distribution · kernel density estimation · univariate regression

1 Introduction

As one of the most frequent natural disasters in California, wildfires have caused great damage to the environment, economy and society in recent years (Duffield et al., 2013; Shvidenko et al., 2011). Especially in the past two decades, changes in climate and land utilization caused by human activities have not only extended the wildfire season, but also significantly increased the severity and burned areas of wildland fires (Nauman, 2020). At the same time, the expansion of the wildland-urban interface (WUI) areas caused by rapid social development and sustained population growth has greatly increased the number of residents and buildings affected by wildfires, which has further aggravated the damage imparted to the human society from wildfires (Miranda et al., 2020; Radeloff et al., 2018). According to the data from the wildfire Redbooks published by CAL FIRE, despite significant administrative investments in wildfire suppression and management in recent years, the property loss caused by wildfires has not been significantly reduced in California (Evarts, 2019).

The development and implementation of proactive fire prevention policies can effectively reduce the probability of wildfire ignition, the risk of extreme fires, and the social and economic losses caused by wildfires. The formulation of effective policies entails a full understanding of the spatial and temporal distribution of different types of wildfires (natural and human-caused), the differences in their impact on human communities, and their various influential factors. To this end, the dominant causes and drivers of California wildfires in different periods and regions have already been analyzed by several researchers. Faivre et al. (2014) used the logistic and Poisson regression models to analyze fires in Southern California National Forests from 1980 to 2009. The results indicated that the distance from wildfire ignition points to houses and highways, and the terrain slope were the leading factors that explain ignition frequency. Nevertheless, the study was limited to Southern California and focused only on the spatial distribution of wildfires.

Keeley and Syphard (2018)'s analysis of the spatial distribution of wildfires over the past 100 years in California found that the frequency of wildfires declined greatly after 1980, but there has been no corresponding significant change in the total annual burned area. Prior to 1980, the main cause of wildfires in most parts of California was human activity. However, in recent decades, most man-made ignition sources other than power lines have become less frequent, and the positive correlation between wildfire frequency and population distribution has been less pronounced in recent years than it was in the last century. Therefore, the relative importance of relevant variables in influencing wildfire occurrence varies over time. However, this study only focused on the spatial distribution of wildfires with different causes, and did not analyze other factors affecting the spatial distribution of wildfires in detail.

Williams et al. (2019) demonstrated that from 1972 to 2018, the drying of forest fuels due to human-induced climate warming has greatly increased the area of California's forest-fires, especially in the summer months. Thus, wildfire management not only needs to reduce and prevent direct anthropogenic fire sources, but also needs to deal with changes in environmental risks such as human-induced climate change. Effective fire management therefore requires a comprehensive and near-real-time analysis of fire risks in the local natural environment, the scope and intensity of human activities, and the distribution of combustible fuels (Jazebi et al., 2019).

Nevertheless, most of the current literature has been found to be focused on the historical distribution of wildfires, with the study periods ranging from 1910 to 2019. However, in the last two decades, the climate and the distribution of human communities in CA have changed greatly, which should have a significant impact on the ignition, spread and distribution of wildfires. The behavior and patterns of wildfires in California over the past two decades have not been adequately explored. Also, the current studies lack a detailed analysis of wildfires across California and their seasonality. From the perspective of wildfire management, the statistical analysis procedures, classification techniques, and analyses criteria are not consistent among different fire management agencies, administrative units, and relevant government departments, which makes it difficult to coordinate firefighting and prevention. Moreover, due to the complexity of the anthropogenic ignition causes, human-caused wildfires need to be further classified to formulate more targeted policies.

Considering the causalities, economic losses and environmental pollution caused by the combustion and spread of wildfires, it can be more cost-effective to pay more attention to preventing human-caused wildfires than putting them out (Martínez et al., 2009) (it is also worth noting that managed prescribed fires and low intensity natural wildfires are actually beneficial from an ecological perspective for particular landscapes). To develop proactive wildfire prevention measures, it is necessary to conduct a detailed analysis of the current spatiotemporal distribution of wildfires with different causes, especially for the large wildfires. Furthermore, preventing human-caused wildfires at source requires more detailed classification of how, where and why these wildfires start, and identifying the the social factors behind them (Short, 2017). Aiming at this gap, the research scope was expanded to the entire State of California in this study, and CAL FIRE's multi-agency integrated wildfire records were selected as the original data to conduct a unified temporal and spatial distribution analysis. For the sake of eliminating the inconvenience caused by the differences in the classification of wildfires between various agencies, the administrative units covering the whole of California and the wildfire causes classification records provided by CAL FIRE were used as the basis of spatial analysis.

The aforementioned publications have established the close relationships among some environmental and social factors and the probability of wildfires occurrence, which are critical in the formulation of wildfire prevention and management policies. However, the contribution of these external factors to the risk of wildfires is not entirely consistent across time, region, and cause of wildfires. In order to achieve a better understanding of California's current wildfire situation, it is necessary to investigate the distribution of different causes of wildfires in the past two decades in detail. The importance of various external factors in explaining California's wildfire occurrences needs to be analyzed as well.

In light of this context, this study mainly answers the following questions: 1) what is the temporal distribution trend of wildfire frequency and burned area between different time scale within the last two decades? 2) What is the spatial distribution characteristics of wildfires with different causes in the last two decades, and whether the wildfire distribution is related to the distribution of population, housing, and land cover? 3) The correlation and importance of the explanatory natural and social variables with the risk of wildfire occurrence.

2 Methods

2.1 Study Area

California (CA) is located in the western United States and has a Mediterranean climate. Summers in CA are hot and dry, and rainfall is concentrated in winter. The vegetation coverage in CA is about one-third of the total area, and according to the United States National Land Cover Database (NLCD), the main vegetation types are shrubs, evergreen forest and herbaceous (39.03%, 18.59%, and 13.47%) (Jin et al., 2019). In addition, over 147 million trees have died since 2010 across the state (USFS, 2019). Dead vegetation accumulated in forests could be easily ignited by lightning, thunderstorms or sparks left by human activities. In addition, each year from September to May, the dry Santa Ana wind, with high desiccating potential and high wind speed, arrives from the Great Basin and the Mojave Desert in the southwestern inland crossing the mountains. This addition of strong wind forces means that even small ignition sources have the risk of developing into extreme wildfires. The natural environmental conditions of CA make it a high-risk area for wildfires.

2.2 Data

The historical wildfire records along with start time, burned area, fire perimeter and the causes of ignition from 2000 to 2019 were extracted from the CAL FIRE database (https://frap.fire.ca.gov/mapping/gis-data). Their latest fire dataset was updated in May 2020. There is a minimum burned area requirement for wildfires to be included in this database, which is 10 acres for timber fires, 30 acres for brush fires, and 300 acres for grass fires. Using the burned area of 500 acres as a criterion (Holmes et al., 2008), this study divided the wildfires into large and small fires, and counted their frequency and burned area by year according to the occurrence time. The Spearman Rank correlations were utilized to quantify temporal trends in the annual number of wildfires and mean fire size. In terms of the spatial analysis, all the wildfires were grouped by their burned area and ignition causes, and were analyzed separately.

To explore the relationships among different environmental conditions, human activities and the wildfires, a series of explanatory variables were selected to implement a univariate analysis. In terms of the environmental conditions, two to three representative variables were selected from each aspect in the wildfire behavior triangle (topography, fuels and weather) (Werth et al., 2016). The human-related variables were selected according to Faivre et al. (2014)'s and Ruffault and Mouillot (2017)'s research. The list of variables and their sources were shown in Table 1. In order to quantify the variation of each variable over a 20-year period to assess its importance, this study calculated the rate of change of the average of all variables in California from 2000 to 2019, and conducted a univariate analysis in conjunction with the frequency of wildfires.

2.3 Statistical Analysis

Several statistical methods were used to obtain the distribution of wildfires and the relationship between environmental variables and wildfire occurrences. For the sake of indicating the role of each variable in describing the distribution of wildfires, the logistic regression and Poisson regression methods were implemented referring to Faivre et al. (2014)'s study. The logistic regression method was used to analyze the occurrence, which is the presence or absence of ignitions in each administrative unit while the Poisson model was used to analyze the number of ignitions in each unit. To correlate the occurrence of the wildfires with the explanatory variables, California was divided into $3\,km \times 3\,km$ cells, with a total number of 73,455, and 5,177 cells were marked as having been on fire during the past 20 years. Then the variable data in each cell were collected and integrated.

As a generalized linear regression model, logistic regression can dichotomize the dependent variables through the attributes of multiple independent variables (Rodrigues et al., 2014). The equation of the logistic regression is shown in equation (1):

$$ln(\frac{P}{1-P}) = w_0 + w_1 x_1 + \ldots + w_n x_n \tag{1}$$

where P represents the probability of the wildfire occurrence, x represents various characteristics of the samples and w represents the weight of the x.

In Poisson's regression model, the number of wildfires in each cell was taken as the dependent variable, and the regression analysis was carried out for all variables to obtain the contribution of each variable to the results (Rodrigues et al., 2016). The equation of the Poisson regression is shown below:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\, k = 0, 1, \ldots \tag{2}$$

Where λ is the wildfire frequency.

After the training of the logistic regression and the Poisson regression, the weight of each variable in the determination process, that is, the influence of various natural or human factors on the occurrence of wildfires, was obtained.

Moreover, Kernel density estimation (KDE) was implemented on the fire occurrence points in ArcGIS to indicate the hot spot of wildfires. KDE calculated the density of ignition points in a neighborhood around those points, and assigned the density values to cells to make up an intensity map. Conceptually, a smoothly curved surface is fitted over each point. The surface value is highest at the location of the point and diminishes with increasing distance from the point, reaching zero at the Search radius distance from the point (Silverman, 1986). The equation of KDE is shown below:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\{\frac{1}{h}(x - x_i)\} \tag{3}$$

where f(x) represents the density, K is the kernel, h is the bandwidth (Silverman, 1986). The kernel density estimation makes full use of the input data itself and avoids the subjective introduction of prior knowledge, so as to achieve the maximum approximation of the sample data.

3 Results and Discussions

3.1 Temporal distribution of wildfires in CA from 2000 to 2019

Based on the wildfire history records provided by CAL FIRE, the frequency and burned area of wildfires in CA from 2000 to 2019 were extracted, and their Z-score were calculated to reflect their relative standard distance from the mean. The equation for calculating the Z-score is as follows:

$$Z = \frac{x - \mu}{\sigma} \tag{4}$$

Variables	Source
Distance to major roads	Census Bureau
Distance to coast	Census Bureau
Distace to powerline	California Energy Commission
Population density	Census Bureau
Housing density	Microsoft Building Footprint
Elevation	USGS (National Elevation Database)
Slope	USGS (National Elevation Database)
Aspect	USGS (National Elevation Database)
Tree	LANDFIRE (Fuel Vegetation cover)
Shrub	LANDFIRE (Fuel Vegetation cover)
Herb	LANDFIRE (Fuel Vegetation cover)
Max Temperature	PRISM
Max vapor pressure deficit	PRISM

Table 1: List of wildfire-related variables

where x is the fire frequency or burned area each year, μ is the mean and σ is the standard deviation. California has seen an average of 317 wildfires a year over the past 20 years, burning an average of 674,410 acres. As shown in Figure 1 (a) and (b), there were spikes of annual fire frequency and burned area occurring every few years. The significant peaks of wildfire occurrence were in 2008 and 2017 both in terms of large and small wildfires. At the same time, after peak years, the frequency of wildfires and especially the burned area, decreased significantly. According to the National Wildland Coordinating Group's (NWCG) definition of the large fire and Holmes et al. (2008)'s research, fires measuring more than 500 acre (higher than class E) were simply classified as large fires in this study. Based on the historical record from CAL FIRE, the frequency of large wildfires accounted for 19.68 % of the total (1247 out of 6336 wildfires), while the burned area of large wildfires accounted for 97.04 % of the total burned area (13,089.68 out of 13,488.19 thousand acre) in the past two decades.

Figure 1 (c)-(f) shows the annual frequency and total burned area of large and small wildfires in California in the last 20 years. The Z-score reflects the relative position and relative variation trend of fire frequency and burned area after the normalization. As shown in the figure, the peak frequency of the large wildfires appeared in 2008 and 2017 as well, which was consistent with the total wildfire frequency, but the annual frequency of small wildfires only had one obvious peak value which appeared in 2017. However, from the overall trend distribution, the annual frequency distribution of small wildfires was more similar to that of total wildfires (Pearson R = 0.9763), while the correlation between the annual frequency of large wildfires and total wildfires was relatively weak (Pearson R = 0.6811). In terms of the temporal distribution of burned areas, the trend of large wildfires and total wildfires' burned areas was basically the same (Pearson R = 0.9999), while the correlation between the small and the total wildfires' burned areas was very weak (Pearson R = 0.5310). It can be inferred that the occurrence of small wildfires accounts for the majority of the annual frequency of wildfires, and the annual total wildfire burned areas mainly come from the contribution of large wildfire burned area.

The histogram of wildfire burned area (Figure2) shows that the distribution is heavy tailed, which effectively follows the Pareto distribution. Also, the cumulative fire frequency proportion in the figure shows obvious Pareto distribution characteristics: the area limit of large and small wildfires, 500 acre, is a clear turning point of this distribution; the number of small wildfires accounts for 80%; and according to the results above, the burned area of large wildfires

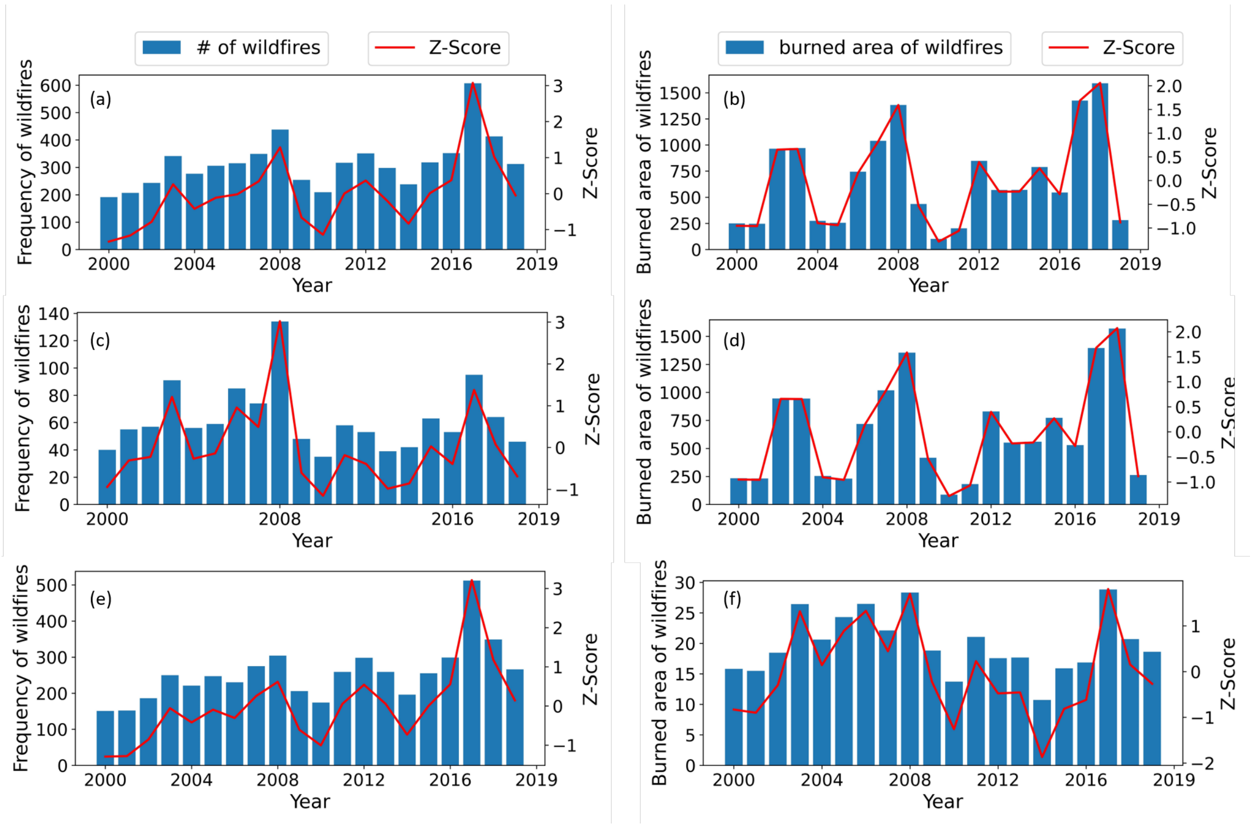


Figure 1: Temporal distribution of wildfire occurrence (blue bars) and their Z-score (red lines) in CA from 2000 to 2019. (a) Wildfire frequency; (b) Wildfire burned area; (c) Large wildfire frequency; (d) Large wildfire burned area; (e) Small wildfire frequency; (f) Small wildfire burned area

accounts for more than 80%. Therefore, the generalized Pareto distribution was applied to fit the burned area distribution, and the probability density function is shown below:

$$y = f(x|k, \sigma, \theta) = \left(\frac{1}{\sigma}\right)\left(1 + k\frac{(x-\theta)}{\sigma}\right)^{-1-\frac{1}{k}} \tag{5}$$

where k represents the shape, σ represents the scale and θ represents the threshold (Embrechts et al., 2013; Kotz and Nadarajah, 2000). The fitting parameters are shown in Table 2. The shape parameter is larger than 1, which means the distribution has a heavy tail with 'infinite mean and variance'. According to Holmes et al. (2008)'s study, the infinite mean and variance indicate that fires burning well over 500 acres within the 20 years are certainly possible. This result confirms the great uncertainty of large wildfires and the challenges of predicting and managing them.

Parameter	Value	Std.Error	Log likelihood
Shape	1.999	0.0391	2320.06
Scale	0.0346	0.0011	

Table 2: Parameter estimates of the Generalized Pareto distribution

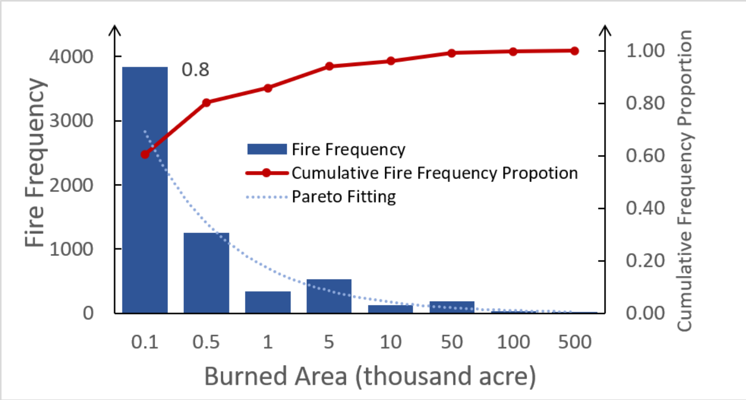


Figure 2: Pareto fitting of the frequency distribution and cumulative proportion of wildfire burned area in CA, 2000-2019

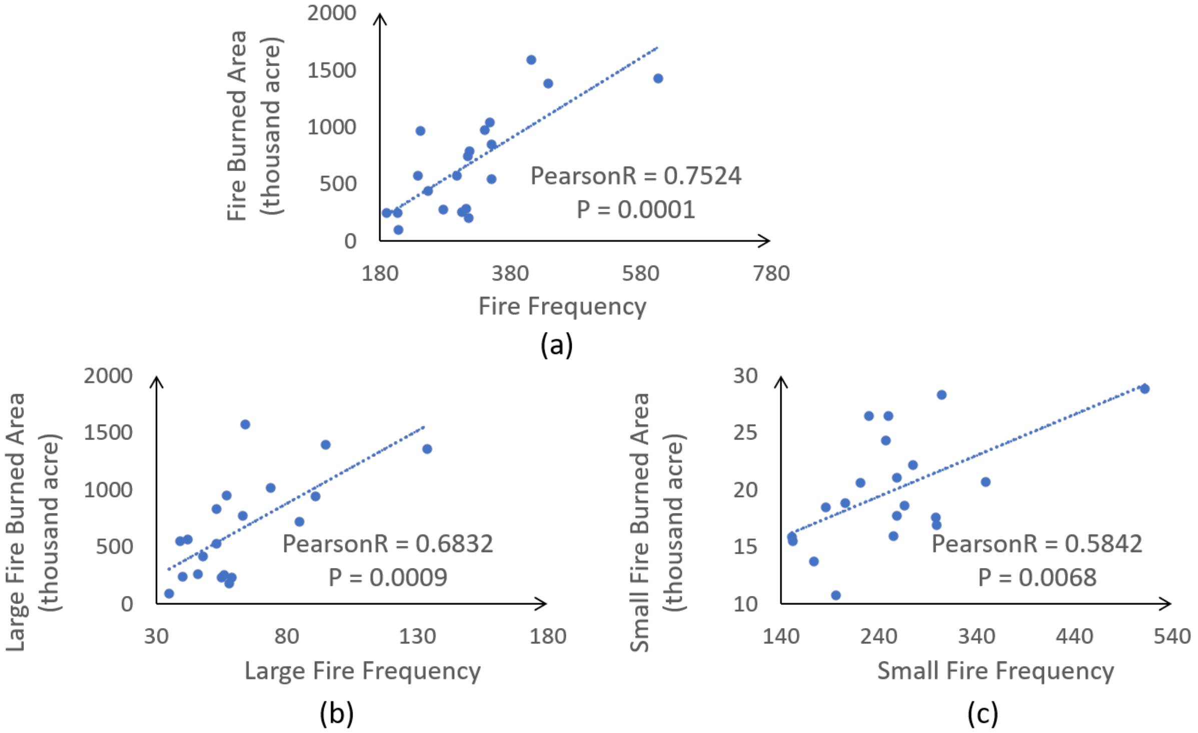


Figure 3: Relationship between the frequency and burned area of wildfires from 2000 to 2019. (a)All the wildfires; (b) Large wildfires; (c) Small wildfires

In addition, the correlation analysis of the frequency and burned area of wildfires in the last 20 years are shown in Figure 3. The Pearson correlation coefficient between total wildfire frequency and burned area is 0.75 ($P < 0.001$), which indicates that they are significantly positively correlated. In addition, the Pearson correlation coefficient between the large wildfire frequency and the burned area of all wildfires is 0.68 ($P < 0.001$). Therefore, even though the burned area is mainly related to large fires, the argument that the more the fire occurred, the greater the burned area is still valid in this study.

The frequency and burned area of wildfires in each month from 2000 to 2019 are shown in Figure 4. It can be observed that wildfires mainly occurred in the second half of the year in California. The peak period was from June to September each year. The number of wildfires in the remaining eight months essentially remains at 20%-30% of the total annual number of wildfires. There is no apparent trend in the distribution of burned areas from June to December, and the peak of the burned area was mainly related to the month in which the extreme fires occurred, which were distributed in all seven months. However, from January to May, there have never been extreme wildfires that reached 20% of the total annual wildfire burned area. The total wildfire areas during these five months were generally less than 10% and have never exceeded 20% of the annual wildfire burned areas in the recent 20 years.

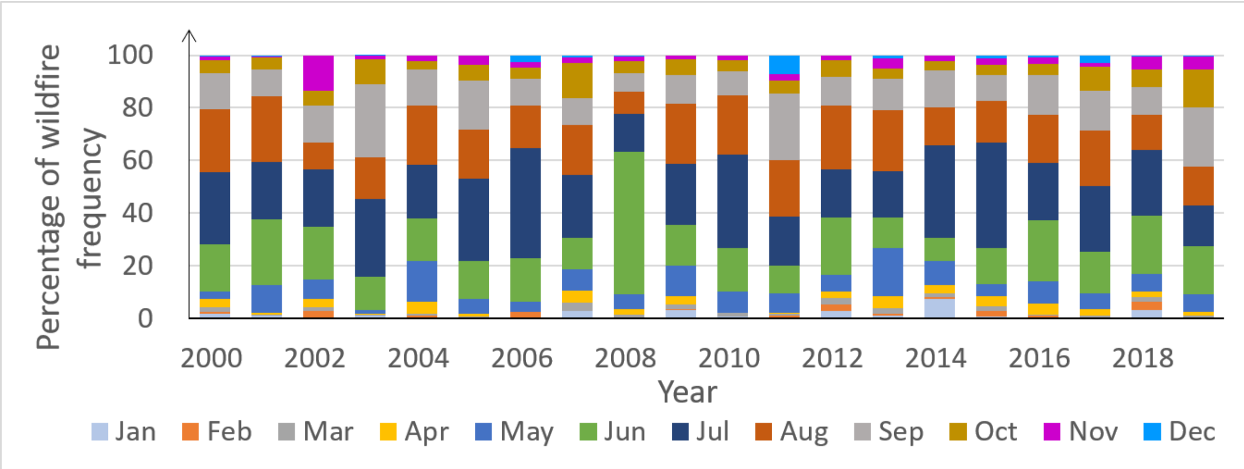
California's Mediterranean climate is characterized by hot and dry summers, which leads to a high wildfire ignition risk (Bryant and Westerling, 2014; Fried et al., 2004). Also, the hot and dry Santa Ana wind accelerated the spread of wildfires each fall (Westerling et al., 2004). The precipitation in California was concentrated in the winter, and the temperature was moderate (Westerling et al., 2003), allowing wildland vegetation to grow fast and storing fuel for next year. These climate factors have maintained a steady trend in the frequency of wildfires in different months over the past two decades, with autumn being the riskiest season for wildfires ignitions, which was consistent with the monthly distribution of wildfire frequency. The temporal distribution of fire frequency and burned area reveals that although they are similar in the general trend, there is no consistency between them. For example, the fire frequency in June and July in 2018 was comparable, but the total burned area in July is about six times that of June. The conclusion that small wildfires contribute more to the total wildfire frequency and large fires contribute more to the total wildfire burned area is valid in the annual and monthly distributions of wildfires. From the perspective of wildfire management, the number of potential wildfires and the probability of extreme fires is relatively high from summer to early autumn in California, which requires closer wildfire monitoring and additional human and material resources investment.

3.2 Spatial distribution of wildfires in CA from 2000 to 2019

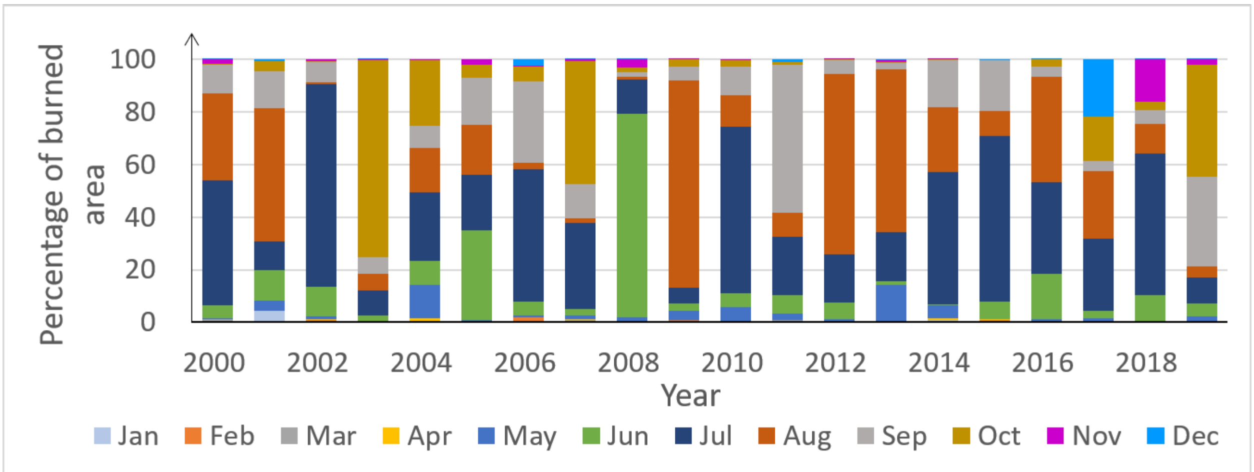
CAL FIRE has 21 operational units throughout the state that are designated to address fire suppression over a certain geographic area and six 'Contract Counties' (Kern, Los Angeles, Marin, Orange, Santa Barbara and Ventura) for fire protection services. The fire frequency and the burned area in each unit from 2000 to 2019 are shown in Figure 5.

It is obvious that the three administrative units in northern California, Lassen-Modoc Unit (LMU), Siskiyou Unit (SKU), and Shasta-Trinity Unit (SHU), had high frequency and burned areas of wildfires. Although wildfires occurred frequently in central and southern California as well, the total burned area was relatively small. This distribution is related to land use in California. In northern California, the evergreen and deciduous forests are the dominant vegetation, the forests are dense and less developed by human, and the population density is relatively low. Wildfires are difficult to be detected early-on in these remote areas and there is enough fuel to keep them burning and spreading. On the other hand shrubs are the dominant vegetation in southern California. Also, most of the area in southern CA have been developed and associated with higher degree of human activity, making it easier for wildfires to be detected and controlled at an early stage. Correspondingly, wildfires in southern CA, though smaller in size, have had a greater social and economic impact on human lives and society.

Due to the complex environmental and terrain conditions in California, the risk of wildfires varies significantly from region to region, and the causes of extreme wildfires are also completely different. In order to provide fire managers



(a)



(b)

Figure 4: Seasonal variation of wildfire frequency and burned area in CA from 2000 to 2019: (a) Percentage of wildfire frequency each month; (b) Percentage of burned area each month.

with more effective fire suppression measures, this study combined human geographic information such as population density in CAL FIRE administration units and land use types in California, and used KDE to analyze hot spot regions of all the wildfires from 2000 to 2019. The resolution of KDE analyses was 3km. The results are shown in Figure 6. This study also indicates the hot spots of wildfires with different causes through KDE analysis. The results are shown in 7 and 8.

A comprehensive analysis of all types of wildfires in Figure 6 shows that the concentrated regions of fires ignition points were near Los Angeles County and along the Sierra Nevada Mountain ranges. The ignition points were most sparse in the east of Riverside Unit (RRU) and San Diego Unit (MVU), the San Bernardino Unit (BDU) in southern California, as well as the western region of the Sierra Nevada Mountains. The land cover in southeastern California is of shrub and wasteland type, and the dominant vegetation type is consistent with that in southwestern California. Yet ignition points are concentrated in Southwestern California and not in Southeastern California. The main differences between these

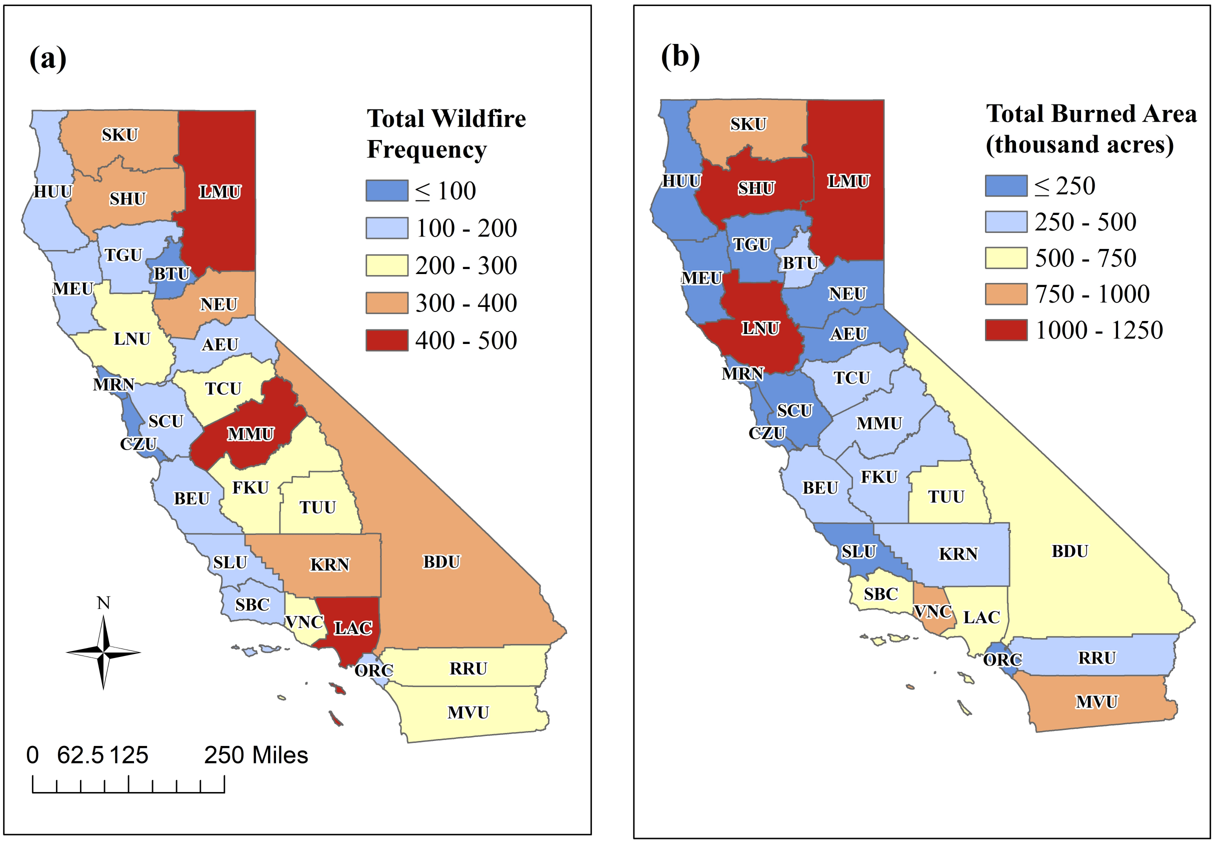


Figure 5: Wildfire frequency and burned area by CAL FIRE operational administration units in CA from 2000 to 2019. (a) total wildfire frequency in each unit in 20 years; (b) total wildfire burned area in each unit in 20 years;

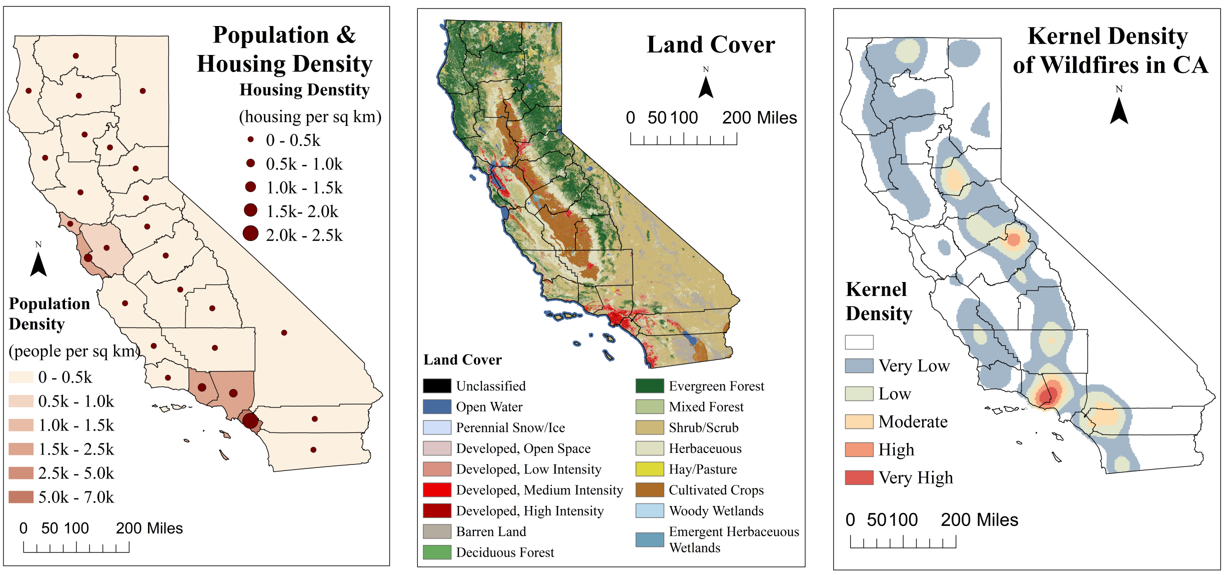


Figure 6: Population, land cover and wildfire records in CA from 2000 to 2019: (a) Population and building density; (b) Land cover; (c) KDE analysis of all the wildfires in CA

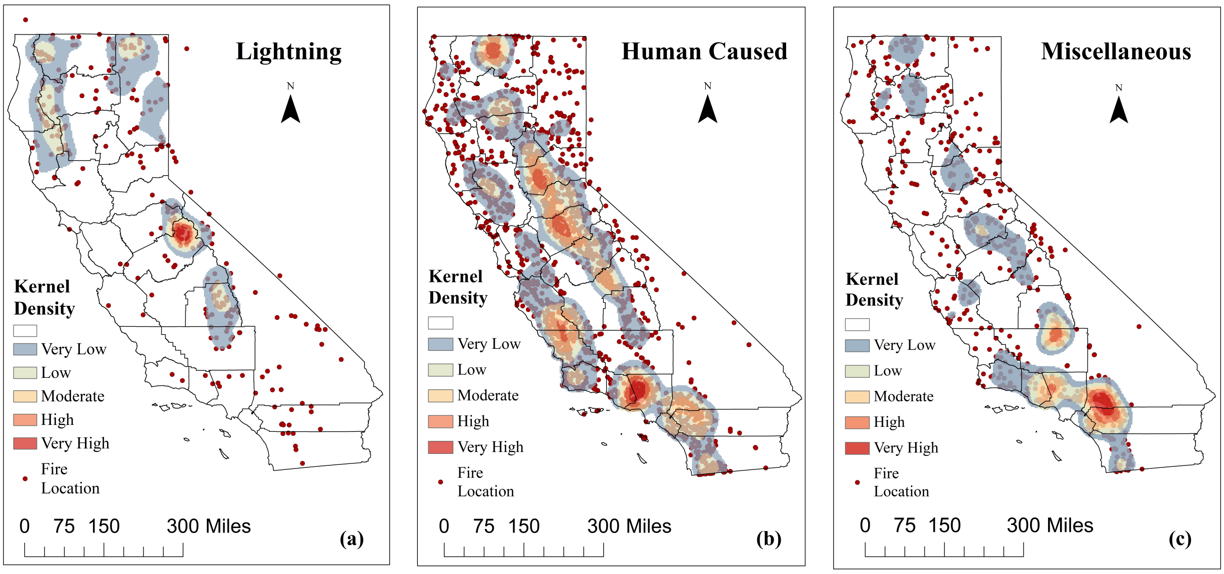


Figure 7: KDE Analysis of wildfires in CA by causes from 2000 to 2019 : (a) Lightning; (b) Human caused; (c) Miscellaneous

two regions are the area of the developed region and population density. This result proves the importance of human activities on the ignition of wildfires. On the other hand, the land cover in the western region of the Sierra Nevada mountain is mainly cultivated crop. Surrounding by herbaceous vegetation, the area of cultivated crop is separated from deciduous forests and evergreen forests on the mountains, which reduce the risk of wildfire spreading to here.

The causes and risks of wildfires vary widely across CA, therefore, Figure 7 separated wildfires in CA by ignition causes: natural (lightning) and human caused. Wildfires with multiple causes were classified as miscellaneous. The statistical results of wildfire ignition causes are shown in Table 3. Lightning-caused wildfires occurred mostly in northern California, with hot spots in the central Sierra Nevada Mountains. Anthropogenic ignition points were distributed from north to south throughout California and were relatively uniform in density. The anthropogenic causes were subdivided by CAL FIRE into 15 types. The spatial distribution of wildfires with different causes are shown in the appendix. In this study, human-caused wildfires were classified into three categories: transportation (railroad, vehicle, aircraft), human activity (equipment use, smoking, campfire, debris, arson, playing with fire, firefighter training, non-firefighter training, escaped prescribed fire, illegal alien campfire) and construction (powerline, structure). The results are depicted in Figure 8. The map shows that these three broad types of human-caused wildfires have a similar general distribution throughout California, with hot spots in the Sierra Nevada Mountains and southern California. It is worth noting that there are a few hot spots in the SKU unit, northernmost of the state, where wildfires have been triggered by human activity, while the population density and area of the developed region were not significantly different from other administrative units in northern California. The main cause of wildfires was equipment use, accounting for 40.88% of the total (717 out of 1,754). The distributions of human-caused wildfires in other units were closely related to the density of the traffic network and the human community.

Causes	Lightning	Human-caused			Miscellaneous	Unknown
		Transportation	Human Activity	Construction		
Frequency	1530	419	1754	302	746	1585
Percentage	24.15	6.61	27.68	4.77	11.77	25.02

Table 3: Statistical results of wildfire ignition causes in CA from 2000 to 2019

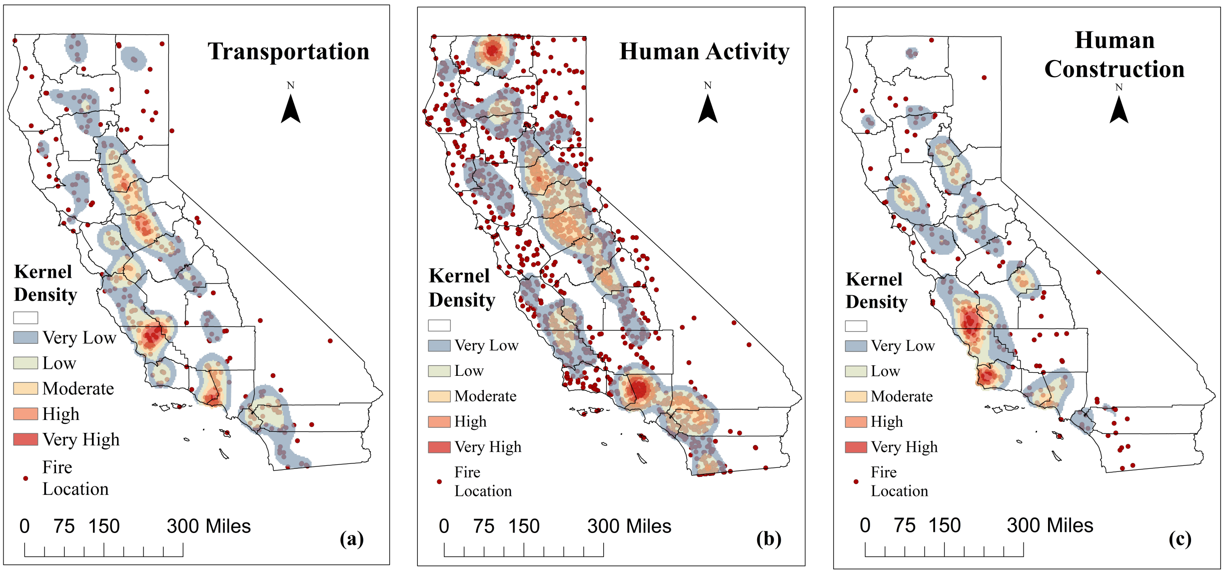


Figure 8: KDE Analysis of human-caused wildfires in CA from 2000 to 2019. (a) Transportation (railroad, vehicle, aircraft); (b) Human Activity (equipment use, smoking, campfire, debris, arson, playing with fire, firefighter training, non-firefighter training, escaped prescribed fire, illegal alien campfire); (c) Human Construction (power line, structure).

3.3 Univariate analysis in California wildfires

As mentioned before, Faivre et al. (2014) analyzed how the 15 explanatory variables affect the occurrence and the frequency of wildfires. Similarly, the relationship between 13 selected variables and the fire occurrence and frequency from 2000 to 2019 recorded in the CAL FIRE database were analyzed by logistic and Poisson regression in this study. The relationships are shown in Table 4, and the Logistic regression and Poisson regression have very similar results. The significance, P-value, indicates whether the variable is correlated with the occurrence of wildfires. The sign of the coefficient indicates the direction of the correlation between variables and the occurrence of wildfires, and the value of the coefficient represents the contribution. Logistic regression was used to fit the occurrence of the fires, while Poisson regression was used to analyze the number of wildfires or the probability of future wildfire occurrence. In general, if the number of samples is large enough, the binomial distribution tends to become a Poisson's distribution. On the other hand, Poisson distribution can fit rare events more easily. Therefore, the table shows the fitting results of the two distributions for mutual comparison and verification.

The results show that all human factors are significantly correlated with the occurrence of wildfire with negative correlation. On a statewide scale, one of the wildfire hot spot - the Sierra Nevada mountain ranges have low population density, while the other hot spot, Great Los Angeles Region, has high population density; therefore the regression analysis cannot accurately describe the contribution of population density. In addition, the probability of wildfire occurrence increases with the distance from major roads and coastlines. When it comes to power lines, even though studies have shown that 'wires in wildland' have become the main cause of destructive human-caused wildfires in recent years (Nauslar et al., 2018), for California as a whole, wildfire risk is greater in places farther from the wires. This is because the effects of variables on wildfires are different at different spatial scales. The area with the highest density of power lines and roads are the human community, but wildlands are where the wildfire start. Although the density of human-constructions and wildfires are consistent in terms of the spatial scale of administrative units, the distances between the ignition points of wildfires and the developed area are evident when the spatial scale is zoomed in to the human community and surrounding wildland. Of the three terrain variables, elevation has the greatest impact on wildfire with an obvious negative correlation, followed by aspect and slope. From the fitting results of vegetation distribution, it is clear that the risk of wildfires is higher in areas dominated by grass, while area with higher cover of trees are less likely to ignite wildfires. As for the last two climate-related variables, there was no significant correlation between temperature and the occurrence of wildfires, while the higher the VPD, which means the dryer the air, the greater the risk of wildfire occurrence. This may be because the impact of temperature on wildfires is indirect, the increase in temperature will increase the VPD and thus increase the risk of wildfires.

Variables	Logistic Regression		Poisson Regression	
	Coef.	P	Coef.	P
Distance to roads (km)	-0.0020	<0.0001	-0.0023	<0.0001
Distance to coast (km)	-0.0341	<0.0001	-0.0398	<0.0001
Distance to power line (km)	-0.0161	<0.0001	-0.0163	<0.0001
Population (persons/km2)	0.0001	0.0018	0.0002	<0.0001
Housing Density (houses/km2)	-0.0025	<0.0001	-0.0025	<0.0001
Elevation (km)	-0.4965	<0.0001	-0.4462	<0.0001
Slope (°)	0.0057	<0.0001	0.0056	<0.0001
Aspect (°)	0.0164	<0.0001	0.0178	<0.0001
Tree (%)	-0.0113	<0.0001	-0.0126	<0.0001
Shrub (%)	0.0005	<0.0001	-0.0014	<0.0001
Grass (%)	0.0205	<0.0001	0.0206	<0.0001
Max Temperature	0.1833	<0.1830	0.1715	<0.1870
Max vapor pressure deficit	0.0512	<0.0001	0.0504	<0.0001

Table 4: Logistic and Poisson regression results for all the explanatory variables in California wildfires (2000-2019)

The variation of each variable over the 20 years was also calculated to assess the variation trend of each variable and its potential impact on the occurrence of wildfires. The topography did not change significantly, while all other variables increased during the 20 years. The results in Table 5 shows that the number of power lines and buildings in California increased rapidly, and the average maximum temperature in California in 2019 has increased by nearly 1°C compared with that in 2000. Combined with the results above, it can be inferred that variables related to the development of human society, such as housing or population density, did not have a direct correlation with the occurrence of wildfires, even though they changed significantly over the 20-year period. However, due to the increase of exposure, the destructive intensification of wildfires to the adjacent residential communities is predictable. Similarly, the increase in maximum temperature does not directly trigger a wildfire, but it does affect the behavior of the fires, such as the intensity and duration, making it more destructive.

Variables	Variations
Major roads density	0.114
Powerline density	0.193
Housing density	0.153
Population density	0.157
Shrub	0.013
Grass	0.070
Max Temperature	0.216
Max vapor pressure deficit	0.182

Table 5: Variation of wildfire-related variables in California from 2000 to 2019

4 Conclusion

This study investigated the temporal distribution of the frequency and burned area of large and small wildfires, the spatial distribution of wildfires with different causes in California from 2000 to 2019, as well as the factors influencing the occurrence of wildfires. In terms of the temporal distribution, the total number of wildfires in California varied little from year to year. However, the total burned area peaked every few years and then decreased significantly over the next two to three years. The distribution and contribution of wildfires frequency and burned area conformed to the Pareto distribution. Regarding the seasonal variation of California wildfires, summer to early autumn has been the most frequent time for wildfire occurrence and the most likely time for extreme wildfires. In the spatial distribution, the frequency of wildfires was not consistent with the area of wildfires. Although the total burned area was relatively small in southern California, the frequency of wildfires was high. Therefore, with the high population and housing density, the wildfires in southern California were still destructive. In the aspect of ignition sources, the primary cause of natural wildfires in California was lightning, which was concentrated in the central Sierra Nevada mountain and the far north of California. Human-caused wildfires were widespread in California, with the hot spots being the Sierra Nevada Mountain range and the developed and densely populated areas along the west coast. Surprisingly, among the variables that affect the occurrence of wildfires, the risk of wildfires is significantly negatively correlated with the distance to the construction sites, but their overall density distribution is consistent. These differences are mainly caused by the different spatial scales selected in the analysis. Terrain elevation, fuel type, and coverage, and VPD were significantly correlated with the risk of wildfire occurrence, while the effect of temperature was not significant. This may be due to the fact that the impact of temperature on the wildfires is indirect. Therefore, to inform targeted wildfire management strategies, geographical analysis in smaller spatial scale and units are necessary. The case studies of representative extreme wildfire events are also necessary to improve the understanding of fire risks.

Acknowledgements

Li and Banerjee acknowledge the funding support from the University of California Laboratory Fees Research Program funded by the UC Office of the President (UCOP), grant ID LFR-20-653572. Banerjee acknowledges the new faculty start up grant provided by the Department of Civil and Environmental Engineering, and the Henry Samueli School of Engineering, University of California, Irvine.

References

B. P. Bryant and A. L. Westerling. Scenarios for future wildfire risk in california: links between changing demography, land use, climate, and wildfire. *Environmetrics*, 25(6):454–471, 2014.

J. W. Duffield, C. J. Neher, D. A. Patterson, and A. M. Deskins. Effects of wildfire on national park visitation and the regional economy: A natural experiment in the northern rockies. *International journal of wildland fire*, 22(8): 1155–1166, 2013.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

B. Evarts. Fire loss in the united states during 2018. *NFPA National Fire Protectio n Association, Quincy*, 2019.

N. Faivre, Y. Jin, M. L. Goulden, and J. T. Randerson. Controls on the spatial pattern of wildfire ignitions in southern california. *International Journal of Wildland Fire*, 23(6):799–811, 2014.

J. S. Fried, M. S. Torn, and E. Mills. The impact of climate change on wildfire severity: a regional forecast for northern california. *Climatic change*, 64(1-2):169–191, 2004.

T. P. Holmes, R. J. Huggett, and A. L. Westerling. Statistical analysis of large wildfires. In *The Economics of Forest Disturbances*, pages 59–77. Springer, 2008.

S. Jazebi, F. De Leon, and A. Nelson. Review of wildfire management techniques—part ii: Urgent call for investment in research and development of preventative solutions. *IEEE Transactions on Power Delivery*, 35(1):440–450, 2019.

S. Jin, C. Homer, L. Yang, P. Danielson, J. Dewitz, C. Li, Z. Zhu, G. Xian, and D. Howard. Overall methodology design for the united states national land cover database 2016 products. *Remote Sensing*, 11(24):2971, 2019.

J. E. Keeley and A. D. Syphard. Historical patterns of wildfire ignition sources in california ecosystems. *International journal of wildland fire*, 27(12):781–799, 2018.

S. Kotz and S. Nadarajah. *Extreme value distributions: theory and applications*. World Scientific, 2000.

J. Martínez, C. Vega-Garcia, and E. Chuvieco. Human-caused wildfire risk rating for prevention planning in spain. *Journal of environmental management*, 90(2):1241–1252, 2009.

A. Miranda, J. Carrasco, M. González, C. Pais, A. Lara, A. Altamirano, A. Weintraub, and A. D. Syphard. Evidence-based mapping of the wildland-urban interface to better identify human communities threatened by wildfires. *Environmental Research Letters*, 2020.

B. C. Nauman. *Variability in California's Fire Activity during the Holocene, across Space and Time*. PhD thesis, UCLA, 2020.

N. J. Nauslar, J. T. Abatzoglou, and P. T. Marsh. The 2017 north bay and southern california fires: a case study. *Fire*, 1 (1):18, 2018.

V. C. Radeloff, 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, et al. Rapid growth of the us wildland-urban interface raises wildfire risk. *Proceedings of the National Academy of Sciences*, 115(13):3314–3319, 2018.

M. Rodrigues, J. de la Riva, and S. Fotheringham. Modeling the spatial variation of the explanatory factors of human-caused wildfires in spain using geographically weighted logistic regression. *Applied Geography*, 48:52–63, 2014.

M. Rodrigues, A. Jiménez, and J. de la Riva. Analysis of recent spatial–temporal evolution of human driving factors of wildfires in spain. *Natural Hazards*, 84(3):2049–2070, 2016.

J. Ruffault and F. Mouillot. Contribution of human and biophysical factors to the spatial distribution of forest fire ignitions and large wildfires in a french mediterranean region. *International journal of wildland fire*, 26(6):498–508, 2017.

K. C. Short. Spatial wildfire occurrence data for the united states, 1992-2015 [fpa_fod_20170508]. 2017.

A. Shvidenko, D. Shchepashchenko, E. Vaganov, A. Sukhinin, S. S. Maksyutov, I. McCallum, and I. Lakyda. Impact of wildfire in russia between 1998–2010 on ecosystems and the global carbon budget. In *Doklady Earth Sciences*, volume 441, pages 1678–1682. Springer, 2011.

B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

USFS. Survey finds 18 million trees died in california in 2018. Available at `https://www.fs.usda.gov/detail/catreemortality/toolkit/?cid=FSEPRD609121`, 2019. Accessed: 2020-08-26.

P. A. Werth, B. E. Potter, M. E. Alexander, C. B. Clements, M. G. Cruz, M. A. Finney, J. M. Forthofer, S. L. Goodrick, C. Hoffman, W. M. Jolly, et al. Synthesis of knowledge of extreme fire behavior: volume 2 for fire behavior specialists, researchers, and meteorologists. *Gen. Tech. Rep. PNW-GTR-891. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 258 p.*, 891, 2016.

A. L. Westerling, A. Gershunov, T. J. Brown, D. R. Cayan, and M. D. Dettinger. Climate and wildfire in the western united states. *Bulletin of the American Meteorological Society*, 84(5):595–604, 2003.

A. L. Westerling, D. R. Cayan, T. J. Brown, B. L. Hall, and L. G. Riddle. Climate, santa ana winds and autumn wildfires in southern california. *Eos, Transactions American Geophysical Union*, 85(31):289–296, 2004.

A. P. Williams, J. T. Abatzoglou, A. Gershunov, J. Guzman-Morales, D. A. Bishop, J. K. Balch, and D. P. Lettenmaier. Observed impacts of anthropogenic climate change on wildfire in california. *Earth's Future*, 7(8):892–910, 2019.

A Appendix

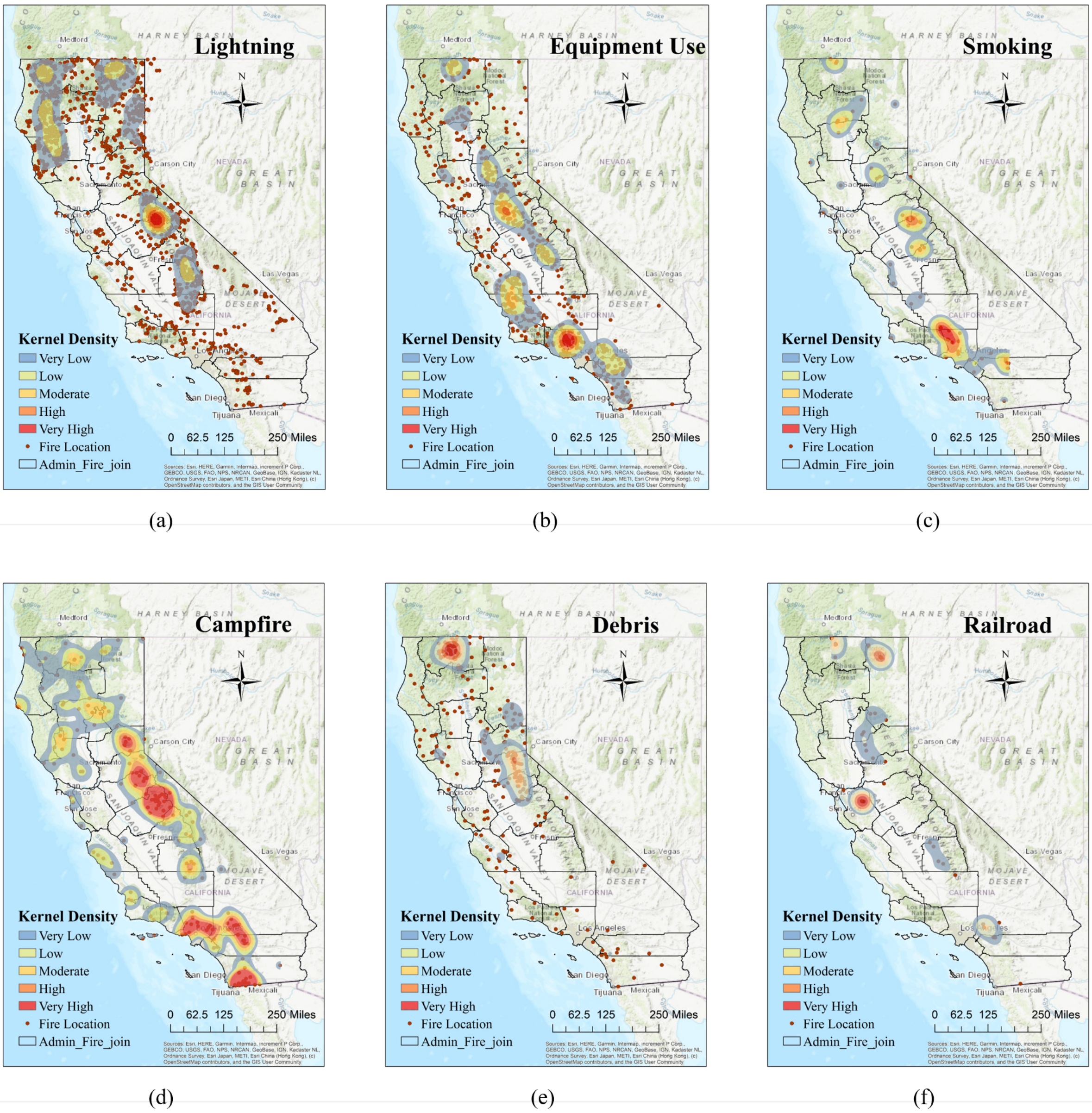
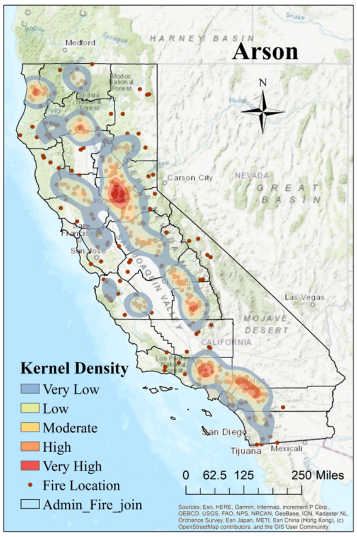
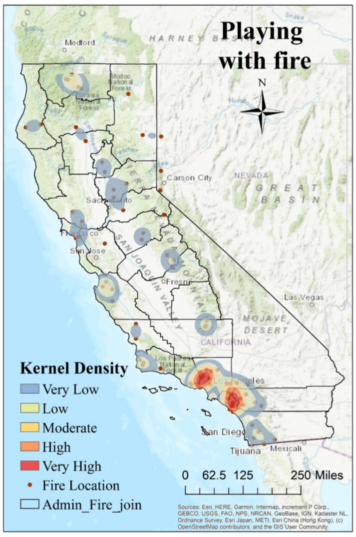


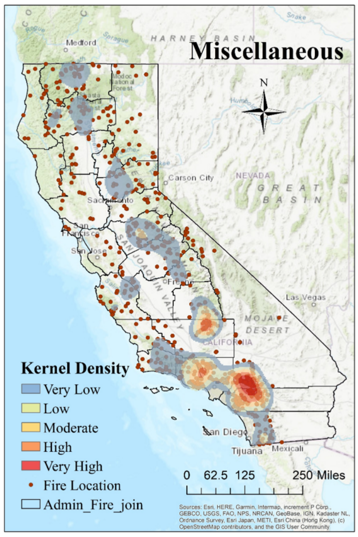
Figure A1: KDE results for wildfires with different causes



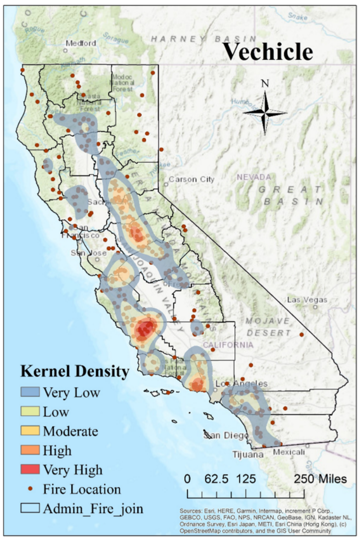
(g)



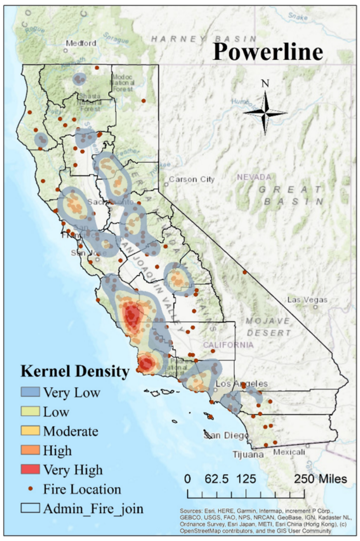
(h)



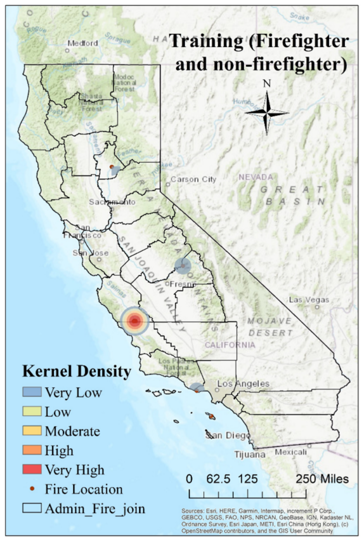
(i)



(j)

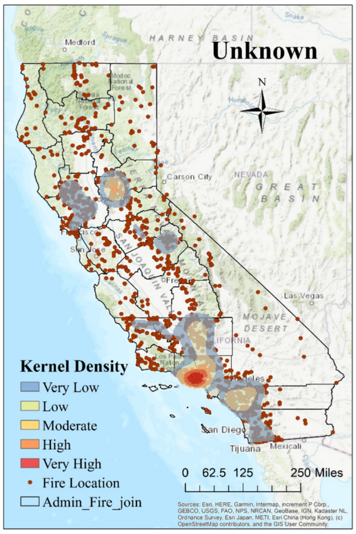


(k)

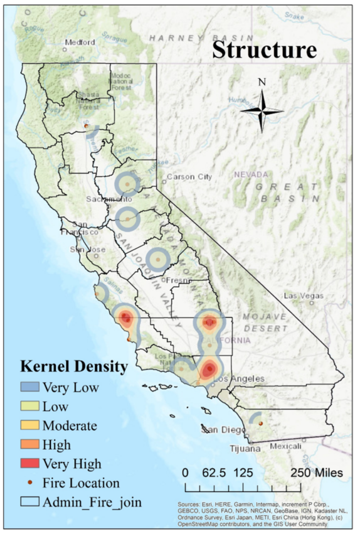


(l)

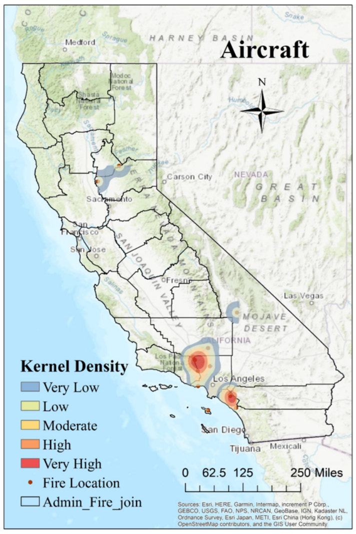
Figure A1: KDE results for wildfires with different causes (cont.)



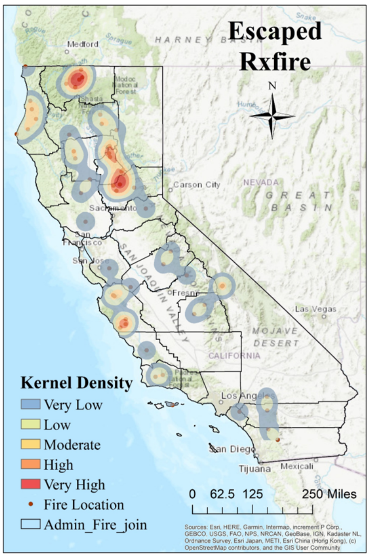
(m)



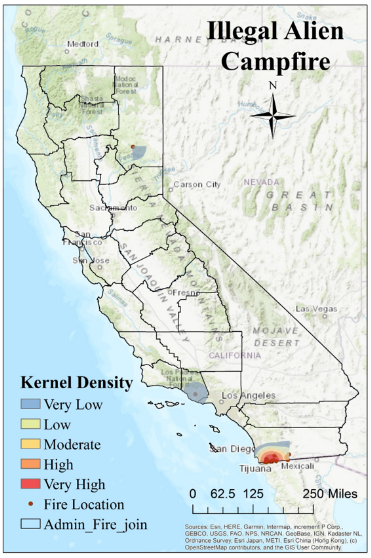
(n)



(o)



(p)



(q)

Figure A1: KDE results for wildfires with different causes (cont.)