Estimation of Wildfire Hazard using Spatial-Temporal Fire History Data

Roger Peng  Frederic Schoenberg*

Abstract

The study of wildfire has long been important to the residents of Los Angeles County, California. Of particular interest is the risk of a location burning given the time since the location’s last burn. Previous authors have investigated the impact of fuel age on the susceptibility of different areas to burning using techniques which are not optimal in the case where detailed spatial and temporal information on fire history is available. Here, a new method is proposed for estimating wildfire hazard using spatial-temporal fire history data. We use this estimator to show that fuel age appears to have a nonlinear threshold-type relationship with fire hazard. Our estimator is shown to be more stable than previous estimators and to have good large sample properties.

Keywords: Hazard estimation; Spatial-temporal modeling; Time-since-fire; Fire risk; Threshold relationship.

1 Introduction

The history of wildfire in Los Angeles County and other parts of Southern California is well documented (Hanes, 1971; Minnich, 1983; Pyne et al., 1996). Fires are responsible for significant amounts of property damage in Los Angeles County and are a subject of research for fire managers, fire scientists, and ecologists. Because of the enormous destructive capabilities of fires there is significant interest in developing methods for estimating hazard. Here we think of hazard (or danger) as the proportion of fuel expected to burn in a given area over the course of a year. Ultimately, we would like to predict the risk of a location burning in a particular year.

There are a number of possible variables that can be used to estimate fire danger. Researchers have examined the relationships between area burned and different meteorological variables such as wind, precipitation, fuel moisture, and temperature (Flannigan and Harrington, 1988; Renkin and Despain, 1992; Viegas and Viegas, 1994). Another important factor is the fuel age, also called time-since-fire, which is defined for each location as the time since that location last burned. Implicit in the definition of fuel age is the notion that once an area has burned, essentially all of the available fuel in that area is expended.

There is some disagreement over the influence of fuel age on fire hazard. Minnich (1983) suggested that the largest fires in recent Southern California history are linked to the increased availability of older fuels. Using Landsat imagery from Southern California and Northern Baja California, Minnich claimed that the policy of total fire suppression created extensive stands of very old age classes. These older stands, he argued, had accumulated fuels over the years and were therefore ripe for burning. Minnich also claimed that fuel age is the most important variable affecting the spatial properties of fires, in that fires tend to burn up to the boundary of another (recent) fire and then stop for lack of fuel. This reasoning was used as support for prescribed burning policies.

Minnich’s paper contradicted some previous thinking concerning fuel age and fire hazard. Van Wagner (1978) suggested that fuel age has little or no effect on hazard and that it may be reasonable to assume uniform flammability of forest stands with age. Van Wagner also noted that since large fires typically burn through stands of many different ages, fuel age is irrelevant when looking at the larger and more destructive fires.

More recently, Keeley et al. (1999) provided evidence showing that the mean fire size in Southern California has not increased over time and that large fires are not necessarily dependent on old age classes of fuels. They went further to suggest that age class manipulation (i.e. prescribed burning) is unlikely to prevent catastrophic fires. The authors examined some of the largest fires in the California Statewide Fire History Database and showed that for those fires there was no apparent relationship between the proportions of fuels burned and the age classes of the fuels (e.g. see their Figure 4).

The lack of full agreement over the precise role of fuel age in determining fire hazard is not surprising. In general, it is difficult to make quantitative statements about fuel age. For decades researchers have been using time-since-fire maps, which show the time of the most recent fire for every location in the study area (Johnson and Gutsell, 1994). But these maps alone are very difficult to interpret in terms of hazard. Johnson and Gutsell (1994) described a useful method for summarizing and quantifying the information stored in these maps, and for producing numerical estimates of hazard. They used survivor curves to estimate the probability of an area surviving without fire beyond a certain age, given that it has already survived to the current day.
Often, time-since-fire maps are the only information available to researchers. However, when more detailed information on the fire history of an area is available, one can expect to obtain more accurate estimates of hazard. In this situation the survivor curve method is not optimal. First, the time-since-fire maps only show the most recent fires and do not contain information on the pattern of overburning that occurs over time. This makes the estimates more dependent on recent observations. Second, the survivor curves, which depend critically on the year of observation, tend to be statistically unstable. Finally, the parametric models suggested by Johnson and Gutsell (1994) place some restrictions on the nature of the relationship between fuel age and hazard.

In this paper we develop a new technique which quantifies concisely the relationship between hazard and fuel age. Our method is nonparametric, stable, and takes advantage of all of the available spatial-temporal information on fire history. The technique therefore addresses some of the problems associated with existing methods. We can discern that for Los Angeles County over the past century, the hazard of burning gradually increases up to fuel ages of 30 to 40 years and remains nearly constant thereafter.

In Section 2 we will describe the data used for the current analysis. Section 3 outlines our method for assessing the relationship between fire hazard and fuel age and Section 4 discusses its statistical properties. Section 5 shows the results of applying our method to data on fire history in Los Angeles County and shows that our estimator tends to be more stable than the estimator described by Johnson and Gutsell (1994). Section 6 outlines some subjects for future research.

2 Fire Data

The data on Los Angeles County fires were obtained from the Los Angeles County Department of Public Works (DPW) and the Los Angeles County Fire Department. Fire information was recorded with the geographic information systems software package ArcInfo and stored in coverage files. In those files each fire is stored as a polygon outlining the fire boundary and the date on which the fire originated.

The data from DPW consist of approximately 2000 fires occurring between the years 1878 and 1996. Figure 1 shows all of the fires in the dataset. Each point represents the centroid of its respective fire boundary. Figure 2 shows the fire boundaries for 1980, a typical year in the dataset. Fire department officials estimate that the polygon boundaries are accurate to within about 16 meters. For the purpose of hazard estimation we regard the errors in the polygon boundaries as negligible. Figure 3 shows the total area burned in each year of the dataset. Although there appears to be an increase in total area burned over the years, that is at least partly due to underreporting in the earlier years. For example, in the early part of the century it was common to record only the occurrence of very large fires. Fire department officials believe that the data are complete and fairly accurate from 1950 on. A few years in the dataset are simply missing, a fact that will be ignored in the current application. The possible effects of missing data on hazard estimation will be discussed in Section 5.

Figure 4 shows how fuel age may affect the spatial configuration of fires. For the fires in 1963 and 1964, a one year interval, the 1964 fires burn up until the border of the 1963 fire and stop (Figure 4a). However, when looking at the years 1928 and 1968, a 40 year interval, we see that the 1968 fire burns right over the 1928 fire (Figure 4b). It should be clarified that in the intervening years between 1928 and 1968, there was almost no overburning of the 1928 fires in panel (b) — the 1968 fire is the first instance of significant overburning at this location.

3 Methodology

Given a fuel age $u$, a quantity of interest is $h(u)$, defined as the proportion of $u$-year-old fuel expected to burn each year. Our estimator of $h(u)$, given $n$ years of fire history in our study area, is constructed as follows. For each year $i$, where $i = 1, \ldots, n$, we calculate $Y_i(u)$, the amount of $u$-year-old fuel that was available in year $i$, and $p_i(u)$, the proportion of $u$-year-old fuel that burned in year $i$. The estimate of $h(u)$ that we use is

$$
\hat{h}(u) = \frac{\sum_{i=1}^{n} p_i(u) Y_i(u)}{\sum_{i=1}^{n} Y_i(u)}.
$$

After computing this proportion for many different fuel ages $u$, we use a local linear smoother to construct a time-since-fire curve and highlight the overall trend. In order to assess the variability of the time-since-fire curve without assuming a parametric model, we constructed 95% confidence bands for the curve using the bootstrap. For each fuel age $u$, we sample with replacement the pairs $\{p_1(u), Y_1(u)\}, \ldots, \{p_n(u), Y_n(u)\}$ to get $\{p_1^*(u), Y_1^*(u)\}, \ldots, \{p_n^*(u), Y_n^*(u)\}$. From the bootstrap samples we compute $\hat{h}^*(u) = (\sum p_i^*(u) Y_i^*(u)) / (\sum Y_i^*(u))$. After computing $\hat{h}^*(u)$ for all fuel ages $u$ we refit the local linear smoother. This procedure is then repeated 1000 times and confidence bounds are constructed using the percentile method (Efron and Tibshirani, 1993).

Computing $p_i(u)$ and $Y_i(u)$ involves intersecting and differenting the fire boundaries, which are very large polygons, each with many vertices. Hence, the computational cost can be substantial, but it is by no means prohibitive. We used the R statistical computing environment (Ihaka and Gentleman, 1996) to write most of the software needed to construct the estimator. For the polygon manipulations we used the very fast General Polygon Clipper (GPC) software library written by Alan Murta (see http://www.cs.man.ac.uk/~amurta/software).

4 Statistical Properties of $\hat{h}(u)$

The estimator $\hat{h}(u)$ may be seen as an estimator of the true hazard $h(u)$, defined via $h(u) := E[p(u)|Y(u)]$. In order to
derive the main statistical properties of the estimator \( \hat{h}(u) \), we make the following assumptions. First, fix \( u \). Given a sample \( p_1, \ldots, p_n \) and \( Y_1, \ldots, Y_n \), we assume

1. \( Y_i \geq 1 \) for all \( i \).
2. \( \text{Var}(p_i|Y_i) = \sigma^2/Y_i \), for some unknown constant \( \sigma^2 \).
3. \( \mathbb{E}[p_k^i|Y_1, \ldots, Y_n] = \mathbb{E}[p_k^i|Y_i] < \infty \), for \( k = 1, 2 \).
4. \( \text{Cov}(p_i, p_j|Y_1, \ldots, Y_n) = 0 \), for all \( i \neq j \).

Assumption 1 can be interpreted in the following manner: if there is some minimal observable over-burn size \( B > 0 \) and \( Y_i \) is measured in units of \( B \), then \( Y_i \) is always greater than or equal to 1. This assumption is necessary in order to prevent the variances of the \( p_i \)'s (and \( \hat{h} \) itself) from being potentially undefined. However, the assumption is not too restrictive because one can typically change the units of measurement so that it holds.

The intuition behind Assumption 2 is that if the entire study area were divided into small 1-unit pieces, and each unit burned independently of the others, then \( Y_i \) represents the number of pieces that are available and \( p_i \) represents the proportion that burn. Therefore, \( \text{Var}(p_i|Y_i) \propto 1/Y_i \). While Assumption 2 is generally difficult to verify in practice, we believe it may provide a reasonable approximation for the variance behavior of the \( p_i \)'s. For example, if \( Y_i \) is very small, then it is more likely that either we will observe total reburning or no reburning. Therefore, \( p_i \) is 1 or 0 and \( \text{Var}(p_i|Y_i) \) is high. If \( Y_i \) is large, then the distribution of \( p_i \) is spread more uniformly between 0 and 1 and will have lower variance.

Assumption 3 states that given the amount of \( u \)-year-old fuel that is available in a year, the proportion expected to burn does not depend on the availability in other years. Finally, while we do not want to restrict the proportions burned to be independent from year to year, we do assume that they are uncorrelated (Assumption 4).

**Proposition 1.** The estimator \( \hat{h} \) is an unbiased estimate of \( h \).

**Proof.** The unbiasedness can be shown through direct calculation:

\[
\mathbb{E} \left[ \hat{h} \right] = \mathbb{E} \left[ \frac{\sum p_i Y_i}{\sum Y_i} \right] = \mathbb{E} \left\{ \mathbb{E} \left[ \frac{\sum p_i Y_i}{\sum Y_i} \bigg| Y_1, \ldots, Y_n \right] \right\} = \mathbb{E} \left\{ \frac{1}{\sum Y_i} \sum Y_i \mathbb{E}[p_i|Y_1, \ldots, Y_n] \right\} = h.
\]

**Proposition 2.** \( \hat{h} \) is a consistent estimator of \( h \).
Figure 2: Fire boundaries for the year 1980; (a) all fires and (b) closeup of some fires.

**Proof.** The variance of $\hat{h}$ can be expressed as follows:

$$
\text{Var} \left( \hat{h} \right) = \mathbb{E} \left[ \text{Var} \left( \frac{\sum p_i Y_i}{\sum Y_i} \Big| Y_1, \ldots, Y_n \right) \right] \\
= \mathbb{E} \left[ \frac{1}{(\sum Y_i)^2} \sum \text{Var} \left( p_i Y_i \Big| Y_1, \ldots, Y_n \right) \right] \\
= \mathbb{E} \left[ \frac{1}{(\sum Y_i)^2} \sum Y_i^2 \text{Var} \left( p_i Y_i \Big| Y_1 \right) \right] \\
= \mathbb{E} \left[ \frac{1}{(\sum Y_i)^2} \sum Y_i^2 \sigma^2 \right] \\
= \frac{\sigma^2}{n} \mathbb{E} \left[ \frac{1}{\bar{Y}} \right]
$$

where $\bar{Y} = (1/n) \sum Y_i$. By Assumption 1, the $Y_i$'s are always $\geq 1$ so $\mathbb{E} \left[ 1/\bar{Y} \right] \leq 1$ for any $n$. Therefore, since $\hat{h}$ is unbiased and has variance of order $1/n$, $\hat{h}$ is consistent.

Equation (1) shows that $\hat{h}$ is simply an average of the observed proportions of overburning weighted by the amount of available fuel. Years for which there was more fuel available get more weight. This scheme makes some intuitive sense — if in a given year, 1% of 100,000 hectares burns, this year should influence our estimation of hazard more than an observation of a year where 1% of 10 hectares burns.

There are of course other possible estimators of $h$. For example, we could choose to use the estimator $\bar{h} = (1/n) \sum p_i$, which gives equal weights to all years. Then $\bar{h}$ is also unbiased for $h$ (and even consistent), but it is possible to show that under Assumptions 1–4, $\bar{h}$ has a larger variance than $\hat{h}$.

**Proposition 3.** Under the same assumptions as Propositions 1 and 2, $\text{Var} \left( \hat{h} \right) \leq \text{Var} \left( \bar{h} \right)$.

**Proof.** A simple calculation yields

$$
\text{Var} \left( \bar{h} \right) = \mathbb{E} \left[ \frac{1}{n} \sum p_i \right] \\
= \frac{1}{n^2} \mathbb{E} \left[ \sum \text{Var} \left( p_i \Big| Y_i \right) \right] \\
= \frac{1}{n^2} \mathbb{E} \left[ \sum \sigma^2 Y_i \right] \\
= \frac{\sigma^2}{n} \mathbb{E} \left[ \frac{1}{n} \sum Y_i \right]
$$

We only need to show that $\mathbb{E} \left[ 1/\bar{Y} \right] \leq \mathbb{E} \left[ \frac{1}{n} \sum \frac{1}{Y_i} \right]$. Now $1/\bar{Y}$ is simply the harmonic mean of $1/Y_1, \ldots, 1/Y_n$. Since $Y_i \geq 1$ for all $i$ and the harmonic mean of a set of positive numbers is always less than or equal to the arithmetic mean, we have $\mathbb{E} \left[ 1/\bar{Y} \right] \leq \mathbb{E} \left[ \frac{1}{n} \sum \frac{1}{Y_i} \right]$ and hence $\text{Var} \left( \hat{h} \right) \leq \text{Var} \left( \bar{h} \right)$.

In this section we have shown that under the given assumptions, the estimator $\hat{h}(u)$ possesses good finite and large sample properties. Furthermore, we have shown that although there exist other unbiased and consistent estimators for $h$, they do not all perform with as much efficiency as $\hat{h}$.

5 Discussion

The result of applying our method to the Los Angeles County DPW data is shown in Figure 5. The time-since-fire curve serves as a concise quantification of the dependence of fire hazard on fuel age. Using data from the past century, it is apparent that as fuel age increases from 1 year to 30 years, the proportion of fuel that burns steadily increases. However, for fuel ages greater than 30 to 40 years, the hazard is nearly
constant. Thus, the relationship between fire hazard and fuel age appears to be nonlinear. There is considerable scatter around the estimated hazard curve. However, the statistical significance of this overall increase and leveling off of the estimated curve is confirmed by the bootstrap 95% confidence bands for the curve.

The shape of the time-since-fire curve agrees with our general knowledge of the vegetation in Los Angeles County. Typically, chaparral, the dominant vegetation, does not burn easily until it has reached 20 to 30 years of age, while older chaparral will burn readily (Pyne et al., 1996). However, the time-since-fire curve indicates that after an area reaches a certain age, it does not necessarily become more flammable or hazardous. This type of nonlinear threshold relationship is distinctly different from the relationships proposed previously. One possible interpretation is that large wildfires occur when conditions are ripe, i.e. when fuel age is at least 30 to 40 years, but that there is little distinction, with regard to risk, between conditions that are sufficient and conditions that are extreme.

5.1 Stability

In order to compare the stability of \( \hat{h}(u) \) to the estimator of Johnson and Gutsell (1994) we used the following empirical method. First, we remove the most recent \( K \) years of our dataset. We chose \( K = 15 \) in order to achieve a balance between having enough years set aside to assess variability and having enough remaining years so that all of the estimates have comparable stability. Using the remaining years in the dataset we recompute the hazard estimator for each \( u \), calling it \( h_1(u) \). Then we add a year (the earliest of those removed) to the dataset and compute \( h_2(u) \), again for each \( u \). We continue this process until we have restored the entire dataset, obtaining the values \( h_1(u), \ldots, h_K(u) \). Then, for each \( u \), our estimate of the variation is simply the sample standard deviation of the set \( \{ h_i(u); i = 1, \ldots, K \} \). We believe this method provides a reasonable summary of the year-to-year variation of \( \hat{h}(u) \). The same method of evaluation was then applied to the survivor curve method of Johnson and Gutsell (1994). The results are shown in Figure 6.

The estimator \( \hat{h}(u) \) appears to exhibit significantly less variation than the survivor curve estimate. The reason behind this is simple: \( \hat{h}(u) \) uses all of the data up to the current year of observation (year \( n \)). When data from year \( n + 1 \) is added, its effect on \( \hat{h}(u) \) is counterbalanced by all of the reburn intervals recorded from years 1 to \( n \). If \( \hat{h}_n(u) \) is the current estimate of \( h(u) \), then given data from year \( n + 1 \), the updated estimate for each \( u \) is

\[
\hat{h}_{n+1}(u) = \hat{h}_n(u) + \frac{Y_{n+1}(u)}{\sum_{j=1}^{n+1} Y_j(u)} \left[ p_{n+1}(u) - \hat{h}_n(u) \right].
\]

Hence, the estimate moves from its old value \( \hat{h}_n(u) \) toward the new observation \( p_{n+1}(u) \), but only by the fraction \( Y_{n+1}(u) / \sum_{j=1}^{n+1} Y_j(u) \). By contrast, the survivor curve method relies only on the most recent burn in each location and hence is more heavily dependent on the most recent observations.

5.2 Missing Data

Perhaps the simplest case of missing data is when the missing fires do not burn over previous fires and are subsequently not burned over. Then the missing fires contribute only to the denominators of our proportions. That is, they represent available area that never gets reburned. If the missing fires were large, the effect of their deletion would be to raise the hazard estimates for all fuel ages. Other than this simple case, it is difficult to say how missing data affect the estimation of \( \hat{h}(u) \), since we know nothing about the spatial configuration of the missing fires.
Figure 4: Overburning for different fuel ages. (a) Fires from the years 1963 (gray) and 1964 (white). Notice the relatively small amount of overburning (black) — approximately 50 hectares. (b) Fires from 1928 (gray) and 1968 (white). The overburning here is much more extensive — 237 hectares.

If the missing years may be assumed to be missing completely at random, one possible method for examining the effect of missing data on $\hat{h}(u)$ could be to use a cross-validation type of procedure. First, select a subset of the data which is believed to be complete and index the years by $j = 1, \ldots, C$. Then denote $\hat{h}_{(j)}(u)$ as the estimate of $h(u)$ computed with the $j$th year removed from this new dataset. For each $u$, a simple measure of the increased variability due to missing years of data could be

$$\gamma_1(u) = \frac{1}{C-1} \sum_{j=1}^{C} \left[ \hat{h}_{(j)}(u) - \frac{1}{C} \sum_{k=1}^{C} \hat{h}_{(k)}(u) \right]^2,$$

or perhaps

$$\gamma_2(u) = \frac{1}{C-1} \sum_{j=1}^{C} \left[ \hat{h}_{(j)}(u) - \hat{h}_c(u) \right]^2,$$

where $\hat{h}_c(u)$ is the estimate of $h(u)$ using all the years $1, \ldots, C$. For the Los Angeles County DPW dataset the assumption that years are missing at random is doubtful, especially since most of the missing years are from the early 1900’s. In general, the availability of detailed information on fires continues to be a major problem and the subject of missing data should be studied more intensely in the future.

6 Conclusions

In this paper we have presented a new technique for estimating wildfire hazard using detailed spatial-temporal fire history data. Using our estimator, we have shown that wildfire hazard and fuel age appear to have a nonlinear threshold-type relationship. The hazard increases steadily for fuels less than 30 to 40 years old, but remains nearly constant thereafter. The estimator presented here is flexible and nonparametric and can capture more complex relationships than previous methods. In addition, our estimator possesses good statistical properties such as stability and consistency.

There are several limitations to the interpretability of our results. First, there are numerous other variables that may interact in important ways with fuel age. Such variables include land use policies, population density, and fire prevention policies, as well as meteorological variables such as wind, temperature, precipitation, and fuel moisture. For example, the expansion of the urban-wildland interface has introduced a major proliferation of fires in previously uninhabited areas. Also, wind is a major factor affecting the size of wildfires. Large catastrophic fires are often driven by high winds and are generally immune to fire suppression (Keeley et al., 1999). The examination of the interactions between these variables and their effect on burn area is an important direction for future research.

Nevertheless, an analysis of fire hazard based solely on antecedent fire can provide a good first-order summary of the data. We hope that such an analysis will provide a quantitative standard against which more complex models can be compared (e.g. see Jackson and Kagan, 1999). While Los Angeles County represents a significant fire regime, an important direction for future research is to investigate the application of our method to wildfire data from other regions. In particular, differences in vegetation life cycles and spatial configurations of fuels may alter considerably the observed relationship between burn area and fuel age.
Figure 5: Time-since-fire curve for the Los Angeles County data.

References


Figure 6: Standard deviations for the survivor curve method of Johnson and Gutsell and for \( \hat{h}(u) \).