Title
Estimation of the Fire Interval Distribution for Los Angeles County, California

Authors
Peng, Roger D.
Schoenberg, Frederic P

Publication Date
2002

Estimation of the Fire Interval Distribution for Los Angeles County, California

Roger D. Peng Frederic Paik Schoenberg

University of California, Los Angeles

Corresponding Author: Roger D. Peng (rpeng@stat.ucla.edu)

8142 Math Sciences

UCLA Department of Statistics

Los Angeles CA 90095-1554

USA

Abstract

This paper describes a new method for estimating the fire interval distribution of a region using historical wildfire boundary data. The new estimator does not assume a specific model and is shown to have good statistical properties. In Los Angeles County, California, detailed information on wildfires has been made available through the use of geographic information systems (GIS) technology. The proposed estimator is applied to GIS data covering the years 1878–1996 and reveals an apparent nonlinear threshold-type relationship with burn area.

Key words: Coverage process data; Time-since-fire distribution; Wildfire risk.

1 Introduction

Wildfire incidence is known to depend critically on numerous covariates including wind, precipitation, fuel moisture, temperature, topography, and many others (see e.g. Flannigan and Harrington, 1988; Renkin and Despain, 1992; Viegas and Viegas, 1994). One particularly important component for understanding a wildfire regime is the fire interval distribution, defined as the distribution of the time until a given location reburns. Estimation of the fire interval distribution involves examining the relationship between burn area and *fuel age* (also called *time-since-fire*), which is specified for each location and time as the time expired since that location last burned.

This paper examines the problem of estimating the fire interval distribution for Los Angeles County, California. Los Angeles County is one of the most active wildfire regimes in the United States and its history is well documented (Hanes, 1971; Minnich, 1983; Pyne et al., 1996; Keeley and Fotheringham, 2001). Wildfires are responsible for significant amounts of property damage in Los Angeles County and are a subject of research for fire managers, fire scientists, urban planners, and ecologists.

Accurate estimation of the fire interval distribution is of interest because there is some disagreement over the relationship between fuel age and burn area in Southern California. Minnich (1983) suggested that the largest fires in recent Southern California history are linked to the increased availability of older fuels. Based on Landsat imagery from Southern California and Northern Baja California, Minnich claimed that the policy of total fire suppression in Southern California 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.

Despite the fact that Minnich's paper was highly influential and was used as support for modern prescribed burning policies, many other works contradict his findings. Van Wagner (1978) and Johnson and Larsen (1991) suggested that fuel age has little or no effect on risk 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), using data from the California Statewide Fire History Database, 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 in Southern California. The authors examined some of the largest fires in the 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).

1.1 Previous Work with Fire Interval Distributions

The lack of full agreement over the precise role of fuel age in contributing to fire risk is not surprising. In general, it is difficult to make precise quantitative statements about fuel age without imposing assumptions. 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) described a useful method for summarizing and quantifying the information stored in these maps, and for producing numerical estimates of quantities such as the fire cycle and average fire interval. They used survivor curves to estimate the probability of an area surviving without fire beyond a certain age. This probability is estimated by computing the proportion of unburned fuel which is beyond a certain fuel age. For example, the probability of an area surviving 30 years (or more) without reburning is estimated using the proportion of the study area which is currently 30 years old or older. The empirical survival function is then fit to a theoretical model using regression. Similar work and applications of this methodology can be found in Johnson and Van Wagner (1985) and Johnson and Larsen (1991).

Often, time-since-fire maps are the only information available to researchers. However, when more detailed information on the fire history is available, one can hope to obtain a more accurate picture of reburn activity in the area. In this situation the survivor curve method is not optimal for several reasons. 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 estimates of average time until reburn more dependent on recent observations. Second, the survivor curves, which depend critically on the particular 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 risk.

Fire interval distributions are often studied by modelling the associated hazard functions. There have been a number of parametric models proposed for describing

such hazard functions. Choosing a particular model depends on specific knowledge of the study area and can be a potentially difficult task. Johnson and Gutsell (1994) use exponential and Weibull models for their survivor curves. The exponential model assumes a constant risk of burning as fuel age increases. This kind of relationship seems implausible for Los Angeles County given the previous research done on the growth cycle of chaparral (the dominant vegetation) and the rate of accumulation of fuels (see e.g. Minnich and Chou, 1997, and references therein). However, the Weibull model, which assumes a linear increase in risk with fuel age, seems equally unrealistic. We have no reason to believe that the risk of burning should increase to infinity as fuel age increases.

Li et al. (1996) and McCarthy et al. (2001) suggest alternative theoretical forms for the fire interval distribution such as Olson, logistic, and fuel moisture models. These models encompass a wide range of different relationships between fuel age and wildfire risk. In research similar to that conducted in this paper, Gill et al. (2000) estimate the "probability of ignition at a point function" using 16 years of Australian satellite image data and compare their results to standard models.

Even with a portfolio of parametric models available, it is still likely that the fire interval distribution falls outside the range of existing options. This paper presents a flexible nonparametric method for estimating the fire interval distribution from detailed wildfire boundary data. The proposed estimator relies on relatively few mathematical assumptions and has good statistical properties. In Section 2 we begin with a description of the data used for the current analysis. Section 3 outlines our method for estimating the fire interval distribution and for obtaining estimates of

variability of the curve. This section elaborates on ideas described in Schoenberg et al. (2002). Section 4 shows the results of applying our method to data on fire history in Los Angeles County and discusses some statistical properties of the estimator. In Section 5 we summarize our results and outline some subjects for future research.

2 Los Angeles County Wildfire Data

Los Angeles County has a long history of collecting detailed information on wildfires. Maps of Los Angeles County wildfires have been recorded by the Los Angeles County Department of Public Works (DPW) and the Los Angeles County Fire Department for roughly the past century. These maps, originally recorded on paper, have since been transferred to the geographic information systems (GIS) software package ArcInfo and stored in coverage files. In those coverage 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 maps of approximately 2000 fires occurring between the years 1878 and 1996. Figure 1 shows the frequency with which different parts of Los Angeles County have burned in the years 1878–1996. This map was constructed by placing a fine grid over the county and counting the number of times each grid square intersected with a polygon in the dataset. One can see that much of the fire activity in Los Angeles County occurs in a band stretching from the northwest to the eastern part of the county. The major exception is in the Malibu area (the region protruding from the western part of the county) where there are some of the highest levels of fire activity in the whole county. Additional information on topography was obtained from the U.S. Geological Survey and the UCLA Institute

of the Environment. In Figure 2 we see that the formation of the mountains largely determines the pattern of fire activity in the County. Much of the burning occurs in the higher elevations and around the mountains.

After a wildfire has occurred the Fire Department sends a team of engineers to the location to trace the boundary of the fire on a map. These maps are then converted into a machine readable format for use in a GIS program. Figure 3 shows the fire boundaries for 1980, a typical year in the dataset. The measurement of the fire boundaries is quite precise: Fire Department officials estimate that the polygon boundaries are accurate to within about 16 meters on each side of the boundary. For the purpose of measuring burn area we regard the errors in the polygon boundaries as negligible. Figure 4 shows the total area burned in each year of the dataset. Although there appears to be a slight increase in total area burned over the years, that is at least partly due to the fact that in the early part of the century smaller fires were significantly underreported. Fire department officials believe that the data from years after 1950 form a complete listing of wildfires burning more than 1 acre.

Figure 5 shows two groups of fires from the northwestern part of the county. This is a typical example of how fuel age might affect the spatial configuration of the fire boundaries. For the fires in 1963 and 1964, a one year interval, the 1964 fires burn around the border of the 1963 fire and stop (Figure 5a). 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 5b). It should be clarified that in the intervening years between 1928 and 1968, there was almost no overburning of the 1928 fires in Figure 5(b) — the 1968 fire is the first instance of significant overburning at this location.

3 Methodology

Our method of estimating the fire interval distribution from the boundary data has two steps. First, for each year in the dataset and for each fuel age u of interest, we compute the amount of u-year-old fuel which was available for burning in that year and the proportion of u-year-old fuel which actually burned. Because of the size of the dataset and the complexity of the polygons, this is a computationally intensive task. Using the computed proportions, we then construct the fire interval curve by applying local smoothing techniques.

3.1 Processing the Wildfire Boundaries

We first define three quantities:

1. B_i = the total area burned in year i.

2. $y_i(u)$ = the total amount of u-year-old fuel available for burning in year i.

3. $p_i(u)$ = the total amount of *available* u-year-old fuel which burned in year i.

All three quantities are defined for the entire County. Let B_0 be the total amount of burning which occurred in the first (earliest) year of the dataset. For a particular year $i > 0$, we compute $y_i(1) = B_{i-1}$, the total amount of burning which occurrred in the previous year. Then $p_i(1)$ is simply the proportion of $y_i(1)$ which burned. In general, for $i > 1$ and $u = 2, 3, \ldots,$

$$y_i(u) = y_{i-1}(u-1)\left[1 - p_{i-1}(u-1)\right].$$

For $u > i$, $y_i(u)$ cannot be computed from the data and is not defined.

Consider a small example: suppose a study area has experienced only three fires in its history — a small fire in 1920, one fire in 1940, and one in 1960. This hypothetical study area is shown in Figure 6. The region labeled INT is the interesection between the 1940 and 1960 fire boundaries; i.e. the reburn region. In this case, the amount of 20-year-old fuel available in 1940 is $Area$(1920 Fire) and the amount of 20-year-old fuel available in 1960 is $Area$(1940 Fire). Similarly, the amount of 20-year-old fuel which burned in 1940 is 0 and the amount of 20-year-old fuel which burned in 1960 is equal to $Area$(INT). Therefore,

$$y_{1940}(20) = Area(\text{1920 Fire}) \quad ; \quad p_{1940}(20) = \frac{0}{Area(\text{1920 Fire})} = 0.$$
$$y_{1960}(20) = Area(\text{1940 Fire}) \quad ; \quad p_{1960}(20) = \frac{Area(\text{INT})}{Area(\text{1940 Fire})}$$

Note that computing $p_i(u)$ and $y_i(u)$ for each u involves intersecting and differencing 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 software library written by Alan Murta (see `http://www.cs.man.ac.uk/~amurta/software`).

3.2 Estimating the Fire Interval Curve

We define the quantity of interest, $h(u)$, as the proportion of u-year-old fuel expected to burn each year. If $p_i(u)$ and $y_i(u)$ are defined as in Section 3.1, then our estimate for each u is

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

Another possibility would have been a naive estimator such as the average burn proportion $\bar{h}(u) = (1/n)\sum_{i=1}^{n} p_i(u)$. Note, however, that $\hat{h}(u)$ has a nice interpretation in the context of the current application. Since $\hat{h}(u)$ is 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 intuitive sense — if in a given year, 1% of $1,000$ available u-year-old hectares burns, this year should influence our estimation of hazard more than an observation of a year where 1% of 10 available hectares burns.

After computing $\hat{h}(u)$ for many different fuel ages u, we use a kernel smoother to construct the estimated *fire interval curve* and highlight the overall trend. The function $h(u)$ is sometimes referred to as the *flammability function* (see e.g. McCarthy et al., 2001). For the estimator in (1) we make the implicit assumption that for each i and u,

$$\mathrm{Var}\left(p_i(u) \mid y_i(u)\right) = \sigma^2 \frac{h(u)}{y_i(u)} \tag{2}$$

where σ^2 is an unknown constant of proportionality independent of u. The intuition behind (2) is that if the entire study area were divided into small 1-unit pieces, and each unit burned independently of the others, then $y_i(u)$ would represent the number of u-year-old pieces that are available in year i and $p_i(u)$ would represent the proportion that burn. In this case, $\mathrm{Var}\left(p_i(u) \mid y_i(u)\right) \propto 1/y_i(u)$. We believe that relation (2) provides a reasonable approximation for the variance behavior of $p_i(u)$. For example, if $y_i(u)$ is very small, then it is more likely that either we will observe total reburning or no reburning. Therefore, $p_i(u)$ will be 1 or 0 and $\mathrm{Var}\left(p_i(u) \mid y_i(u)\right)$ will be high. If $y_i(u)$ is large, then the distribution of $p_i(u)$ is spread more uniformly

between 0 and 1 and will have lower variance. In Figure 7 the log of the sample conditional variance is plotted against $\log[\hat{h}(u)/y(u)]$. The superimposed dotted line has a slope of 1 and an intercept fitted to the data. The linear relationship appears consistent with (2). A variant of Figure 7 was constructed using $\bar{h}(u)$ in place of $\hat{h}(u)$, however there was little difference between the resulting figure and Figure 7.

Given (2), $\hat{h}(u)$ is simply the weighted least squares estimate of $h(u)$, where the weights are proportional to the inverses of the variances. Therefore, $\hat{h}(u)$ is the best linear unbiased estimate of $h(u)$ for each u. Alternatively, one could formulate $\hat{h}(u)$ as a maximum quasi-likelihood estimate of $h(u)$, using (2) for the variance relationship. Since quasi-likelihoods behave much like regular likelihoods, estimates derived from quasi-likelihoods share many of the desirable properties of standard maximum likelihood estimates (Wedderburn, 1974; McCullagh, 1983). For example, under general conditions, estimates of the type in (1) are consistent and asymptotically normal (Jennrich, 1969).

In order to assess the variability of the fire interval curve we construct approximate 95% confidence bands for the curve using the bootstrap. For each fuel age u, we sample with replacement the pairs $\{p_i(u), y_i(u)\}$ to get $\{p_i^*(u), y_i^*(u)\}$. From the bootstrap samples we compute $\hat{h}^*(u) = \left(\sum p_i^*(u) y_i^*(u)\right) / \left(\sum y_i^*(u)\right)$. After computing $\hat{h}^*(u)$ for all fuel ages u we refit the kernel smoother. This procedure is then repeated 1000 times and confidence bounds are constructed using the percentile method (Efron and Tibshirani, 1993).

4 Results and Discussion

The result of applying our method to the Los Angeles County DPW boundary data is shown in Figure 8. The estimated fire interval curve serves as a concise quantification of the dependence of burn area 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 available fuel that burns steadily increases. However, for fuel ages greater than 30 years or so, the proportion remains nearly constant. Thus, the relationship between fire risk and fuel age appears to be nonlinear. There is considerable scatter around the estimated fire interval curve. However, the statistical significance of this overall increase and leveling off of the estimated curve is supported by the bootstrap 95% confidence bands for the curve.

The type of nonlinear threshold relationship detected in Figure 8 is distinctly different from the linear models in common use. The fit resembles the model described by Olson (1963) where the hazard of burning increases to an asymptote at a rate determined by properties of the vegetation. The shape of the fire interval curve is in general agreement with our knowledge of the vegetation in Los Angeles County. Typically, chaparral, the dominant vegetation, does not burn easily until it has reached about 30 years of age, while older chaparral will burn readily (Pyne et al., 1996). The estimated fire interval curve indicates that after an area reaches a certain age, it does not necessarily become more flammable or hazardous. 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 based on fuel age, between conditions that are sufficient and conditions that are extreme.

It should be noted that small fires burning less than 1 acre were not included in our dataset. Ed Johnson (personal communication, October, 2001) has pointed out that an apparent decrease in hazard in locations having recently burned could perhaps be partially attributed to an increase in the rate of very small undetected fires during the vegetative regeneration cycle. However, as such small fires are thought to account for only a tiny fraction of total burn area, it is unlikely that these fires are solely responsible for the apparent decrease in hazard in Figure 8.

In order to take into account some of the spatial inhomogeneity of the fires, the fire interval curve was also estimated for separate sub-regions of Los Angeles County. These sub-regions are shown in Figure 9 and their estimated fire interval curves are shown in Figure 10. Each of the curves displays a threshold similar to the estimated curve for the entire County; for each region there is no significant increase in the estimated hazard curve after 25–30 years. However, only in region 3 do we see a significant increase in hazard for fuel ages between 1 and 15 years. For regions 1, 2, and 4 it is difficult to say whether there is simply not enough data to detect an increase in hazard or whether there is genuinely no increase.

4.1 Stability of the Estimates

We compare the stability of $\hat{h}(u)$ to the estimate corresponding to the survivor function in (3) proposed by Johnson and Gutsell (1994) by computing each estimator for consecutive sample sizes. Johnson and Gutsell (1994) model the the survivorship of a location based on fuel age information in time-since-fire maps. Specifically, if T is the "lifetime" of a location (i.e. the time until it reburns), then they propose, as an

estimator of $\mathbf{P}(T > u)$, the empirical survivor function defined via

$$\hat{S}(u) = \sum_{k \geq u} y_n(k) \Big/ \sum_{k \geq 1} y_n(k) \; . \tag{3}$$

Here, $y_n(k)$ has the same meaning as in Section 3.1 and n is the year in the dataset closest to the present. From (3) it is possible to get the corresponding hazard function via the relation $h(u) = -\frac{d}{du} \log S(u)$.

We start with a small subset of the dataset and progressively increase the sample size by one year, each time computing both estimators, until the entire dataset is used. Let $h_k^{JG}(u)$ and $\hat{h}_k(u)$ denote the Johnson and Gutsell (1994) estimator and the estimator from (1), respectively, estimated from a sample of size k. Figure 11 shows both estimators for $u = 1$ and for increasing values of k. It is apparent that as k increases, $\hat{h}_k(u)$ tends to stabilize and converge while $h_k^{JG}(u)$ continues to vary. Note that the y-axes on the Figures 11(a) and 11(b) are different; the weighted least squares estimate in 11(b) varies on a much smaller scale.

This process is then repeated for all u. Rather than show $h_k^{JG}(u)$ and $\hat{h}_k(u)$ for each u, we take the sample standard deviation of the set $\{\hat{h}_k(u) : k = 1, 2, \dots\}$ for values of u between 1 and 50. The results are shown in Figure 12. 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 k). When data from year $k + 1$ is added, its effect on $\hat{h}(u)$ is counterbalanced by all of the reburn intervals recorded from years 1 to k. If $\hat{h}_k(u)$ is the current estimate of $h(u)$, then given data from year $k + 1$, the updated estimate for each u is

$$\hat{h}_{k+1}(u) = \hat{h}_k(u) + \frac{y_{k+1}(u)}{\sum_{j=1}^{k+1} y_j(u)} \Big[p_{k+1}(u) - \hat{h}_k(u) \Big] \, .$$

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

4.2 Bootstrap Methodology

Applications of bootstrap methodology to spatial data is still an active area of research and clear prescriptions for conducting resampling or simulation are rare. As with any situation involving the use of the bootstrap, the difficulty lies in approximating the probability mechanism which generates the observed data. In Section 3.2 we proposed one possibility for using the bootstrap to assess the variability of the estimated fire interval curve. Resampling the $p_i(u)$'s for each u assumes that given a fuel age, the proportion of that fuel age which burns is independent from year to year. Our exploratory analysis of the data indicated that the data do not appear to be in gross violation of this assumption. For example, plots of the sample autocorrelation function for the sequences $p_1(u), \ldots, p_n(u)$ did not indicate any significant intra-sequence dependence.

4.3 Extension to Wildfire Risk Estimation

For the general problem of wildfire risk estimation, one could use the methodology described in Section 3 in a number of ways. Given a location $\mathbf{z}$ and the fuel age $u_{\mathbf{z}}$ at that location, it seems feasible to model the hazard of fire as a function of $\hat{h}(u_{\mathbf{z}})$ and other meterological, topographical, and socio-economic covariates. Hence, we could

have hazard($\mathbf{z}$) $= f(\hat{h}(u_{\mathbf{z}}), x_{\mathbf{z}})$, where $x_{\mathbf{z}}$ represents a vector of covariates for location $\mathbf{z}$. A simple example for f would be a linear model where hazard($\mathbf{z}$) $= \hat{h}(u_{\mathbf{z}}) + \beta' x_{\mathbf{z}}$ and β is a vector of parameters. For an example of the use of linear models in fire prediction, see Mandallaz and Ye (1997).

5 Conclusions

This paper presents a new technique for estimating the fire interval distribution of a region using historical wildfire boundary data. It has already been noted by other authors that fuel age and the fire interval distribution are important to understanding the overall behavior of wildfire. The estimator presented here can capture more complex relationships than previously developed methods because it does not impose a parametric model. As a weighted least squares estimate $\hat{h}(u)$ possesses good statistical properties such as consistency and asymptotic normality and is the best linear unbiased estimator. Johnson and Gutsell (1994) proposed a useful estimator when the available data consist of time-since-fire maps. In Section 4.1 we showed that this estimator can be highly variable from year to year since it does not efficiently incorporate historical fire data. The estimator proposed here is better suited for the situation when historical data on wildfires are available.

Using our estimator, we demonstrated that for Los Angeles County, the proportion of area burned increases steadily for fuels less than 30 years old, but remains nearly constant thereafter. The data suggest that the proportion of fuel burned and age of the fuel have a nonlinear threshold-type relationship. This relationship is distinctly different from the commonly used exponential and Weibull models.

The characterization of the relationship between burn area in Los Angeles County and fuel age is intended to be useful to fire hazard modelers. The focus on fuel age by no means is meant to underemphasize the importance of other factors in influencing fire risk. These other factors include land use policies, population density, and fire prevention policies, as well as meteorological and topographic variables. 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). Examining the interactions between these variables and their effect on burn area is still an important direction for future research.

While Los Angeles County represents a significant wildfire regime, an important subject 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 considerably alter the observed relationship between burn area and fuel age. Finally, properties of the bootstrap procedure described in Section 3.2 for estimating confidence bounds for renewal densities of coverage processes require further attention. Modifications to the procedure would be needed for the case where the $p_i(u)$'s display significant serial autocorrelations. The development of useful bootstrapping methods for coverage data is an issue of fundamental importance for the problems discussed in this paper and should continue to be actively investigated.

6 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 9978318. The authors thank the Los Angeles County Fire Department and the Los Angeles County Department of Public Works (especially Mike Takeshita and Denise Kamradt) for their generosity in sharing their data, Patricia Kwon for assisting with the data collection, and Yafang Su for generously volunteering her help in the initial data processing.

References

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Flannigan, M. and Harrington, J. (1988) A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953-1980). *Journal of Applied Meteorology*, **27**, 441–452.

Gill, A. M., Ryan, P. G., Moore, P. H. R. and Gibson, M. (2000) Fire regimes of World Heritage Kakadu National Park, Australia. *Austral Ecology*, **25**, 616–625.

Hanes, T. L. (1971) Succession after fire in the chaparral of southern California. *Ecological Monographs*, **41**, 27–52.

Ihaka, R. and Gentleman, R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Jennrich, R. I. (1969) Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, **2**, 633–643.

Johnson, E. and Gutsell, S. (1994) Fire frequency models, methods and interpretations. *Advances in Ecological Research*, **25**, 239–287.

Johnson, E. and Larsen, C. (1991) Climatically induced change in fire frequency in the southern canadian rockies. *Ecology*, **72**, 194–201.

Johnson, E. and Van Wagner, C. (1985) The theory and use of two fire history models. *Canadian Journal of Forest Research*, **15**, 214–220.

Keeley, J. E. and Fotheringham, C. J. (2001) Historic fire regime in Southern California shrublands. *Conservation Biology*, **15**, 1536–1548.

Keeley, J. E., Fotheringham, C. J. and Morais, M. (1999) Reexamining fire suppression impacts on brushland fire regimes. *Science*, **284**, 1829–1832.

Li, C., Ter-Mikaelian, M. and Perera, A. (1996) Temporal fire disturbance patterns on a forest landscape. *Ecological Modelling*, **99**, 137–150.

Mandallaz, D. and Ye, R. (1997) Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research*, **27**, 1685–1694.

McCarthy, M. A., Gill, A. M. and Bradstock, R. A. (2001) Theoretical fire-interval distributions. *International Journal of Wildland Fire*, **10**, 73–77.

McCullagh, P. (1983) Quasi-likelihood functions. *The Annals of Statistics*, **1**, 59–67.

Minnich, R. A. (1983) Fire mosaics in southern California and northern Baja California. *Science*, **219**, 1287–1294.

Minnich, R. A. and Chou, Y. H. (1997) Wildland fire patch dynamics in the chaparral of Southern California and northern Baja California. *International Journal of Wildland Fire*, **7**, 221–248.

Olson, J. S. (1963) Energy storage and the balance of producers and decomposers in ecological systems. *Ecology*, **44**, 322–331.

Pyne, S. J., Andrews, P. L. and Laven, R. D. (1996) *Introduction to wildland fire*. Wiley.

Renkin, R. A. and Despain, D. G. (1992) Fuel moisture, forest type, and lightning-caused fire in Yellowstone National Park. *Canadian Journal of Forest Research*, **22**, 37–45.

Schoenberg, F. P., Peng, R. D., Huang, Z. and Rundel, P. (2002) Detection of non-linearities in the dependence of burn area on fuel age and climatic variables in Los Angeles County, California. *International Journal of Wildland Fire*, to appear.

Van Wagner, C. E. (1978) Age class distribution and the forest fire cycle. *Canadian Journal of Forest Research*, **8**, 220–227.

Viegas, D. X. and Viegas, M. T. (1994) A relationship between rainfall and burned area for Portugal. *International Journal of Wildland Fire*, **4**, 11–16.

Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **3**, 439–447.

A Appendix: Figure Captions

1. Frequency with which different areas of Los Angeles County have burned between 1878 and 1996.

2. Centroids of fire boundaries for years 1878–1996 with elevation (meters).

3. Fire boundaries for the year 1980, with elevation.

4. Total area burned in each year of the dataset (ha).

5. Overburning for different fuel ages. (a) Fires from the years 1963 (gray) and 1964 (white). There is relatively little overburning (black) — approximately 50 hectares. (b) Fires from 1928 (gray) and 1968 (white). The overburning here is much more extensive — 237 hectares.

6. Hypothetical study area with three fires.

7. A plot of the log conditional variance of $p(u) \mid y(u)$ vs. $\log h(u)/y(u)$. Given n years of fire history data $\{(p_1(u), y_1(u)\}, \ldots, \{(p_n(u), y_n(u)\}$, the values of $p_i(u)$ are binned according to values of $\log \tilde{h}(u)/y_i(u)$. In each bin, the variance of the $p_i(u)$'s is used as an estimate of $\mathrm{Var}\,(p(u) \mid y(u))$.

8. Estimated fire interval curve for the Los Angeles County data.

9. Four sub-regions of Los Angeles County. The points represent the centroids of the fire boundaries.

10. Estimated fire interval curve for each of the sub-regions of Los Angeles County.

11. Estimated values of $h(1)$ using (a) the estimate based on (3) and (b) the weighted least squares estimate (1).

12. Standard deviations for $\hat{h}(u)$ and $h^{JG}(u)$.

B Appendix: Figures

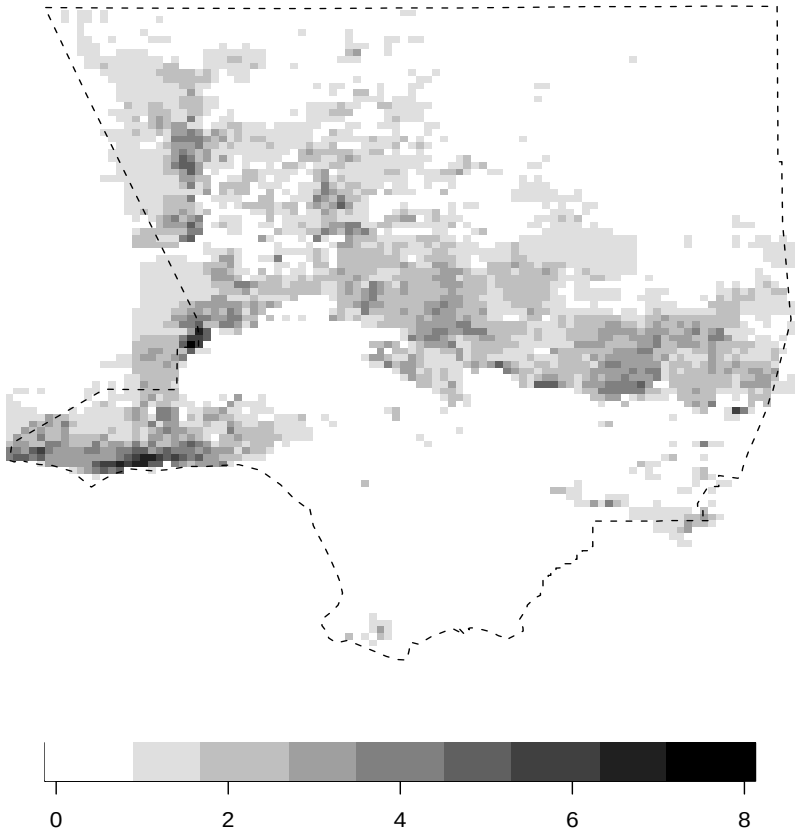

Figure 1: Frequency with which different areas of Los Angeles County have burned between 1878 and 1996.

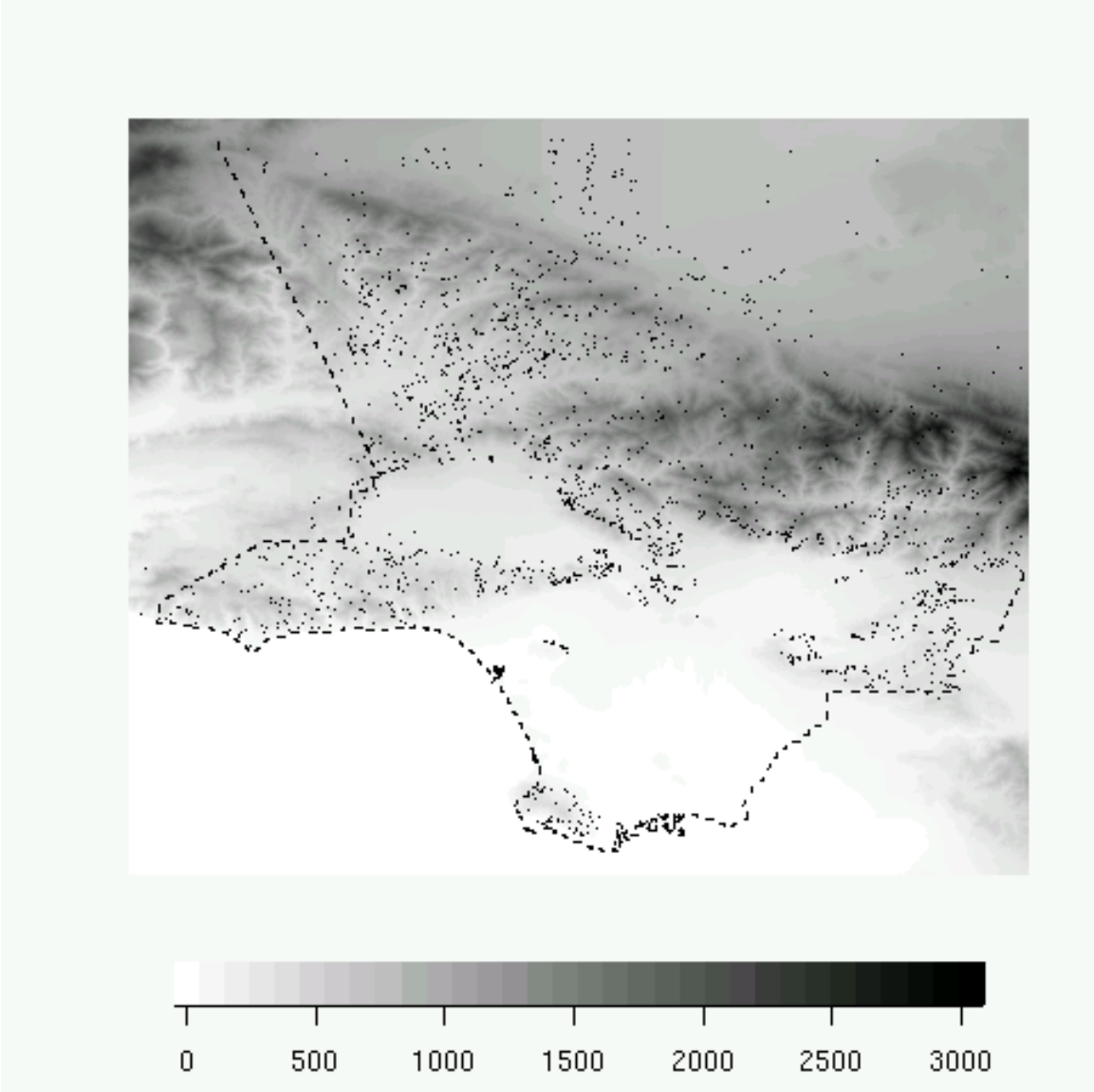

Figure 2: Centroids of fire boundaries for years 1878–1996 with elevation (meters).

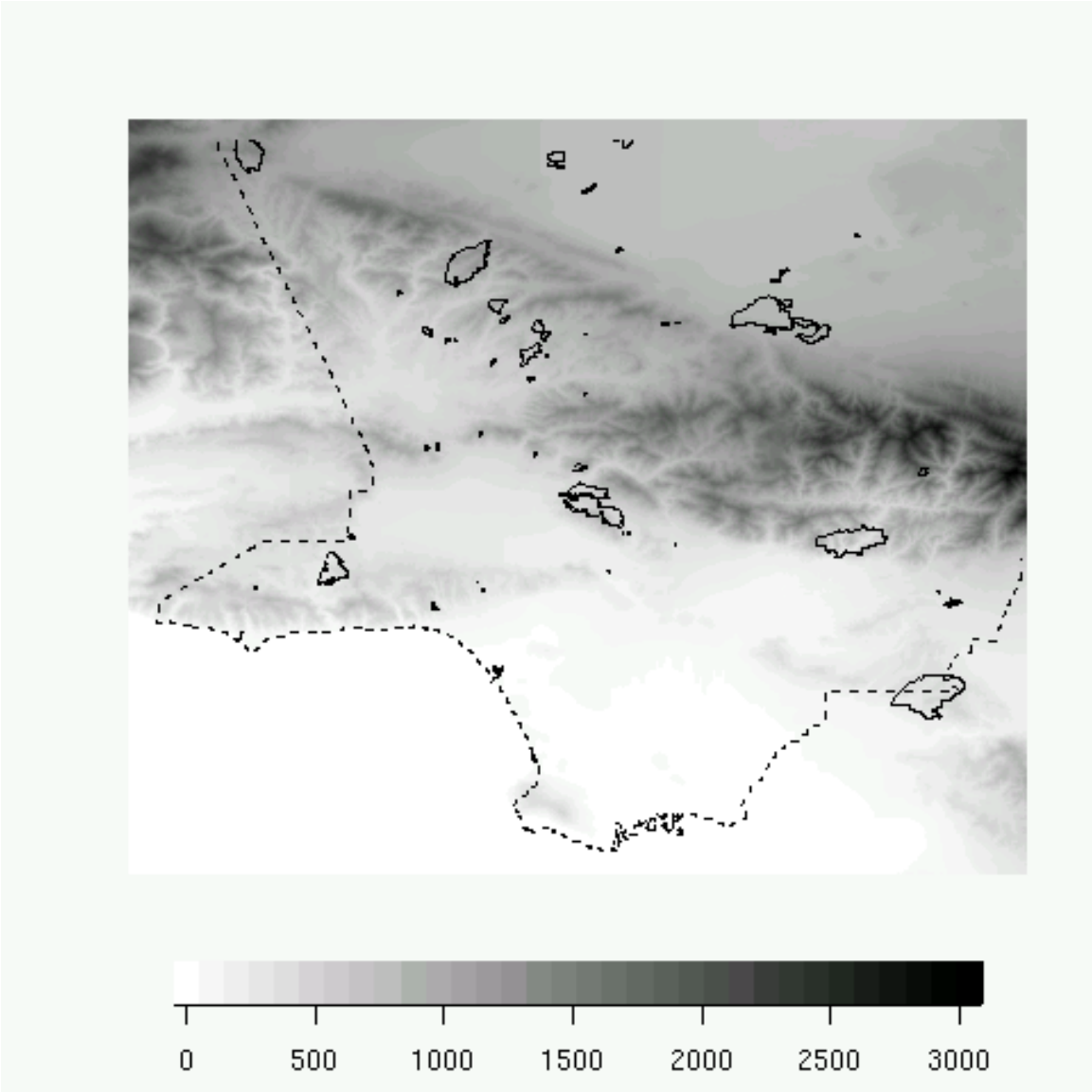

Figure 3: Fire boundaries for the year 1980, with elevation.

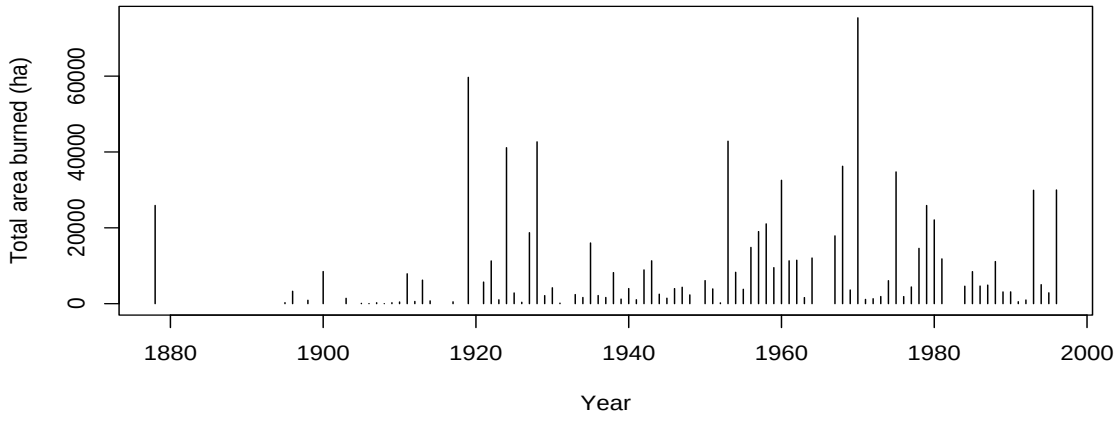

Figure 4: Total area burned in each year of the dataset (ha).

Figure 5: Overburning for different fuel ages. (a) Fires from the years 1963 (gray) and 1964 (white). There is relatively little overburning (black) — approximately 50 hectares. (b) Fires from 1928 (gray) and 1968 (white). The overburning here is much more extensive — 237 hectares.

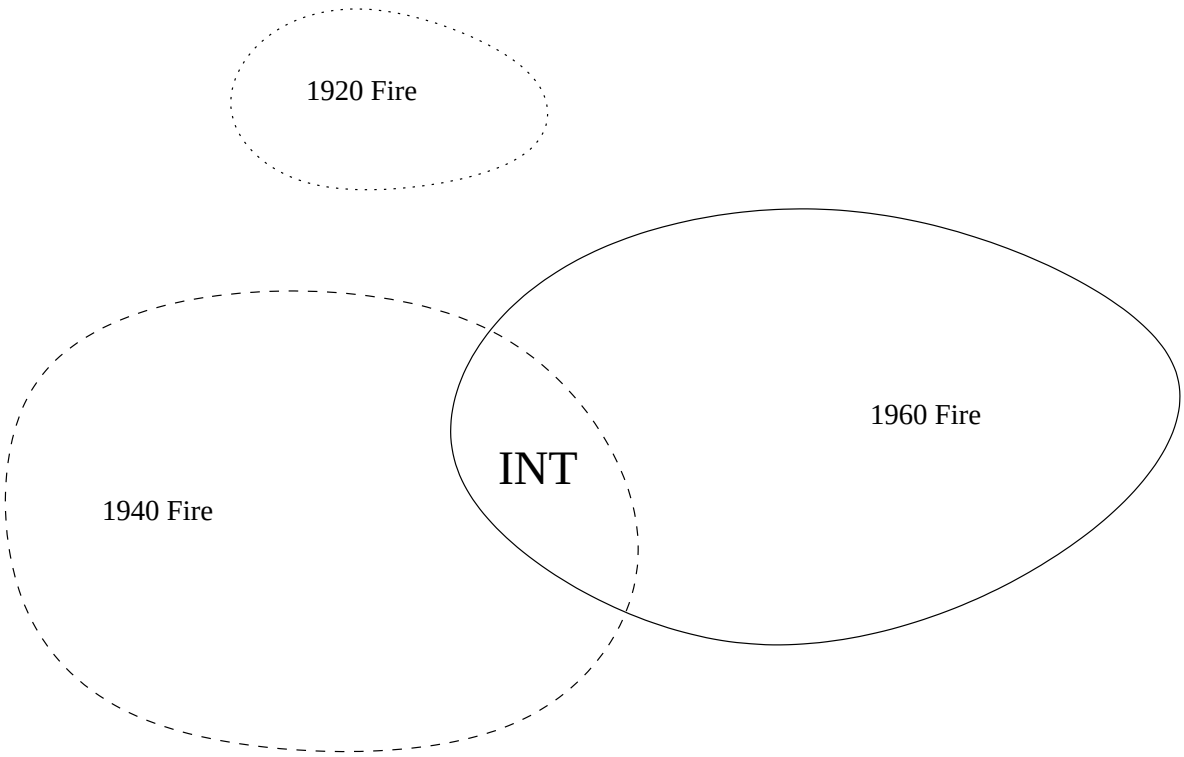

Figure 6: Hypothetical study area with three fires.

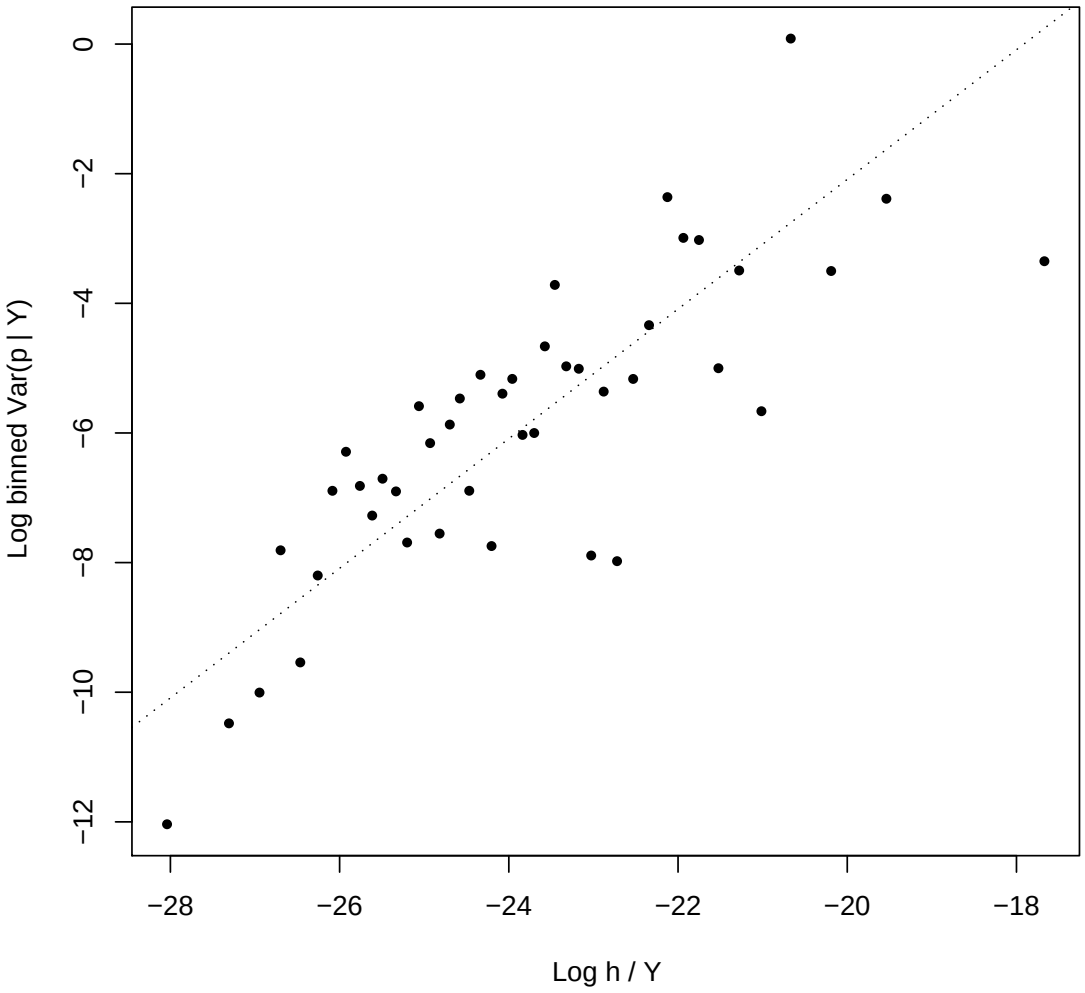

Figure 7: A plot of the log conditional variance of $p(u) \mid y(u)$ vs. $\log h(u)/y(u)$. Given n years of fire history data $\{(p_1(u), y_1(u)\}, \ldots, \{(p_n(u), y_n(u)\}$, the values of $p_i(u)$ are binned according to values of $\log \tilde{h}(u)/y_i(u)$. In each bin, the variance of the $p_i(u)$'s is used as an estimate of $\mathrm{Var}\,(p(u) \mid y(u))$.

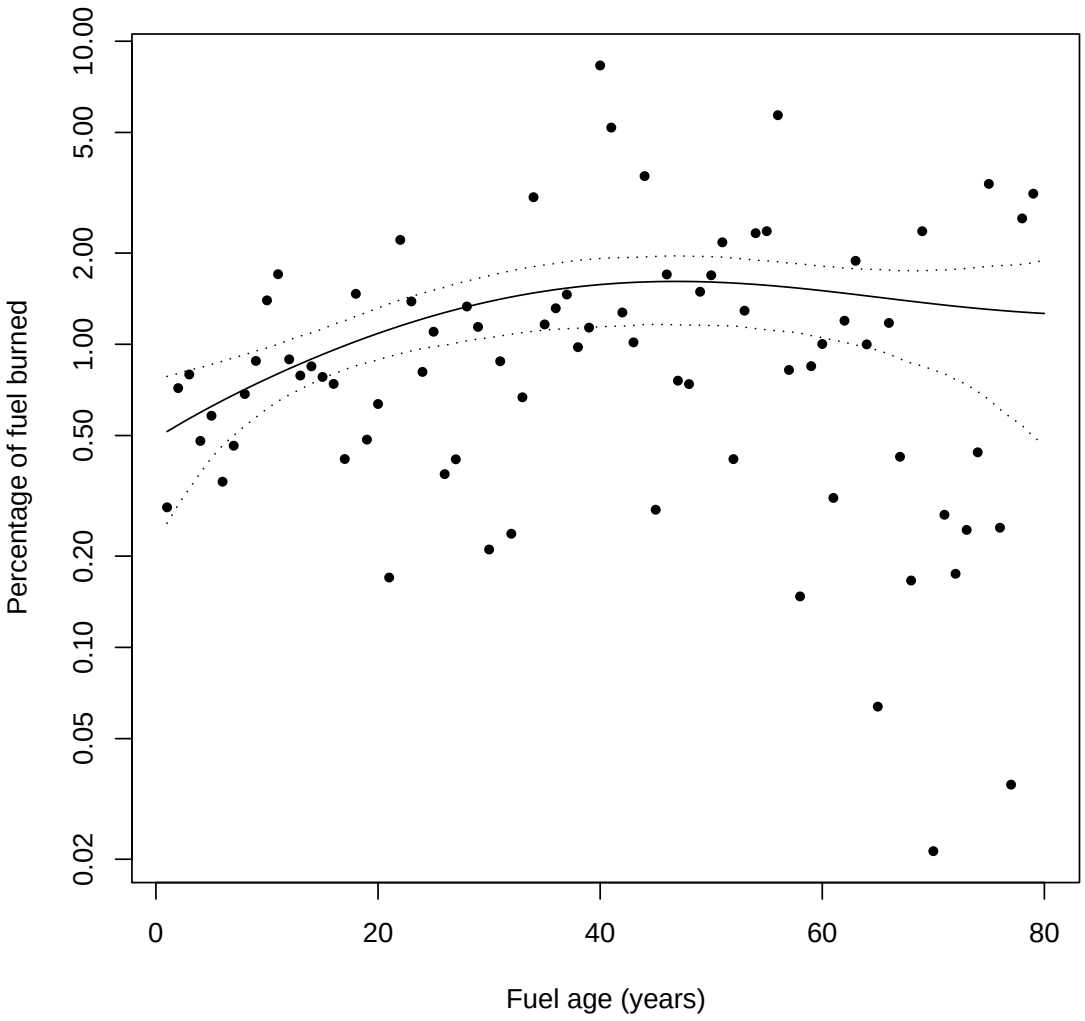

Figure 8: Estimated fire interval curve for the Los Angeles County data.

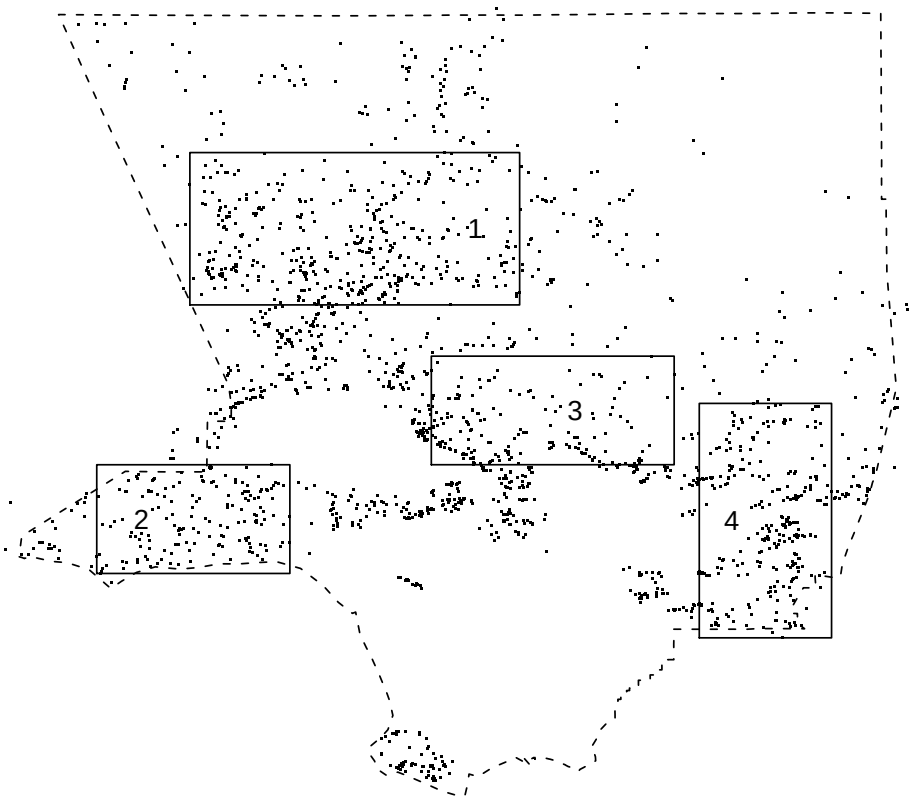

Figure 9: Four sub-regions of Los Angeles County. The points represent the centroids of the fire boundaries.

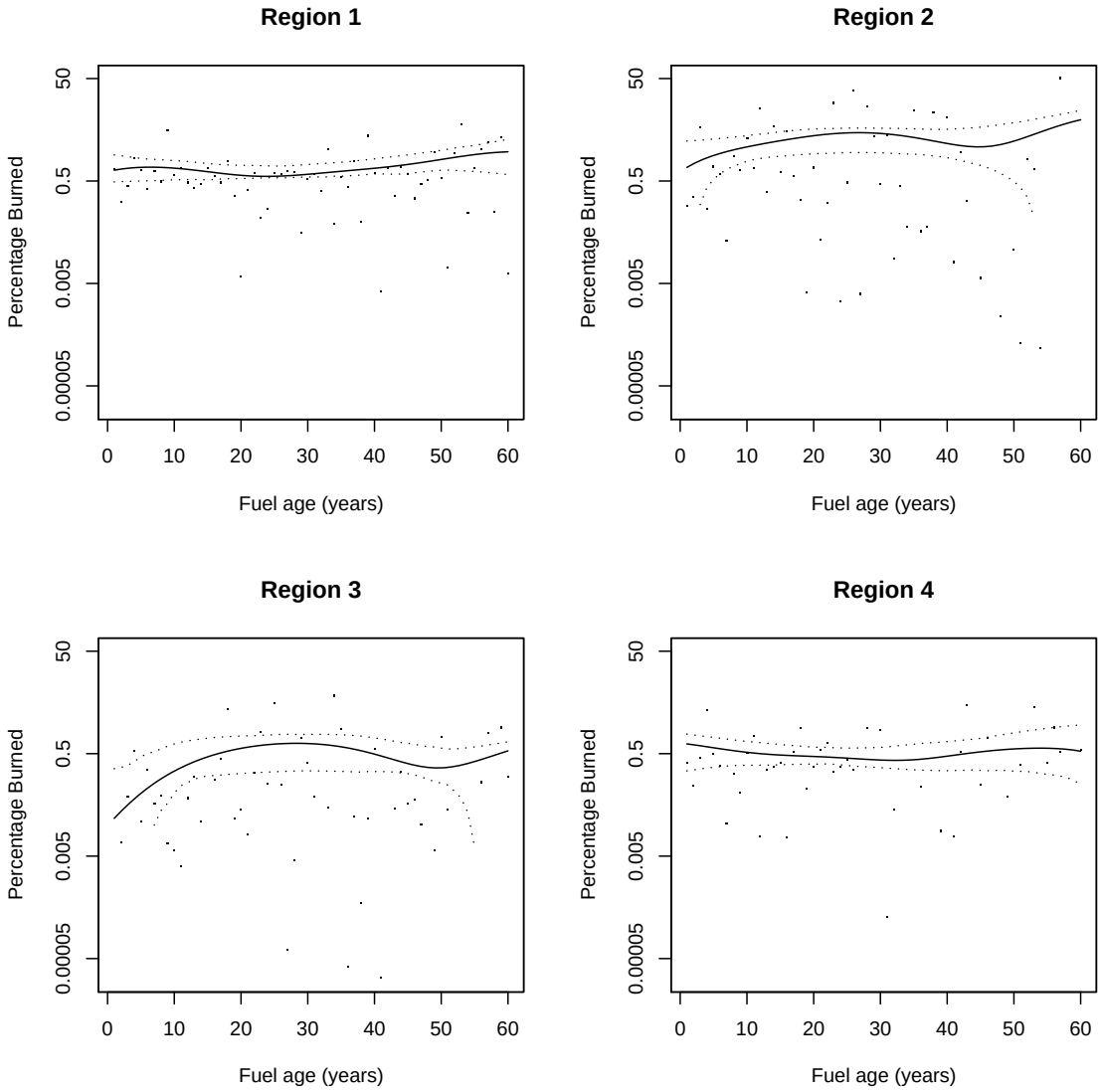

Figure 10: Estimated fire interval hazard curve for each of the sub-regions of Los Angeles County.

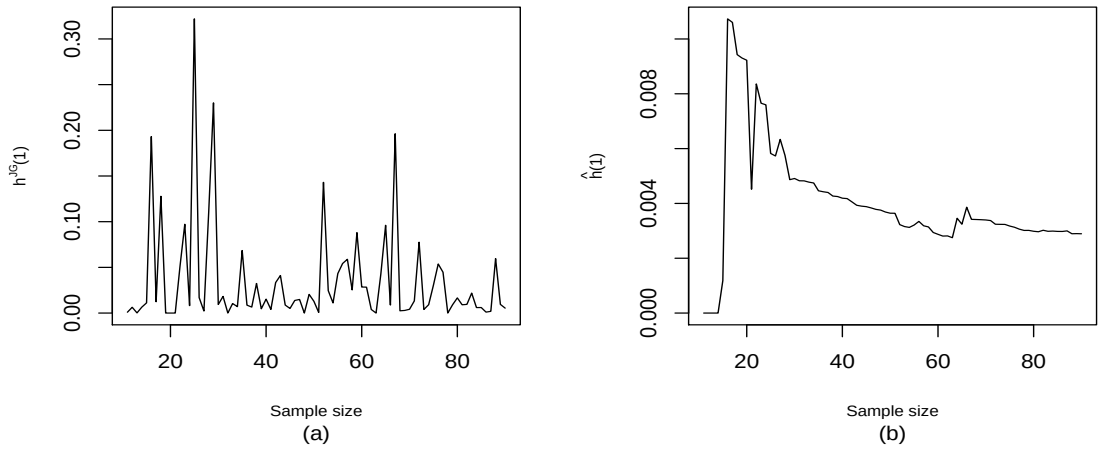

Figure 11: Estimated values of $h(1)$ using (a) the estimate based on (3) and (b) the weighted least squares estimate (1).

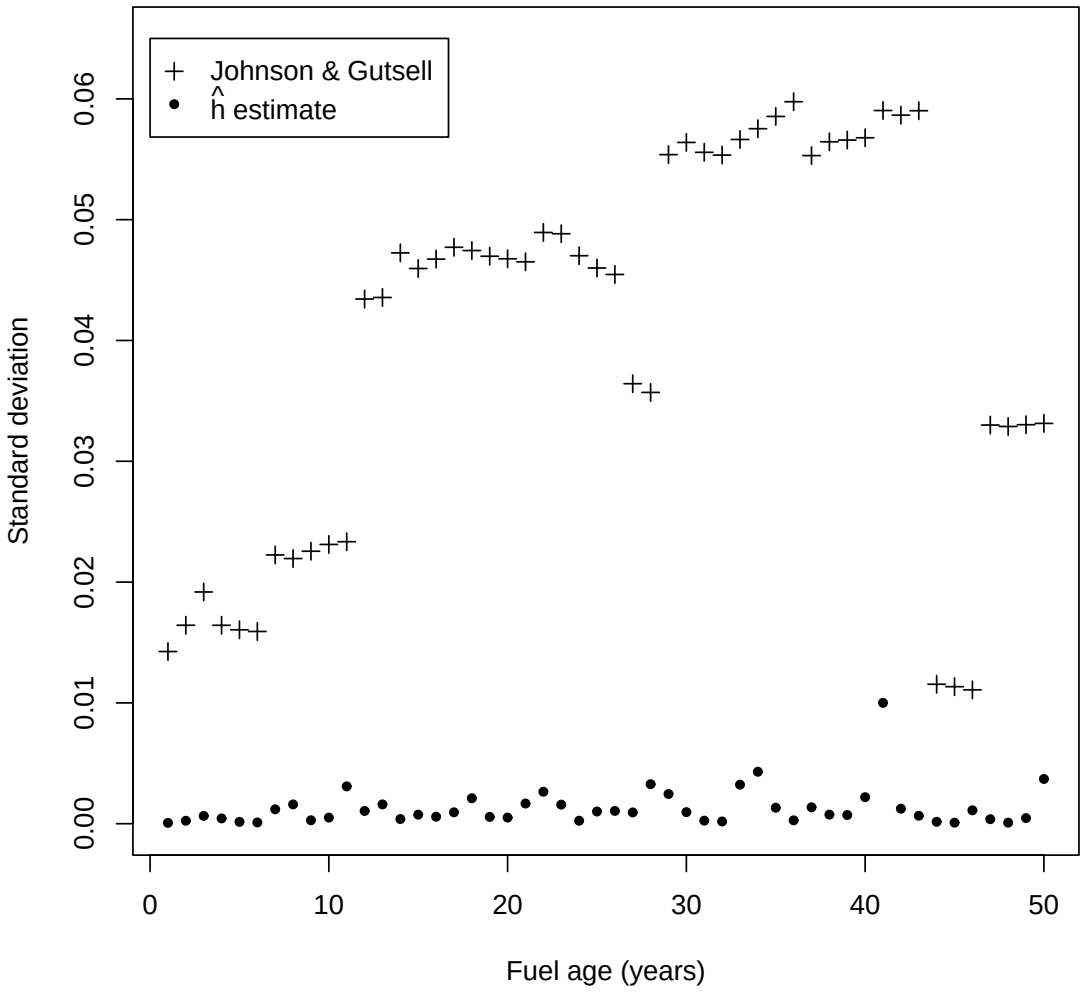

Figure 12: Standard deviations for $\hat{h}(u)$ and $h^{JG}(u)$.