

Estimating the Fire Interval Distribution for Los  
Angeles County, California using Spatial-Temporal  
Fire History Data

Roger Peng      Frederic Schoenberg

*University of California, Los Angeles*

Corresponding Author: Roger Peng ([rpeng@stat.ucla.edu](mailto:rpeng@stat.ucla.edu))

8142 Math Sciences

UCLA Department of Statistics

Los Angeles CA 90095

USA

## Abstract

In this paper a new method is described for estimating the fire interval distribution of a region using spatial-temporal fire history data. In Los Angeles County, California, detailed information on fires has been available through the use of geographic information systems (GIS) technology. The proposed estimator is applied to GIS data covering the years 1878–1996 and it is shown that fuel age appears to have a nonlinear threshold-type relationship with burn area. The estimator is shown to be more stable than previous estimators and to have good finite and large sample properties.

Keywords: Spatial-temporal modeling; Time-since-fire; Fire risk; Threshold relationship; Maximum quasi-likelihood.

## 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 risk.

Fire incidence is known to depend critically on numerous covariates. These covariates include factors such as wind, precipitation, fuel moisture, temperature, topography, and many others. Much work has already been done examining the marginal

relationships between burn area and wind, precipitation, fuel moisture, and temperature (see e.g. Flannigan and Harrington, 1988; Renkin and Despain, 1992; Viegas and Viegas, 1994). One particularly important component of fire risk assessment 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 as the time since that location last burned.

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. 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. Minnich's paper was highly influential and its conclusions were used as support for modern prescribed burning policies.

Minnich's paper contradicted some previous thinking concerning fuel age and fire risk. 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).

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. 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, given that it has already survived to the current day. Similar work can also be found in Johnson and Larsen (1991) and Johnson and Van Wagner (1985).

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 reburn time

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.

In Section 2 we begin with an introductory exploration of the data used for the current analysis. Section 3 outlines our method for estimating the fire interval distribution and discusses its statistical properties. Section 4 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 corresponding to Johnson and Gutsell (1994). Section 5 summarizes our results and outlines some subjects for future research.

## **2 Los Angeles County Fire Data**

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. Fire information is 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 the frequency with which different parts of Los Angeles County have burned in the years 1878–1996. 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 and precipitation were obtained from the U.S. Geological Survey and the UCLA Institute of the Environment, respectively. In Figure 2 we see that the topography of the County closely matches the areas of fire activity. Much of the burning occurs in the higher elevations and around the mountains. Figure 3 shows the average yearly precipitation in different areas of Los Angeles County. One can see that the county is somewhat spatially homogeneous with respect to levels of precipitation.

Figure 4 shows the fire boundaries for 1980, a typical year in the dataset. The measurement of the fire boundaries is very precise. Fire department officials estimate that the polygon boundaries are accurate to within about 16 meters. For the purpose of measuring burn area we regard the errors in the polygon boundaries as negligible. Figure 5 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 inaccuracies in the earlier years. For example, in the early part of the century smaller fires were significantly underreported. Fire department officials believe that the data from years after 1950 are complete.

From Figure 6 one can see 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 6a). 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 6b). It should be clarified that in the intervening years between 1928 and 1968, there was almost no overburning of the 1928 fires in Figure 6(b) — the 1968

fire is the first instance of significant overburning at this location.

### 3 Methodology

First, we define the quantity of interest,  $h(u)$ , as the proportion of  $u$ -year-old fuel expected to burn each year. If  $p_i(u)$  is the proportion of  $u$ -year-old fuel which burns in year  $i$ ,  $i = 1, \dots, n$  (where  $n$  is the most recent year), and  $y_i(u)$  is the amount of  $u$ -year-old fuel available in year  $i$ , then we assume

$$\mathbf{E}[p_i(u) \mid y_i(u)] = h(u). \quad (1)$$

There are a few different methods one could use to estimate the function  $h$ . For instance, Johnson and Gutsell (1994) model 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

$$\hat{S}(u) = \frac{\sum_{k \geq u} y_n(k)}{\sum_{k \geq 1} y_n(k)}. \quad (2)$$

From (2) it is possible to get corresponding hazard estimates. It is also possible to use a parametric model to estimate  $h$ ; however, it may be preferable not to assume a specific model. In contrast, given fire history data  $\{p_i(u), y_i(u)\}$ ,  $i = 1, \dots, n$ , one could simply use a naive estimator such as the average burn proportion

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

or perhaps the weighted average

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

Note that  $\tilde{h}$  has a nice interpretation in the context of the current application. Since  $\tilde{h}$  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.

Alternatively, it is possible to find estimates for  $h$  via a maximum quasi-likelihood procedure. This type of method is useful when one does not wish to specify an entire distribution for the data but rather just the first two moments (Wedderburn, 1974). First, fix  $u$  at a particular fuel age. Suppose we have uncorrelated observations  $(p_i, y_i)$ ,  $i = 1, \dots, n$ , where the conditional mean is  $\mathbf{E}[p_i | y_i] = h$  and the conditional variance  $\text{Var}(p_i | y_i)$  is some function  $V(h; y_i)$ . Then the quasi-loglikelihood function  $q$  is defined by the differential equation

$$\frac{\partial}{\partial h} q(h; p_1, \dots, p_n, y_1, \dots, y_n) = \sum_{i=1}^n \frac{p_i - h}{V(h; Y_i)}. \quad (3)$$

Solving (3) for  $q$  and setting

$$\hat{h} = \arg \max_h q(h; p_1, \dots, p_n, y_1, \dots, y_n) \quad (4)$$

yields the maximum quasi-likelihood estimate (MQLE) of  $h$ .

Wedderburn (1974) and McCullagh (1983) note that quasi-likelihoods behave much like regular likelihoods and estimates derived from quasi-likelihoods share many of the desirable properties of standard maximum likelihood estimates. For example, under general conditions, estimates of the type in (4) are consistent and asymptotically normal (Jennrich, 1969).

The maximum quasi-likelihood procedure requires that the first two conditional moments of  $p_i$  be specified. For fixed  $u$ , the conditional mean is constant from (1). Therefore, we only need to specify a model for the conditional variance. For the current application, it seems reasonable to use

$$\text{Var}(p_i | y_i) = \sigma^2 \frac{h}{y_i} \quad (5)$$

where  $\sigma^2$  is an unknown constant of proportionality independent of  $u$ . The intuition behind (5) is that if the entire study area were divided into small 1-unit pieces, and each unit burned independently of the others, then  $y_i$  would represent the number of  $u$ -year-old pieces that are available in year  $i$  and  $p_i$  would represent the proportion that burn. In this case,  $\text{Var}(p_i | y_i) \propto 1/y_i$ . We believe that relation (5) provides a reasonable approximation for the variance behavior of  $p_i$ . For example, if  $y_i$  is very small, then it is more likely that either we will observe total reburning or no reburning. Therefore,  $p_i$  will be 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. In Figure 7 the log of the sample conditional variance is plotted against  $\log[\tilde{h}(u)/y(u)]$ . The superimposed dotted line has a slope of 1 and an intercept fitted to the data. The linear relationship appears consistent with (5). A variant of Figure 7 was constructed using  $\bar{h}(u)$  in place of  $\tilde{h}(u)$ , however there was little difference between the resulting figure and Figure 7.

### 3.1 Estimation of $h(u)$

A nice property of  $\tilde{h}$  (defined in the previous section) is that if we assume the data are conditionally uncorrelated, then  $\tilde{h}$  coincides exactly with the MQLE of  $h$ . To see

this, note that if  $\text{Cov}(p_i, p_j | y_1, \dots, y_n) = 0$ , for all  $i \neq j$ , then the quasi-loglikelihood for  $h$  (using (5) for the variance) is defined by the differential equation

$$\frac{\partial}{\partial h} q(h; p_1, \dots, p_n, y_1, \dots, y_n) = \sum_{i=1}^n y_i \frac{p_i - h}{h \sigma^2}.$$

Solving the differential equation, we get for the quasi-loglikelihood

$$q(h; p_1, \dots, p_n, y_1, \dots, y_n) = \sum_{i=1}^n \frac{y_i}{\sigma^2} (p_i \log h - h). \quad (6)$$

Maximizing (6) over  $h$  gives us the maximum quasi-likelihood estimate (MQLE)

$$\hat{h} = \frac{\sum_{i=1}^n p_i y_i}{\sum_{i=1}^n y_i}. \quad (7)$$

In addition to the already mentioned large sample properties of MQLE's, we can also readily obtain finite sample properties for  $\hat{h}$ . First, it is clear that  $\hat{h}$  is also the weighted least squares estimate of  $h$ . Therefore,  $\hat{h}$  is the best unbiased estimate among linear combinations of the  $p_i$ 's. Second, if we assume that

$$\mathbf{E}[p_i^k | y_1, \dots, y_n] = \mathbf{E}[p_i^k | y_i] \quad (i = 1, \dots, n)$$

for  $k = 1, 2$ , the variance of  $\hat{h}$  for a sample of size  $n$  can be expressed as

$$\text{Var}(\hat{h}) = \frac{\sigma^2}{n} \mathbf{E}[1/\bar{y}].$$

For larger samples, this variance formula could be used in conjunction with a normal approximation to get confidence bounds for the individual estimates.

After computing  $\hat{h}$  for many different fuel ages  $u$ , we use a local linear smoother to construct the estimated *fire interval hazard curve* and highlight the overall trend. In order to assess the variability of the fire interval hazard curve without imposing a more restrictive model, we construct 95% confidence bands for the curve using the bootstrap. For each fuel age  $u$ , we sample with replacement the pairs

$(p_1, y_1), \dots, (p_n, y_n)$  to get  $(p_1^*, y_1^*), \dots, (p_n^*, y_n^*)$ . From the bootstrap samples we compute  $\hat{h}^* = (\sum p_i^* y_i^*) / (\sum y_i^*)$ . After computing  $\hat{h}^*$  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, ch. 13).

For the larger problem of wildfire risk estimation, one could generalize the above approach 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 meteorological, topographical, and socio-economic covariates. Hence, we could have

$$\text{hazard}(\mathbf{z}) = f\left(\hat{h}(u_{\mathbf{z}}), x_{\mathbf{z}}\right)$$

where  $x_{\mathbf{z}}$  represents a vector of covariates for location  $\mathbf{z}$ . A simple example for  $f$  would be a linear model where

$$\text{hazard}(\mathbf{z}) = \hat{h}(u_{\mathbf{z}}) + \beta' x_{\mathbf{z}}$$

and  $\beta$  is a vector of parameters. For an example of the use of linear models in fire prediction, see Mandallaz and Ye (1997).

Note that computing  $p_i$  and  $y_i$  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>).

## 4 Results and Discussion

The result of applying our method to the Los Angeles County DPW data is shown in Figure 8. The estimated fire interval hazard 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 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 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 type of nonlinear threshold relationship detected in Figure 8 is distinctly different from the linear models in common use. However, the shape of the fire interval hazard 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 hazard 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 hazard 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 hazard curves are shown in Figure 10. Each of the curves seem to agree with 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

In order to compare the stability of  $\hat{h}(u)$  to the estimate corresponding to the survivor function in (2) proposed by Johnson and Gutsell (1994) we compute each estimator for consecutive sample sizes. 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 time-since-fire estimator and the estimator from (7), 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 MQLE 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

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

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  $y_{k+1}(u) / \sum_{j=1}^{k+1} y_j(u)$ . 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 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 inflate the estimates of  $h(u)$  for all fuel ages. Other than this simple case, it is difficult to say how missing data affect the estimation of  $h(u)$  since little is known about the spatial configuration of the missing fires.

One possible method for examining the effect of missing data on  $\hat{h}(u)$  could be to use a cross-validation type of procedure. Let  $\hat{h}_{(j)}(u)$  represent the estimate of  $h(u)$  computed with the  $j$ th year removed from the dataset. For each  $u$ , a simple measure of the increased variability due to missing years of data is

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

or alternatively

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

where  $\hat{h}(u)$  is the estimate of  $h(u)$  using all the years  $1, \dots, n$ . If the fire patterns in the missing years are essentially similar to those in non-missing years, then  $\gamma_1$  (or  $\gamma_2$ ) may provide a decent measure of how missing data affects  $\hat{h}(u)$ . For the Los Angeles County DPW dataset the assumption that the missing years of data are similar to the non-missing years is doubtful, especially since most of the missing years are from the early 1900's. Increases in development as well as changes in land use and fire suppression policies since the turn of the century likely have changed the fire burn patterns over time. 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.

## 5 Conclusions

In this paper we have presented a new technique for estimating fire interval distribution of a region using detailed spatial-temporal fire history 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 fire. The estimator presented here can capture more complex relationships than previous methods because it does not impose a parametric model. As a maximum quasi-likelihood estimate, we know that  $\hat{h}(u)$  possesses good statistical properties such as consistency and asymptotic normality. Furthermore,  $\hat{h}(u)$  exhibits good finite sample behavior such as being the best linear unbiased estimator and having a variance of order  $1/n$ . Using our estimator, we have shown 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.

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.

Finally, 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 considerably alter the observed relationship between burn area and fuel age.

## 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.

- 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., Fotheringham, C. J. and Morais, M. (1999) Reexamining fire suppression impacts on brushland fire regimes. *Science*, **284**, 1829–1832.
- Mandallaz, D. and Ye, R. (1997) Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research*, **27**, 1685–1694.
- 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.

- 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.
- 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. Precipitation in Los Angeles County (average inches per year). Each number on the plot represents the location and average yearly precipitation at each weather station.
4. Fire boundaries for the year 1980, with elevation.
5. Total area burned in each year of the dataset (ha).
6. 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.
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  $\{(p_1(u), y_1(u)), \dots, (p_n(u), y_n(u))\}$ , 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  $\text{Var}(p(u) \mid y(u))$ .
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 hazard curve for each of the sub-regions of Los Angeles County.
11. Estimated values of  $h(1)$  using the (a) estimate based on (2) and (b) the MQLE (7).
12. Standard deviations for the  $\hat{h}(u)$  and the estimates derived from the survivor curve method of Johnson and Gutsell.

## B Appendix: Figures

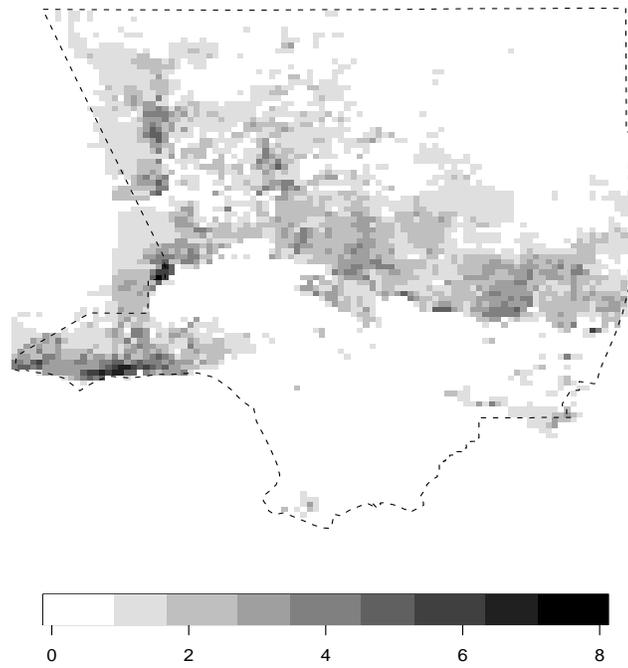


Figure 1:

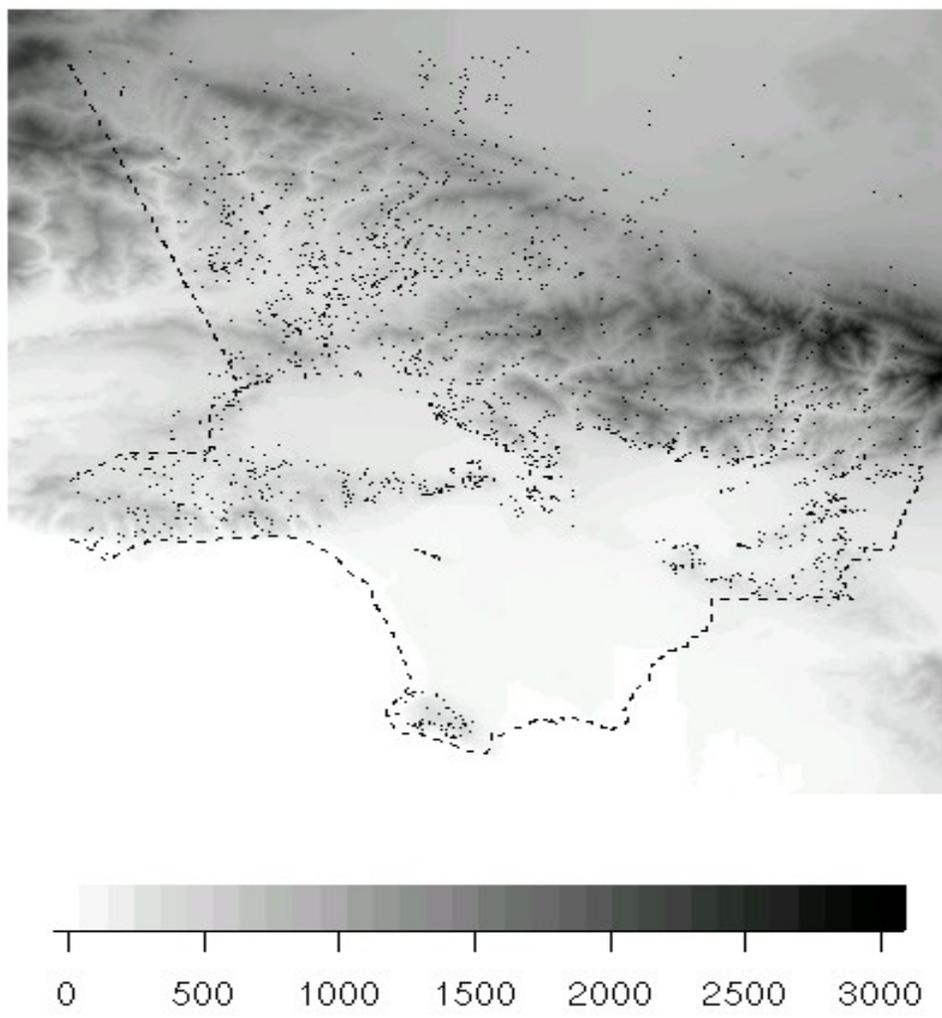


Figure 2:

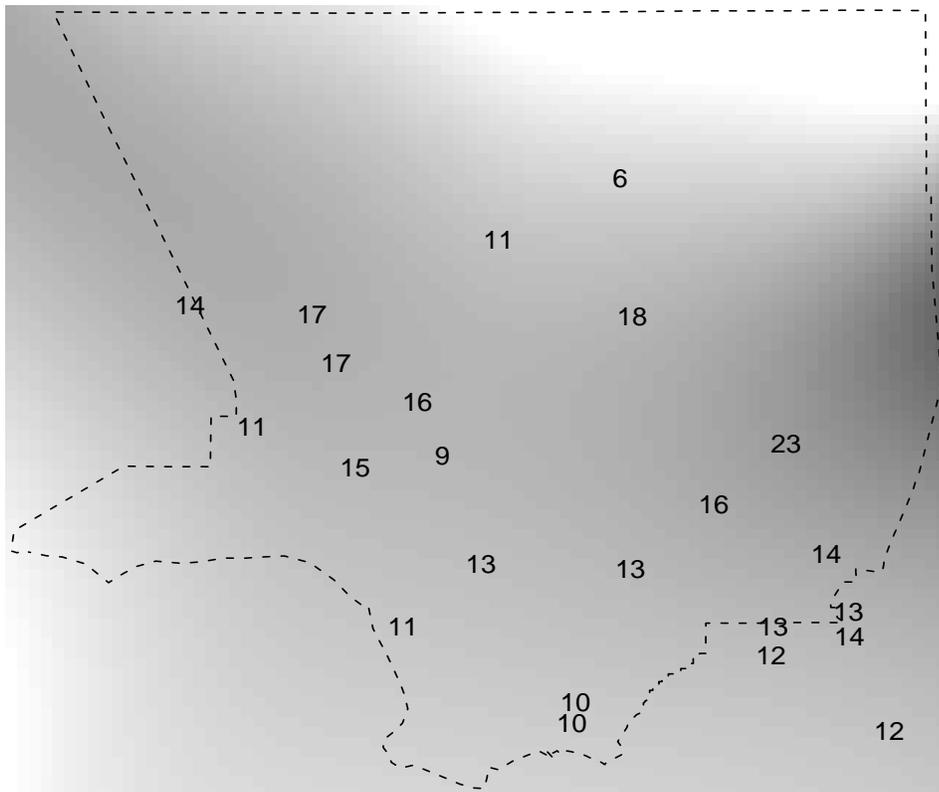


Figure 3:

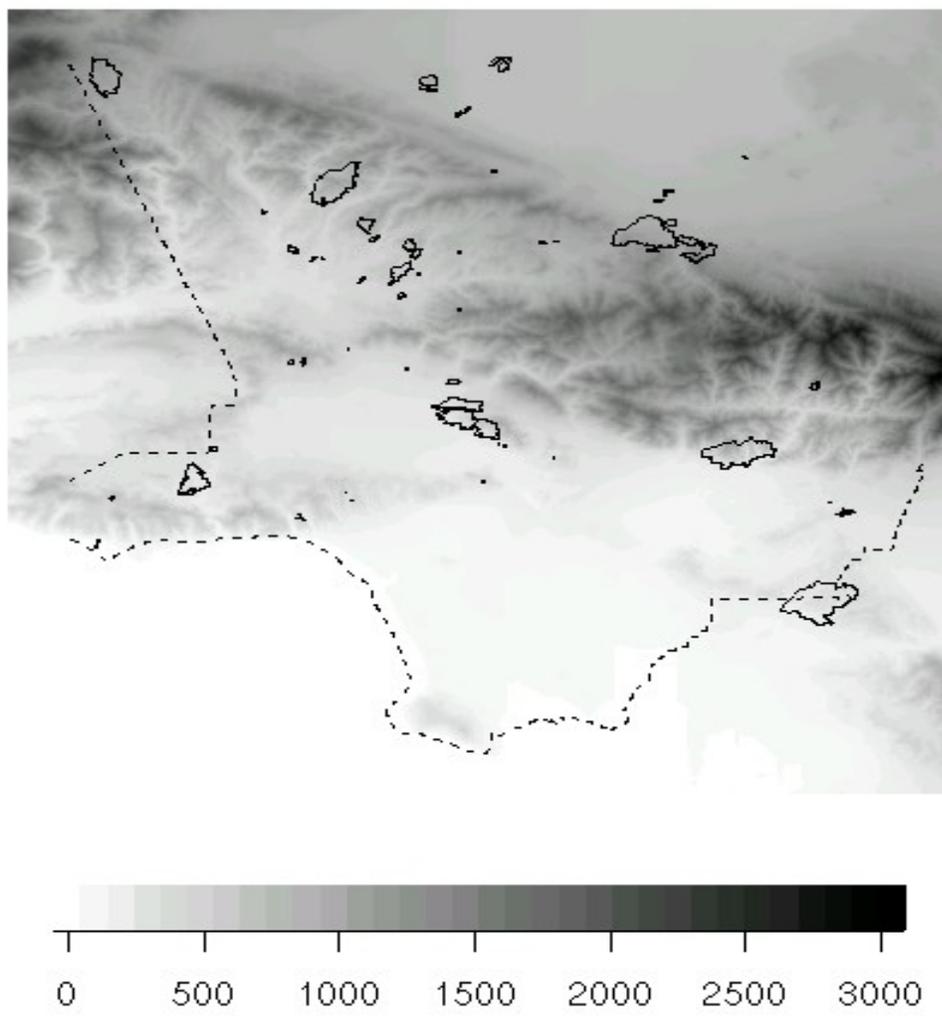


Figure 4:

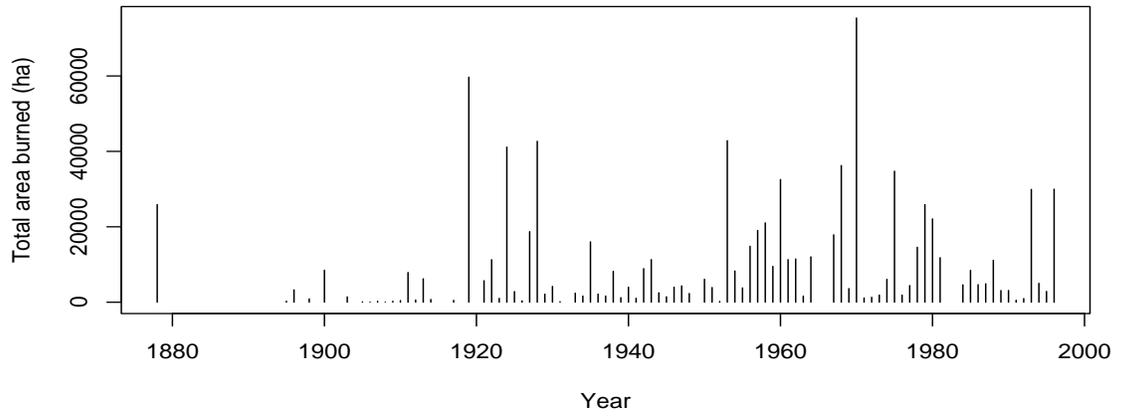
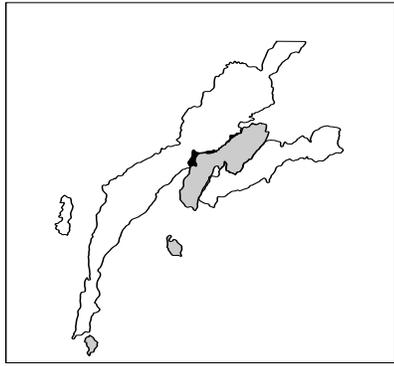
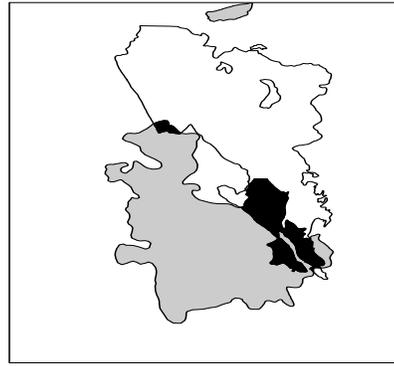


Figure 5:



(a)



(b)

Figure 6:

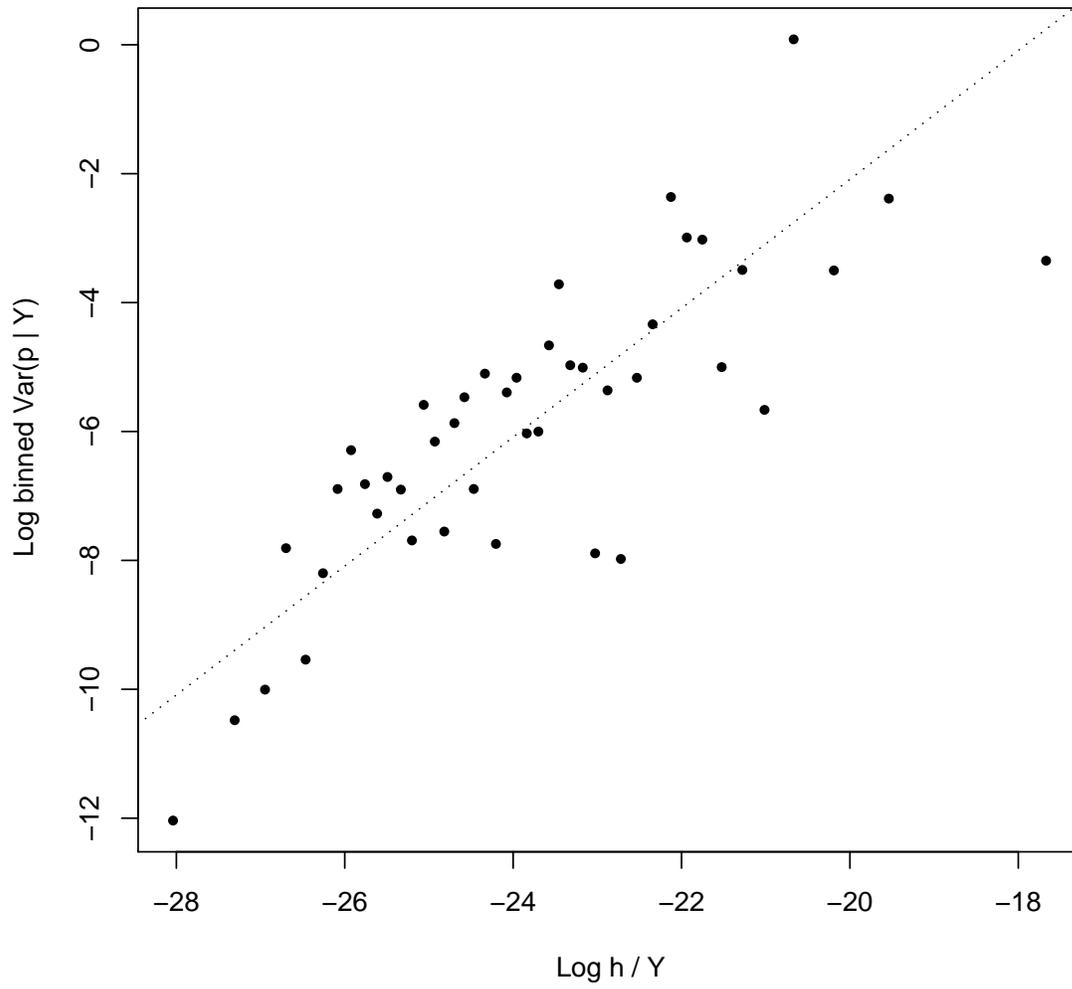


Figure 7:

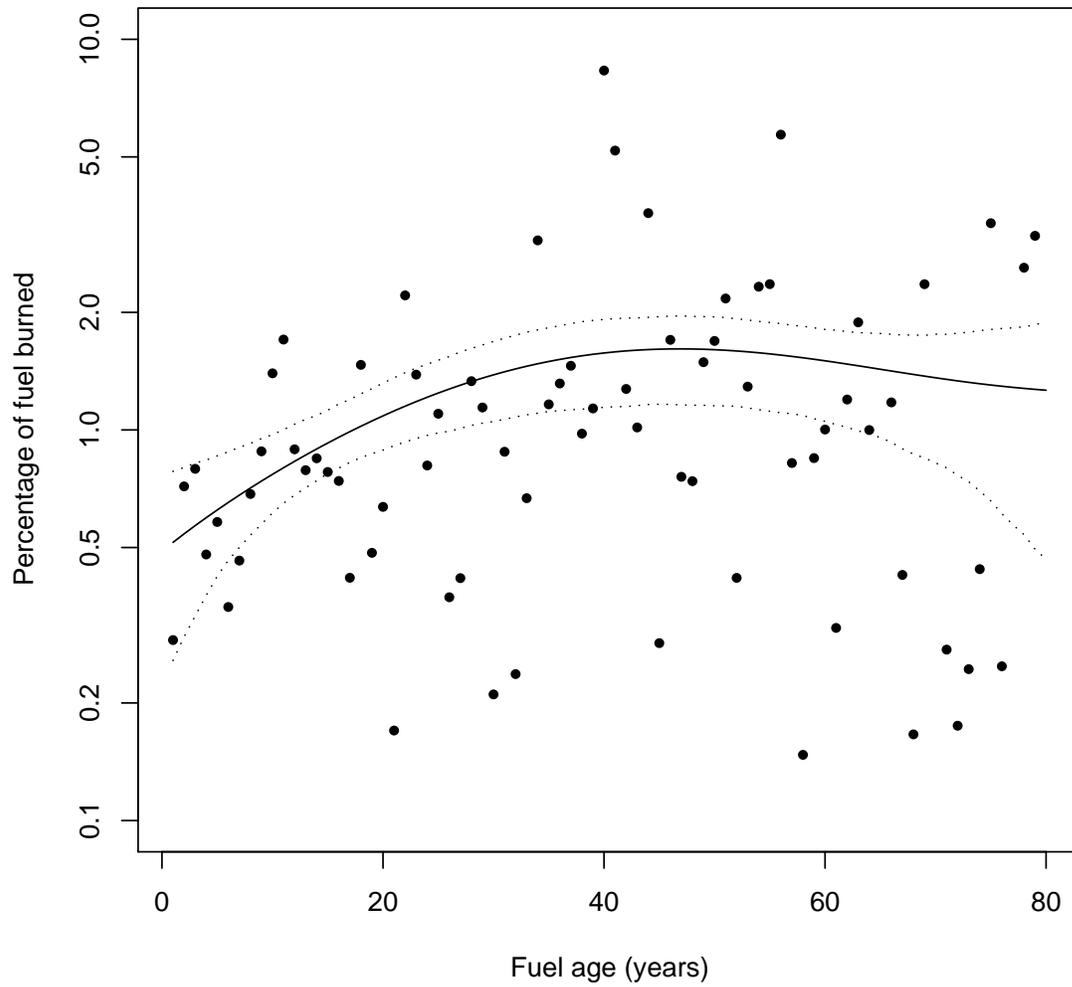


Figure 8:

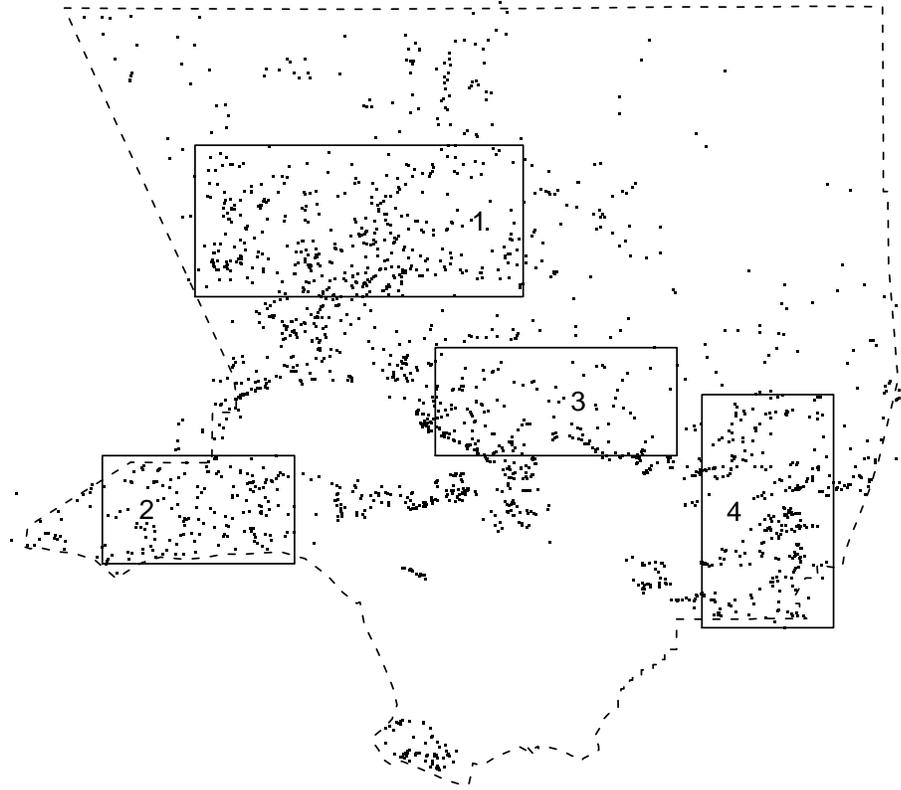


Figure 9:

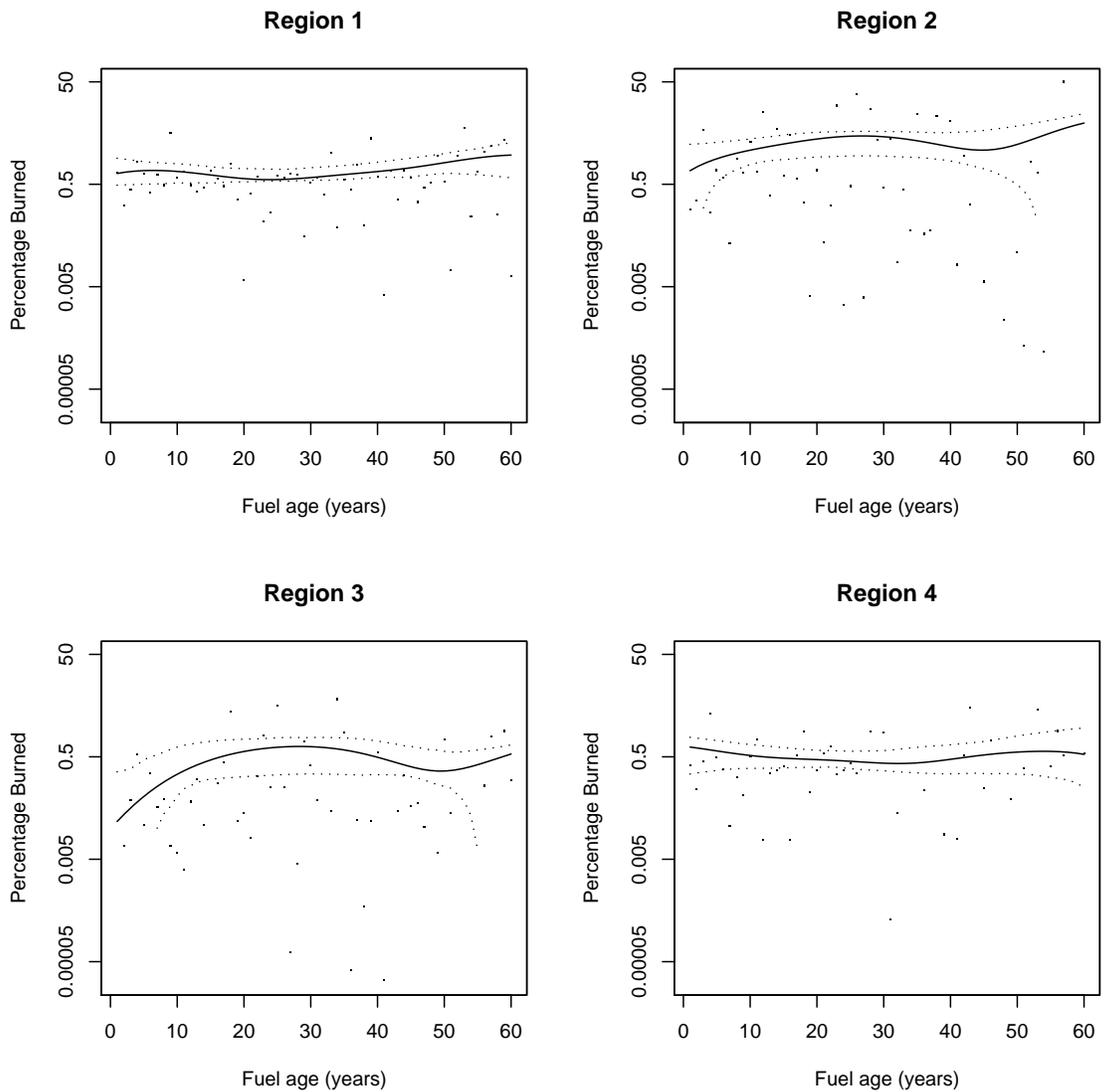


Figure 10:

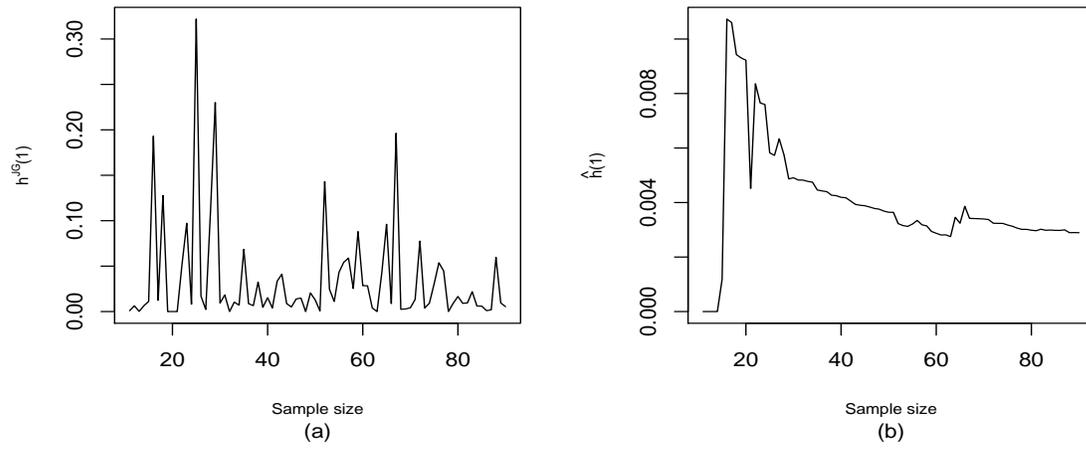


Figure 11:

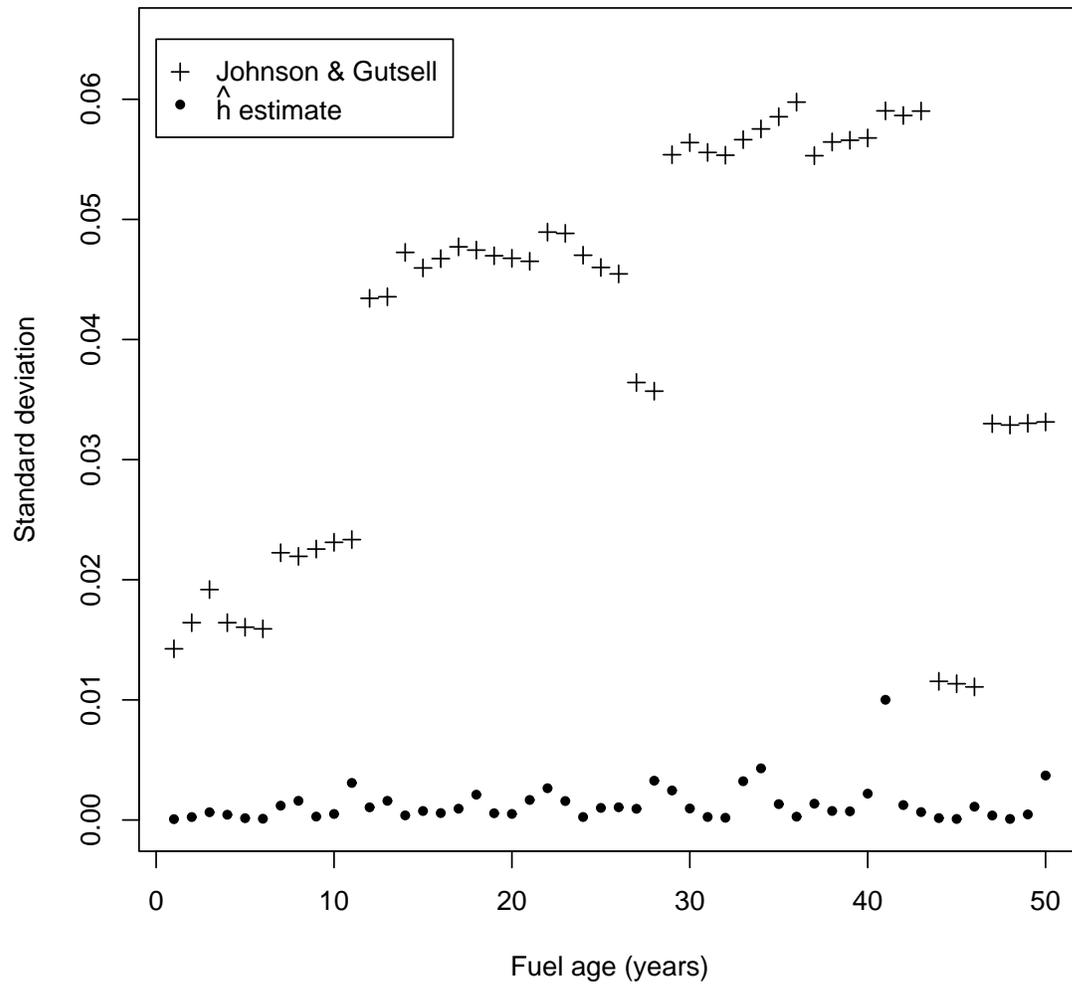


Figure 12: