Reconstructing Forest Fire History -- Identifying Hazard Rate Change Points Using the Bayes' Information Criterion

William J. Reed

DMS-783-IR June 1997

RECONSTRUCTING FOREST FIRE HISTORY – IDENTIFYING HAZARD RATE CHANGE POINTS USING THE BAYES' INFORMATION CRITERION.

William J. Reed*

Department of Mathematics and Statistics
University of Victoria
PO Box 3045
Victoria B.C.
Canada V8W 3P4.

June 1997

Abstract

Graphical analysis of data from a time-since-fire map often suggests temporal variations in historical forest-fire frequency. In this article it is assumed that changes in the fire hazard rate, if they ever occurred, happened at distinct change points, separating epochs during which the hazard rate was constant. A methodology for determining the most plausible number of change points, using the Bayes' Information Criterion, is developed. It is based on an overdispersed model (with corresponding quasi-likelihood function) for the burning or survival of units of the forest. The method is applied to two datasets from time-since-fire maps in the Canadian Rockies. In both

*Research supported by NSERC grant OGP 7252.

examples a single most plausible model stands out. In each case maximum likelihood estimates of change points and of the hazard rates which prevailed between them are calculated.

Keywords: BIC, change points, fire hazard rate, model selection, quasi-likelihood.

1 Introduction

This article discusses a problem in model selection, namely the identification of multiple change points for a piece-wise constant hazard rate, in an overdispersed survival model. The context in which the problem arises is that of identifying changes in the historical frequency of forest fire. This is a question of interest to ecologists and forest-fire researchers and may also provide information for more general studies on climate change. Furthermore the estimation of historical fire frequencies is important in attempts to establish preserves of natural forest for purposes of biodiversity conservation. Such preserves are required to mimic, at the landscape level, the age distribution of "natural" undisturbed forest (Anon, 1995). At least for boreal forests where the main agent of stand regeneration is fire, the natural age distribution will depend on the "natural" fire frequency, *i.e.* the fire frequency which prevailed before intervention by people of European origin.

While the methodology developed in the article is described in terms of the fire history problem, there is no reason why it should not be adapted for application to other problems in survival analysis in which the identification of change points for the hazard rate is of major interest (Matthews and Farewell, 1982; Worsley, 1988; Loader, 1991).

The role of fire in the ecology of boreal forests has been recognised since the pioneering work of Heinselman (1974), who identified fire as a main factor in forest succession dynamics and consequent landscape vegetation mosaic. Heinselman's research involved the construction of a time-since-fire map from a complete survey of the 215,000 ha. wilderness area of the Boundary Waters Canoe Area, Minnesota. From this he was able to estimate the fire frequency in historical (pre European settlement) times. In the twenty-odd years since Heinselman's paper appeared there have been many other fire-history studies of boreal wilderness regions (see Johnson and Gutsell, 1994, and references therein). In some of these studies the time since last fire has been determined at all points in the study area (time-since-fire *map*); in others it has been determined only at a number of randomly sampled points in the study area (time-since-fire *sample*). One of the main concerns of fire-history studies has been to estimate the fire frequency prevailing historically, and to determine when significant changes in that frequency have occurred.

The statistical methodology used in analyzing fire history data has, until recently, been largely informal, based on a graphical analysis of the semi-log plot of the cumulative time-since-fire area distribution. The expository article of Johnson and Gutsell (1994) provides a description of such methods. After first checking for spatial heterogeneity, and if necessary partioning the region into spatially homogeneous sub-regions, the recommended method for identifying temporal changes is to identify, by eye, changes in slope in the semi-log plot of the cumulative time-since-fire distribution.

In recent articles likelihood inference for the hazard rates prevailing between *pre-specified* change points for fire-map (Reed *et al.*, 1997) and fire sample data (Reed, 1996) has been developed using overdispersed survival models. The problem of identifying change points from the data alone has been addressed by Reed (1997), where a test for a homogeneous hazard (no change points) against the alternative of a single change point (partitioning the past into two homogeneous epochs) is developed. However when multiple change points are possible the problem is essentially one of model selection, analagous to the variable selection problem in regression. While the maximum likelihood paradigm is not well suited to such problems, other procedures (backwards elimination, forward selection, stepwise etc.) have been suggested in the regression context and the Reed (1996) paper explores the use of these for the inclusion or removal of change points.

In this paper an alternative approach to model selection problems based on the *Bayes Information Criterion* (BIC) (Raftery, 1995; Kass and Raftery, 1995 and references therein) is employed for identifying change points in historical fire frequency.

The method is outlined in Section 3, following the formulation of the problem in Section 2. Two examples are given in Section 4, using time-since-fire map data for two areas of wilderness in the Rocky Mountain region of Western Canada.

2 Preliminaries.

A time-since-fire map records the time since the most recent fire at every geographical point in a study area. Often such a map is equivalent to a stand-age map since in many boreal regions stands originate only after a destructive fire. Since fires at different times may overlap spatially, for most fires in the past one cannot determine from a time-since-fire map, the exact extent of the fire, only place a lower bound on it. In fact it is only for recent fires that a time-since-fire map contains much information on spread. For a fire which occurred a long time ago there may be very little or even no evidence of it recorded in a time-since-fire map. In consequence in modelling historical fire frequency, it is difficult to explicitly incorporate spatial aspects of forest fires. The standard procedures used to date (Johnson and Gutsell, 1994) essentially ignore spatial factors, at least once a partition of the region into what are considered to be spatially homogeneous areas has been accomplished. For each homogeneous area, analysis is confined to the time-since-fire area distribution examples of which are displayed in Figs. 1 & 2. Also shown in these figures are plots of the cumulative area distribution of time-since-fire on a logarithmic scale. As in previous studies the analysis presented here will be based on the time-since-fire distribution for a presumed homogeneous area. However as in Reed *et al.* (1997) and Reed (1996) but unlike in most previous studies the fact that fires spread spatially will be acknowledged via the incorporation of overdispersion effects in this distribution

Assume that the time-since-fire distribution has classes of equal width (save for the oldest class). Let $A_1, A_2, \ldots, A_{m-1}$ denote the areas falling into classes with time since fire in the half open intervals $((j-1)T - jT]$ $(j = 1, 2, \ldots m-1)$; and let A_m denote the area in the open-ended "collector" class with time since fire greater than $(m-1)T$. In practice the resolution of dating fires is no finer than one year, so T will be a positive integer. Let

$$y_i = \frac{A_j}{\sum_{i=1}^{m} A_i}, \quad j = 1, \ldots, m \tag{1}$$

denote the proportional areas in the time-since-fire classes.

Suppose, to begin, that (possibly distinct) fire hazard rates[1] $\lambda^{(j)}$ prevailed in each of the time periods $j = 1, 2, \ldots, m$, where period j is defined as being between $(j-1)T$ and jT years ago $(j = 1, 2, \ldots, m-1)$ while period m is defined as more than $(m-1)T$ years ago. Because of the grouped nature of the data it is more convenient to deal with the the per-period conditional

[1] In this paper the term *fire hazard rate* will be used as in the statistical survival analysis (*i.e.* the instantaneous probability of fire per unit time, conditional on no previous fire). To foresters, fire hazard has a different meaning *viz.* the potential of fire based on fuel structure but not fuel moisture. Johnson and Gutsell (1994) refer to the fire hazard rate as used in this paper as *the hazard of burning* to distinguish it from the foresters' more customary usage. The forestry term equivalent to the fire hazard rate (or hazard of burning) is the *fire frequency* (defined in Johnson and Gutsell, 1994, Sec III.E as the probability of an element burning per unit time). The reciprocal of the hazard rate (or fire frequency) is known as the *fire cycle*. To statisticians, of course, this represents the expected time between fires if the current hazard rate prevails. Foresters and ecologists however define the fire cycle as "the time required to burn an area equal in size to the study area" (Johnson and Gutsell, 1994, Sec III.E), a concept based on equating the per annum probability of burning with the proportional annual area burned. For the most part estimation results in this paper concerning fire hazard rates (fire frequencies) will be expressed in terms of the fire cycle.

survival probabilities (complement of the discrete hazard)

$$q^{(j)} = e^{-\lambda^{(j)}T} \quad \text{for} \quad j = 1, \cdots, m-1 \tag{2}$$

The marginal probability, θ_j, that the forest at a particular site belongs to time-since-fire class j, for $j = 1, 2, \ldots, m-1$, is then simply the probability that there was a fire in time period j with no subsequent fire *i.e.*

$$\theta_j = (1 - q^{(j)}) \prod_{i=1}^{j-1} q^{(i)}, \tag{3}$$

while for time-since-fire class m the corresponding probability is

$$\theta_m = \prod_{i=1}^{m-1} q^{(i)}. \tag{4}$$

If fires ocurred in a spatially independent fashion, the distribution of areas in various time-since-fire classes would be multinomial with parameters $\theta_i, i = 1, \ldots, m$. To incorporate the effect of spatial correlation assume instead that the proportional areas $y_1, y_2, \ldots, y_m$ behave like proportions in an overdispersed multinomial model *i.e* that they have covariance matrix equal to that of the multinomial, multiplied by a (constant) *overdispersion parameter*, σ^2. One can thus contemplate a *quasi (log) likelihood function* (McCullagh & Nelder, 1989, Ch. 9) of the form

$$\begin{aligned} Q &= \frac{1}{\sigma^2} \sum_{j=1}^{m} y_j \log \theta_j \\ &= \frac{1}{\sigma^2} \sum_{j=1}^{m-1} [s_j \; ln \; q^{(j)} + y_j \; ln \; (1 - q^{(j)})]. \end{aligned} \tag{5}$$

where

$$s_j = \sum_{i=j+1}^{m} y_i, \tag{6}$$

is the cumulative proportional area of forest at least jT years old (for $j = 1, 2, \ldots, m-1$). The overdispersion parameter σ^2 will lie between 0 and 1 and its magnitude relect the degree of contagion in forest fires. Large values of σ^2 correspond to the situation where fires are likely to spread a great deal, and vice-versa. The limiting case $\sigma^2 \to 1$ corresponds to the situation in which all fires extend throughout the whole study area (one of the y_i equal to 1, the rest equal to zero); while $\sigma^2 \to 0$ corresponds to the artificial situation in which every point is independent, so that over the infinity of points, by the Law of Large Numbers, the proportions exhibit no randomness. While it is possible to use a more explicit contagious distribution model for $y_1, y_2, \ldots, y_m$, there seems to be little advantage in doing so. For example the Dirichlet distribution (Reed,1994) exhibits contagion, but since its covariance matrix is proportional to that of a multinomial distribution it can be adequately represented by the overdispersed multinomial. A similar situation prevails for the multivariate Polya-Eggenberger (beta-binomial) distribution, which can be used for time-since-fire sample data (Reed, 1997).

Before proceeding we note that the quasi-likelihood function (5) is identical to that which would have arisen if time had moved backward and succesively in each time period areas of forest had either burned (with marginal probability $1 - q^{(j)}$) or survived (with marginal probability $q^{(j)}$), with those that burned in time period j ending up in time-since-fire class j, (for $j = 1, 2, \ldots, m-1$) and with there being an overdispersion effect (with constant overdispersion factor σ^2) in the way units burn or survive. It is not difficult to show that the first and second order moments of the proportions y_i generated in this reverse-time fashion, coincide with those of the overdispersed multinomial distribution, which of course is not surprising given the equivalence of the two quasi-likelihoods. From the point of view of inference the important point is that, by the strong likelihood principle, we can consider the time-since-fire distribution of areas as arising in the reverse time fashion. Viewed like this the problem is closer to more familiar problems in survival analysis, where units survive until they fail (as opposed to surviving from a failure time until the present).

If each of the $q^{(j)}$ $(j = 1, 2, \ldots, m-1)$ is regarded as a free parameter to be fitted (saturated model) it is clear that the maximum likelihood estimates (MLEs) are $\hat{q}^{(j)} = \frac{s_j}{s_{j-1}}$, $j = 1, \ldots, m-1$. Substituting the MLEs into (5) gives the quasi-likelihood of the saturated model as

$$Q_s = \frac{1}{\sigma^2} \sum_{j=1}^{m-1} [s_j \log(\frac{s_j}{s_{j-1}}) + y_j \log(\frac{y_j}{s_{j-1}})] = \frac{1}{\sigma^2} \sum_{j=1}^{m} y_j \log y_j \qquad (7)$$

For any other model M with parameters Θ (*i.e* in which the probabilities $q^{(j)}$, $j = 1, \ldots, m-1$ are expressed in terms of the parameters Θ) the *scaled quasi-deviance* (see McCullagh & Nelder 1989, Sec. 9.2) is defined as

$$D_M = -2\sigma^2[Q(\hat{\Theta}) - Q_s] \qquad (8)$$

where $Q(\hat{\Theta})$ is the quasi-likelihood maximized over the parameters Θ of the model M. Inferences concerning the parameters Θ can be based on changes in scaled quasi-deviance divided by an estimate of the overdispersion parameter

$$\hat{D}_M = -2\frac{\sigma^2}{\hat{\sigma}^2}[Q(\hat{\Theta}) - Q_s] \qquad (9)$$

in the way discussed in McCullagh and Nelder (1989). The usual estimators for σ^2 are those based on the Pearson chi-squared statistic, $\hat{\sigma_P}^2$ and on the residual deviance, $\hat{\sigma_D}^2$.

For a model with k change points at specified times $p_1T < p_2T < \ldots < p_kT$ years ago, with intervening hazard rates λ_i (between $p_{i-1}T$ and p_iT years ago), it is easily shown (Reed,1996) that the MLEs of the hazard rates are $-\frac{1}{T}\log\hat{q}_i$ where

$$\hat{q}_i = \frac{\sum_{j=p_{i-1}+1}^{p_i} s_j}{\sum_{j=p_{i-1}+1}^{p_i} s_{j-1}} \tag{10}$$

(where p_{k+1} is defined as m and p_0 as 0) and that the scaled quasi-deviance (8) is

$$\hat{D}_k = \frac{2}{\hat{\sigma}^2}\left(\sum_{j=1}^{m} y_j \log y_j - \sum_{i=1}^{k+1}\left[\Big(\sum_{j=p_{i-1}+1}^{p_i} s_j\Big)\log\hat{q}_i + \Big(\sum_{j=p_{i-1}+1}^{p_i} y_j\Big)\log(1-\hat{q}_i)\right]\right) \tag{11}$$

While one could use these statistics to test for the significance of any of the change points, in practice in most situations it will not be known when changes occurred. One is thus faced with estimating the number and location of change points, along with the intervening hazard rates. To this end one can consider a hierarchy of models:

H_0: No change points (constant hazard rate),

H_1: One change point (separating two distinct hazard rates)

H_2: Two change point (separating three distinct hazard rates)

etc.

Under any model H_k there are $2k+1$ undetermined parameters (k change points separating $k+1$ hazard rates or parameters q_i). ML estimates of the change points can be found by comparing the likelihoods, maximized over the $k+1$ hazard parameters, for all $\binom{m-1}{k}$ possible choices of k change points.

In principle one could use a likelihood ratio test (LRT) to test H_k *vs.* H_l ($l > k$). However it is well-known that even for testing H_0 *vs.* H_1 (and when observations are real-valued exact failure times), the asymptotic chi-square approximation for the null distribution of the log-likelihood ratio statistic does not hold (Worsley, 1988; Henderson, 1990; Loader, 1991). Reed (1996) determined an approximation for computing P-values for testing H_0 *vs.* H_1, but the method does not readily extend for higher order nulls. Instead Reed used iterative procedures to add change points one at a time. However there are difficulties with the "forward selection" procedure, in that the MLEs of the change points of a higher model do not necessarily include those of a lower order model. Alternative procedures such as "backward elimination" and "stepwise" suffer from problems in determining the exact null distribution for tests for eliminating change points, even though for the data considered by Reed (1996) a somewhat *ad hoc* application of the backwards elimination procedure seemed to provide the best final model.

As has been well-recognized in other areas the significance testing paradigm does not lend itself very well to model selection problems such as this (see *e.g.* Raftery, 1995 and references therein). Rather than test one "null" model against another "alternative", one really wants to identify plausible models from a number of alternatives. In the following section it is shown how the Bayes' Information Criterion (BIC) can be used to this end.

3 Using the Bayes' Information Criterion to identify plausible models.

Suppose that *a priori* one assigns probabilities $P(H_k)$ to model H_k ($k = 0, 1, \ldots, K$, $\sum_{l=0}^{K} P(H_l) = 1$), where K is the greatest number of change points contemplated. After observing the data Y the posterior probability of model H_k is (from Bayes' Theorem)

$$P(H_k|Y) = \frac{P(Y|H_k))P(H_k)}{\sum_{l=0}^{K} P(Y|H_l))P(H_l)} \tag{12}$$

The probabilities $P(Y|H_l)$ are, in a full Bayesian formulation, obtained by integrating the likelihood under H_l with respect to a prior distribution for the $2l+1$ unknown parameters of that model. However a large sample approximation to $P(Y|H_l)$ can be obtained using a Laplace approximation for the appropriate integral (Schwarz, 1978; Tierney and Kadane, 1986). Raftery (1995) shows that to an $O(1)$ approximation

$$P(Y|H_l) \propto \exp(-\frac{1}{2}\text{BIC}_l) \tag{13}$$

where BIC_l is the Bayes' Information Criterion for model H_l. For n observations from model H_l with an overdispersion parameter σ^2 the BIC is

$$\text{BIC}_l = \frac{D_l}{\sigma^2} - (\text{df})_l \log(\frac{n}{\sigma^2}) \tag{14}$$

where D_l is the deviance (8) associated with model H_l and $(\mathrm{df})_l$ is the associated residual degrees of freedom. The reason for presence of σ^2 in the denominator of the logarithm is because it appears in the denominator of the (quasi) Fisher information of the overdispersed model. When σ^2 is unknown it can be estimated by $X^2/(n-p_l)$ where X^2 is the Pearson statistic for model H_l and p_l is the number of parameters estimated under H_l, leading to an an approximation to the BIC

$$\mathrm{BIC}_l \approx \hat{D}_l - (\mathrm{df})_l \log(\frac{n}{\hat{\sigma}^2}). \tag{15}$$

For a time-since-fire map the data is in the form of areas, not a sample of n observations, and so formula (15) does not directly apply. However one can conceive of the total study area $\sum_{j=1}^{m} A_j$ as being divided into a large number N of small sub-areas (or units) of equal size (in a grid for example). If N is sufficently large so that each small unit either burns or survives in its entirety through each period, one can think of the data as being generated in the reverse-time fashion, with units of forest facing successive overdispersed (survive or fail) Bernoulli trials in periods $1, 2, \ldots$ with marginal survival probabilites $q^{(j)}$ and marginal failure probabilities $1 - q^{(j)}$. The number of units facing the trial in period 1 would clearly be $N = Ns_0$; the number facing the trial in period 2 would be $N(1-y_1) = Ns_1$; the number in period 3 would be $N(1-y_1-y_2) = Ns_2$, *etc.* Thus the total number of units facing trials in periods $1, 2, \ldots, m-1$ would be $n = N\sum_{j=1}^{m-1} s_{j-1}$.

Rather than expressing the quasi-likelihood function, overdispersion parameter and deviances *etc.* in terms of the *proportions* y_j, they could be expressed in terms of the *counts* of units $x_j = Ny_j$ in the various age classes. An examination of (5) and (8) will reveal that the quasi-deviance for such count data would be N times that in the formula (8). However the estimate of the overdispersion parameter used in (9) would also be increased by a factor N, regardless of whether the residual deviance or the Pearson statistic is used. Thus statistics $\hat{D}_k$ of the form (11) would be the same regardless of whether proportions or counts of units were used. Furthermore since, for count data, both n and $\hat{\sigma}^2$ are proportional to N and since the residual degrees of freedom for model H_l are

$$(\mathrm{df})_l = (m-1) - (2l+1) = m - 2l - 2 \tag{16}$$

the BIC for model H_l can be expressed approximately as

$$\mathrm{BIC}_l \approx \hat{D}_l - (m-2l-2)\log\left(\frac{\sum_{j=1}^{m-1} s_{j-1}}{\hat{\sigma}^2}\right) \tag{17}$$

with $\hat{D}_l$ given by (11), and $\hat{\sigma}^2$ the Pearson estimate of the overdispersion parameter for proportional data (*i.e.* of σ^2 in model (5)).

Bayes' Theorem (12) can be used with the BICs (17) to obtain posterior probabilities for the various models, for any specified prior distribution. In particular for a uniform prior the posterior probabilities are

$$P(H_k|Y) = \frac{\exp(-\frac{1}{2}\mathrm{BIC}_k)}{\sum_{l=0}^{K}\exp(-\frac{1}{2}\mathrm{BIC}_l)} k = 1, \ldots, K \tag{18}$$

Clearly with this prior the model with the smallest BIC will be the single most plausible model.

One could contemplate other priors *e.g.* binomial probabilities based on the assumption that a change either did or did not occur between each time period, independently and with equal probability, or more conveniently as probabilities from a truncated Poisson distribution. However both of these priors involve the specification of an additional (hyper-) parameter.

In implementing the above procedure, there is the problem of specifying a maximum number of change points, K. While, in principle one could set $K = m - 1$, corresponding to different hazards in each period, there are difficulties with doing this. On the one hand it would lead to an considerable computational load, since under a given model H_k, the MLEs of the k change points are found by direct search. In addition to this, there is the problem of estimating the overdispersion parameter σ^2, which is customarily estimated under the largest model contemplated. If K is set too large there will be few degrees of freedom for estimating σ^2. In the examples presented in the following section a maximum of $K = 6$ change points was used. In neither example was the BIC minimum with either 5 or 6 change points and the relative values of the BICs did not change much when K was reduced from 6 to 5, giving some comfort to the assumption that the models $H_0, \ldots, H_6$ cover all realistic possibilities.

4 Applications.

In this section the methods described above are applied to fire map data from two studies of regions of boreal forest in the Canadian Rockies.

EXAMPLE 1. *Kananaskis River Watershed.* Johnson and Larsen (1991) present results of a fire history study of the 495 km^2 area of the Kananaskis watershed on the eastern side of the southern Rocky Mountains in Alberta, with a climate "transitional between plains and cordilleran types". Attempts by Johnson and Larsen to divide the map into spatial sub-units with distinct fire hazard rates were unsuccessful. The time-since-fire distribution for the whole study area is displayed in Fig. 1. There are $m = 40$ age classes of width $T = 10$ years. The lower panel (a plot of cumulative frequency against time since fire), suggests a number of possible change points (*e.g.* at 40, 60, 130, 230 and 280 years ago). Table 1 presents the MLEs of change points for models $H_0, \ldots, H_6$, along with the BICs and posterior probabilities

assuming a uniform prior on $H_0, \ldots, H_6$. The overdispersion parameter was estimated under H_6. The only plausible models appear to be H_2, H_3 and H_4, with H_3 being by far the most plausible. In fact the Bayes factor (Raftery, 1995) for H_3 against H_2 is 4.6, and for H_3 against H_4 it is 20.5. Since the change point at 4 (1940) and 6 (1920) appear as MLEs for all of these models one can conclude with a very high degree of certainty that there were indeed changes in fire frequency at around those times. Furthermore there is very strong support of an additional change at 23 (1750).

Of course by varying the prior probabilities of the various models, one can change the posteriors. However a fairly substantial skewness in the prior is required to shift the highest posterior probability from H_3. For example using as prior a Poisson distribution truncated at 6, with parameter λ, it is

Model	MLEs of change points	BIC	Posterior probability
H_0	-	-55.52	0.000
H_1	4	-149.06	0.004
H_2	4, 6	-156.35	0.171
H_3	4, 6, 24	-159.39	0.780
H_4	4, 6, 13, 23	-153.35	0.038
H_5	4, 6, 7, 13, 23	-149.42	0.005
H_6	4, 6, 7, 13, 19, 27	-136.71	0.000

Table 1: Maximum likelihood estimates of change points in various models; associated BICs; and posterior probabilities of the various models, assuming *a priori* that all seven models are equally probable (for Kananaskis Watershed time-since-fire data).

Epoch	Date	Fire Cycle (years)	
i		MLE	95% Con. Int.
1	1940 - 1980	6409	969 – 715,000
2	1920 - 1940	49	34 – 73
3	1750 - 1920	136	101 – 189
4	pre 1750	48	30 – 85

Table 2: Maximum likelihood estimates and 95% LR confidence intervals for the fire cycle in the four epochs between the three estimated change points in model H_3 (for Kananaskis Watershed time-since-fire data).

only when λ is less than about 0.6, that the posterior mode shifts from H_3 (to H_2). This requires the prior probability on H_3 to be less than 0.02. On the other hand it is only when λ is greater than about 55 that the posterior mode shifts from H_3 (to H_5). This requires the prior probability on H_3 to be less than 0.001. Thus one can conclude that the data contain strong support for the three change point model, with the estimated changes occurring around 1940, 1920 and 1750.

The ML estimates of the fire cycle (inverse of the hazard rate) and 95% likelihood ratio confidence intervals (Reed *et al.*, 1997) in the four epochs separated by the three identified change points are displayed in Table 2. The MLEs are also shown as line segments superimposed on the semi-log cumulative frequency plot in Fig.3 (top panel).

As one would expect, the confidence intervals for the fire cycle in adjacent epochs do not overlap. The estimated hazard for the post-1940 epoch is neglible. In fact less than one tenth of one percent of the whole study area burned in the forty year period 1940-1980.

Johnson and Larsen (1991) graphically identified a change point around 1730, and estimated the pre-1730 fire cycle at about 50 years, which agrees well with the above results. However they failed to identify the more recent change points identified above.

EXAMPLE 2. *Glacier National Park.*

Fig.2 presents data obtained by Johnson *et al.* (1990) from stand-origin maps for Glacier National Park (600 km^2 of forested land) in the Rocky Mountains of British Columbia. The vegetation is classified as of the Interior Wet Belt Forest type. The data were presented in 20-year age classes ($T = 20$, $m = 21$). Table 3 gives details of MLEs, BICs and posterior probabilities under a uniform prior for models $H_0, H_1, \ldots, H_6$.

The data provide support only for models with four or more change points, with the model H_4 standing out with a very large posterior probability. With a truncated Poisson prior the posterior mode does not shift up from H_3 until the Poisson parameter λ exceeds about 20 (at which point it shifts to H_5). However this requires the prior probability of H_5 to be about 0.72, and that of H_3 to be about 0.05. It is only by putting virtually all of the prior probability at zero (λ less than about 5 x 10^{-6}) that the posterior mode can be shifted downwards from H_3.

In consequence the four change point model seems by far the most plausible, with estimated change points at 2, 5, 10 and 16 *i.e.* around 1940, 1880, 1780 and 1660. There is some possibilty of a fifth change point, estimated at 1740. ML estimates and 95% likelihood ratio confidence intervals for the fire cycle (inverse of the hazard rate) in the five epochs separated by the four estimated change points under H_4 are displayed in Table 4. The MLEs are also shown as line segments superimposed on the semi-log cumulative frequency plot in Fig.3 (bottom panel).

It is worthy of note that a very similar model was identified for these data using backward elimination methods (Reed, 1996), the only difference

Model	MLEs of change points	BIC	Posterior probability
H_0	-	201.63	0.000
H_1	16	74.38	0.000
H_2	12, 16	42.79	0.000
H_3	5, 10, 16	28.08	0.000
H_4	2, 5, 10, 16	-9.27	0.797
H_5	2, 5, 10, 12, 16	-6.98	0.202
H_6	2, 5, 10, 12, 16,18	1.76	0.002

Table 3: Maximum likelihood estimates of change points in various models; associated BICs; and posterior probabilities of the various models, assuming *a priori* that all seven models are equally probable (for Glacier National Park time-since-fire data).

Epoch	Date	Fire Cycle (years)	
i		MLE	95% Con. Int.
1	1940 - 1980	1980	565 – 16700
2	1880 - 1940	156	40 – 181
3	1780 - 1880	1827	673 – 8102
4	1660 - 1780	151	106 – 224
5	pre 1660	25	17 – 42

Table 4: Maximum likelihood estimates and 95% LR confidence intervals for the fire cycles in the five epochs between the four estimated change points in model H_4 (for Glacier National Park time-since-fire data).

being the change point in the eighteenth century was estimated at 12 (1740) rather than 10 (1780). Furthermore the change points identified correlate fairly well with known facts of the history of the region. The abrupt decrease in fire hazard in the 18*th.* century corresponds to the onset of the Little Ice Age, when conditions cooler and wetter than those existing earlier prevailed - a fact established from major advances in glaciers in the area (Johnson *et al.*, 1990). The 1880's correspond to a period of European activity in the region. The Canadian Pacific Railroad, which passes through the study area, was constructed in this decade, and Glacier National Park was established in 1888. The marked drop in fire frequency occurring around 1940 coincides roughly with the onset of increased fire protection and suppression programs during and immediately after World War II.

The results for the Kananaskis watershed are broadly compatible with those for Glacier N.P. (see Fig. 4), with the Glacier fire frequency on the whole lower than that for Kananskis. Not too much should be made of the difference between the two regions in the earliest epochs, since the data on which these estimates are based are less reliable than those for more recent epochs. (Finney (1995) gives a number of reasons for this, including death of trees through causes other than fire, fragmentation of very old forest into patches smaller than the minimum mapping unit *etc.*). Like Glacier, Kananskis exhibited a big drop in fire frequency around 1940, and a high frequency before the middle of the eighteenth century. For Kananaskis the drop in frequency corresponding to the onset of the Little Ice Age occurred around 1750, somewhat earlier than that estimated for Glacier in the four change point model (1780). Note however that in the five change point model for Glacier there is an additional change estimated at 1740, which agrees very well with the Kananaskis results. The Little Ice Age period of low fire frequency seems to have ended sooner for Glacier (around 1880) than for Kananaskis (around 1920). While European incursion in the 1880s provides a tempting explanation for the abrupt increase for Glacier, European incursion into the Kananaskis region also occurred in this decade (Johnson and Larsen, 1991), and it is not until about 1920, that fire frequency there seems to have increased. Without going into historical details of European activities (mining, logging, tourism *etc.*) it is difficult to arrive at an explanation for the apparent differences in timing in the two regions.

It should be noted that Johnson *et al.*, (1990) report that many of the large fires in the Glacier region post-1880, were due to lightning rather than human activities, even though several large fires occurred in the Rogers Pass corridor in the period of railroad construction through the pass. However they also claim that "there has not been a decrease in the fire frequency since the establishment of Glacier National Park in 1888, despite a fire-supression policy". While the average fire frequency over the whole period 1880-1980 may not be any lower than that prevailing earlier, there is consisderable evidence to indicate the fire frequency increased around 1880 and subsequently decreased around 1940. A similar claim made by Johnson and Larsen (1991) for the Kananskis watershed is likewise not supported by the time-since-fire map data. The precise reasons for these changes cannot of course be established without a detailed analysis of what records remain of the particular fires involved. However *prima facie* the data are in agreement with the supposition that fire frequency at first increased with European activity, and subsequently decreased as fire prevention and suppression measures were introduced.

5 Concluding Remarks.

This paper has addressed the problem of identifying changes in historical forest fire frequency, and of estimating that frequency in epochs between identified changes. Statistically the problem of identifying change points is one of model selection and in the paper this problem has been addressed through use of the Bayes Information Criterion. On the two datasets used the method appears to perform very well, identifying models which provide a very good fit to the data, and also which correlate well with known facts in the history of the regions under study.

Usually in survival time analysis, one deals with data on the time elapsed from the start of a trial until a failure (or censoring) event. Time-since-fire data is different in that the elapsed period *starts* with a fire event and ends at the present. However for independent observations with an age-independent hazard rate, the likelihoods arising respectively from time-since-fire observations and from survival time observations, would be formally identical. With a time-since-fire map, observations are grouped and clearly not independent. However it is established in the paper that the quasi-likelihood of an overdispersed multinomial model for the grouped time-since-fire data is the same as that which would arise with time moving backwards and units of forest failing or surviving in an overdispersed binomial fashion, provided the overdispersion parameter is constant for each trial, and does not vary with failure probability (fire hazard). This assumption is made for the sake of tractability although of course it is quite possible that the contagion effect of fire is greater when the (marginal) hazard is greater. The equivalence of the quasi-likelihoods enables one to treat the data (*via* the likelihood principle) as if it had been generated in the reverse-time fashion. This is in turn enables the calculation of the correct logarithmic term to be subtracted from the deviance in calculating the BIC.

While the paper has employed the BIC for model selection, it falls far short of a full Bayesian analysis, which would require specification of a prior on the number and location of change points and on the intervening hazard rates. The BIC has been used here to identify plausible models. In each of the two examples presented, the subset of plausible models is quite small, with one single model standing out as much more plausible than the rest. Thus it has been possible to adopt a single model which can then be analysed by likelihood methods. It is quite possible that for other datasets, there could be many plausible models.

It may be possible to carry out a full Bayesian analysis using Markov Chain Monte Carlo jump diffusion sampling. Phillips and Smith (1996) describe this method and illustrate its use in identifing change points in a stationary Gaussian time-series. An alternative to BIC for model selection is the Akaike Information Criterion (AIC) suitably adapted to quasi-likelihoods for overdispersed data. Several such adjustments have been proposed (Anderson and Burnham, 1994; Hurvich and Tsai, 1995). However when applied to the two datasets used in the paper, the selected model under each criterion was always the largest model. It is well-known in other contexts that the AIC tends to overestimate the number of parameters needed (see *e.g.* the discussion in Kass and Raftery, 1995, Sec 8.3). In contrast to the AIC, the BIC performed very well, identifying parsimonious models with a very good fit to the data.

One of the datasets used in this paper (Glacier N.P.) has been analyzed using variants of step-wise and backwards elimination model selection procedures (Reed, 1996). The final model identified by backwards elimination was a four change-point model very similar to that identified in this paper as the most plausible using the BIC.

References.

Anderson, D. R. and K. P. Burnham. (1994). AIC model selection in overdispersed capture-recapture data. *Ecology* **75**:1780-1793.

Anonymous (1995). Biodiversity Guidebook (Forest practices code of British Columbia). British Columbia Ministry of Forests.

Finney, M. A. (1995). The missing tail and other considerations for the use of fire history models. *International J. of Wildland Fire* **5**:197-202.

Heinselman, M. L. (1973). Fire in the virgin forests of the Boundary Waters Canoe Area. *Quaternary Research* **3**:329-382.

Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika* **74**:835-843.

Hurvich, C. M. and C.-L. Tsai . (1995). Model selection for extended quasi-likelihood in small samples. *Biometrics* **51**:1077-1084.

Johnson, E. A., Fryer, G. I. and M. J. Heathcott. (1990). The influence of man on frequency of fire in the interior wet belt forest, British Columbia. *Journal of Ecology* **78**:403-412.

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

Johnson, E. A. and C. P. S. Larsen (1991). Climatically induced change in fire frequency in the southern Canadian Rockies. *Ecology* **72**:194-201.

Kass, R. E. and A. E. Raftery (1995). Bayes Factors. *J. American Statistical Association.* **90**:773-795

Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika* **78**:749-757.

McCullagh, P. and J. A. Nelder (1989) *Generalized Linear Models,* 2nd. Edition. London: Chapman and Hall.

Phillips, D. B. and A. F. M. Smith. (1996) *Bayesian model comparison via jump diffusions.* in Gilks, W. R. *et al.* (eds.) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Raftery, A. E. (1995) *Bayesian model selection in social research.* in Marsden, P. V. (ed.) *Sociological Methodology*, Cambridge, Mass.: Blackwells.

Reed, W. J. (1994). Estimating the historic probability of stand-replacement fire using the age-class distribution of old-growth forest. *Forest Science* **40**:104-119.

Reed, W. J. (1996). Determining changes in historical forest fire frequency from a time-since-fire map. Submitted to *J. Agricultural, Biological and Environmental Statistics.*

Reed, W. J. (1997). Estimating historical forest fire frequency from time-since-fire sample data. *I.M.A. Journal of Applications in Medicine and Biology.* **14**:71-83.

Reed, W. J., Larsen C. P. S., Johnson, E. A. and G. M. MacDonald (1997). Estimation of temporal variations in fire frequency from time-since-fire data. Submitted to *Forest Science.*

Schwarz, G (1978) Estimating the dimension of a model. *Annals of Statistics.* **6**:461-464.

Tierney, L and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *J. American Statistical Association.* **81**:82-86.

Worsley, K. J. (1988). Exact percentage points of the likelihood ratio test for a change-point hazard rate model. *Biometrics* **44**:259-263.

Acknowledgements.

I would like to thank Profs. Chris Larsen and Ed Johnson. Besides introducing me to change point problems in fire history research and supplying the data they both have made many useful comments and suggestions. However the conclusions and opinions expressed concerning the analysis of their data, are entirely my own. I also gratefully acknowledge the contribution of Prof. Adrian Raftery who suggested to me the use of the BIC for this problem.

Figure Captions

Fig.1. Time-since-fire area distribution for Kananaskis Watershed. The lower panel shows the cumulative area (log scale) exceeding a given time since fire.

Fig.2. Time-since-fire area distribution for Glacier National Park. The lower panel shows the cumulative area (log scale) exceeding a given time since fire.

Fig.3. Cumulative time-since-fire distributions. The line segments extend over the epochs between estimated change points in the most plausible models, and have slopes determined from the estimated constant epochal hazards.

Fig.4. Estimated hazards and change points for the most plausible models.

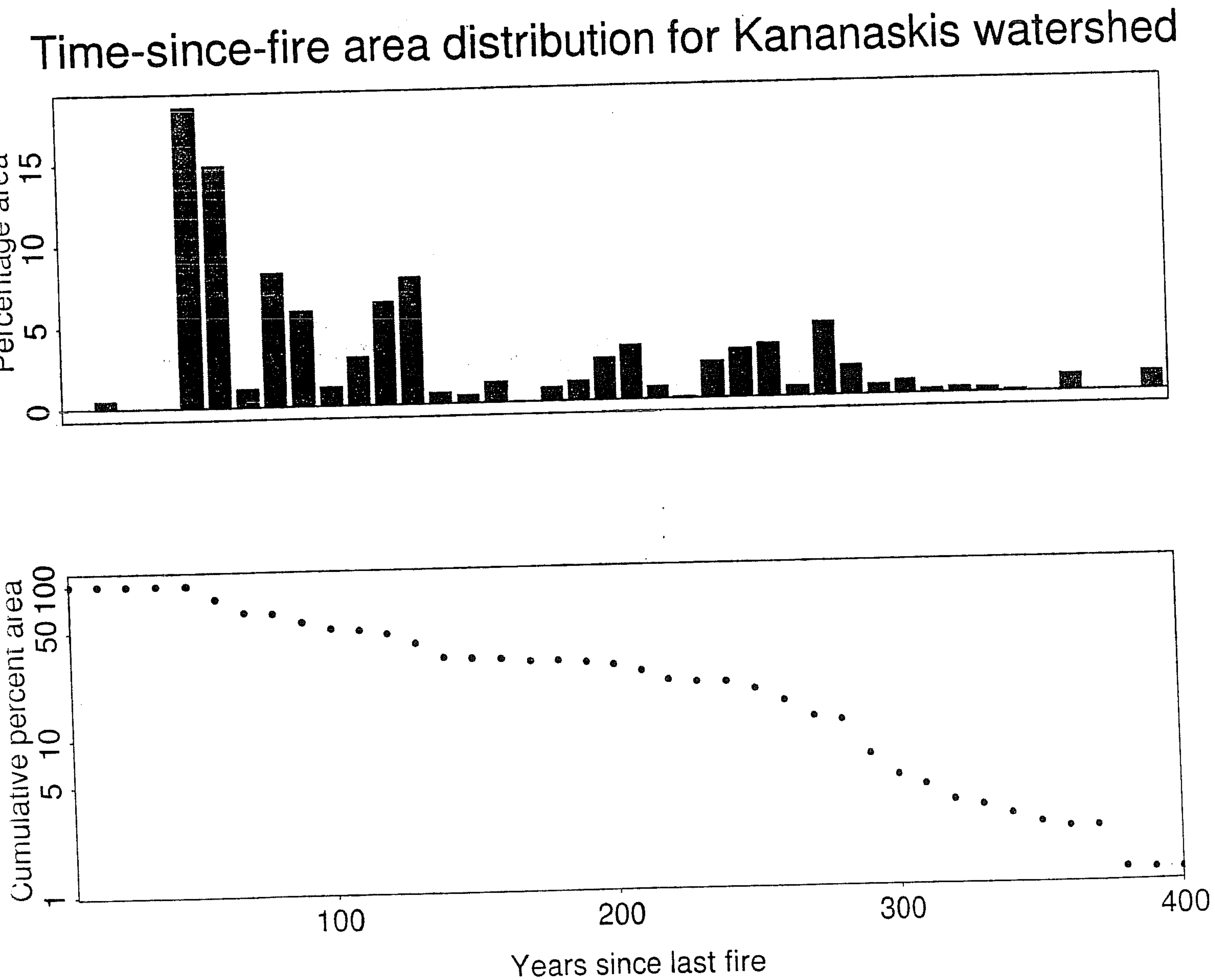
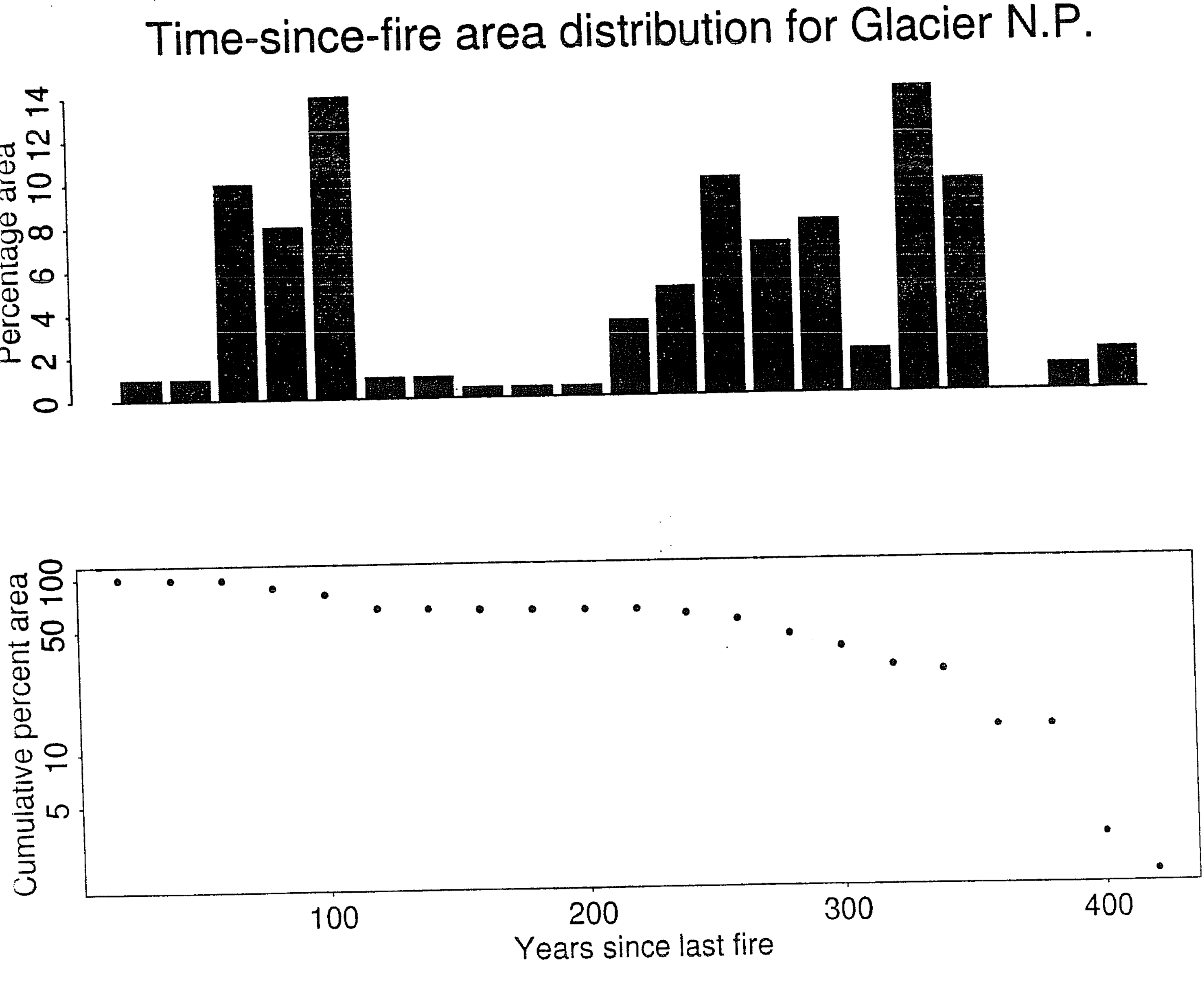
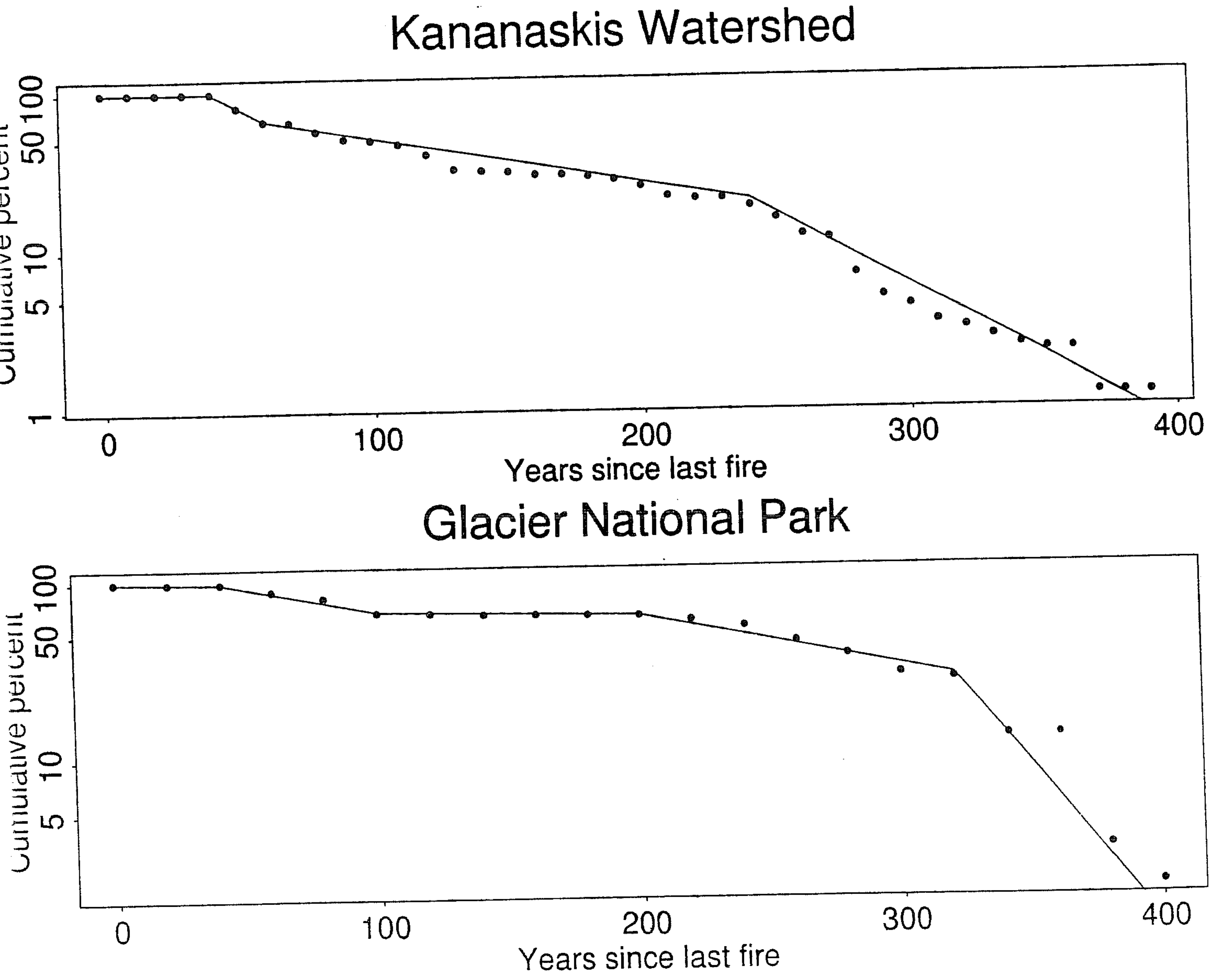

Time-since-fire area distribution for Kananaskis watershed
Percentage area
0
5
10
15
Cumulative percent area
1
5
10
50
100
100
200
300
400
Years since last fire

Figure 1:



Time-since-fire area distribution for Glacier N.P.
Percentage area
0 2 4 6 8 10 12 14
Cumulative percent area
5 10 50 100
100 200 300 400
Years since last fire

Figure 2:



Kananaskis Watershed
Cumulative percent
1
5
10
50
100
0
100
200
300
400
Years since last fire
Glacier National Park
Cumulative percent
5
10
50
100
0
100
200
300
400
Years since last fire

Figure 3:

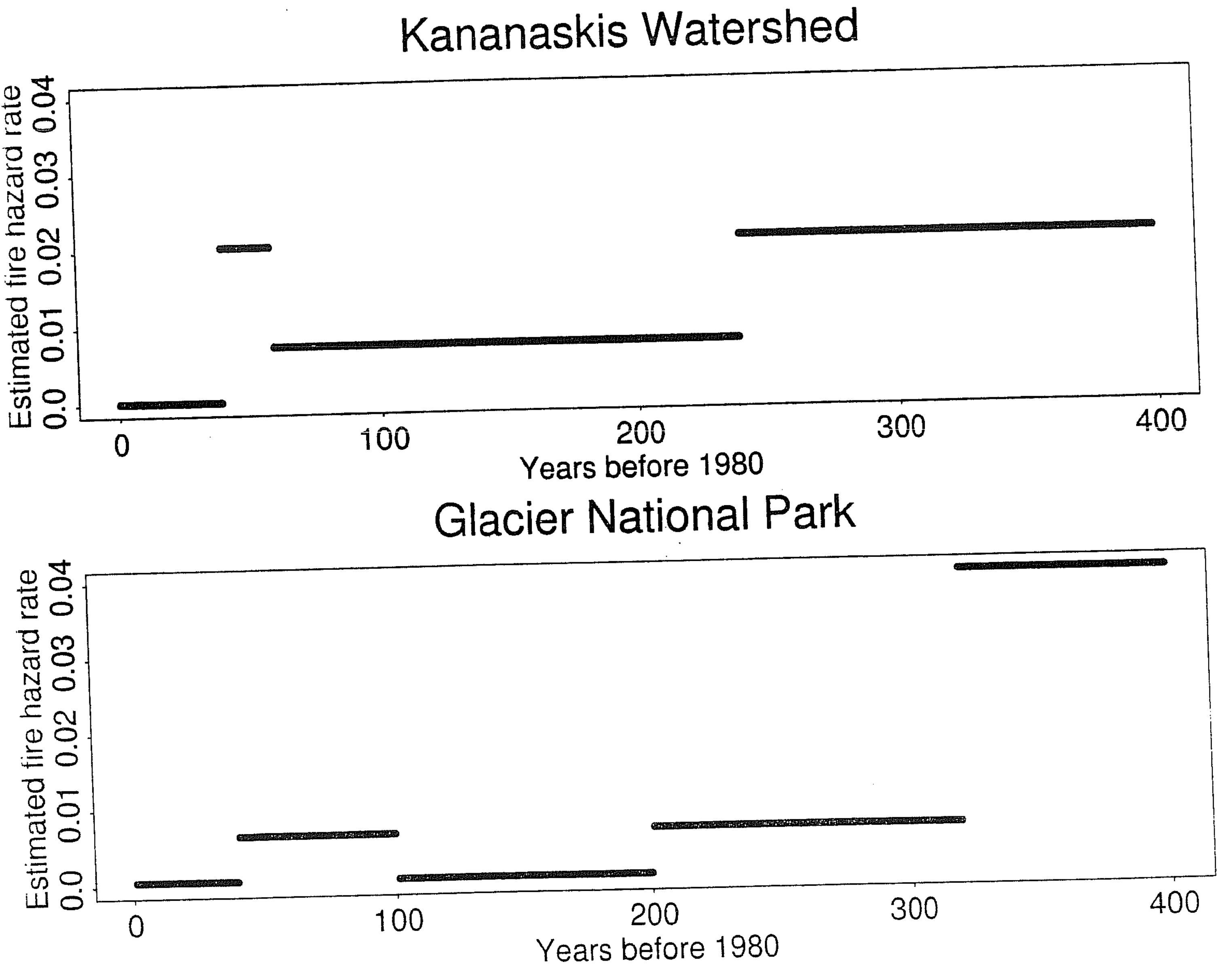

Kananaskis Watershed
Estimated fire hazard rate
0.0
0.01
0.02
0.03
0.04
0
100
200
300
400
Years before 1980
Glacier National Park
Estimated fire hazard rate
0.0
0.01
0.02
0.03
0.04
0
100
200
300
400
Years before 1980

Figure 4: