DETERMINING CHANGES IN HISTORICAL FOREST FIRE FREQUENCY FROM A TIME-SINCE-FIRE MAP

William J. Reed

DMS-741-IR June 1996

DETERMINING CHANGES IN HISTORICAL FOREST FIRE FREQUENCY FROM A TIME-SINCE-FIRE MAP.

William J. Reed*

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

June 1996

Abstract

The paper deals with methods of identifying change points for historical forest fire frequency (hazard rate) using data from a time-since-fire map and an overdispersed survival model with associated quasi-likelihood function. A method of obtaining an approximate P-value for testing the null hypothesis of no change points (homogeneous hazard), against the alternative of one change point, using the likelihood ratio statistic, is presented. It is based on use of the Poisson clumping heuristic for the maximum of a self-normalized Brownian bridge process. Iterative methods of determining multiple change points (analogous to methods of variable selection in regression) are discussed. Once change points have been identified, confidence intervals for the fire frequency in the intervening epochs can be obtained using likelihood ratio methods. Various procedures are applied to time-since-fire

*Research supported by NSERC grant OGP 7252.

map data for Glacier National Park and backwards elimination found to give the most plausible results. Three of the four change points identified correlate well with established historical processes or events.

Keywords: Fire frequency, fire history, time-since-fire map, quasi likelihood, change points, selection bias, Brownian bridge, Poisson clumping heuristic, model selection, backwards elimination.

Introduction

This paper deals with the subject of using time-since-fire maps to determine changes in historical forest-fire frequency, and to estimate fire frequencies prevailing between identified change points.

Time-since-fire maps (which identify the time since last fire at every point in a study area) originated with the work of Heinselman (1973), who used the cumulative distribution of areas in time-since-fire classes to estimate survivorship from wildfire. Van Wagner (1978), Johnson (1979) and Johnson and Van Wagner (1985) formalized fire frequency concepts, using the notions of survival analysis. Much of this material and many other aspects fire-frequency studies, including data collection, sampling design and analysis are described in the expository article of Johnson and Gutsell (1994). More recently Reed (1994) and Reed, Larsen, Johnson and MacDonald (1996) have introduced more sophisticated statistical methodology, based on likelihood methods for overdispersed models of the time-since-fire distribution, reflecting contagion effects in the spread of forest fire. Reed (1996) presents a parallel method for time-since-fire sample (as opposed to map) data. These papers however deal mainly with estimation and with testing for changes in fire frequency at

pre-specified potential change points (for example at the time of construction of a railroad, or at the onset of a systematic fire suppression program *etc.*), and not with the problem of using time-since-fire map or sample data to determine when changes in fire frequency occurred. This latter problem is often one of considerable interest to ecologists, who wish to identify factors affecting fire frequency.

The 'established' method of accomplishing a temporal partition of historical fire frequency is a graphical one based on identifying changes in slope in the semi-log plot of the cumulative area distribution of time-since-fire (Johnson and Larsen, 1991; Johnson and Gutsell, 1994), and then testing their significance. While the method has much to recommend it, and probably should always be incorporated in a preliminary data analysis, there are some statistical problems associated with it. One, described in Reed *et al.* (1996), relates to the recommended way of replotting two separate time-since-fire distributions for the epochs either side of a change point. An implicit assumption in the method (as described in Johnson and Gutsell (1994)) is that trees whose last fire occurred in the earlier epoch are not vulnerable to the risk of fire in the later epoch. In many cases this is an unreasonable assumption, and will lead to overestimating the fire frequency in the later epoch, since the survival of older trees through that later epoch is ignored. Nevertheless it is straightforward to adjust the method to include the vulnerability of trees to fire throughout their whole lifetimes. A second difficulty lies in testing the significance of a change point graphically identified. Because of

the fact of *selection bias* it is incorrect, from a statistical point of view, to test for the significance of such a change point in the same way as one would for a change point determined without reference to the data. By testing for a change point where there is an apparent change in slope in the semi-log plot of the cumulative distribution of time-since-fire, one is, in essence, choosing a change point for which the test statistic is large. Its null distribution will therefore not be the same as that at a potential change point chosen without reference to the data. One of the aims of this paper is to rectify this situation, and to obtain (approximate) tail probabilities of the null distribution of such a test statistic.

If in fact there is only one change point separating two epochs with different fire frequencies, the method provides a statistically sound way of identifying the change point and estimating the fire frequencies in the two epochs. However there is always the possibility that there is, in fact, more than one change point. Although, after identifying a first change point, one could proceed to examine each of the two resulting epochs for further change points using the same method, and thence continue iteratively (a procedure analogous to the *forward selection* procedure for determining which regressor variables to include in a regression analysis) there is no guarantee, when there are multiple change points present, that such a procedure will end up identifying them (just as forward selection in regression may end up with an inappropriate set of regressor variables). Alternatives are to use sequential methods of identifying change points analogous to the *backwards elimination*

and *stepwise* model selection procedures. While there are some distributional problems associated with the implementation of these methods (relating to significance levels for dropping a change point), in the paper a plausible *ad hoc* method of testing is used. The procedures are applied to time-since-fire map data for Glacier National Park in the Canadian Rocky Mountains (Johnson, Fryer and Heathcott, 1990). Backwards elimination seems to provide the best partition, partioning the past into five distinct fire-frequency epochs with change points around 1660, 1740, 1880 and 1940 CE.

Preliminaries.

Consider the situation in which a time-since-fire map is available *i.e.* the time of the last occurring fire is identified everywhere within the specified study region. Since the resolution for dating fires is at best one year, and often much larger (*e.g.* one decade), one can consider the data in the form of the areas belonging to disjoint classes defined by the time since last fire. Of course this ignores geographical aspects, but it will be assumed that first any spatial or topographic partitioning into component survival distributions has already taken place (Johnson and Gutsell 1994, Sec. III.F).

The precise meaning of time-since-fire depends on the nature and objective of the study (*e.g.* one can be dealing with stand-replacement fires, for which the age of the currently growing forest stand provides an indication of the time since fire; or with surface or understorey fires, which can be dated by fire scars on surviving trees, using dendrochronological methods, *etc.* For

the sake of brevity we shall refer to these 'time-since-last-fire' classes as *age classes*, corresponding to the former example above, although the methods are valid for other kinds of fires and methods of dating.

Before proceeding a clarification of terminology is in order. 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 herein 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.

Suppose that the age classes are all of equal width, T years say, save for the oldest which is open-ended. Specifically define age classes $1, 2, \ldots, m-1$

with time since last fire in the intervals $[0, T)$, $[T, 2T)$,....., $[(m-2)T, (m-1)T)$ respectively, and age class m with time since last fire greater than or equal $(m-1)T$. In practice the resolution of dating fires is no finer than one year, so T will be a positive integer. Denote the areas of forest in the respective age classes by $A_1, A_2, A_3, \ldots, A_m$ and let

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

denote the corresponding proportional areas.

Because of the grouped nature of the data the resolution for determining fire hazard rates is the width T of the age classes. Thus suppose that (possibly distinct) hazard rates $\lambda^{(j)}$ prevailed in each of the time periods $j = 1, 2, \ldots, m$, where period j is defined as being between $(j-1)T$ and $jT - 1$ years ago *i.e.* for period 1 (between the present (time 0) and $T - 1$ years ago) the hazard rate was $\lambda^{(1)}$; for period 2 (between T years ago and $2T - 1$ years ago) it was $\lambda^{(2)}$,... etc., while during period m ($(m-1)T$ years ago and earlier) it was $\lambda^{(m)}$.

It is more convenient, because of the grouped nature of the data, to deal with the per-period survival probabilities

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

rather than the hazard rates.

The marginal probability that the time since fire at a particular site belongs to age 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 age class m the corresponding probability is

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

Forest fires spread, and in consequence there will be spatial autocorrelation in time-since-fire observations. One way of modelling this (Reed, 1994) is to assume that the proportional areas $y_1, y_2, \ldots, y_m$ in the various age classes follow a Dirichlet distribution

$$g(y_1, y_2, \ldots, y_m) = \frac{\Gamma(1/\rho)}{\prod_{i=1}^{m} \Gamma(\theta_i/\rho)} y_1^{(\theta_1/\rho - 1)} y_2^{(\theta_2/\rho - 1)} \cdots y_m^{(\theta_m/\rho - 1)} \tag{5}$$

on the simplex $y_1 + y_2 + \cdots + y_m = 1,\ 0 \leq y_i \leq 1$, where $\Gamma(.)$ is the usual gamma function. The following properties of the distribution (5) can easily be derived (see e.g. Johnson and Kotz, 1972).

$$\mathrm{E}(y_i) = \theta_i; \quad \mathrm{var}(y_i) = \theta_i(1 - \theta_i)\frac{\rho}{1+\rho}; \quad \mathrm{cov}(y_i, y_j) = -\theta_i\theta_j\frac{\rho}{1+\rho} \tag{6}$$

Thus under the Dirichlet model the proportional areas $y_1, y_2, \ldots, y_m$ behave like proportions in an overdispersed multinomial model. One can thus contemplate a *quasi (log) likelihood function* (McCullagh & Nelder, 1989, Ch. 9) of the form

$$\begin{aligned} Q &= \frac{1}{\sigma^2} \sum_{j=1}^{m} y_j \log \theta_j \\ &= \frac{1}{\sigma^2} \sum_{j=1}^{m} [s_j \; ln \; q^{(j)} \; + \; y_j \; ln \; (1 - q^{(j)})]. \end{aligned} \tag{7}$$

where

$$s_j = \sum_{i=j+1}^{m} y_i, \tag{8}$$

is the cumulative proportional area of forest at least jT years old (for $j = 1, 2, \ldots, m-1$), $s_m = 0$ and σ^2 is an overdispersion parameter.

While this quasi-likelihood was suggested from the Dirichlet model, it is not confined to it. In other words (7) can be thought of as a quasi-likelihood arising from proportional areas $y_1, y_2, \ldots, y_m$ from a whole class of over-dispersed multinomial-type distributions (McCullagh & Nelder, 1989, Sec. 5.4). Thus it provides a more general model formulation than that based explicitly on the Dirichlet distribution. Furthermore, unlike with a Dirichlet likelihood, empty age classes cause no difficulty with the quasi-likelihood formulation.

Before proceeding we note that the quasi-likelihood function (7) 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 age-class j, (for $j = 1, 2, \ldots, m$) and with there being an overdispersion effect in the way units burn or survive. Thus by the strong likelihood principle (*e.g.* Cox and Hinkley, 1974) we can consider, without loss as far as inference is concerned, the age class distribution of areas as arising in this way. 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).

The overdispersion parameter σ^2 represents the degree of 'contagion' in the fire process, with large values of σ^2 corresponding to situations in which fires tend to be of large extent (strong dependence between proximate sites and consequent large spatial autocorrelation of time-since-fire observations) and small values corresponding to small fires and less autocorrelation.

If each of the $q^{(j)}$ $(j = 1, 2, \ldots, m)$ is regarded as a free parameter to be fitted (saturated model) it is clear that the maximum likelihood estimates (MLEs) are $\hat{q}^{(j)} = \frac{s_j}{s_{j-1}}$, $j = 1, \ldots, m-1$, while $\hat{q}^{(m)} = 0$. Substituting these MLEs into (7) gives the quasi-likelihood of the saturated model as

$$Q_s = \frac{1}{\sigma^2} \sum_{j=1}^{m-1} [s_j \log(\frac{s_j}{s_{j-1}}) + y_j \log(\frac{y_j}{s_{j-1}})] \tag{9}$$

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

$$D_M = -2\sigma^2[Q(\hat{\Theta}) - Q_s] \tag{10}$$

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 (the log quasi likelihood ratio), in the way discussed in McCullagh and Nelder (1989).

Testing for the significance of a pre-specified change point.

Suppose for reasons exogenous to the data, the possibility of a change in the fire hazard rate pT years ago is suspected (e.g. at the time of the first European intervention; or of the construction of a highway or railroad; or of the commencement of a fire suppression regime etc.), and one wishes to test whether the putative change point p is significant or not. To address this question one can test

$$\begin{aligned} H_0 \ : \quad & q^{(1)} = q^{(2)} = \ \ldots = q^{(m)} \ \ (= q_0, \ \text{say}) \quad \textit{vs.} \\ H_p \ : \quad & q^{(1)} = q^{(2)} = \ \ldots = q^{(p)} \ \ (= q_1, \ \text{say}) \quad \text{and} \\ & q^{(p+1)} = q^{(p+2)} = \ \ldots = q^{(m)} \ \ (= q_2, \ \text{say}) \end{aligned}$$

This problem is discussed in Reed *et al.* (1996).

Under H_0 the MLE of q_0 is

$$\hat{q}_0 \ = \ \frac{\sum_{j=1}^{m-1} s_j}{\sum_{j=1}^{m-1} s_{j-1}} \tag{11}$$

while under H_p the MLEs of q_1 and q_2 are

$$\hat{q}_1 \ = \ \frac{\sum_{j=1}^{p} s_j}{\sum_{j=1}^{p} s_{j-1}} \quad \text{and} \quad \hat{q}_2 \ = \ \frac{\sum_{j=p+1}^{m-1} s_j}{\sum_{j=p+1}^{m-1} s_{j-1}} \tag{12}$$

Using these values in (7) the scaled deviances under H_0 and H_p can be determined, leading to the (quasi) likelihood ratio (LR) statistic

$$R_p = \frac{\nabla D}{\hat{\sigma}_P^2} = 2\frac{\sigma^2}{\hat{\sigma}_P^2}[Q(\hat{\Theta}_1) - Q(\hat{\Theta}_0)] \tag{13}$$

where $\hat{\Theta}_1 = (\hat{q}_1, \hat{q}_2)'$ (under H_p) and $\hat{\Theta}_0 = \hat{q}_0$ (under H_0), and $\hat{\sigma}_P^2$ is the Pearson estimate[1] of the overdispersion parameter

$$\hat{\sigma}_P^2 = \frac{1}{m-3} \sum_{i=1}^{m} \frac{(y_i - \hat{\theta}_i)^2}{\hat{\theta}_i}, \tag{14}$$

with the fitted values $\hat{\theta}_i$ calculated under H_p (*i.e.* using estimates $\hat{q}_1$ and $\hat{q}_2$). Specifically (using (12)) the likelihood ratio statistic is

$$\begin{aligned} R_p &= \frac{2}{\hat{\sigma}_P^2}\Big(\sum_{j=1}^{p}[s_j \log\frac{\hat{q}_1}{\hat{q}_0} + y_j \log\frac{1-\hat{q}_1}{1-\hat{q}_0}] \\ &\quad + \sum_{j=p+1}^{m-1}[s_j \log\frac{\hat{q}_2}{\hat{q}_0} + y_j \log\frac{1-\hat{q}_2}{1-\hat{q}_0}]\Big) \end{aligned} \tag{15}$$

For a large study area (many area units) the null distribution of R_p should be approximately $F_{1,(m-3)}$, enabling the computation of an approximate P-value for testing H_0 *vs.* H_p.

Testing for the significance of a change point determined from the data.

The test in the previous section is not appropriate for a change point suggested from exploratory data analysis because of the problem of selection bias. If, for example, one identified a possible change point as one at which there were an apparent change in slope in the semi-log plot of cumulative area, one would, in effect, be choosing as a change point one at which the LR statistic was large. Thus because of this selection of the change point to

[1] Alternatively one can estimate σ^2 by the residual deviance divided by its degrees of freedom. In all numerical examples in this paper both estimates were calculated. They were always very close, and resulted in essentially the same conclusions. In the following reference is only made to the Pearson estimate.

test, under the null hypothesis of no changes the test statistic R_p would not follow an $F_{1,(m-3)}$ distribution but rather a distribution shifted to the right of $F_{1,(m-3)}$. In consequence the computed P-value would be too small, and too many non-existent change points would be found as significant.

To take the selection effect into account one can consider as test statistic a scaled version of the largest (over all possible change points) R_p statistic,

$$R^\star = \left(\frac{m-4}{m-3}\right)\max\{R_p; p = 1, 2, \ldots, m-2\}. \tag{16}$$

where R_p is the LR statistic for testing H_0 *vs.* H_p. Note that the possibility of a change point at $p = m-1$ is not included. In this case the estimate $\hat{q}_2$ in (12) would be zero, and that of $\hat{q}_1$ would be the same as $\hat{q}_0$ in (11). This corresponds to estimating the hazard rate as infinity at all times earlier than $(m-1)T$ years ago. A similar phenomenon was identified by Matthews & Farewell (1985) for the change point problem with ungrouped survival time data, leading to an unbounded likelihood. Although for our problem, because of the grouping into a collector class, the likelihood remains bounded when there is a change point at $m-1$, the LR statistic for such a change point is always zero, and it is impossible to detect a change in hazard at this time.

The statistic $R^\star$ is the LR statistic for testing

H_0: No change points (constant hazard rate) *vs.*

H_A: One change point (separating two distinct hazard rates) between T and $(m-2)T$ years ago.

The dimensionality of the parameter space under H_A is three (two hazard

rates and a change point) while under H_0 it is one (a single hazard rate). Under H_A the estimate of σ^2 has $m-4$ degrees of freedom while under H_p, for any fixed p, it has $m-3$. This is the reason for the inclusion of the factor $\left(\frac{m-4}{m-3}\right)$ in the definition of $R^\star$. Even though the difference in the dimensionality of the parameter space under H_A and H_0 is two because of the fact that the parameter space is not a continuum (p can assume only discrete values) the standard asymptotics do not hold – the null distribution of $R^\star$ is not asymptotically χ^2_2. This fact is well-known from change point studies of survival time data (see e.g. Worsley, 1988; Henderson, 1990; Loader, 1991)).

The change point p which maximizes R_p over $p = 1, \ldots, m-2$ will maximize its numerator

$$\nabla D = -2(\sum_{j=1}^{p}[s_j \log \frac{\hat{q}_1}{\hat{q}_0} + y_j \log \frac{1-\hat{q}_1}{1-\hat{q}_0}] + \sum_{j=p+1}^{m-1}[s_j \log \frac{\hat{q}_2}{\hat{q}_0} + y_j \log \frac{1-\hat{q}_2}{1-\hat{q}_0}]) \quad (17)$$

(since the denominator $\hat{\sigma}^2_P$ decreases with the residual deviance). One can show (see Appendix) by considering the data as arising from successive fire trials in reverse time, as discussed above, that for a given p under H_0 (with $q_1 = q_2 = q_0$)

$$\frac{\nabla D}{\sigma^2} = \frac{1}{N_1+N_2}\left[\sqrt{N_2}\frac{X_1 - N_1 q_0}{\sqrt{N_1\sigma^2 q_0(1-q_0)}} - \sqrt{N_1}\frac{X_2 - N_2 q_0}{\sqrt{N_2\sigma^2 q_0(1-q_0)}}\right]^2 + o_P(1) \quad (18)$$

where $o_P(1)$ indicates random terms for which the probability of exceeding 0 in absolute value goes to zero as $N_1, N_2 \to \infty$; N_1 and N_2 are the numbers of area units facing fire trials between 0 and pT years ago and greater than pT years ago, respectively, and X_1 and X_2 are the corresponding numbers

surviving those fire trials. Specifically

$$N_1 = \sum_{j=1}^{p} s_{j-1} \quad X_1 = \sum_{j=1}^{p} s_j \tag{19}$$

and

$$N_2 = \sum_{j=p+1}^{m-1} s_{j-1} \quad X_2 = \sum_{j=p+1}^{m-1} s_j. \tag{20}$$

For a fixed p it is clear from the Central Limit theorem for Bernoulli trials, that the asymptotic distribution of $\nabla D/\sigma^2$ is χ^2_1, a result used in determining the asymptotic distribution of the LR statistic for testing the significance of a pre-specified change point.

Another way of approaching the asymptotics is to consider the diffusion approximation for large N_1, N_2. It can be shown (see Appendix) that asymptotically

$$\frac{\nabla D}{\sigma^2} \sim \frac{1}{\tau(1-\tau)} \left(B(\tau) - \tau B(1)\right)^2 = \left(\tilde{B}(\tau)\right)^2 \tag{21}$$

where $B(t)$ is a *standard Brownian motion*, $\tau = \frac{N_1}{N_1+N_2}$ and $\tilde{B}(t)$ is a *self-normalized Brownian bridge*, i.e a standard Brownian bridge normalized by dividing by its standard deviation. For a fixed p the asymptotic χ^2_1 distribution of $\nabla D/\sigma^2$ is again obvious from the fact that the marginal distribution of $\tilde{B}(t)$ is standard normal.

Thus to find the distribution of the maximum of $\nabla D/\sigma^2$ we can consider the distribution of

$$M = \max_{t \in \{a_1, a_2, \ldots, a_{m-2}\}} [\tilde{B}(t)]^2 \tag{22}$$

where

$$a_p = \frac{\sum_{j=1}^{p} s_{j-1}}{\sum_{j=1}^{m-1} s_{j-1}} \quad \text{for} \quad p = 1, 2, \ldots, m-2 \tag{23}$$

correspond to the possible change points. An approximation to the tail of the distribution of the random variable (22) can be found using the *Poisson clumping heuristic* (PCH) (Aldous, 1989). In the Appendix it is shown that

$$\Pr(M \geq x|H_0) \approx 2\left(1 - \exp\left[-\Phi^c(\sqrt{x})\sum_{p=1}^{m-2}\frac{1}{\mu_x(p)}\right]\right) \tag{24}$$

where Φ^c is the complement of the c.d.f of a standard normal r.v and $\mu_x(p)$ is the number of points a_i (defined in (23)) in the interval

$$a_p \pm a_p(1-a_p)\frac{\Phi^c(\sqrt{x})}{\sqrt{x}\phi(\sqrt{x})} \tag{25}$$

where ϕ is the p.d.f of a standard normal r.v. This PCH approximation appears to be very good (see Appendix) for small values of m (few age classes) for which the exact distribution of M can be calculated using low dimensional Gaussian integrals. For larger values of m the performance of the approximation can be determined by Monte Carlo methods. Again, it appears that the PCH provides good approximations to the tail probabilities, although will tend to overestimate probabilities for values of x not in the right hand tail. In contrast the P-values computed from a χ^2_2 distribution appear always to be too small.

The distribution of the LR statistic [2] $R^\star = \max\left(\frac{\nabla D/\sigma^2}{\hat{\sigma}^2/\sigma^2}\right) = \frac{M}{\hat{\sigma}^2/\sigma^2}$ involves variation from both M and $\hat{\sigma}^2$. While it seems difficult to determine precisely the asymptotic distribution of $R^\star$, one can include a correction term in the formula (24), to account for the variation in $\hat{\sigma}^2$ which improves the

[2] By considering $\hat{\sigma}^2$ as the Pearson estimate of σ^2 with divisor and degrees of freedom $m-4$, it is not necessary to include the factor $\frac{m-4}{m-3}$ which occurs in (16).

approximation of the P-value in the tail. The correction is of the form

$$\frac{\sqrt{x}\ \phi(\sqrt{x})\ \Sigma_x\ \exp\left[-\Phi^c(\sqrt{x})\Sigma_x\right]}{2(m-4)}\ \left[1+x-\sqrt{x}\ \phi(\sqrt{x})\ \Sigma_x\right]. \qquad (26)$$

where

$$\Sigma_x = \sum_{p=1}^{m-2} \frac{1}{\mu_x(p)} \qquad (27)$$

The derivation is given in the Appendix, along with a comparison of the approximate P-values and those computed from an $F_{2,m-4}$ distribution with values obtained by Monte Carlo simulation. The simulations indicate that for the most part, the PCH approximations err on the side of conservatism, while those based on the F-distribution err in the other direction. A safe compromise seems to be to regard the values computed by PCH and F procedures as providing approximate upper and lower bounds on the P-value.

EXAMPLE. *Glacier National Park.*

Fig.1 presents data obtained by Johnson *et al.* (1990) from stand-origin maps for Glacier National Park (600 km^2 of forested land) in British Columbia. The data were presented in 20-year age classes ($T = 20$, $m = 21$). Johnson et al. identified, graphically, from the semi-log plot of cumulative area (Fig. 1, lower panel) a change point around 1760 CE (i.e. at $p = 11$). The R_p statistics for all possible single change points are presented in Table 1 (second row). The maximum value is 14.4 at $p^\star = 16$, with $R^\star = 13.6$. The P-value computed from the PCH approximation ((24) with correction (26)) is 0.018 (and P = .007 from an $F_{2,17}$ distribution). Thus we identify a significant

change at $p = 16$ and partition the study period into two epochs: from 1660 - 1980 CE (corresponding to $j = 1, \ldots, 16$), and before 1660 CE (corresponding to $j = 17, \ldots, 21$). It remains to be determined whether either of these epochs can be further divided.

Identifying further change points.

If the ML estimated change point $p^\star$ is found to be significant, one can continue to look for further change points in each of the two epochs identified by the establishment of the first change point, starting with the earlier epoch ($j = p^\star + 1, \ldots, m$). The procedure is essentially the same, only now the LR statistic (16) involves the maximum change in deviance (10) between a model with one change point (at the previously established $p^\star$) and one with two change points (at $p^\star$ and one new one). In using the PCH approximation the number of potential change points (23) changes as do the degrees of freedom in the estimate $\hat{\sigma}_P^2$. If a further change point is identified as significant one can continue investigating the newly identified epochs for further possible partitioning.

EXAMPLE. *Glacier National Park (continued).*

After making the identification of a change point at $p^\star = 16$ subsequent investigation of the earlier epoch ($j = 17, \ldots, 21$) indicated no significant change point, while the LR test statistic for a change point in the more recent epoch $j = 1, \ldots, 16$ has a value 6.36 corresponding to a second change point at $p = 12$ (see Table 1, bottom row). The corresponding P-value is 0.38

computed from the PCH (and 0.09 from an F distribution). Regarding the true P-value as being bounded by these extremes one could conclude that there is no evidence of a second change point.

However plots of (Pearson, deviance and Anscombe) residuals for the two epoch (single change point) model are not totally satisfactory in that residuals for periods $j = 6, \ldots, 10$ are all negative, and those for periods $j = 11, \ldots, 15$ all positive. In fact a runs test yields a P-value of 0.05, suggesting in fact that there is some indication of model lack of fit. Adding a second change point (at $p = 12$) makes the runs test no longer significant (P= .30), breaking up run $j = 11, \ldots, 15$ of positive residuals (but not the run $j = 6, \ldots, 10$ of negative residuals). However adding the second change point makes the run $j = 6, \ldots, 10$ no longer appear untoward to visual inspection. Furthermore adding further change points (at 5 and 2) does not break up the run. Thus there seems to be some grounds for adding a second change point at $p = 12$ or 1740 CE.

It should be recognized that there could be some problems associated with using the procedure described above iteratively to identify multiple change points. For example consider the situation in which there really are two change points (at p_1 and p_2 say). It is possible that the LR statistic for testing for a single change point as opposed to the null hypothesis of no change points would result from an MLE of the single change point at some value $p^\star$ between p_1 and p_2. This means that the two change point

model identified by the iterative procedure must have one change point at $p^\star$ even though the MLEs in a two change point model could be at p_1 and p_2. Unless there is a mechanism for dropping change points already identified the iterative method based on partitioning established epochs will always include in the final model the first change point identified. In fact if there are multiple change points the LR test on the first iteration may be of little meaning. With it one tests a single change point model against a no change point model, when in fact both may be very poor representations of reality. In consequence a significant P-value may have little meaning in such a situation.

As should be apparent from the above comments, the problem of identifying change points is analogous to the model selection problem of deciding which of a set of candidate regressor variables should be included in a regression model. The iterative procedure discussed above is analogous to a forward selection procedure. A possible improvement would be to develop it into something closer to a *stepwise* procedure in which change points could be dropped as well as added. Another possibility is to consider a *backwards elimination* procedure (see below). Yet another approach is to use the Bayes Information Criterion as a method of model selection (see Raftery, 1995) – a method described and employed in a forthcoming paper.

Backward elimination and stepwise procedures.

To initiate a backwards elimination procedure one could start with change points at all points where they appear feasible (apparent changes in slope in

the semi-log cumulative area plot). One would then check the significance of the one with the smallest LR statistic. If it were not significant it would be removed, and the procedure iterated again with one fewer change point; if it were significant it would be retained and the procedure would stop. The difficulty in implementing such a procedure however lies in establishing the appropriate reference distribution for determining significance.

The approach followed here is somewhat *ad hoc*. If initially change points are identified at $p_1^\star, p_2^\star, \ldots, p_r^\star$ and the one with the smallest LR statistic (for testing the null of $r-1$ change points at the remaining $p_j^\star$ against the alternative of r change points) is $p_u^\star$, say, we test its significance by testing for no change point on $p_{u-1}^\star + 1, \ldots, p_{u+1}^\star$ using the PCH approximation for the distribution of the maximum LR statistic value in the range $p_{u-1}^\star + 1, \ldots, p_{u+1}^\star$. If it is significant it is retained and the procedure stops with r change points; if not it is removed and the whole procedure re-iterated with $r-1$ change points.

A stepwise procedure would work exactly as in regression. After the inclusion of a new change point at any iteration, those already included would be re-examined for significance as in the backward elimination procedure above.

EXAMPLE. *Glacier National Park (continued).*

Firstly for backwards elimination, possible change points at $p = 2, 5, 12$ and 16 were identified from the semi-log plot in Fig. 2, dividing the study

period into 5 epochs. The corresponding LR statistics for their removal were 11.68, 11.04, 14.96 and 11.54. Thus the smallest corresponds to the change point at 5. Using the PCH approximation yields a P-value of 0.03 (for testing the hypothesis of no change point between 2 and 12 - *i.e.* change points only at 2, 12 and 16) and a similar P-value (0.02) obtained from an $F_{2,11}$ distribution[3]. Regarding these as bounds on the true P-value leads to the conclusion that there is evidence of a change point between 2 and 12 (when one at 16 is also included). Thus the change point at 5 is retained, along with the three others. Residual plots for this model gave no indication of lack of fit. Thus the final model by backwards elimination includes 4 change points (at around 1660, 1740, 1880 and 1940 CE) separating 5 epochs.

Since the second change point (at 12) by forward selection was not truly significant, the stepwise procedure terminates in the same place as forward selection (with one change point at 16). However residual analysis lends some support to the inclusion of the second change point at 12. If this point is included the LR statistic for the other change point (at 16) is 7.05. The bounds on the P-value are 0.12 (PCH approximation) and 0.05 (from F distribution). Depending on the significance level employed one could either argue for the removal of the change point 16 or its retention. If it is removed one is left with a model with a single change point at 12; no addition of any further change point is significant. Furthermore this model yields satisfactory

[3]The denominator degrees of freedom are those used in computing $\hat{\sigma}_P^2$ – here $m-1-9 = 11$, the nine corresponding to the 4 change points and 5 hazard rates estimated from the data

residual plots (however with larger residuals than the 4 change point model). This is in contrast to the model with a single change point at 16. If the change point at 16 is retained one can seek to add further change points. However none are significant, the largest LR statistic being 3.65 for a change point at 5. Thus the conclusion from the stepwise procedure is either: a single change point at 12 (around 1740 CE) dividing the study period into two epochs; or two change points at 12 and 16 (1740 and 1660 CE) resulting in three distinct epochs.

The results of the the stepwise and forward selection procedures are quite different to those of backwards elimination. The latter suggest that unless the effects of the other three change points are first removed, that of the fourth will not show up. This may be because the estimated hazard rate for the five epochs (see below) alternate as one moves backwards in time – a low hazard for Epoch 1, followed by a higher one in Epoch 2, a lower one in Epoch 3 *etc.*. The stepwise and forward selection procedures appear to be poor at picking up such a pattern of changes. Once change points at 12 and 16 are in the model, these procedures do not allow the inclusion of a single extra change point to partition the pattern of low-high-low hazards over the period $1, 2, \ldots, 12$. This is a weakness in the procedures, which possibly could be ameliorated by allowing change points to enter two at a time. We do not investigate this further here. Instead we put more faith in the backwards elimination procedure and settle with the four change point model (five epochs separated by change points at 2, 5, 12 and 16) identified

by that procedure. This five-epoch model exhibits a very good fit with the data. The estimate of the overdispersion parameter is $\hat{\sigma}_P^2 = 0.022$ compared with estimates of 0.058, 0.050 and 0.040 for respectively the models with single change points at 12, 16 and for two change points at 12 and 16.

Thus we conclude that there are four change points (at around 1660, 1740, 1880 and 1940 CE) defining five epochs with distinct hazard rates. ML estimates and 95% likelihood ratio confidence intervals (Reed *et al.*, 1996) for the fire cycles in the five epochs are

Epoch	Date	Fire Cycle (years)	
i		MLE	95% Con. Int.
1	1940 - 1980	1980	373 – 73053
2	1880 - 1940	156	37 – 209
3	1740 - 1880	728	333 – 2209
4	1660 - 1740	111	63 – 223
5	pre 1660	25	14 – 53

Notice how there is no overlap in confidence intervals in adjacent epochs, confirming the conclusions drawn from the P-values in backwards elimination. For the most part the epochs identified correlate well with known facts in the history of the region. The abrupt decrease in hazard around the middle of the $18th.$ 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, and from documented lower fire incidence in other areas of the Main and Front Ranges of the Rocky Mountains (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. While undoubtedly there was an increase in fire frequency at this time, whether the increase can be attributed to the European incursion into the region is open to question. Johnson *et al.*, 1990 point out that many of the large fires occurring at this time were caused by lightning rather than human activity. The marked drop in fire frequency occurring around 1940 also occurs in other areas of the Canadian Rockies *e.g.* the Kananaskis River Watershed (Johnson and Larsen, 1991; Reed *et al.*, 1996). It is tempting to conclude that this drop is due to the onset of increased fire protection and suppression programs during and immediately after World War II. However further detailed analysis of causes and sizes of fires which have occurred in this epoch is required before such a conclusion can be firmly established.

Conclusions.

The main result of the paper is a method of obtaining approximate P-values for testing the null hypothesis of no change points in historical forest fire frequency *vs.* the alternative of one change point with two distinct fire frequencies. The method explicitly accounts for selection bias in deciding which potential change point to test. If indeed there was in fact at most one change point the method is sound. However one cannot be sure that this is a reasonable assumption and that there was not in fact more than one change point. Thus one needs a method of deciding, once a first change point has

been identified, whether or not others should be considered. The methods pursued in the paper follow the iterative methods of model selection widely used in regression analysis – forward selection, stepwise and backwards elimination.

With all three methods there are difficulties. For backwards elimination and stepwise the main problem is the determination of the reference distribution for eliminating change points. In the article a somewhat *ad hoc* approach has been used which ignores the fact that one is testing, for elimination, the change point with the smallest LR statistic, but does nonetheless recognize the fact that with no change point on an interval the distribution of the maximum LR statistic for a change point on that interval will differ from that for a change point at a pre-specified location.

For forward selection the main problem is that of never being able to remove a change point once it has been identified. This can cause difficulties if the MLE in a one change point model is, in fact, some sort of compromise between two or more real change points.

All three methods have been applied to real time-since-fire map data from Glacier National Park, and give somewhat divergent results. However of the three, the backwards elimination results seem most plausible. The other two methods cannot reach the stage of including a third change point, even though a four change point model seems very plausible.

Just as in model selection for regression, it is probably wise to apply all three procedures and compare the results, along with residual analysis *etc.*

and use statistical judgement to decide on a most plausible model.

Another approach to the problem, to be presented in a forthcoming article, is to consider a hierarchy of models – no change points, one change point, two change points, ..., *etc.*, with an associated prior distribution. After calculating the MLEs and the maximized quasi log likelihood for each model, one can calculate *via* the Bayes Information Criterion, a posterior distribution for the number of change points. Thus rather than reaching a conclusion involving a single model (with say n change points) the conclusion will include various models with various numbers of change points and associated posterior probabilities.

References.

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristsic.* New York: Springer-Verlag.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics.* London: Chapman and Hall.

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.

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.

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

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.

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). Estimating historical forest fire frequency from time-since-fire sample data. Submitted to *I.M.A. Journal of Applications in Medicine and Biology.*

Reed, W. J., Larsen C. P. S., Johnson, E. A. and G. M. MacDonald (1996). Estimation of temporal variations in fire frequency from time-since-fire data. Submitted to *Canadian Journal of Forest Research.*

Van Wagner, C. E. (1978). Age-class distribution and the forest fire cycle. *Canadian Journal of Forest Research* **8**:222-227.

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 Chris Larsen, Ed Johnson and Adrian Raftery for their many useful comments and suggestions.

Appendix. Derivation of the method of obtaining the approximate P-value for testing for the significance of a change point suggested from the data.

For testing

H_0: No change points (constant hazard rate) *vs.*

H_A: One change point (separating two distinct hazard rates) between T and $(m-2)T$ years ago.

the likelihood ratio statistic is

$$R^\star = \Big(\frac{m-4}{m-3}\Big)\max\{R_p; p = 1, 2, \ \ldots, m-2\}. \tag{28}$$

where R_p is the LR statistic (13) for testing H_0 *vs.* the alternative H_p, say, of one change point pT years ago. The change point $p^\star$ which maximizes the r.h.s. of (28) will maximize the numerator of R_p, *i.e.* will maximize the difference in scaled deviances

$$\nabla D = 2\sigma^2[\hat{Q}_p - \hat{Q}_0] \tag{29}$$

where $\hat{Q}_0$ is the maximized value of the quasi-likelihood under H_0, *i.e.* of

$$Q_0(q_0) = \frac{1}{\sigma^2}\sum_{j=1}^{m-1}[s_j \log q_0 + y_j \log(1-q_0)] \tag{30}$$

and $\hat{Q}_p$ is the maximized value of the quasi-likelihood under H_p, *i.e.* of

$$Q_p(q_1, q_2) = \frac{1}{\sigma^2}\{\sum_{j=1}^{p}[s_j \log q_1 + y_j \log(1-q_1)] + \sum_{j=p+11}^{m-1}[s_j \log q_2 + y_j \log(1-q_2)]\} \tag{31}$$

We wish to find the distribution of the statistic (28) under H_0, *i.e.* when $q_1 = q_2 = q_0$.

To order $o([q_0 - \hat{q}_0]^2)$ we can write

$$\hat{Q}_0 = Q_0(\hat{q}_0) \doteq Q_0(q_0) + \frac{(q_0 - \hat{q}_0)^2}{2/I_0(q_0)} \tag{32}$$

where $I_0(q_0)$ is the Fisher information under model H_0. Similarly

$$\hat{Q}_p \doteq Q_1(q_0, q_0) + \frac{(q_0 - \hat{q}_1)^2}{2/I_1(q_0)} + \frac{(q_0 - \hat{q}_2)^2}{2/I_2(q_0)} \tag{33}$$

where $I_1(q_1)$ and $I_2(q_2)$ are the diagonal entries of the Fisher information matrix under H_p. Substituting into (29) gives

$$\nabla D \doteq \sigma^2[\frac{(q_0 - \hat{q}_1)^2}{1/I_1(q_0)} + \frac{(q_0 - \hat{q}_2)^2}{1/I_2(q_0)} - \frac{(q_0 - \hat{q}_0)^2}{1/I_0(q_0)}]. \tag{34}$$

Now from (11) and (12)

$$\hat{q_0} = \frac{X_1 + X_2}{N_1 + N_2}, \quad \hat{q_1} = \frac{X_1}{N_1}, \quad \hat{q_2} = \frac{X_2}{N_2} \tag{35}$$

where

$$N_1 = \sum_{j=1}^{p} s_{j-1} \quad X_1 = \sum_{j=1}^{p} s_j \tag{36}$$

and

$$N_2 = \sum_{j=p+1}^{m-1} s_{j-1} \quad X_2 = \sum_{j=p+1}^{m-1} s_j. \tag{37}$$

As noted in the text N_1 and N_2 are the numbers of area units facing fire trials between 0 and pT years ago and greater than pT years ago, respectively, and X_1 and X_2 are the corresponding numbers surviving those fire trials. Also

$$I_0(q) = \frac{N_1 + N_2}{q(1-q)}, \quad I_1(q) = \frac{N_1}{q(1-q)}, \quad I_2(q) = \frac{N_2}{q(1-q)} \tag{38}$$

Substituting these into (34) and simplifying yields

$$\frac{\nabla D}{\sigma^2} \doteq \frac{1}{N_1 + N_2} \left[\sqrt{N_2} \frac{X_1 - N_1 q_0}{\sqrt{N_1 \sigma^2 q_0 (1-q_0)}} - \sqrt{N_1} \frac{X_2 - N_2 q_0}{\sqrt{N_2 \sigma^2 q_0 (1-q_0)}} \right]^2, \tag{39}$$

which is equation (18).

Now X_1 and X_2 can be thought of as overdispersed binomial random variables with, with expected values (under H_0) $N_1 q_0$ and $N_2 q_0$ and variances $N_1 \sigma^2 q_0 (1-q_0)$ and $N_2 \sigma^2 q_0 (1-q_0)$. For a fixed p and large N_1 and N_2, from the Central Limit Theorem, the two terms in the square brackets in (39) are approximately zero-mean normal random variables with respective variances N_2 and N_1, and thus as $N_1, N_2 \to \infty$, the distribution of $\frac{\nabla D}{\sigma^2}$ in (39) tends to χ^2_1, the standard asymptotic null distribution of the LR statistic for testing H_0 *vs.* H_p

Another way of reaching the same result, which extends to the case of testing H_0 *vs.* the alternative H_A, of exactly one (unspecified) change point is the following: if $N_1, N_2 \to \infty$ with

$$\frac{N_1}{N_1 + N_2} \to \tau, \quad \frac{N_2}{N_1 + N_2} \to 1 - \tau \tag{40}$$

then under H_0

$$\frac{X_1 - N_1 q_0}{\sqrt{(N_1+N_2)\sigma^2 q_0(1-q_0)}} \to B(\tau); \quad \frac{X_2 - N_2 q_0}{\sqrt{(N_1+N_2)\sigma^2 q_0(1-q_0)}} \to B(1) - B(\tau), \tag{41}$$

where $\{B(t),\ 0 \le t \le 1\}$ is a *standard Brownian motion* process. After some re-arranging one can then show that

$$\frac{\nabla D}{\sigma^2} \to \frac{1}{\tau(1-\tau)}[B(\tau) - \tau B(1)]^2 = [\tilde{B}(\tau)]^2 \tag{42}$$

where $\{\tilde{B}(t)\}$ is a *self-normalized Brownian bridge* process, *i.e.* a standard Brownian bridge process, $\{B(t) - tB(1)\}$, divided by its standard deviation. From this it is clear that, for a fixed p, (since then τ is fixed), $\frac{\nabla D}{\sigma^2}$ has approximately a χ^2_1 distribution, as was shown above.

However for testing H_0 *vs.* the alternative H_A of exactly one (unspecified) change point, the LR statistic depends on the r.h.s. of (39) maximized over the choice of change point p, or in other words over the choice of N_1 and N_2 (with $N_1 + N_2 = \sum_{j=1}^{m-1} s_j$ fixed), and $N_1 = \sum_{j=1}^{p} s_{j-1}$ for some integer p between 1 and $m-1$. Using the limiting Brownian bridge as an approximation, the distribution of the maximum of $\frac{\nabla D}{\sigma^2}$ is then, for large N_1, N_2, approximately the same as that of

$$M = \max_{\tau \in \{a_1, a_2, \ldots, a_{m-2}\}} [\tilde{B}(\tau)]^2 \tag{43}$$

where

$$a_p = \frac{\sum_{j=1}^{p} s_{j-1}}{\sum_{j=1}^{m-1} s_{j-1}} \quad \text{for } p = 1, 2, \ldots, m-2. \tag{44}$$

Now suppose $r^\star$ is the observed value of the LR statistic (28) for testing H_0 *vs.* H_A. The P-value for the test is then $P(R^\star \geq r^\star | H_0)$, which using the Brownian bridge approximation, is approximately $P(\frac{M}{\hat{\sigma}^2/\sigma^2} \geq r^\star | H_0)$, where $\hat{\sigma}^2$ is the Pearson estimate of the overdispersion parameter. Asymptotically the distribution of the denominator $\hat{\sigma}^2/\sigma^2$ is χ^2_ν, where ν is the residual degrees of freedom. Thus it can be expressed $\hat{\sigma}^2/\sigma^2 = 1 + \epsilon$ where ϵ is a random variable, with support on $(-1, \infty)$, with $\mathrm{E}(\epsilon) = 0$, $\mathrm{var}(\epsilon) = 2/\nu$. The P-value can thus be expressed as

$$P(R^\star \geq r^\star | H_0) \approx P(\frac{M}{1+\epsilon} \geq r^\star | H_0) \approx \int_{-1}^{\infty} P(M \geq r^\star(1+\epsilon))\, dF_\epsilon(\epsilon) \quad (45)$$

where $F_\epsilon(\epsilon)$ is the distribution function of ϵ, and the second approximation assumes the asymptotic independence of M and ϵ. Denoting the complementary distribution function of the random variable M by $G_M(x) = P(M \geq x)$, and expanding $G_M(r^\star(1+\epsilon))$ about $r^\star$, one arrives at following the approximation for the P-value

$$P(R^\star \geq r^\star | H_0) \approx G_M(r^\star) + \frac{r^{\star 2}}{\nu} G_M{}''(r^\star) \quad (46)$$

where the error involves third and higher order moments of ϵ

To find $G_M(x)$, one can use the Poisson Clumping Heuristic (PCH) of Aldous (1989) to obtain the tail probabilities of the maximum of a self-normalized Brownian bridge process. Using the symmetry of the Brownian bridge, one has

$$G_M(x) = P(\max_{t \in A} [\tilde{B}(t)]^2 \geq x) = 2[1 - P(\max_{t \in A} \tilde{B}(t) < \sqrt{x})], \quad (47)$$

where A is the set $A = \{a_1, a_2, \ldots, a_{m-2}\}$. To approximate the probability on the r.h.s we apply the PCH, first letting $S_x = \{t \in A : \tilde{B}(t) \geq \sqrt{x}\}$. Then $P(\max_{t\in A} \tilde{B}(t) < \sqrt{x}) = P(S_x \cap A \text{ empty})$. Aldous (1989, Sec. D.12) establishes the "clump rate" for a self-normalized Brownian bridge process. Using this one gets that the expected clump size in the finite set A is the cardinality of the set

$$C_j = \{a_i \in A : a_j - a_j(1-a_j)\frac{\Phi^c(\sqrt{x})}{\sqrt{x}\phi(\sqrt{x})} \leq a_i \leq a_j + a_j(1-a_j)\frac{\Phi^c(\sqrt{x})}{\sqrt{x}\phi(\sqrt{x})}\} \tag{48}$$

which we denote in the text by $\mu_x(j)$. It then follows from the PCH, since the marginal probability that $\tilde{B}(t)$ exceeds $\sqrt{x}$ is $\Phi^c(\sqrt{x})$ that for large x

$$G_M(x) \approx 2\left(1 - \exp\left[-\Phi^c(\sqrt{x})\sum_{j=1}^{m-2}\frac{1}{\mu_x(j)}\right]\right) \tag{49}$$

which is the formula (24) in the text. The correction formula (26) is obtained by applying (46) to (49).

Many approximations have been applied to arrive at the results (49) and (46) used in the paper to obtain approximate P-values. It would certainly not be prudent to put much faith in the method without some attempt to ascertain how well the approximations perform. The approximation (49) can be checked for small m by numerically evaluating Gaussian integrals to determine exact probabilities. This is not possible for larger m. However simulation of a Brownian bridge process can be used for larger m, and further in this case the "correction" (46) to account for variability in $\hat{\sigma}^2/\sigma^2$ can be included.

Table 2 presents tail probabilities using the PCH approximation (49); using a χ^2_2 distribution, and the exact probabilities, calculated by numerically evaluating the appropriate Gaussian integral, in the case $m = 4$ with $a_1 = 0.5$, $a_2 = 0.75$. It can be seen that the PCH approximation is very good especially in the extreme tail, and that when it errs it is always on the side of conservatism. For other values of a_1 and a_2 the PCH approximation appears equally good. With $m = 3$ they are even better. The χ^2_2 distribution (which would be the correct asymptotic approximation for a regular problem, but is incorrect here because of the nature of the change point parameter p) appears to yield tail probabilities which are too conservative (*c.f.* Worsley, 1988; Henderson, 1990).

Table 3 presents simulation results (10,000 replications) for $m = 10, 20, 30$ and 40. In each case the points $a_1, a_2, \ldots, a_{m-2}$ were computed on the unit interval assuming that the age-class distribution of areas was geometric following (3) and (4) with constant survival probability $q = 0.90$, corresponding to a constant hazard rate. Along with P-values estimated from the simulations (95% confidence intervals) those determined from the PCH formula (49) corrected by (46) and those from respectively an $F_{2,m-4}$ distribution and a χ_2 distribution are displayed. The χ^2_2 distribution which ignores variation in the estimate of the overdispersion parameter, as well as being incorrect for the reason stated above, yields P-values which are much too small. Its use is not recommended. The $F_{2,m-4}$ approximation includes variation in the overdispersion parameter, and performs better than the χ^2_2 approximation,

but still tends to give P-values which are too small. The PCH approximation does well in the tail, (as one would expect since it derived on the basis of sparse "clumps") but errs on the side of conservatism (except for $m = 10$ and small P-values). For $m \geq 20$ the PCH and $F_{2,m-4}$ approximations bracket the Monte Carlo values, and would thus seem to provide useful bounds for the true P-value. For smaller values of m it should be possible to improve the PCH approximation by including third and higher order terms in the Taylor series expansion leading to (46). We do not investigate this further here.

p	1	2	3	4	5	6	7	8	9	10	11	12
R_p	1.24	2.91	1.32	1.04	0.21	.61	1.28	2.43	4.18	6.92	9.47	11.99
$\tilde{R}_p$	2.04	5.22	1.23	0.56	0.06	0.02	0.26	0.95	2.22	4.34	5.73	6.36

13	14	15	16	17	18	19
9.48	9.72	8.43	14.40	3.50	0.06	0.19
2.64	1.31	$\sim$0	-	.03	.75	.07

Table 1: Values of the likelihood ratio statistic R_p (13) for testing the null hypothesis of no change points *vs.* the alternative of one change point at p (*i.e.* pT years ago); and of the likelihood ratio statistic, $\tilde{R}_p$, for testing the null of one change point at 16 against the alternative of two change points at 16 and p.

$\sqrt{x}$	1	1.5	2.0	2.5	3.0	3.5
exact $P(M \geq x)$	.49	.22	.08	.02	.01	.001
PCH approx.	.54	.25	.09	.02	.01	.001
χ^2_2 approx.	.61	.33	.14	.04	.01	.002

Table 2: Performance of the PCH approximation and χ^2_2 approximation for the probability that the maximum of a self-normalized Brownian bridge over $\{a_1 = 0.5, a_2 = 0.75\}$ (corresponding to $m = 4$ age classes) exceeds values x.

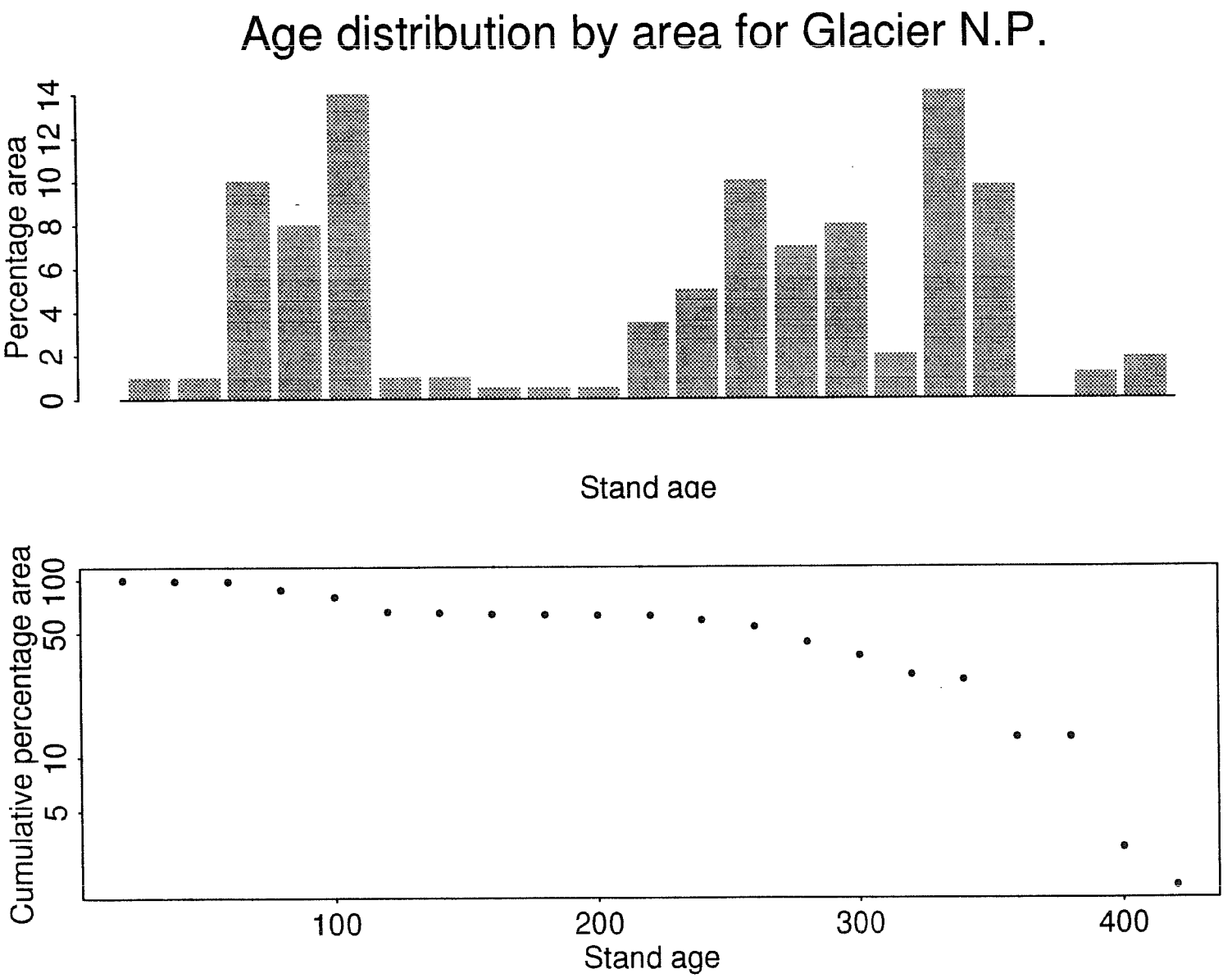
	Observed LR statistic $r^\star$				
$m = 10$	5.0	7.5	10.0	12.5	15.0
PCH approx.	.477	.209	.083	.031	.011
$F_{2,6}$ approx.	.163	.088	.053	.034	.023
χ^2_2 approx.	.082	.024	.007	.002	.001
Monte Carlo	.25-.27	.13-.14	.07-.08	.04-.05	.023-.029

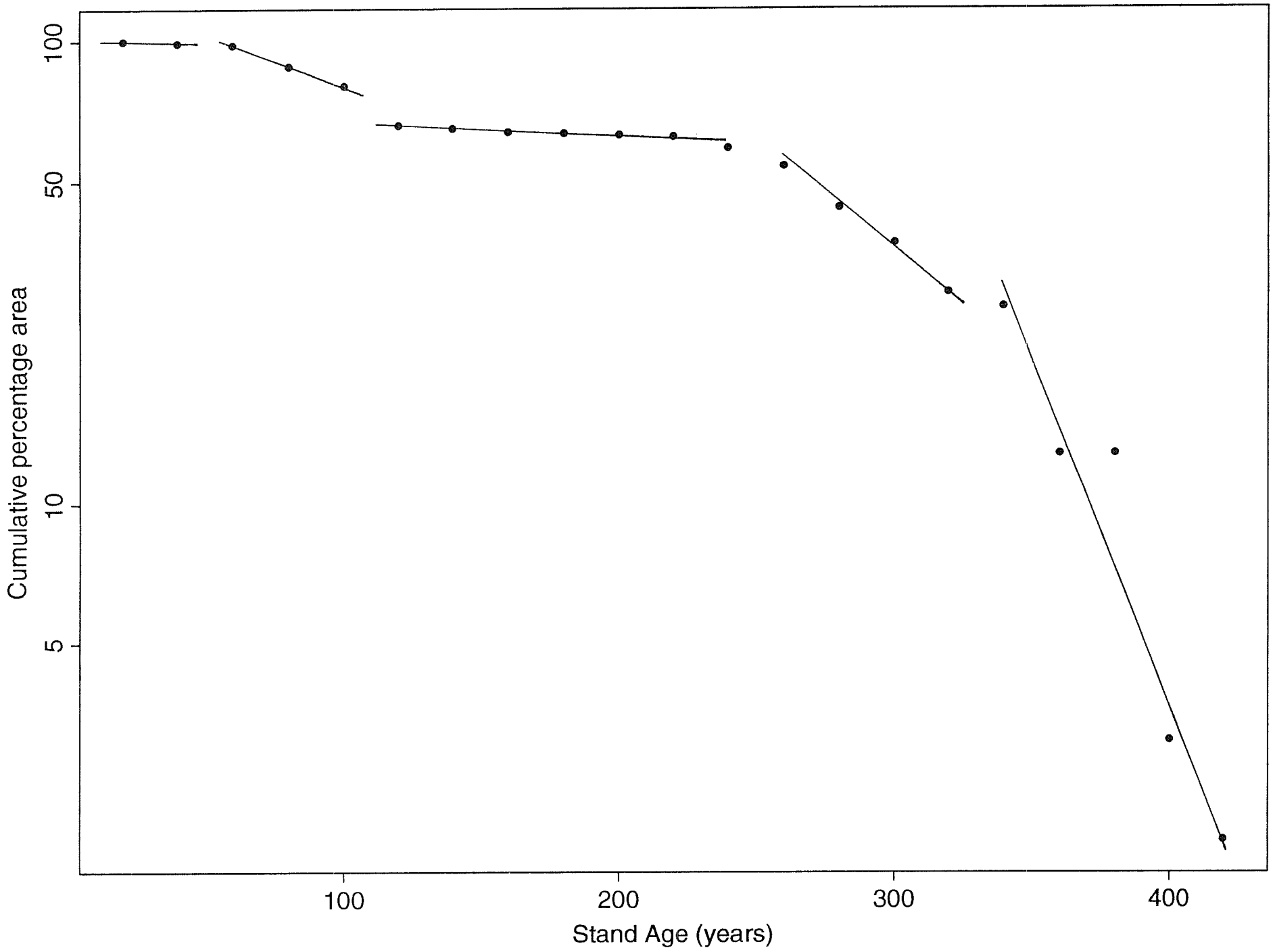
	Observed LR statistic $r^\star$				
$m = 20$	5.0	7.5	10.0	12.5	15.0
PCH approx.	.585	.229	.083	.029	.010
$F_{2,16}$ approx.	.114	.046	.021	.010	.005
χ^2_2 approx.	.082	.024	.007	.002	.001
Monte Carlo	.25-.27	.10-.11	.04-.05	.02-.03	.010-.014

	Observed LR statistic $r^\star$				
$m = 30$	5.0	7.5	10.0	12.5	15.0
PCH approx.	.603	.273	.095	.032	.011
$F_{2,26}$ approx.	.102	.037	.015	.006	.003
χ^2_2 approx.	.082	.024	.007	.002	.001
Monte Carlo	.30-.31	.11-.12	.04-.05	.015-.021	.006-.009

	Observed LR statistic $r^\star$				
$m = 40$	5.0	7.5	10.0	12.5	15.0
PCH approx.	.607	.319	.108	.035	.012
$F_{2,36}$ approx.	.096	.033	.012	.005	.002
χ^2_2 approx.	.082	.024	.007	.002	.001
Monte Carlo	.32-.34	.12-.13	.05-.065	.018-.023	.007-.010

Table 3: Approximate P-values calculated using: PCH approximation (49)) with correction (46); an $F_{2,m-4}$ approximation; a χ^2_2 approximation, along with 95% confidence interval for the P-value calculated from a Monte Carlo simulation with 10,000 runs. The four panels correspond to different numbers (m) of age classes.


Age distribution by area for Glacier N.P.
Percentage area
0 2 4 6 8 10 12 14
Stand age
Cumulative percentage area
5 10 50 100
100 200 300 400
Stand age


Semi-log plot of cumulative area vs. age. Glacier National Park
Cumulative percentage area
100
50
10
5
100
200
300
400
Stand Age (years)