CHAPTER 12

# Statistical Inference for Historical Fire Frequency Using the Spatial Mosaic

W. J. Reed
*Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada*



## I. INTRODUCTION

In forested regions where the principal agent of stand regeneration is fire (e.g., in boreal forests), the spatial mosaic of stands depends on the frequency and extent of fires in the past. To some extent, knowledge of the spatial mosaic can be used to make inference about past fire frequency. This chapter deals with statistical methodology for such inference. Given complete knowledge of fire history, one could, in principle, reconstruct the spatial mosaic. In contrast, knowledge of the current spatial mosaic is not sufficient for complete knowledge of past fires. The reason for this is that evidence of fires in the past can be obliterated by subsequent fires burning over all or part of the extent of the earlier fires. The present spatial mosaic provides complete evidence of only the most recent fire or fires; it provides decreasing information about fires more distant in time; and of very early fires, little or no evidence may remain.

A *time-since-fire map* is closely related to the spatial mosaic and gives the time since the most recent fire at every point in a study area. In contrast a *time-since-fire sample* contains only the time since the most recent fire at a sample of

*Forest Fires*

points (drawn according to some well-defined sampling plan) from the study area. Usually because of the limitations of techniques of dating stand origins or past fires, the time-since-fire observations are grouped into classes (often of width one decade). In this chapter statistical methodologies for analyzing time-since-fire map and sample data will be discussed. The main issues addressed are:

- methods of graphical data analysis of such data;
- estimation of fire frequency at distinct epochs in the past. Likelihood methods will be used to obtain point estimates and confidence intervals.
- testing for the significance of changes in fire frequency at specified times;
- determination of when (possibly multiple) changes in fire frequency may have occurred.

The methods presented collect and summarize recent research in this area.

## II. GRAPHICAL ANALYSIS

A great deal of information about historical fire frequency can be obtained by simple graphical analysis. For time-since-fire map data, one can plot cumulative proportional areas, with time since last fire $\geq t$, on a logarithmic scale, against time $t$ (on a natural scale). The lower panels of Figures 1, 3 and 4 are examples of such plots. For sample data, the same procedure can be followed using the cumulative proportion of sample points with time since last fire $\geq t$ (Figure 2). If the hazard of burning has been both age-independent and time-homogeneous, the points on the plot should lie close to a straight line with negative slope (equal to the hazard of burning), since at any point the probability that the most recent fire occurred $t$ or more years ago is $e^{-\lambda t}$, where $\lambda$ is the hazard of burning. Abrupt changes in the plot suggest temporal changes in the hazard of burning. For example, if the hazard of burning changed at some distinct change point (as might possibly have occurred for example at the time of European intervention), then one would expect a sharp elbow in the plot at that point. Thus a plot which naturally divides into a number of straight line segments suggests a number of distinct historical epochs of roughly constant fire frequency, with the changes in fire frequency occurring at times determined by the points of intersections of adjacent segments. By comparing such plots for different subsets of the study area (e.g., broken up spatially or by aspect) one can get a good idea of whether historical fire frequency was homogeneous or not with respect to these categories.

Even though the vertices and slopes of line segments determined graphically provide rough estimates of change points and intervening fire frequencies, it is desirable to have more objective statistical methodologies for accomplishing this. Subsequent sections of this chapter describe such methodologies.

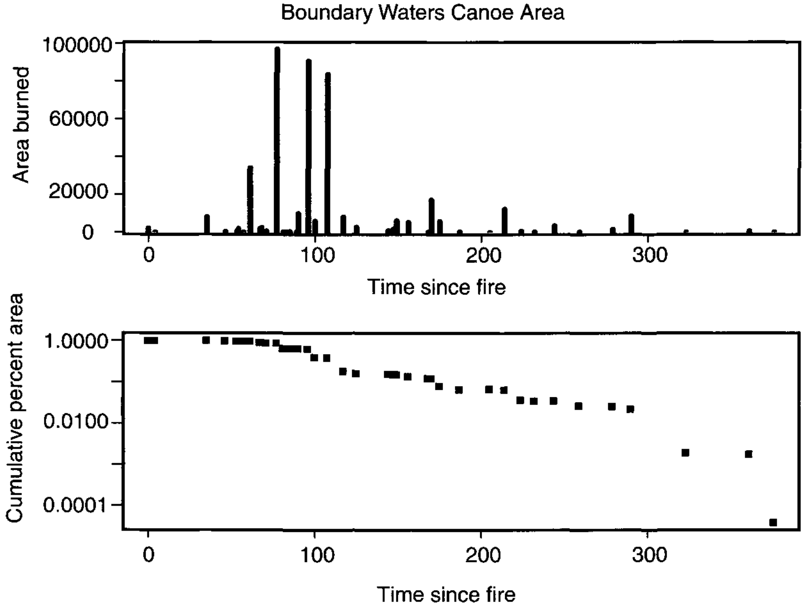



FIGURE 1 Time since last fire for Boundary Waters Canoe Area (Heinselman, 1973). The top panel shows the areas (on vertical axis in hectares) with given time since last fire (horizontal axis in years). The bottom panel plots the cumulative proportional area (on logarithmic scale) with time since fire exceeding the time (in years) on the horizontal axis.

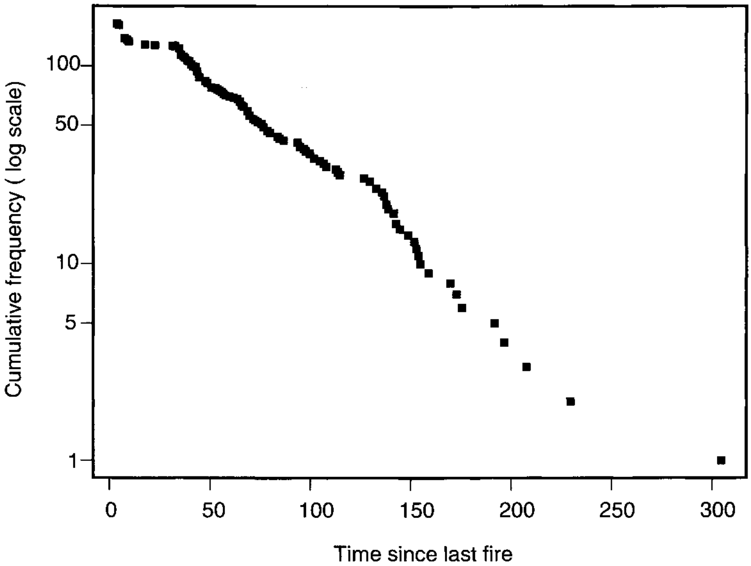



FIGURE 2 Cumulative percent of sample points (on logarithmic scale) with time since fire exceeding the time (in years) on the horizontal axis for time-since-fire sample data of Wood Buffalo National Park (Larsen, 1996).

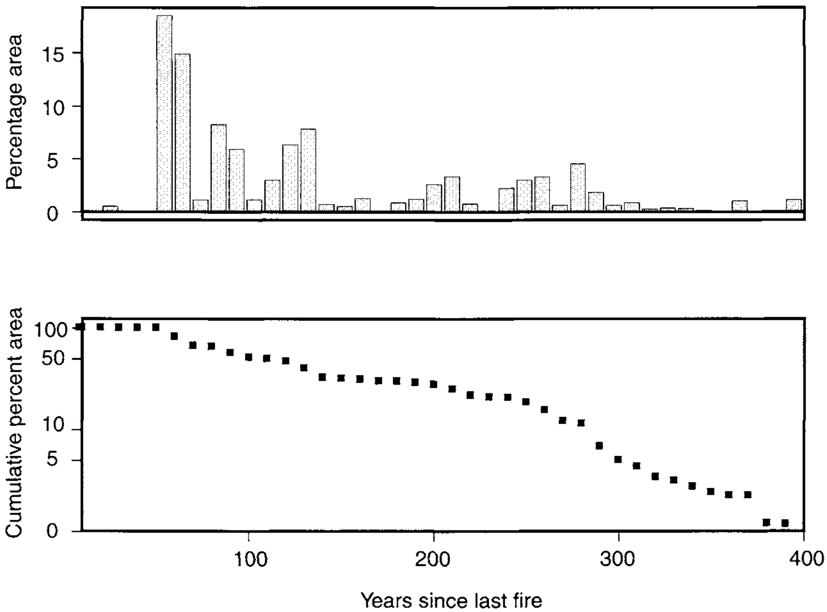



FIGURE 3 Time since last fire for Kananskis Watershed (Johnson and Larsen, 1991). The top panel shows the percentage area (on vertical axis) in given time-since-fire class (horizontal axis in years). The bottom panel plots the cumulative percentage area (on logarithmic scale) with time since fire exceeding the time (in years) on the horizontal axis.

However, the problem is approached in stages. First, the problem of obtaining maximum likelihood estimates (MLEs) of the hazards of burning prevailing between prespecified change points is addressed, along with how to obtain approximate confidence intervals and tests for the significance of the change points. This is done for both map and sample data. Following this, the question of determining change points from the data alone is addressed.

## III. STATISTICAL INFERENCE WITH PRESPECIFIED CHANGE POINTS

We assume that the observations whether for map or sample data are binned into time-since-fire classes $1, 2, \ldots, m$, which are of equal width, say $T$ years, save for the oldest which is open-ended. Specifically, assume that the classes $1, 2, \ldots, m-1$ are defined by time since last fire in the intervals $[0, T)$, $[T, 2T), \ldots, [(m-2)T, (m-1)T)$, respectively and that class $m$ is defined by time since last fire greater than or equal to $(m-1)T$. In practice, the resolution

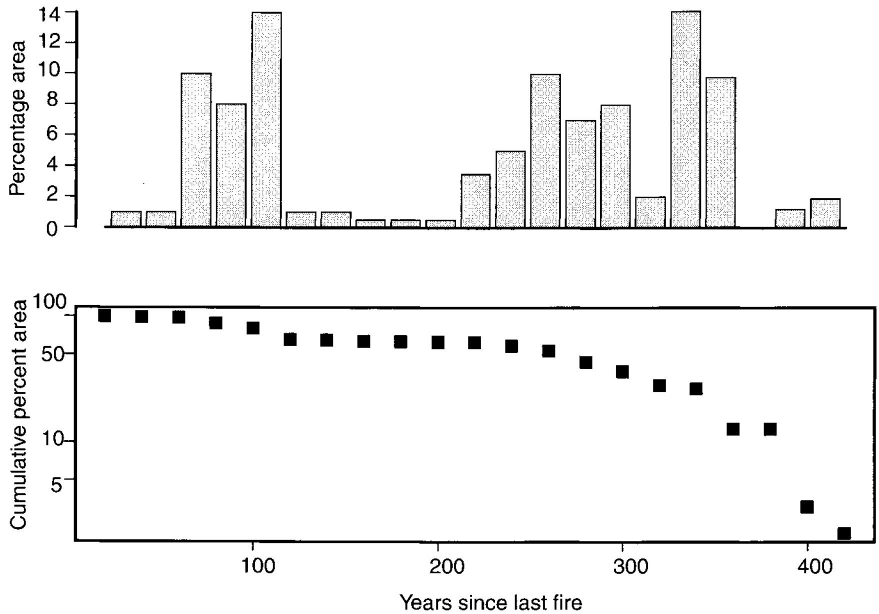



FIGURE 4 Time since last fire for Glacier National Park (Johnson *et al.*, 1990). The top panel shows the percentage area (on vertical axis) in given time-since-fire class (horizontal axis in years). The bottom panel plots the cumulative percentage area (on logarithmic scale) with time since fire exceeding the time (in years) on the horizontal axis.

of dating fires is no finer than one year, so $T$ will be a positive integer (often $T = 10$ or 20).

Under the assumption of homogeneity, at any given location in the study area, the probability that the time since last fire belongs to class $j$ can be determined in terms of the hazards of burning prevailing at and since that last fire date, by observing that to belong to class $j$, there must have been a fire between $(j-1)T$ and $jT$ years ago, and subsequently no fire. Specifically, it is $e^{-\sum_{i=1}^{j-1}\lambda^{(i)}T}(1 - e^{-\lambda^{(j)}T}) = \theta_j$, say, where $\lambda^{(i)}$ is the instantaneous hazard of burning between $(i-1)T$ and $iT$ years ago. With known change points, the parameters $\lambda^{(i)}$, for $i = 1, 2, \ldots, m$, can be expressed in terms of parameters representing the hazards of burning in the epochs between the change points. For example, with no change points, all the $\lambda^{(i)}$ are equal to the assumed constant hazard $\lambda_0$, say. If there were one change point $pT$ years ago, with a hazard of burning $\lambda_2$, say, prevailing before that time and $\lambda_1$, since that time, then $\lambda^{(i)} \equiv \lambda_1$ for $i = 1, 2, \ldots, pT$ and $\lambda^{(i)} \equiv \lambda_2$ for $i = p + 1, p + 2, \ldots, m$, etc.

To estimate by maximum likelihood the parameters representing the hazards of burning in the epochs between the change points, one needs to construct a likelihood function [i.e., determine the probability of the observed areas (or frequencies of sample points) falling in the various classes] in terms of these parameters. If one could reasonably assume independence between points in the study area, with respect to time since last fire, the likelihood function could be obtained directly from the multinomial probabability distribution. However, such an assumption is not reasonable because of the fact that fires spread spatially. To reflect this fact, one needs a model for the distribution of areas in time-since-fire classes which exhibits a contagion effect. Two approaches have been used to date to accomplish this. One is to use a parametric model (the *Dirichlet distribution*) for the areas in time-since-fire classes. This provides a likelihood directly for map data (Reed, 1994), whereas a likelihood for sample data can be obtained by computing the (marginal) probability of observing the observed frequencies in the various classes, assuming that the proportional areas follow the Dirichlet distribution. This leads to a likelihood of the form of the *Dirichlet compound multinomial* distribution, which is the multivariate analogue of the widely used *beta-binomial* distribution (Reed, 1998).

The other approach, which in many ways is more straightforward, is to incorporate the contagion effect by using an *overdispersed multinomial* model with corresponding *quasi likelihood* (Reed *et al.*, 1998). This is based on the assumption that the contagion effect, due to spatial spread of fires, has no effect on the expected areas (or frequencies) in the time-since-fire classes, but it does have an effect on the variances–covariances, inflating them by a constant factor. This second approach is more general in the sense that the Dirichlet distribution for proportional areas, and the resulting distribution for frequencies of sample points, are both overdispersed multinomial distributions themselves (Reed, 1998). Which of the two approaches is superior in any given context is not known. This issue could probably be resolved by studying the performance of the two methods using simulated data. However, to date such a study has not been undertaken. In this chapter, we concentrate on the overdispersed multinomial approach. One advantage is that analytic expressions can be obtained for maximum likelihood estimates of epochal hazards of burning and for their standard errors. There is a simple interpretation of the MLEs, and furthermore they agree well with the graphical estimates. In contrast, using the Dirichlet model approach, numerical maximization is required to obtain MLEs of the epochal hazards of burning, and it is not possible to obtain a simple interpretation of the estimates.

## A. Constructing a Quasi Likelihood for an Overdispersed Multinomial Model

For data from an overdispersed multinomial distribution in which there is a constant inflation factor (*overdispersion parameter*, $\sigma^2$) for all variances and covariances, one can construct a *quasi likelihood* function of the form

$$Q = \frac{1}{\sigma^2} \sum_{j=1}^{m} y_j \log(\theta_j) \tag{1}$$

where for map data $y_j = A_j/(\sum_{j=1}^{m} A_j)$, the proportional area in class $j$, and for sample data, with frequencies $f_j$, $y_j = f_j/(\sum_{j=1}^{m} f_j)$, the proportion of the observations in class $j$. The overdispersion parameter $\sigma^2$ will differ for map and sample data. In most respects regarding model parameters (other than $\sigma^2$), the quasi likelihood behaves like an ordinary log likelihood. Thus, one can find MLEs by maximizing the quasi likelihood and obtain approximate standard errors from the observed (quasi) *Fisher information matrix* (Hessian matrix of second derivatives of quasi likelihood at the maximum).

Simple analytic expressions are available for the MLEs of the hazards of burning in distinct epochs. For example if there is only one epoch (i.e., the same hazard of burning $\lambda_0$ is assumed to have prevailed at all times in the past), then from differentiation of the quasi-likelihood it is easy to show that the MLE is

$$\hat{\lambda}_0 = (-1/T)\log(\hat{q}_0)$$

where

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

and

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

is the proportional area (or proportion of the observations, for sample data) with time since last fire at least $jT$, for $j = 1, 2, \ldots, m-1$. Furthermore, the approximate variance of the estimator $\hat{q}_0$ is

$$\mathrm{var}(\hat{q}_0) \approx \sigma^2 \frac{(\sum_{j=1}^{m-1} s_j)(\sum_{j=1}^{m-1} y_j)}{(\sum_{j=1}^{m-1} s_{j-1})^3}$$

If it is assumed that there are several epochs with distinct hazards, say $\lambda_1$ between 0 and $p_1T$ years ago, $\lambda_2$ between $p_1T$ and $p_2T$ years ago, etc., then the MLEs of the separate hazards of burning can be found in a similar way. For example, if there are two epochs separated by a single change point $pT$ years ago, then for the more recent epoch

$$\hat{\lambda}_1 = (-1/T)\log(\hat{q}_1)$$

and for the earlier epoch

$$\hat{\lambda}_2 = (-1/T)\log(\hat{q}_2)$$

where

$$\hat{q}_1 = \frac{\sum_{j=1}^{p} s_j}{\sum_{j=1}^{p} s_{j-1}} \qquad \hat{q}_2 = \frac{\sum_{j=p+1}^{m-1} s_j}{\sum_{j=p+1}^{m-1} s_{j-1}}$$

Furthermore, these estimates are independent and have approximate variances given by

$$\mathrm{var}(\hat{q}_1) \approx \sigma^2 \frac{(\sum_{j=1}^{p} s_j)(\sum_{j=1}^{p} y_j)}{(\sum_{j=1}^{p} s_{j-1})^3}; \qquad \mathrm{var}(\hat{q}_2) \approx \sigma^2 \frac{(\sum_{j=p+1}^{m-1} s_j)(\sum_{j=p+1}^{m-1} y_j)}{(\sum_{j=p+1}^{m-1} s_{j-1})^3}$$

To use these variance formulas to obtain standard errors and confidence intervals for the hazard of burning, etc., one needs to estimate the overdispersion parameter $\sigma^2$. This is usually done using the Pearson estimate, which for map data is

$$\hat{\sigma}^2 = \frac{1}{m - r - 1} \sum_{j=1}^{m} \frac{(y_j - \hat{\theta}_j)^2}{\hat{\theta}_j(1 - \hat{\theta}_j)}$$

where $\hat{\theta}_j$ is the MLE of $\theta_j$ and $r$ is the number of epochs. With a single hazard of burning assumed to have prevailed at all times past, the MLEs of the cell probabilities are $\hat{\theta}_j = (1 - \hat{q}_0)\hat{q}_0^{j-1}$, $j = 1, 2, \ldots, m - 1$ and $\hat{\theta}_m = \hat{q}_0^{m-1}$. With two epochs, they are $\hat{\theta}_j = (1 - \hat{q}_1)\hat{q}_1^{j-1}, j = 1, 2, \ldots, p$, $\hat{\theta}_j = (1 - \hat{q}_2)\hat{q}_1^{p}\hat{q}_2^{j-p-1}$, $j = p + 1, 2, \ldots, m - 1$ and $\hat{\theta}_m = \hat{q}_1^{p}\hat{q}_2^{m-p-1}$. Similar formulas hold when there are more than two epochs.

An approximate $100(1 - \alpha)\%$ confidence interval for the hazard of burning in epoch $r$ can be obtained from the confidence interval for $q_r$, which is

$$\hat{q}_r \pm z_{\alpha/2}\sqrt{(\hat{\sigma}^2/\sigma^2)\mathrm{var}(\hat{q}_r)}$$

where $z_{\alpha/2}$ is a percentage point from the standard normal distribution. The transformation $q \rightarrow (-1/T)\log(q)$ will convert this to a confidence interval for the hazard of burning and $q \rightarrow (-T)/(\log(q))$ will convert it to a confidence interval for the fire cycle.

A test for the significance of a change point between epochs can be constructed by computing the liklelihood ratio (LR) statistic. Specifically, one computes the test statistic

$$\Lambda = 2\frac{\hat{\sigma}_1^2}{\sigma^2}(\hat{Q}_1 - \hat{Q}_0)$$

where $\hat{Q}_1$ and $\hat{Q}_0$ are the maximized quasi likelihoods assuming the change point present and not, respectively, and $\hat{\sigma}_1^2$ is the Pearson estimate of $\sigma^2$ computed with the change point present. The null distribution of this statistic from which a P-value can be obtained is approximately $F_{1,m-\nu-1}$ where $\nu$ is the number of epochs when the change point is present. An important point to note here is that, because of the problem of selection bias, this procedure is not valid if the change point is suggested from an examination of the data, e.g., from a graphical anlysis (Reed *et al.*, 1998). One can also use a LR statistic to construct confidence intervals for the epochal hazards of burning (or fire cycles) using the method described in Reed *et al.* (1997).

### Example 1. Map Data. Boundary Waters Canoe Area

Figure 1 shows the time-since-fire distribution for the Boundary Waters Canoe Area, from the maps obtained by Heinselman (1973) in his pioneering fire history study. Since fires are dated to the year, the cell width here is $T = 1$. European settlement in the area first occurred in the 1860s, and there appears to be a change in the slope of the semi-log cumulative frequency plot around this time (a little more than 100 years ago). Assuming a single change point in 1865 yields the following maximum likelihood estimates of the hazards of burning and corresponding fire cycle (with approximate 95% confidence intervals):

Epoch 1 (1865–1971): $\hat{\lambda}_1 = 0.00883$ p.a., $FC_1 = 113$ (73–245) yr
Epoch 2 (before 1865): $\hat{\lambda}_2 = 0.0121$ p.a., $FC_2 = 83$ (42–3915) yr

The overdispersion parameter is estimated as 0.0593.

### Example 2. Sample Data. Wood Buffalo National Park

Larsen (1996) reported a detailed study of the fire history of Wood Buffalo National Park, a 45,000 $km^2$ area of boreal forest straddling the border between Alberta and the Northwest Territories. Larsen determined (to the nearest year) the time since last fire at 166 randomly selected sites. These are displayed in Figure 2 with $m = 300$ classes of width $T = 1$ year. The graph suggests a change point at $p = 129$ years ago (1860). Assuming this to be the case, the maximum likelihood estimates of the hazards of burning and corresponding fire cycle (with approximate 95% confidence intervals) are:

Epoch 1 (1861–1989): $\hat{\lambda}_1 = 0.0142$ p.a., $FC_1 = 70$ (58–90) yr
Epoch 2 (1860 and earlier): $\hat{\lambda}_2 = 0.0292$ p.a., $FC_2 = 34$ (23–69) yr

The overdispersion parameter is estimated as 0.0106.

## IV. THE EFFICIENCY OF SAMPLE VS. MAP DATA

Map data corresponds to a complete census of the study area. Since one can obtain estimates of past fire frequencies using observations on time since the last fire at only a random sample of points in the study area, an obvious question is how much precision is foregone by determining time since fire at a sample of points rather than everywhere in the study area. Of course, the reasons for collecting time-since-fire map data extend beyond the estimation of historical fire frequency. These issues are not addresssed here. Rather, attention is confined to assessing the relative statistical efficiency of using sample vs. map data for such estimation.

To this end, suppose that the proportional areas over the whole study area $a_j = A_j/(\sum_{j=1}^{m} A_j)$ follow an overdispersed multinomial distribution with overdispersion parameter $\sigma_0^2 < 1$. Under this assumption,

$$\mathrm{var}(a_j) = \sigma_0^2\theta_j(1 - \theta_j); \qquad \mathrm{cov}(a_i, a_j) = -\sigma_0^2\theta_i\theta_j$$

By first conditioning on the proportional areas $a_j$, one can show that for $n$ time-since-fire observations at a random sample of points in the study area, the proportional frequencies $y_j = f_j/n$ in the $m$ classes follow an overdispersed multinomial distribution with overdispersion parameter

$$\sigma_1^2 = \frac{1}{n}[1 + (n - 1)\sigma_0^2] > \sigma_0^2$$

The *quasi Fisher information matrix I*, which is obtained as the expected value of the Hessian matrix of second derivatives of the quasi likelihood, provides a measure of the precision of the MLEs. It is easy to show that this information matrix for map data, $I_0$ say, differs from that for sample data, $I_1$ say, only by the scalar multiplicative constant, that is,

$$I_1 = \frac{\sigma_0^2}{\sigma_1^2} I_0$$

The relative efficiency (ratio of variances of estimators) for sample data relative to map data is, thus,

$$\text{Rel. Eff.} = \frac{\sigma_0^2}{\sigma_1^2} = \frac{n\sigma_0^2}{1 + (n - 1)\sigma_0^2}$$

which is easily confirmed to be less than one and converging to one as sample size $n \to \infty$ (i.e., estimates based on sample data will have lower precision than those based on map data) but will approach the latter as sample size grows large.

One can substitute an estimate for $\sigma_0^2$ in this expression to assess the loss in precision by using sample data.

**Example 1 (continued). Boundary Waters Canoe Area**

The estimate of the overdispersion parameter for these data was $\hat{\sigma}_0^2 = 0.0593$. Table 1 gives estimates, using this value, of the relative efficiency of sample data

**TABLE 1 Estimated Relative Efficiency of Using Sample Data (sample size $n$) vs. Map Data for Boundary Waters Canoe Area and Ratio of Expected Widths of Confidence Intervals**

| Sample size $n$ | 10 | 20 | 50 | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|---|
| Rel. eff. | 0.39 | 0.56 | 0.76 | 0.86 | 0.93 | 0.97 | 0.98 |
| Ratio of C.I. widths | 1.61 | 1.34 | 1.15 | 1.08 | 1.04 | 1.02 | 1.01 |

(for various sample sizes $n$) vs. map data and the ratio of (expected) widths of corresponding confidence intervals. It is clear that a sample of moderate size (100 to 200) would be almost as good as a complete map, in terms of the precision of estimates. The expected width of confidence intervals would be only 8% greater with $n = 100$ and only 4% greater with a sample of size $n = 200$.

### Example 2 (continued). Wood Buffalo National Park

Larsen (1996) determined time since last fire at $n = 166$ sample points in Wood Buffalo National Park. The overdispersion parameter for a sample of this size was estimated as $\hat{\sigma}_1^2 = 0.0106$. Thus, an estimate of the overdispersion parameter for map data is $\hat{\sigma}_1^2 = (166 \times 0.0106 - 1)/(166 - 1) = 0.0046$. Note that this estimate is smaller by an order of magnitude than the corresponding estimate for the Boundary Waters Canoe Area, suggesting a much smaller contagion effect (and thus a finer patch mosaic) in Wood Buffalo than in the Boundary Waters. Using this estimate, one can assess how much precision was foregone using a sample of points rather than a complete map survey and how the precision of estimates from sample data depends on sample size. Table 2 gives information on this.

The results suggest that confidence intervals would be about one third shorter if map data had been obtained. Unlike the case of the Boundary Waters, where relatively small samples could provide precision comparable with map data, for Wood Buffalo such small samples would not be adequate. For example, for the Boundary Waters, a sample of size 50 would yield confidence intervals

**TABLE 2 Estimated Relative Efficiency of Using Sample Data vs. Map Data for Wood Buffalo National Park and Ratio of Expected Widths of Corresponding Confidence Intervals (second row)**

| Sample size $n$ | 10 | 20 | 50 | 100 | 166 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|---|---|
| Rel. eff | 0.044 | 0.085 | 0.19 | 0.32 | 0.43 | 0.48 | 0.70 | 0.82 |
| Ratio of C.I. widths | 4.76 | 3.44 | 2.31 | 1.78 | 1.52 | 1.44 | 1.20 | 1.10 |
| C.I. width rel. to actual | 3.13 | 2.27 | 1.52 | 1.17 | 1.00 | 0.95 | 0.79 | 0.73 |

The last row is the ratio of width of confidence intervals using sample size $n$ and actual sample size 166.

only about 15% wider than map data; for Wood Buffalo, such a sample size would yield confidence intervals 130% wider than map data (and 50% wider than those obtained with the actual sample size, 166). The reason for the difference lies with the order of magnitude difference in the overdispersion parameter estimates. With greater contagion, and on average larger patches in the time-since-fire mosaic in the Boundary Waters, a relatively small number of sample points can, with high probability, provide information on most patches. On the other hand, with the smaller contagion parameter and finer patch mosaic in Wood Buffalo, one needs many more sample points on average to cover most patches.

## V. DETERMINING EPOCHS OF CONSTANT FIRE FREQUENCY

In the previous sections, it was assumed that the number of epochs, with constant hazard of burning, and the change points dividing them were known. Most often, this will not be the case, and one will face the problem of determining the epochs from the data. This section provides a brief description of a methodology for accomplishing this.

From the statistical point of view, determining the number and location of the change points (i.e., determining the epochs) is a problem in model identification, analagous to deciding which regressor variables should be included in a regression model. There are a number of ways to approach such problems. One is to use some sort of iterative procedure in which change points can be added or removed from the model (analagous to forward selection and stepwise procedures in regression). This approach has been used for identifying fire epochs (Reed, 1998), but there are a number of difficulties associated with it. An alternative approach which avoids these difficulties is to use the *Bayes Information Criterion* to decide on the best single (or several) models.

From one point of view, the Bayes Information Criterion (BIC) can be viewed as a log likelihood, adjusted for the number of parameters in a model (here the number of change points and intervening hazards of burning). It can also be interpreted in a Bayesian context (see later discussion).

Consider a hieararchy of models:

$H_0$: No change points (constant hazard of burning at all times in the past),
$H_1$: One change point (separating two epochs with distinct hazards of burning),
$H_2$: Two change points (separating three epochs with distinct hazards of burning),
etc.

$H_k$: $k$ change points (separating $k + 1$ epochs with distinct hazards of burning).

Under any model $H_r$ there are $2r + 1$ undetermined parameters ($r$ change points separating $r + 1$ hazards of burning $\lambda_i$ or parameters $q_i = e^{-\lambda_i T}$) in addition to the overdispersion parameter. MLEs of the change points can be found by comparing the quasi likelihoods, maximized over the $(r + 1)q_i$ parameters, for all $\binom{m-1}{r}$ possible choices of $r$ change points. A comparison of models $H_0$, $H_1, \ldots$, etc., can be achieved by comparing the BIC for each model. It can be shown (Reed, 2000) that an approximation for the BIC is

$$\mathrm{BIC}_r \approx \hat{D}_r - (\mathrm{df})_r \log\left(\frac{n}{\hat{\sigma}^2}\right) \tag{2}$$

where $\hat{D}_r$ is the minimum scaled quasi deviance for model $H_r$ i.e.,

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

minimized over the choice of $r$ change points $p_0, p_1, \ldots, p_r$ (with $p_0 = 0$ and $p_{r+1} = m$). The terms $\hat{q}_i$ are the MLEs of the $q_i$ parameters for epochs $i = 1, 2, \ldots, r + 1$). Precisely,

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

In Eq. (2), the term $(\mathrm{df})_r$ is the residual degrees of freedom for model $H_r$ [i.e., $\mathrm{df}_r = (m - 1) - (2r + 1)$ and $n = \sum_{j=1}^{m-1} s_{j-1}$]. The estimate $\hat{\sigma}^2$ is computed under the "biggest" model contemplated, $H_k$. Precisely,

$$\hat{\sigma}^2 = \frac{1}{m - 2r - 2}\sum_{j=1}^{m}\frac{(y_j - \hat{\theta}_j)^2}{\hat{\theta}_j(1 - \hat{\theta}_j)}$$

where $\hat{\theta}_j$ is the MLE of $\theta_j$ under $H_k$.

The model with the smallest BIC can be thought of as the one with the best fit, adjusting for the number of parameters. Also one can give a Bayesian interpretation to the BIC (see Raftery, 1995). If one associates prior probabilities $\pi_1, \pi_2, \ldots, \pi_k$ to the models $H_1, H_2, \ldots, H_k$, respectively, then the posterior probabilities, after incorporating the data, are

$$P(H_r|\text{data}) = \frac{\exp(-\frac{1}{2}\mathrm{BIC}_r)\pi_r}{\sum_{i=0}^{k}\exp(-\frac{1}{2}\mathrm{BIC}_i)\pi_i}, \qquad r = 1, \ldots, k$$

In particular for a uniform prior (i.e., all models having same credibility *a priori*), the $\pi$s drop out of this expression, and then clearly the model with the smallest BIC is the one with the largest posterior probability.

In implementing this procedure, there is the problem of specifying a maximum possible number of change points, $k$. While, in principle, one could set $k = m - 1$, corresponding to different hazard of burning in each period, there are difficulties with doing this. First, it would lead to a considerable computational load, since under a given model $H_r$, the MLEs of the $r$ change points are found by direct search. Second, there is the problem of estimating the overdispersion parameter $\sigma^2$, which is estimated under the largest model contemplated. If $k$ is set too large, there will be few degrees of freedom for estimating $\sigma^2$. In the following examples, 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.

### Example 3. 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 subunits with distinct fire hazard rates were unsuccessful. The time-since-fire distribution for the whole study area is displayed in Figure 3. There are $m = 40$ age classes of width $T = 10$ yr. The lower panel (a logarithmic 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 3 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

**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 Kananaskis Watershed time-since-fire data)**

| 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 4 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)

| Epoch $i$ | Date | Fire cycle (yr) MLE | Fire cycle (yr) 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 |

the most plausible. Since the change point at 4 (1940) and 6 (1920) appear as MLEs for all 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).

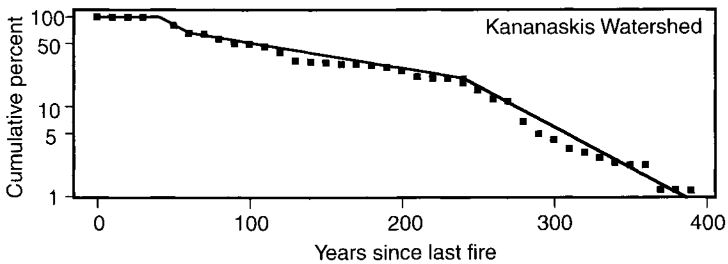
Of course by varying the prior probabilities of the various models, one can change the posteriors. However, a substantial skewness in the prior is required to shift the highest posterior probability from $H_3$. Thus, one can conclude that the data contain strong support for the three-change-points model, with the estimated changes occurring around 1940, 1920, and 1750.

The MLEs of the fire cycle (inverse of the hazard rate) and 95% likelihood ratio confidence intervals (Reed *et al.*, 1998) in the four epochs separated by the three identified change points are displayed in Table 4. The MLEs are also shown as line segments superimposed on the semilog cumulative frequency plot in Figure 5 (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 negligible. In fact, less than one tenth of one percent of the whole study area burned in the 40-yr 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 preceding results. However, they failed to identify the more recent change points identified here.

### Example 4. Glacier National Park

Figure 4 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 being of the Interior Wet Belt Forest type. The data were presented in 20-yr age classes ($T = 20$, $m = 21$). Table 5 gives details of MLEs, BICs, and posterior probabilities under a uniform prior for models $H_0, H_1, \dots, H_6$.



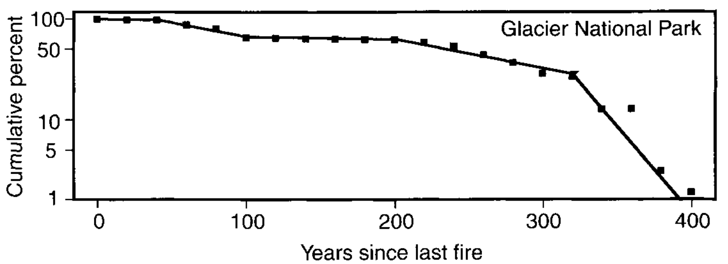




FIGURE 5 Epochs of constant hazard of burning as determined by the analysis in Section IV for Kananskis Watershed and Glacier National Park. The points are the cumulative percentage areas (on logarithmic scale) with time since fire exceeding the time (in years) on the horizontal axis. The ranges of the line segments correspond to distinct epochs of constant hazard of burning; the slopes of the segments correspond to the maximimum likelihood estimates of the hazards.

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. To shift the posterior mode from $H_4$ requires a very substantial skewness in the prior distribution. Consequently 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,

TABLE 5 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)

| 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 6 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)**

| Epoch | | Fire cycle (yr) | |
|---|---|---|---|
| $i$ | Date | 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 |

1880, 1780, and 1660). There is some possibility of a fifth change point, estimated at 1740. MLEs 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 6. The MLEs are also shown as line segments superimposed on the semilog cumulative frequency plot in Figure 5 (bottom panel).

It is worth noting that a very similar model was identified for these data using backward elimination methods (Reed, 1998), the only difference being the change point in the 18th century was estimated at 12 (1740) rather than 10 (1780).

## REFERENCES


Heinselman, M. L. (1973). Fire in the virgin forests of the Boundary Waters Canoe Area. *Quat. Res.* 3, 329–382.

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

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

Larsen, C. P. S. (1996). Spatial and temporal variations in boreal forest fire frequency in northern Alberta. *J. Biogeog.* 24, 663–673.

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

Reed, W. J. (1997). Estimating historical forest fire frequency from time-since-fire sample data. *IMA J. Appl. Med. Biol.* 14, 71–83.

Reed, W. J. (1998). Determining changes in historical forest fire frequency from a time-since-fire map. *J. Agri. Biol. Environ. Stat.* 3, 430–450.

Reed, W. J. (2000). Reconstructing the history of forest fire frequency—Identifying hazard rate change points using the Bayes' Information Criterion. *Can. J. Stat.* 28, 353–365.

Reed, W. J., Larsen C. P. S., Johnson, E. A. and MacDonald, G. M. (1998). Estimation of temporal variations in fire frequency from time-since-fire map data. *For. Sci.* 44, 465–475.