Parametric models of the fire-size distribution

S. G. Cumming

Department of Renewable Resources, University of Alberta

August 5, 1999

Abstract

This paper characterises the size distribution of wildfires in the boreal mixed-wood forests of Alberta Canada, for the interval 1980–1998. For fires exceeding 2.5 ha, the logarithm of fire size is exponentially distributed, according to standard tests. This result is motivated by an apparent exponential distribution of fire durations (in days) and an empirical power law relating duration to size. However, computer simulations using the estimated distribution would seriously over-predict the frequency of large fires. A truncated exponential distribution can be used to model fire with bounded sizes. This upper bound can not be estimated from the data by standard methods, including the statistical analysis of extremes. A non-parametric bootstrap method combined with least squares estimation produces a satisfactory estimate. Parametric modelling of the fire size data using forest inventory data as covariates shows that the expected size of a fire is positively related to the abundance of pine forest in the vicinity of the point of detection, and negatively related to the abundance of recently logged or burnt areas. Variation in forest structure and disturbance history impose marked spatially variability on the fire size distribution.

1 Introduction

In this this paper, I present a simple parametric model which describes, for recent decades, the fire size distribution in a large region of the boreal mixedwood forest of Alberta, Canada. I believe this model to be of both theoretical and applied interest. The theoretical interest consists partly in the fact that it has apparently never been done (He and Mladenoff, 1999), and partly in the way in which the model emerges naturally from the distribution of fire durations and a relationship between duration and size. The applied interest lies in the statistical and simulation modelling of fire dominated landscapes. With a parametric model, spatial and temporal covariates that effect fire size can be studied more rigorously than has been possible. Also, with such a model, the fire component of the growing number of spatial dynamic models of large forested landscapes can be placed on a firm empirical foundation. The latter application was the principal motivation for this study. My methods should be applicable at the least to any part of the circumpolar boreal forest for which minimal fire history data exist. Specifically, a few decades of empirical fire size data will probably suffice.

There have been few statistical analyses of fire size distributions. Strauss et al. (1989) introduced the extreme proportion function $EP(p), 0.0 \leq p \leq 1.0$ which they define as the "proportion of all the area burnt that is attributable to the largest $100\,p\%$ of the fires." They present expressions for $EP(p)$ for a variety of underlying size distributions, and propose the truncated Pareto distribution as a model of fire size. This truncation places an upper bound on fire size. However, they do not fit their model to empirical size data, noting that the task is complicated by the large number of imprecisely measured small fires. On the basis of visual inspection of empirical EP functions, they conclude that "the degree of size inequality" between fires in southern California and Baja California chaparral systems is much the same. Chou et al. (1993) take issue with these informal methods, introduce an alternate functional form for the EP function, and resort to ANOVA to conclude that fire size distributions in fact differ between the two area. I have used non-parametric statistics to demonstrate temporal trends in the fire size distribution in Alberta, consistent with increases in fire suppression effort or effectiveness (Cumming et al., 1995, Cumming, 1997). Most recently, Moritz (1997) has applied the statistical analysis of extremes to annual maxima of fire sizes. He tested for differences in size distributions between two regions of the Los Padres National Forest, another southern California chaparral system.

A taxonomy of fire modelling strategies used in spatial dynamics models is beyond the scope of this paper, but two common methods are based on percolation or cellular automata models. A fire burning in a given model cell can spread to adjacent cells according to a probability, which may depend upon cell characteristic such as forest type or age. In some formulations, for each simulated ignition, a fire size is drawn from an empirical distribution (Andison, 1998) or from a continuous density function with the first moment fit to empirical data (He and Mladenoff, 1999). The fire then spreads randomly until the specified size is reached. In other cases, a size distribution emerges from the simulation, varying as hypotheses regarding *e.g.,* age-specific flammability are tested(Ratz, 1995, Li et al., 1997). It is also possible to tune the spread parameter(s) so as to approximate an empirical distribution, although this is seldom done.

These models have certain common features. Fires smaller than the model resolution don't count, which I will argue is an advantage. There is insufficient care given to matching the actual size distribution. For example, percolation models imply an exponential size distribution, while the lognormal distribution adopted by He and Mladenoff (1999) is judged suitable merely because most fires are small, but some large fires emerge: they prevent fires from becoming too large by censoring at an historical maximum size. However, getting the size distribution right is important, especially at the tail where individual fires actually matter in structuring the landscape. For example, some models are concerned to match real patch size distributions because of their presumed utility for forest management. Others model distance-limited seed dispersal, so that recruitment of tree species into burns is limited by fire size. Finally, no approach is amenable to studying the effects of changing conditions such as climate, fire suppression strategy, or forest pattern, as the landscape is fragmented, or at least structurally altered, by harvesting.

The present research is directed at simulating the dynamics of natural and managed boreal forests, at large spatial and temporal scales. The target modelling platform runs at ≈ 100 km^2 spatial resolution and annual or multi-annual time steps. The time horizon is 200 yr, which reflects long-term forest management planning in Alberta. The spatial extent of the simulations is $\approx 1 \times 10^6$ ha or more. Each model cell maintains a detailed description of forest age and size structure and species composition, similar to the spatially aggregated model of Armstrong et al. (1999). Although fires of any size may be simulated, there is reason to set a lower bound. This is because the size distribution of small fires is probably governed by several different random processes: the onset of crowning, at which point the fire becomes, so to speak, autonomous, and by fire suppression efforts, which are most effective when the fire is very small (Hirsch et al., 1998). These factors present a difficult statistical modelling problem, which vanishes when only fires exceeding a certain minimum size are considered. Since these small fires, though numerous, have little aggregate effect on the landscape, they may be ignored in most applications.

I will show that, for fires larger than a few ha, the logarithm of fire size follows an exponential distribution. To address the maximum size problem, I introduce the truncated exponential distribution, and show how the upper limit can be estimated from historical data. This distribution is amenable to parametric modelling of the effects of covariates on expected fire size. As an example, I show that the expected fire size is related to forest composition in the vicinity of the fire's point of detection.

2 Data and methods

My study area is a $\approx 86,000$ km^2 rectangle in northeastern Alberta, Canada, bounded by 55° N, 110° W and 58° N, 115° W (Fig. 1). Most of the study area is contained in the boreal mixedwood ecological region (Rowe, 1972). The mixedwood region, of total extent $\approx 485,000$ km^2, is transitional between colder, conifer-dominated forests to the north and warmer, dryer aspen parklands to the south, which are now mostly farmland. In Alberta, $\approx 270,000$ km^2 of the mixedwood is still forested (Strong, 1992). The most abundant tree species are trembling aspen, black spruce (*Picea mariana* (Mill.)

Figure 1: Location of the study area in Alberta, Canada. Reproduced from Armstrong (1999).

B.S.P.), jack pine (*Pinus banksiana* Lamb), white spruce (*Picea glauca* (Moench) Voss) and balsam poplar (*Populus balsamifera* L.). Paper birch (*Betula papyrifera* Marsh.), tamarack (*Larix laricina* (Du Roi) Koch) and balsam fir (*Abies balsamea* (L.) Mill) are widely distributed, but rarely form pure stands. Mature mixed stands containing both aspen and white spruce are characteristic of the region. Peatlands and sparsely treed muskeg cover about half of the study area. The regional forest types are described by Kabzems et al. (1986). Climate, topography and geomorphology are reviewed by Strong (1992).

2.1 Fire history data

The Government of Alberta's Department of Environmental Protection (AEP) maintains databases of fire records from 1961–1998 (AEP 1998). Fire attributes recorded over this interval include the location and date of detection, the date of extinction, the final size and an indication of cause, whether by lightning or human agency. This study considers lightning fires only. Locations are determined by triangulation from a network of fire towers and by a lightning detection system. Prior to 1996, fires are spatially registered to land-survey units called townships, most of which are square regions exactly six miles ($\approx$9.66 km) on a side ($\approx$93.2 km^2). Since 1996, only the latitude and longitude of detection were recorded. I used the coordinates of township centroids to infer in which township these fires started. Tests of fires where both the township and latitude and longitude were recorded showed the method to be almost 100% accurate.

Although the period of record begins in 1961, observational effort has not been uniform. The network of fire towers was not completed until 1968 (Murphy, 1985). Also, fire suppression policy and strategy changed markedly between 1961 and 1983 (Murphy, 1985, Gray and Janz, 1985). The three years 1980–1982 were the most severe on record (Armstrong, 1999), and contained most of the largest recorded fires. Thus, summer climate, or fire weather conditions, have not been uniform either. As a compromise, I select 1980–1998 as my study interval for most of the analyses, and defer the problem of annual variability in fire suppression effort and fire weather to a future study. I will assume however that the largest fire burning in a given year was reliably recorded since 1961. Such fires seem unlikely to have escaped notice if large, and would not effect the analysis if small.

2.2 Forest inventory data

I used the Alberta Phase 3 forest inventory (Alberta Forest Service, 1985) to describe the forest composition in the vicinity of each fire. This inventory was interpreted from 1:15,000 scale aerial photography, flown between 1970 and 1982. Phase 3 data are available as 1:15,000 paper maps and as machine readable extracts from the "Alberta Forest Service Inventory Storage and Maintenance System" (AFORISM) database. Both versions are spatially organised to township resolution. AFORISM data are stand lists, with no representation of the underlying topology. I refer to the township in which a fire was first detected as the fire's *locale*.

Mapped forested polygons are regions of uniform canopy attributes, where the canopy is defined as the tallest strata of trees which has at least 6% canopy cover. Canopy attributes include height and crown-closure, species composition, area, and an estimate of stand age. Deciduous species (aspen, balsam poplar and birch) are not usually distinguished. Canopy species composition is expressed as proportions of estimated merchantable volume for stands > 12 m in height, and by proportion of crown closure otherwise. Various classes of nonforested polygons such as wetland types, bodies of water, burned areas, clearcuts and clearings are also mapped. For most burns and cuts more recent than 1980, the year of the disturbance is recorded. The minimum mapping unit is 2 ha. AFORISM data for the study area were provided by AEP. The only attributes used in the present study are area (in ha), the species composition of forested polygons, and the class attribute of unforested polygons.

I assign mapped polygons to one of six classes: Deciduous, White spruce, Black spruce, Pine and Muskeg and Disturbed. Class Muskeg includes all non-forested areas, predominantly wetlands, excepting open water. Class Disturbed is that portion of a locale which, at the year of a fire, had recently been burnt or logged. The four forested classes are based on the polygon's dominant species (or species type, in the case of the Deciduous class). The characteristic mixed stands are thus classed as either White spruce or Deciduous. To make the classification exhaustive, I consider balsam fir to be equivalent to white spruce, and larch to be equivalent to black spruce. This reflects the successional relationships and/or site associations of these species (Kabzems et al., 1986). A locale is described by a vector of the proportional area in each class. Because forest cover is spatially autocorrelated at a scale of 20 km or more (Cumming et al., 1996), the composition of a locale estimates the pre-fire composition of the surroundings of even very large fires.

In order to model the effect of forest cover on fire size, it is necessary to estimate the composition of a locale at the time the fire started. This is complicated, because, after logging or fire, the AEP databases are updated to reflect the disturbance, and update records are not maintained. The update process splits any polygons partially burnt or cut. For example, the area of polygon 86A which burnt in 1995, can be unambiguously assigned to the same class as unburnt polygon 86 for any fire burning before 1995. The development of the forest industry in the area is such that any stand marked as clearcut prior to 1993 may be assumed to have belonged to class White spruce. My primary data sources were a 1998 snapshot of the inventory for the entire study area, and a 1993 snapshot for most of the study area. In addition, records for 284,000 ha of burnt stands had been recovered from the original map annotations in the course of another study. From these data sources and rules, I reconstituted the pre-fire composition of 291,000 ha of burnt or logged stands. Fires for which than 5% or more of the pre-fire forested area of the locale could not be classified were excluded from this part of the analysis.

2.3 Statistical methods

Statistical methods will mostly be introduced as needed. Here, I establish some notational conventions. With respect to univariate distributions, my definitions and notation will hue closely to Johnson et al. (1994, 1995). Random variables are denoted by X and Z. Probability density functions are written $p_X(x)$ which denotes the given density function for X, evaluated at x. The relational symbol $\sim$ means "is distributed as." Cumulative density functions are denoted

$$F_X(x) = \Pr[X \leq x] = \int_a^x p_X(x)dx,$$

where a is the lower bound of the domain of p, usually 0 for the distributions considered here. To avoid excess notation, it will always be clear from context to which distribution p or F refer.

Distributional parameters are either lower case Greek letters, or in one case, the symbol b. Parameter estimates are hatted (*e.g.*, $\hat{\sigma}$). Sample means are barred (*e.g.*, $\bar{x} = n^{-1} \cdot \sum_{i=1}^{n} x_i$). Where dependency on parameters $\boldsymbol{\sigma}$ needs emphasis, a probability density function will be written $p_X(x; \boldsymbol{\sigma})$. Fitting a distribution to data means estimating the parameters from a sample $\{x_1, x_2, \ldots, x_n\}$. In most cases, I do this by finding the maximum likelihood estimate, which is that value $\hat{\boldsymbol{\sigma}}$ for which

$$\prod_{i=1}^{n} p_X(x_i; \boldsymbol{\sigma})$$

is maximised. This is equivalent to minimising

$$L = \sum_{i=1}^{n} -p_X(x_i; \boldsymbol{\sigma}),$$

which must usually be done numerically. L is the *log-likelihood* of the parameters, given the data. Unless otherwise specified, I use the general function minimisation routine `ms` of Splus version 3.3 (MathSoft Inc., 1995) for estimation. Model 1 is said to be nested in Model 0 if it can be reduced to Model 0 by setting p parameters to zero. Whether Model 1 is a significant better model, given the additional p parameters, is tested by the likelihood ratio statistic $S = -2(L_0 - L_1) \sim \chi^2_p$.

When maximum likelihood estimates are not useful, I use bootstrap estimators. Given a set of observations $\{x_1, x_2, \ldots, x_n\}$, a bootstrap sample is generated by randomly sampling n elements from the set, with replacement. By applying an estimation procedure to a large number of bootstrap samples, one can determine the mean and standard deviation of the estimated parameter(s). My source for this method, and for maximum likelihood methods, is Hilborn and Mangel (1997).

3 Motivation and tests for an exponential form

Between 1980 and 1998, 295 lightning fires ≥ 9 ha started in the study area. The lower threshold of 9 ha is the spatial resolution of a companion landscape modelling initiative beyond the scope of this paper. The effect of varying this threshold is explored below.

The diurnal patterns of fire arrival and the cumulative number of fires burning (N_i) can be reconstructed from the recorded dates of fire detection and extinction. A typical pattern of multiple arrivals and subsequent decline in N_i is illustrated in Figure 2. The pattern suggests a constant fire decay rate or, equivalently, a constant daily probability of extinction, so that fire durations (in days) are exponentially distributed. The natural logarithms of fire duration and size are linearly related (Figure 3). These two observations suggest the exponential distribution as a model for the logarithm of fire sizes.

The random variable X has an exponential distribution if it has a probability density function (pdf) of the form

$$p_X(x) = \sigma^{-1} \exp\left[-\frac{(x-\theta)}{\sigma}\right], \quad x \geq \theta; \ \ \sigma > 0. \tag{1}$$

The cumulative density function (cdf) is

$$F_X(x) = 1 - \exp(-(x-\theta)/\sigma).$$

The maximum likelihood estimate (mle) for the shape parameter σ is the sample mean $\bar{x} = 2.56$. Note that if $z = \exp(x)$ is the fire size in ha, and $t = \exp(\theta)$ is the threshold size ($z \geq t$), then $x - \theta = \log(z/t)$, which is the log of the fire size in units of size t, for example, the grid resolution of a raster-based spatial simulation model. In what follows, I therefore dispense with the location parameter θ and deal in scaled units $x = \log(z/t)$, with $t = 9$ ha in most cases.

The Anderson-Darling goodness-of-fit statistic A^2 is a recommended test for exponentiality, with "generally good power properties over a wide range of alternative distributions" (Kotz et al., 1982). It is defined as

$$A^2 = -\left\{\sum_{i=1}^{n}(2i-1)[\log F_X(x_{i:n}) + \log(1 - F_X(x_{n+1-i:n})]\right\}/n - n.$$

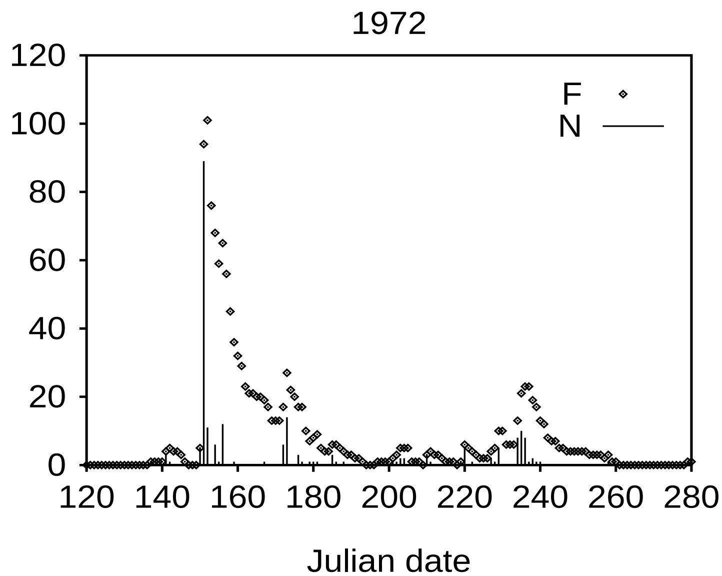

Figure 2: Diurnal arrivals of detected lightning fires (N) and the total number of fires burning (F) in a subregion of the study area, between April and October of 1972. This pattern of multiple arrivals and apparent exponential decay within a year is typical. Reproduced from Cumming (1997).

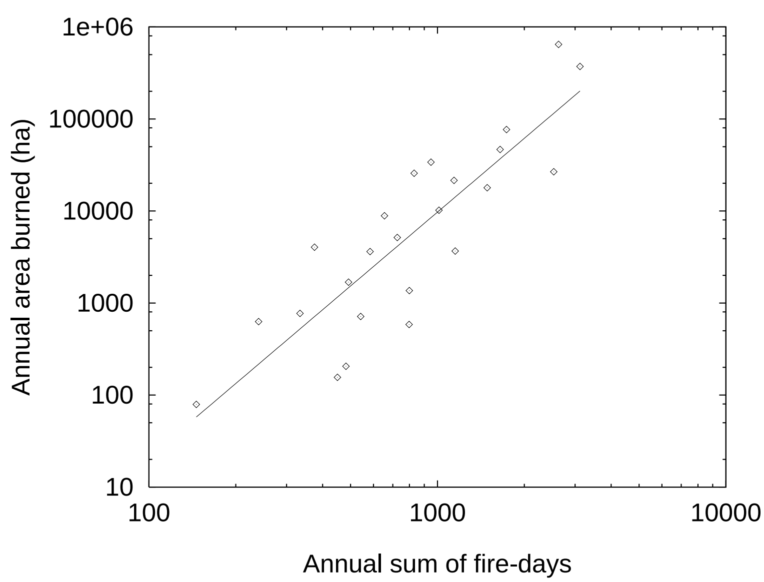

Figure 3: The summed durations, in days, of all fires burning in a given year (x) is related to the annual total area burnt in ha (y). The relation is shown for the interval 1970–1993, as $\log y = -9.22 + 2.66 \log x$, ($r^2 = 0.74, p \ll 0.001, n = 24$). The relation between the duration and size of individual fires also fits a linear model in log-log space, but is markedly heteroscedastic. Reproduced from Cumming (1997).

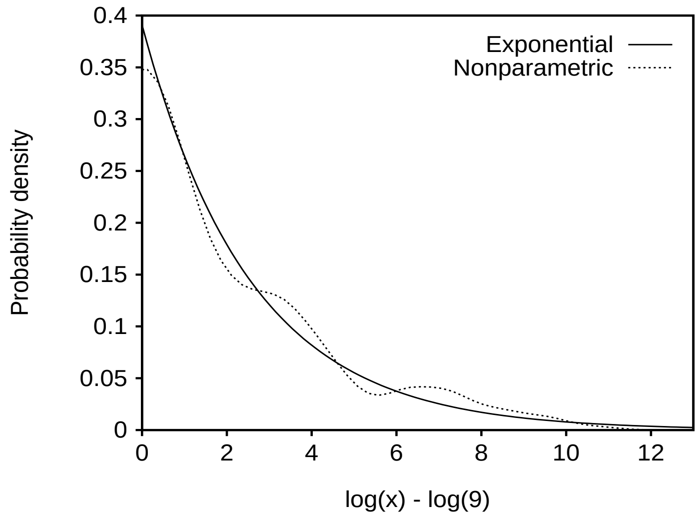

Figure 4: The best fit exponential distribution of the logarithm of fire size, for fires exceeding 9 ha, left shifted to the origin. A smoothed representation of the empirical probability density of the data is shown for comparison—see the text for an explanation.

where the notation $x_{i:n}$ denotes the i-th smallest sample. I computed the test statistic and associated p-values by the methods of Davis and Stephens (1989), with the requisite modification for when σ is estimated from the data. The hypothesis that the log-transformed fire size data are exponentially distributed cannot be rejected ($A^2 = 0.621$, $p = 0.351$.) This result holds when $\exp(\theta) \geq 3$ ha (Table 1). Thus, t may be selected to suite the target application(s) of the analysis. Figure 4 compares the fitted pdf to the empirical distribution, described by a non-parametric density estimate computed using a Gaussian filter kernel of bandwidth 2.010 (Venables and Ripley, 1997, Ch. 5.5). This is essentially a smoothed histogram of optimal bin-width.

3.1 The Gamma distribution

Although the exponential distribution provides an adequate statistical model of the data, I will show that it is unsuitable for the intended application, because it over-predicts the frequency of very large fires. One solution to this sort of difficulty is to adopt a size distribution which resembles the exponential, but which has less weight at the tail, so that *e.g.,* large fires are less probable. The gamma distribution is a commonly recommended candidate.

The random variable X has a gamma distribution if its pdf is of the form

$$p_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha},(\alpha)}, \quad \alpha > 0,\ \beta > 0,$$

where , is the gamma function. The family of gamma distributions includes the exponential as a special case ($\alpha = 1$) but also includes lighter-tailed distributions. I obtained

Table 1: Sample sizes (n_t), Anderson-Darling test-statistics (A_t^2), p-values, and estimated scale parameter $\hat{\sigma}_t$ for the fire size distribution, truncated at various lower bounds ($t = \exp(\theta)$ in Eqn. 1). An exponential distribution fits the data when $t > 2,5$ ha. Fires are all lightning fires ignited in the study area from 1980–1998.

t	n_t	A_t^2	p_t	$\hat{\sigma}_t$
2	557	2.378	0.004	2.47
2.5	508	1.994	0.009	2.47
3	463	0.947	0.137	2.51
4	397	1.031	0.101	2.61
10	280	0.628	0.345	2.59
20	209	0.862	0.174	2.65
40	165	1.044	0.106	2.58
100	123	0.790	0.214	2.39
200	97	0.815	0.199	2.25

the mles $\hat{\alpha} = 1.055$, $\hat{\beta} = 2.428$ by minimising the log likelihood

$$L_g = \sum_{i=1}^{n} -(\alpha - 1)\log x_i + x_i/\beta + \alpha \log \beta + \log, (\alpha)$$

For the gamma distribution $L_g = 471.102$, as compared to $L_e = 471.366$ for the exponential form ($\alpha = 1.0, \beta = 2.56$). The two models are nested, with $S = -2 * (L_e - L_g) = 0.528$ which is distributed as a χ_1^2 ($p = 0.467$). The data give no warrant to prefer the gamma distribution. However, as noted, it might be preferable because of its extremal properties.

4 The problem of large fires

Simulations of forest dynamics over large spatial and temporal scales require repeated sampling from the chosen fire size distribution. The exponential and gamma distributions are defined for all $x \geq 0$, so their use in a stochastic model will result in simulated fires arbitrarily larger than real fires are observed to be. The largest fire in the study area since 1961 was a 1981 fire of $\approx 428,000$ ha, while the largest fire known to have burnt in Alberta is the 1.4×10^6 ha Chinchaga River fire of 1950 (Johnson, 1992), which actually started in British Columbia.

A landscape simulation of the study area would be expected to generate about $295/19 = 15.5$ large fires per year. Assuming for now that fire sizes are independent between and within years, the probability of all simulated fires being smaller than x in a given year is $q(x) = F_X(x)^{15.5}$ and the probability of there being at least one fire larger than x is $p(x) = (1 - q(x))$. Then the interval in yr between fires of size at least

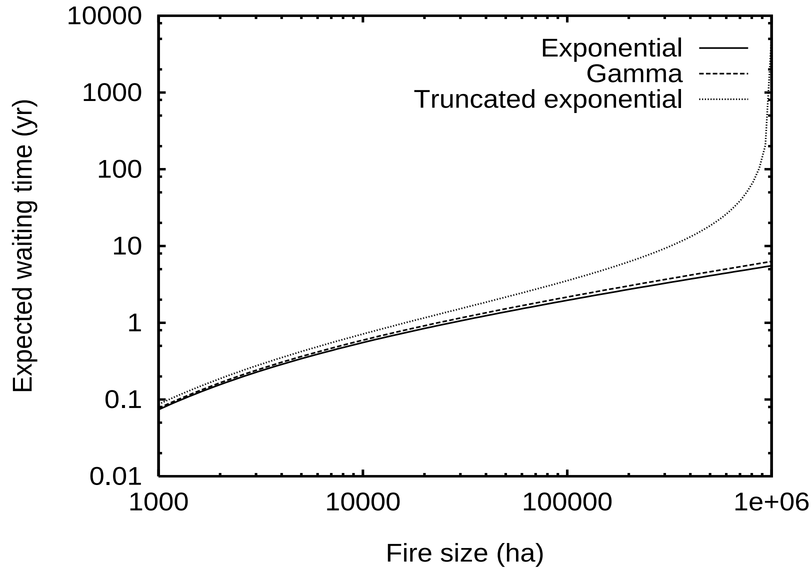

Figure 5: Expected waiting time for fires of size $\geq x$, in yr, plotted for the best fit exponential and gamma distributions. These unbounded distributions over-predict the frequency of large fires. The relation for the truncated form of the exponential is also shown (see Section 4.1 for explanation.)

x follows a geometric distribution with mean $R(x) = q(x)/p(x)$ (see *e.g.,* Fraser, 1996). The function $R(x)$ is shown for the exponential and gamma distributions in Figure 5. Under the exponential distribution, the expected waiting times for fires of 1×10^5 ha, 5×10^5 ha and 1×10^6 ha are 2.0 yr, 4.1 yr and 5.5 yr, respectively. The values are only very slightly greater for the lighter-tailed gamma distribution. The empirical value for fires of $\geq 1 \times 10^5$ ha is 2.8 yr, computed from the 1980–1998 records using the same independence assumptions. However, only one fire exceeded 2×10^5 ha over this interval. Clearly, both exponential forms radically over-predict the frequency of large fires, so neither is suitable for use in landscape simulations. In the remainder of this section, I present a modified form of the exponential distribution which does appear to be suitable.

4.1 The truncated exponential distribution

By definition (Kendall and Stuart, 1973, p. 542), a variate with pdf $p_X(x;\sigma)$, doubly truncated at known points a, b with $a < b$, has pdf

$$p_{X_t}(x) = \frac{p_X(x;\sigma)}{\int_a^b p_X(x;\sigma)}; \quad a \leq x \leq b, \tag{2}$$

where the subscript t indicates truncation. For the case of an exponential distribution with $a = 0$ and b some value exceeding the largest sample size,

$$p_{X_t}(x) = \frac{\sigma^{-1}\exp(-x/\sigma)}{1-\exp(-b/\sigma)} \quad (3)$$

By integration, the cdf for the truncated exponential is

$$F_t(x) = \frac{1-\exp(-x/\sigma)}{1-\exp(-b/\sigma)}.$$

As this has a simple closed-form inverse function (x can be derived from $F_t(x)$ by elementary algebra), random deviates from the distribution are easy to generate from the standard uniform distribution (Press et al., 1988). That is, the distribution is easy to use in simulation models.

The mle for σ conditional on b is obtained by minimising

$$L_t = \sum_{i=1}^{n} \log(\sigma) + \frac{x_i}{\sigma} + \log(1-\exp(-b/\sigma)) \quad (4)$$

In a modelling application, a suitable value of b could be chosen, and the appropriate instance of the truncated exponential distribution used to generate random samples of fire sizes. This seems more elegant than drawing from the unbounded exponential or gamma distribution and then censoring at b. Better still if b could be estimated from the data.

4.2 How big may fires get?

The maximum likelihood estimate of b is just the sample maximum. This is easy to see from Eqn. 4. For $\sigma > 0$, $\exp(-b/\sigma)$ is monotone decreasing in b, so $\log(1-\exp(-b/\sigma))$ becomes smaller (more negative) as b approaches the sample maximum from above. Since this is the only term in L_t in which b appears, the result follows. This estimate is unsatisfactory. The period of record is short relative to the intended length of simulations (200 yr), so fires larger than observed must surely be allowed for.

One way to proceed is to use the extreme value distribution (Johnson et al., 1995, Ch. 22) to estimate the maximum expected fire size over a time horizon of interest. The approach does not actually work in this instance. However, I think the failure is worth documenting, as extreme value methods have been used in a previous study of fire sizes (Moritz, 1997), without a formal evaluation of their suitability. To actually solve the problem at hand, I use a non-parametric bootstrap procedure to generate a distribution of least squares estimates of b and σ jointly.

4.2.1 Estimation from extremal values

From Chapter 22 of Johnson et al. (1995), if $Z_1, Z_2, \ldots, Z_n$ are independent, identically distributed exponential random variables, then the limiting distribution (as $n \to \infty$) for the sample maximum has cdf

$$H(z) = Pr[Z \le z] = \exp\left\{-e^{-(z-\xi)/\theta}\right\} \quad (5)$$

Armstrong (1999) has shown that the annual total area burned in the study area can be modelled as a serially independent log-normal variate. Since the annual maxima of fire sizes are typically a large proportion of annual sums, it seemed plausible that they satisfy the distributional assumptions. Unfortunately, this is not the case. From 38 years of annual maxima (1961–1998), I computed the mles $\hat{\xi} = 4703.9$ and $\hat{\theta} = 18468.1$ by an iterative procedure given in section 22.9.6 of Johnson et al. (1995). I started the iteration with a simple linear estimator of θ based on order statistics (Johnson et al.'s (1995) Eqns. 22.75 and 22.77). The Anderson-Darling test statistic, computed as per Johnson et al.'s (1995) Table 22.19, was $A^2 = 8.15$, which greatly exceeds the 0.99 upper tail percentage value of 1.038 for H. Clearly, at least one distributional assumption fails.

Alberta's fire suppression strategies changed in 1983, in response to the severe 1980–1982 fire years (Gray and Janz, 1985). Inter-annual variation in fire suppression effectiveness or in fire weather conditions may violate the assumption that the Z_i are identically distributed. I therefore also tested the annual maxima for the intervals 1980–1998, 1983–1998 and 1961-1979. The resultant A^2 statistics were 3.60, 3.28 and 1.82 respectively. The distribution of Eqn. 5 just does not fit the data, and cannot be used validly to estimate b. Notably, fires $> 100,000$ ha are predicted to recur every 174.7 yr (determined by $1/(1 - H(1e5; \hat{\xi}, \hat{\theta}))$. In fact, there have been at least 5 such fires in the study area since 1961.

An alternate and reportedly more general form of the cumulative extreme value distribution was introduced to the ecological literature by Gaines and Denny (1993):

$$H'(x) = \exp - \left[\frac{\alpha - \beta x}{\alpha - \beta \epsilon}\right]^{1/\beta} \tag{6}$$

Moritz (1997) has used H' to compare fire size distributions in two southern California chaparral forests, and to test for changes resultant from the introduction of water-bombers in the 1950s. Although he presents informal arguments that the distributional assumptions are satisfied, he does not compare his fitted distributions with the underlying data.

From the likelihood expression for Eqn. 6 (Gaines and Denny, 1993, Appendix A), I obtained mles of $\hat{\alpha} = -4.333$, $\hat{\beta} = -3.108$ and $\hat{\epsilon} = 300.25$. This distribution is also unsatisfactory. Although no appropriate goodness-of-fit test seems to be available, this distribution radically over-predicts the frequency of very large fires (Figure 6). The expected return interval for 1×10^5 ha fires is 7.0 yr, compared to the empirical 6.6 yr over the 38 yr sample interval. However, the return interval for 1×10^6 ha fires is 14.1 yr, yet only 1 has been observed in Alberta since 1940. The "200-year fire" is an excessive $\approx 4.20 \times 10^9$ ha. The lesson from these two trials is that a maximum likelihood parameter estimate for a statistical model is not sufficient. One must also check that the fitted model is reasonable.

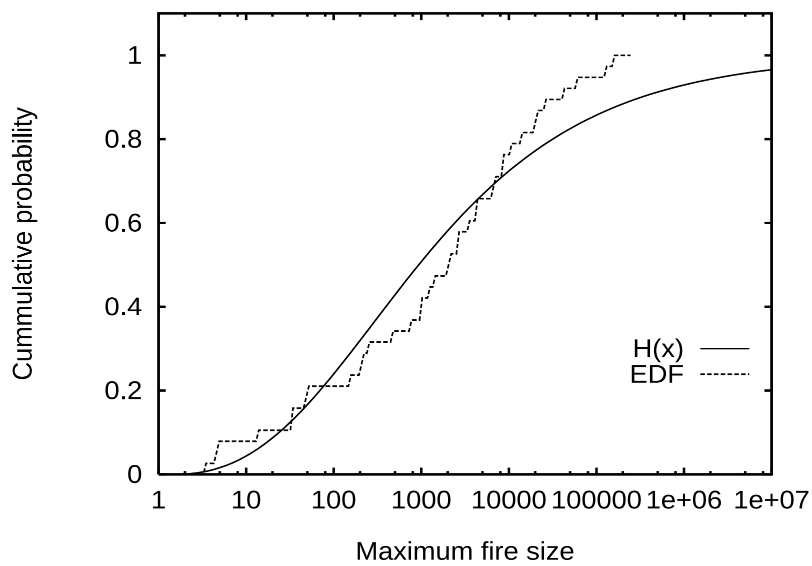

Figure 6: The extreme value function $H'(x)$ of Eqn. 6 with parameters estimated by maximum likelihood, plotted with the empirical cumulative density function of maximum annual fire sizes (EDF). Note the high predicted probability of extremely large fires.

4.2.2 Estimation by sum of squares

By definition, the expected value of the truncated exponential distribution ($p_{X_t}(x)$ of Eqn. 3), conditional on the parameters, is

$$\mu(\sigma, b) = \int_0^b x\, p_{X_t}(x)\, dx \tag{7}$$

$$= (e^{-b/\sigma} - 1)\left[e^{-b/\sigma}(\sigma + b) - \sigma\right] \tag{8}$$

For 10,000 bootstrap samples of the fire size data from 1961–1998 (485 fires $\geq$ 9), I found pairs (σ_j, b_j) which minimised the sum of squared error

$$\sum_{i=1}^{n} (x_i - \mu(\sigma_j, b_j))^2$$

over the j-th sample, using Press *et al.*'s (1988) implementation of the Nelder and Mead downhill simplex method. The bootstrap parameter estimates have mean $(\bar{\sigma}, \bar{b}) = (2.651, 11.621)$ and covariance matrix

$$\begin{bmatrix} 0.013 & 0.042 \\ 0.042 & 0.318 \end{bmatrix}.$$

Table 2: Recurrence times for fires of various sizes, estimated from the empirical counts for the intervals 1961–1998 and 1980–1998, and as computed from the estimated parameters of a truncated exponential distribution, based on 1961–1998 fires.

	Fire size				
	1×10^4 ha	5×10^4 ha	1×10^5 ha	4.3×10^5 ha	9×10^5 ha
1961–1998	0.58	4.4	6.6	37	∞
1980–1998	0	1.7	2.8	18	∞
$R(x; \hat{\sigma}, \hat{b})$	0.71	2.2	3.5	14.6	136.5

The correlation coefficient of the bootstrap estimates is 0.658. The estimated maximum fire size is $\exp(11.621 + \log(9)) = 1.003 \times 10^6$ ha. The mle of σ, conditional on the bootstrap estimate for b, is 2.568. It is unclear which estimate of σ ought to be used. The log-likelihoods (L_t of Eqn. 4) are not significantly different. As the conditional mle provides a slightly better fit to the empirical distribution of large fires, I chose to adopt it. The waiting time function $R(x)$ for the truncated exponential is shown in Figure 5, alongside those for the exponential and gamma distributions. Some values for specific fire sizes are tabulated in Table 2, with the empirical values for the intervals 1961–1998 and 1980–1998, computed from the historical records under the assumptions of a geometric waiting time distribution. The truncated exponential matches the observed frequency of large fires quite well. The difficulty of over-prediction associated with the exponential and gamma distributions does not arise. $R(9.31 \times 10^5 \text{ ha}) \approx 200$ yr, so some very large fires can be expected in the course of long-range simulations.

When the sample is restricted to the years 1980–1998, the many large fires of 1980–1982 inflate the estimate of b, resulting in a maximum fire size of $\approx 1.9 \times 10^6$ ha. I suspect that this sample interval places too much weight on the contingent occurrence of three severe years, and chose to proceed with the estimate based on the full 39 years of data.

5 Modelling spatial covariates

Here, I use the truncated exponential form to model the effect of local forest composition on expected fire size. Of the 295 sample fires from 1980–1998, 242 could be matched with forest cover data. Unfortunately, most of the largest fires could not be so matched. These typically burnt most of the forested area in their respective locales, and most burnt during 1980–1982, before the earlier of the two inventory snapshots. Thus, reconstruction of the pre-fire locales for these fires was not possible from the available data. To avoid biasing parameter estimates by including the effect of these large fires, I re-estimated the parameters of the truncated exponential for this subsample to be $\hat{b} = 10.557$, $\hat{\sigma_0} = 2.458$, by the methods of the previous section.

Table 3: Models of the effect of a single component of local forest composition on expected fire size. a_0 and a_1 are the model coefficients estimated by maximum liklihood; S is the likelihood ratio statistic of the one parameter model relative to the null model where $a_0 = \log(\sigma_0) = 0.899$ and $a_1 = 0$ (see text); and p is the associated significance level of $S \sim \chi_1^2$.

Forest type	a_0	a_1	S	p
Aspen	0.979	-0.456	0.654	0.419
White spruce	0.938	-0.589	0.350	0.554
Black spruce	0.802	0.604	1.285	0.247
Pine	0.731	1.775	5.063	0.024
Muskeg	1.036	-0.279	0.566	0.452
Disturbed	0.960	-2.354	4.353	0.037

I modelled σ in the likelihood expression of Eqn. 4 as $\sigma = \exp(f(\boldsymbol{y};\boldsymbol{a}))$, where f is a linear combination of the vector of forest cover variates $\boldsymbol{y}$ specified by $\boldsymbol{a}$, the vector of parameters to be estimated. The exponential transformation ensures that σ is positive. The results of the one-term models are presented in Table 3. Only two components of forest cover are significant: the proportions of pine and of recently disturbed area. The best two-term model is

$$\sigma = \exp(0.775 + 2.05y_1 - 2.56y_2) \tag{9}$$

($p = 0.022$), where y_1 is the proportional area of pine dominated stands, and y_2 is the proportion of forested area recently disturbed. Pine stands in the vicinity of a fire tend to increase its expected size, while areas of recently disturbed forest tend to decrease the expected size. This model describes marked spatial variability in the fire size distribution (Fig. 7). No three-term models were significant.

6 Discussion

Past a certain minimum, the logarithm of fire size (x) follows an exponential distribution. I have argued that this arises from an exponential distribution of fire durations, and a log-log relation between size and duration. The exponential distribution is said to be memoryless, in the sense that *e.g.,* "the future lifetime of an individual has the same distribution no matter how old it is at present" (Johnson et al., 1994). Ignoring the limitation which winter places on fire survival, this is tantamount to a Poisson arrival of weather conditions sufficient to extinguish the fire as it exists at a particular time. This seems to be the case, as fire durations do appear to follow an exponential distribution. If we conceive of a fire as an expanding flame front which at any given time describes a roughly elliptical pattern of growth, then the area burnt as a function of time will follow a power-law. Then a log-log relation between duration and size follows. This mental model could be elaborated to incorporate growth and shrinkage of the active

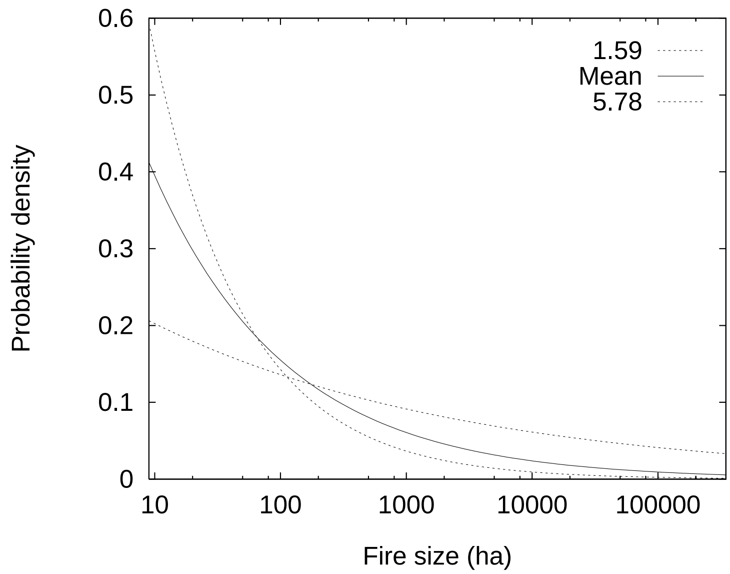

Figure 7: 95% confidence intervals of fire size distributions under the joint variation in abundance of pine and disturbed areas in the sample, based on regression Eqn. 9.

fire front as weather conditions change. However, the exponential characterisation of fire size follows from a simple phenomenological model of fire survival and growth. My subsequent adoption of a truncated exponential form reflects real world limitations on fire survival time and on the area available to be burnt.

The expected size of a fire is positively related to the abundance of pine-dominated stands in the vicinity of ignition, and negatively related to the abundance of previously disturbed areas. This last relation is important, because it implies that either forest management can effect fire behaviour or that burnt areas are refractory to fire for about 10 yr, or both. In either case, such a temporary resistance to fire has a large effect on the behaviour of computer simulations of fire dominated landscapes (Ratz, 1995). This result is the first clear evidence in boreal forests of a relation between patch age as such and the probability of burning: previous reports are contradictory (Antonovski et al., 1992, Johnson, 1992).

The statistical models on which this result is based must be regarded as demonstrative. Many large fires could not be analysed because the pre-fire forest composition of their locales could not be reconstructed from the available data. However, this reconstruction is usually possible, by referencing the original forest inventory maps. This should be done and the data reanalysed, before the relations are incorporated into any landscape models. The difficulties I encountered in this retrospective analysis illustrate the great importance of maintaining audit trails or periodic snapshots of extensive environmental databases.

Another limitation of the models is the low spatial resolution of the independent variables describing forest cover—100 km^2. Experience has shown that this resolution is often sufficient for the statistical analysis of ecological relationships (Vernier and Cumming, 1998, Cumming, 1999). However, the statistical relations in this instance are relatively weak. The data resolution may be too coarse for the problem at hand.

For areas where digital forest cover data exist, covariates could be generated for circle of radius perhaps1 km, centred at the reported origin of each fire. This is just a straightforward application of GIS techniques.

To support the analysis of recurrence times as a function of size, I assumed that fire sizes were independent within and between years. This is clearly not the case. Most of my sample fires exceeding 1×10^5 ha burnt in the three year interval 1980–1982. It seems probable that there is relation between summer weather conditions and fire size, just as total area burnt is related to various indices of weather at monthly and annual time scales (Flannigan and Harrington, 1988). Given sufficient meteorological data, it would be easy to test for and quantify the effect within the modelling framework developed here. The results could be used to extend previous studies of the effect of global warming scenarios on fire, which have focussed on weather indices as a measure of area burnt (Flannigan et al., 1998). It may also be possible to study additional covariates describing fire suppression effort to test for their influence on the expected size of larger fires, once fire weather and fuels are accounted for.

As Moritz (1997) concluded, "quantitative approaches for characterising disturbance regimes are necessary to understand ecological processes and manage disturbance-mediated ecosystems." In the case of fire in boreal regions, there are four processes which require quantification: the spatial pattern of ignitions, the probability that a fire will reach a certain minimum size, the size distribution of these non-trivial fires, and finally, the effect of individual fires, measured as within-fire variability in severity and the forest types consumed. Many long-standing debates in fire ecology, such as the impact of fire suppression, the relative role of fire weather and fuels, and the relation between patch age or type and hazard of burning, could be formulated as testable statistical hypothesis, and resolved with existing data, once parametric models are developed for these processes.

7 Acknowledgments

I thank Drs. D. Kelker and D. Wiens, of the Department of Mathematical Sciences, University of Alberta for their statistical advice, and also H. Rubin, D. Murdoch, J. Miller, P. Hamer and H. Hirose, participants in `sci.stat.consult`. I also thank Dr. M. Flannigan and colleagues of the Canadian Forest Service Fire Management Network for their interest and encouragement. Forest inventory data were provided by courtesy of D. Price of Alberta Environmental Protection, and by Alberta Pacific Forest Industries Inc. This research was funded by the Sustainable Forest Management Network.

References

Alberta Environmental Protection. 1998. Fire history databases. http://www.gov.ab.ca/env/forests/fpd.

Alberta Forest Service. 1985. Alberta phase 3 forest inventory: An overview. ENR Report **I/86**, Alberta Energy and Natural Resources, Edmonton, Alberta.

Andison, D. W. 1998. Temporal patterns of age-class distributions on foothills landscapes in Alberta. Ecography **21**:543–550.

Antonovski, M. Y., M. T. Ter-Mikaelian, and V. V. Furyaev. 1992. A spatial model of long-term forest fire dynamics and its applications to forests in western Siberia. Pages 373–403 *in* H. H. Shugart, R. Leemans and G. B. Bonan, editors. A systems analysis of the global boreal forest. Cambridge University Press, Cambridge, UK.

Armstrong, G. W. 1999. A stochastic characterisation of the natural disturbance regime of the boreal mixedwood forest with implications for sustainable forest management. Canadian Journal of Forest Research **29**:424–433.

Armstrong, G. W., S. G. Cumming, and W. L. Adamowicz. 1999. Timber supply implications of natural disturbance management. The Forestry Chronicle **75**:497–504.

Chou, Y. H., R. A. Minnich, and R. J. Dezzani. 1993. Do fire sizes differ between southern California and Baja California? Forest Science **39**:835–844.

Cumming, S. G. 1997. Landscape dynamics of the boreal mixedwood forest. Ph.D. thesis, University of British Columbia, Vancouver, B. C., Canada.

Cumming, S. G. 1999. Compositional analysis of wildfire in the Alberta boreal mixedwood: what do fires burn? Working Paper **1999–12**, Sustainable Forest Management Network, G-208 Biological Sciences, University of Alberta, Edmonton, Alberta.

Cumming, S. G., P. J. Burton, M. Joy, B. Klinkenberg, F. K. A. Schmiegelow, and J. N. M. Smith. 1995. Experimental habitat fragmentation and simulation of landscape dynamics in the boreal mixedwood: A pilot study. Canada-Alberta Partnership Agreement in Forestry Report **A5016-137**, Canadian Forest Service and the Alberta Land and Forest Service. A digital version of this report is available at ftp://grebe.geog.ubc.ca/pub/Papers/.

Cumming, S. G., P. J. Burton, and B. Klinkenberg. 1996. Canadian boreal mixedwood forests may have no "representative" areas: some implications for reserve design. Ecography **19**:162–180.

Davis, C. S., and M. A. Stephens. 1989. Empirical distribution functions and goodness-of-fit tests. Algorithm AS 248. Applied Statistics **38**:535–543.

Flannigan, M. D., Y. Bergeron, O. Engelmark, and B. M. Wotton. 1998. Future fildfire in circumpolar forests in relation to global warming. Journal of Vegetation Science **9**:467–476.

Flannigan, M. D., and J. B. Harrington. 1988. A study of the relation of meteorological variables to monthly provincial ara burned by wildfire in CANADA 1953–80. J. Appl. Meteor. **27**:441–452.

Fraser, D. A. S. 1996. Probability & statistics. theory and applications. Duxbury Press, North Scituate, Massachusetts.

Gaines, S. D., and M. W. Denny. 1993. The largest, smallest, highest, lowest, longest and shortest: extremes in ecology. Ecology **74**:1677–1692.

Gray, H. W., and B. Janz. 1985. Initial-attack initiatives in Alberta: a response to the 1980s. Pages 25–35 *in* D. E. Dubé, editor. Proceedings of the intermountain fire council 1983 fire management workshop. Information Report **NOR-X-271**, Edmonton, Alberta, Can. For. Serv., Nor. For. Cent.

He, H. S., and D. J. Mladenoff. 1999. Spatially explicit and stochastic simulation of forest-landscape fire disturbance and succession. Ecology **80**:81–99.

Hilborn, R., and M. Mangel. 1997. The ecological detective. confronting models with data. Monographs in Population Biology **28**. Princeton University Press, Princetone, New Jersey.

Hirsch, K. G., P. N. Corey, and D. L. Martell. 1998. Using expert jusgment to model initial attack fire crew effectiveness. Forest Science **44**:539–549.

Johnson, E. A. 1992. Fire and vegetation dynamics: Studies from the North American boreal forest. Cambridge University Press.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. Continuous univariate distributions, volume 1. 2nd edition. John Wiley & Sons, Inc.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1995. Continuous univariate distributions, volume 2. 2nd edition. John Wiley & Sons, Inc.

Kabzems, A., A. L. Kosowan, and W. C. Harris. 1986. Mixedwood section in an ecological perspective. Technical Bulletin **8,** 2nd Edition, Saskatchewan Parks and Renewable Resources, Forestry Division, Regina, Saskatchewan.

Kendall, M. G., and A. Stuart. 1973. The advanced theory of statistics, volume 2. 3rd edition. Hafner Publishing Company.

Kotz, S., N. L. Johnson, and C. B. Read, editors. 1982. Encylopedia of statistical sciences, volume 1. John Wiley and Sons, New York.

Li, C., M. Ter-Mikaelian, and A. Perera. 1997. Temporal fire disturbance patterns on a forest landscape. Ecological Modelling **99**:137–150.

MathSoft Inc. 1995. S-PLUS guide to statistical & mathematical analysis. version 3.3 for WINDOWS and UNIX. StatSci Division, Seattle, Washington.

Moritz, M. A. 1997. Analyzing extreme disturbance events: fire in Los Padres national forest. Ecological Applications **7**:1252–1262.

Murphy, P. J. 1985. History of forest and prairie fire control policy in Alberta. Report **T/77**, Alberta Energy and Natural Resources, Edmonton, Alberta.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1988. Numerical recipes in C. 1st edition. Cambridge University Press, New York.

Ratz, A. 1995. Long-term spatial patterns createdd by fire: a model oriented towards boreal forests. International Journal of Wildland Fire **5**:25–34.

Rowe, J. S. 1972. Forest regions of Canada. Publication No. **1300**, Canadian Forest Service, Ottawa, Ontario.

Strauss, D., L. Bednar, and R. Mees. 1989. Do one percent of forest fires cause ninety-nine percent of the damage? Forest Science **35**:319–328.

Strong, W. L. 1992. Ecoregions and ecodistricts of Alberta. ENR Technical Report **T/244**, Alberta Forestry, Lands and Wildlife, Edmonton, Alberta.

Venables, W. N., and B. D. Ripley. 1997. Modern applied statistics with S–PLUS. 2nd edition. Springer.

Vernier, P., and S. G. Cumming. 1998. Predicting landscape patterns from stand attribute data in the alberta boreal mixedwood. Working Paper **1998-7**, Sustainable Forest Management Network, G-208 Biological Sciences, University of Alberta, Edmonton, Alberta.