

Probability based models for estimation of wildfire risk*

Haiganoush K. Preisler^A, David R. Brillinger^B, Robert E. Burgan^C and J. W. Benoit^D

^AUSDA Forest Service, Pacific Southwest Research Station, 800 Buchanan St., West Annex, Albany, CA 94710, USA. Telephone: +1 510 559 6484; fax: +1 510 559 6440; email: hpreisler@fs.fed.us

^BDepartment of Statistics, 367 Evans Hall, University of California, Berkeley, CA 94720-3860, USA. Telephone: +1 510 642 0611; fax: +1 510 642 7892; email: brill@stat.berkeley.edu

^CRetired, Intermountain Fire Sciences Laboratory.

^DUSDA Forest Service, Pacific Southwest Research Station, Riverside Fire Laboratory.

Abstract. We present a probability-based model for estimating fire risk. Risk is defined using three probabilities: the probability of fire occurrence; the conditional probability of a large fire given ignition; and the unconditional probability of a large fire. The model is based on grouped data at the 1 km²-day cell level. We fit a spatially and temporally explicit non-parametric logistic regression to the grouped data. The probability framework is particularly useful for assessing the utility of explanatory variables, such as fire weather and danger indices for predicting fire risk. The model may also be used to produce maps of predicted probabilities and to estimate the total number of expected fires, or large fires, in a given region and time period. As an example we use historic data from the State of Oregon to study the significance and the forms of relationships between some of the commonly used weather and danger variables on the probabilities of fire. We also produce maps of predicted probabilities for the State of Oregon. Graphs of monthly total numbers of fires are also produced for a small region in Oregon, as an example, and expected numbers are compared to actual numbers of fires for the period 1989–1996. The fits appear to be reasonable; however, the standard errors are large indicating the need for additional weather or topographic variables.

Additional keywords: fire danger indices; fire occurrence probabilities; fire weather; forest fires; non-parametric regression; Oregon; spatial–temporal model.

Introduction

In 1968 the USDA Forest Service started work on the development of a fire danger rating system that would rely on science and engineering principles and on local observations. It was to incorporate basic laws of physics, thus making the system applicable nationwide. The result was the National Fire Danger Rating System (NFDRS). The first version of the NFDRS was released in 1972 (Deeming *et al.* 1972). In 1988, in response to concerns raised by users in the south-eastern USA relative to the accuracy and applicability of NFDRS outputs in their area, a modified version of the 1978 NFDRS was released. It included better recognition of drought and fire response after precipitation (Burgan 1988).

Outputs of NFDRS include fire danger maps. The maps are based on what might be referred to as ‘judgment components’ (Mosteller and Tukey 1977) as opposed to, for example, principal components, linear discriminants, or linear predictors. Judgment components are constructed by combining explanatory variables (in the present case fire weather and

fire danger variables) using knowledge based both on insight and data. Principal components or linear discriminants, on the other hand, are constructed using a mechanical/objective process based on multivariate data values. Regardless of the method used to arrive at danger indices, it is important to assess their usefulness by relating them to probabilities of actual fire occurrences. One method for doing this is with the aid of modern regression techniques, such as generalized linear or generalized additive models. Logistic regression is a special case of generalized additive models (Hastie and Tibshirani 1991).

The purpose of this paper is twofold. Firstly, we demonstrate a method for assessing fire danger and fire weather variables. Second, we present a procedure that may be used to forecast numbers of fire occurrences and numbers of larger fires for a given area and a given time period. We demonstrate the methods using data on fire occurrence and a total of eight fire weather and fire danger variables for a period of 8 years in Oregon. The methods are not limited to this dataset and may

* This paper was written and prepared by U.S. Government employees on official time, and therefore is in the public domain and not subject to copyright.

be used to assess the usefulness of other fire danger indices and other locations.

Previous statistical work

Few studies seem to have been done to assess the relationship between fire danger indices and actual fire occurrences. Andrews and Bradshaw (1997) present programs for rating fire danger indices at a given location. These authors use logistic regression to estimate the linear relationship of each fire index to the logit of the probability of a fire-day, large fire-day, or multiple fire-day. They generate probability curves relating each index to each of the three responses by linking daily fire activity at a given forest to index data from the closest weather station.

In Dayananda (1977) a Poisson model is used to study the effect of one of the fire danger indices (Keetch-Byram Drought Index), Keetch and Byram (1968) on the expected number of fire occurrences. Mandallaz and Ye (1997) also use a Poisson model to evaluate the relationships between various European Dryness indices and a few weather variables on expected numbers of fires. Chuvieco and Salas (1996) use GIS tools for mapping probabilities of fire ignition based on fire danger indices established by the Spanish Forest Service. Anderson *et al.* (2000), Cunningham and Martell (1972), and Martell *et al.* (1987, 1989) use logistic regression to study relationships between indices produced by the Canadian Forest Fire Danger Rating System (Van Wagner 1987) and the probability of fire days. Chou *et al.* (1990, 1993) use logistic regression with weather and other explanatory variables, one of which is a modified Moran's coefficient, to take into account the spatial autocorrelation between nearby fires. Papers that study frequencies of forest fires as a function of time-since-last-fire include Johnson and Gutsell (1994), Peng and Schoenberg (2001), Reed (1998), Reed *et al.* (1998) and Grissino-Mayer (1999).

In the present paper we present models for estimating probabilities of fire on a given day and on a 1 km grid on Federal land. The models are spatially and temporally explicit. Each square-kilometre grid for each day (henceforth referred to as a voxel) can have a different probability value. The model may also include any number of fire danger and weather variables. We demonstrate how the model allows the study of the simultaneous effects of explanatory variables on the probability of fire occurrence. The model employs non-parametric logistic regression. Consequently, no *a priori* assumptions beyond smoothness need to be made about the forms of the relationships involving the various explanatory variables (e.g. the danger indices) and the probability of fire.

In the following sections we discuss the various probabilities needed to assess fire risk. We build on the statistical models developed in Brillinger *et al.* (2003) to estimate the risk probabilities as functions of explanatory variables. The methods are then applied on a sample data from Federal

lands in a region including the State of Oregon. We demonstrate how one can produce risk probability maps for a large region—e.g. the State of Oregon—or how to estimate fire risks for a selected National Forest District.

The data

Typically, data publicly available for fire risk analyses include historical fire occurrences on Federal lands, and daily values of weather variables and fire danger indices from weather stations. Historical wildfire data (1970–present) with locations and dates of all reported fires greater than 0.04 ha (0.1 acre) are stored in the National Interagency Fire Management Integrated Database (NIFMID) at the USDA National Computer Center in Kansas City (<http://famweb.nwecg.gov/kcfast/mnmenu.htm>). Because a large number of incorrect entries are known to exist in this dataset, such as misreported fire locations and sizes, we retrieved a version that has been thoroughly checked and cleaned of errors from <http://www.fs.fed.us/fire/fuelman/fireloc.htm>. This corrected set of historical fire occurrence data is complete for the entire contiguous USA for 1986 through 1996. A map of Federal fire locations for 1996 in a region including the state of Oregon (Region 6) is provided in Fig. 1.

Observed weather data were downloaded from the Kansas City Computer Center. These data consist of daily weather records from over 1800 fire weather stations around the USA. Each station's records include the daily minimum, maximum, and afternoon (1pm) values of dry bulb temperature and relative humidity, wind speed and direction, the state of the

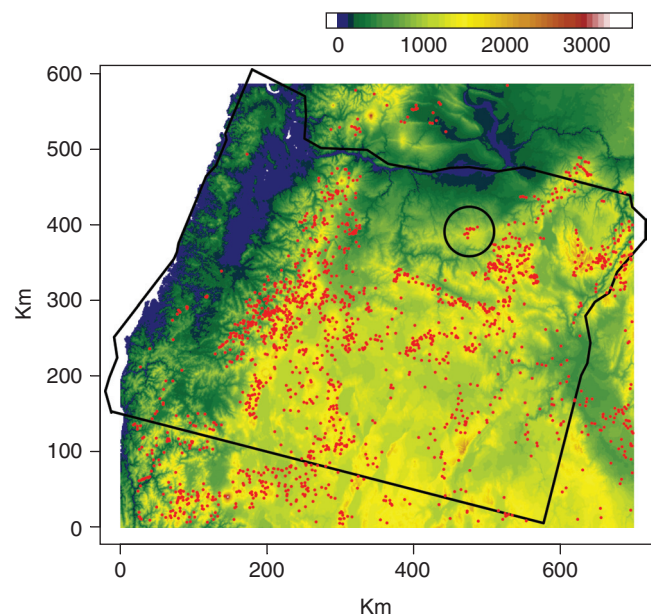


Fig. 1. Elevation map (m) of Oregon region. Red circles are locations of Federal fires in 1996. The black circle shows the location of Heppner Ranger District of the Umatilla Forest. This region is used in the text as an example for demonstrating the uses of the probability models.

weather, 10-h lag fuel moisture, precipitation, and lightning activity level. The Kansas City Computer Center's database was queried for all weather data from 1970 to 2001, though individual stations had varying time periods during which they collected data. There were 294 weather stations located within the state of Oregon, though most states had sparser station coverage.

Gridded fire danger index data was obtained for the USA. The indices were originally computed using observed data from the fire weather stations, and included Burning Index (*BI*), Energy Release Component (*ERC*), Fire Potential Index (*FP*), Keetch-Byram Drought Index (*KB*), Spread Component (*SC*), and Thousand-hour fuel moisture (*TH*). A weighted inverse distance method was used to interpolate the indices to the voxel level. Estimates for two weather variables, relative humidity (*RH*) and dry bulb temperature (*DBT*), for each voxel were obtained from the weather station data using a non-parametric regression equation with location, elevation, day in year and value at nearest station as explanatory variables. The National Fire Danger Rating fuel model categories at 1 km resolution were obtained from the web (Burgan *et al.* 1998). The fuel categories were derived from a combination of satellite imagery used to create a land cover database for the conterminous USA and ground data sampled from across the USA.

For the present study we used data from a region including and surrounding the State of Oregon as shown in Fig. 1 and for the dates 26 April 1989 to 31 December 1996. Four fire danger indices, *BI*, *FP*, *KB*, *TH*, were used in the analysis. The *BI* provides a good seasonal profile of fire danger, particularly for vegetation types that have a significant loading of large dead fuels; the *KB* and *TH* both provide measures of long term drying, and the *FP* is a new index based on satellite imagery of vegetation greenness that reflects the susceptibility of vegetation to fire (Burgan *et al.* 1998). Two other indices, *ERC* and *SC*, were not used because *BI* is calculated from these two indices ($BI = k \times SC \times ERC$, where k is a constant). In addition, we used two interpolated weather variables, relative humidity (*RH*) and dry bulb temperature (*DBT*). Values of two other weather variables, wind speed (*Wspeed*) and state of weather (*Weather*), as recorded at the nearest weather stations were also used. The state of weather is a categorical variable describing 6 states of weather. Clear skies; scattered clouds; broken clouds; overcast; showering, raining or snowing; and thunderstorm. These four weather variables are thought to significantly influence fire danger when the vegetation is susceptible to fire.

Defining fire risk probabilities

Given the data just described, it is possible to estimate the following risk probabilities:

- (1) The probability, p_1 , of a fire of size greater than 0.04 ha occurring at a given location and given day.

This probability might be thought of as the probability of ignition since it includes very small fires (0.04 ha = 20 m × 20 m).

- (2) The conditional probability, p_2 , of a fire (or an ignition) becoming a large fire. Large fires may be defined as those greater than a specified number of hectares when the fire was finally contained. In our Oregon example analysed below we define large fires as those greater than 40.5 ha (100 acres). This conditional probability is useful for decisions such as ceasing prescribed burn activity, implementing specific public use restrictions or deciding on whether or not to continue letting a wild fire burn.
- (3) The unconditional probability, p_3 , of a fire occurring and becoming a large fire. This probability is the product, $p_1 \times p_2$, of the probabilities in 1 and 2 above. Decisions on number of fire personnel and equipment needed might be made with the help of these unconditional probabilities.

It is of interest to note that different fire danger indices might be important to the different probabilities listed above. For example, there were very few fire occurrences in the south-eastern region of Oregon (Fig. 1). However, the number of large fires was disproportionately large in this region (Fig. 2). In other words, while very few fires appear to get started in this regions, those that do reach 0.04 ha size appear to spread and become large fires more often than in, say, the Cascades region where there are many ignitions but only a small fraction of them become large fires. One possible cause for differences in fire sizes in the different regions may be differences in suppression strategies.

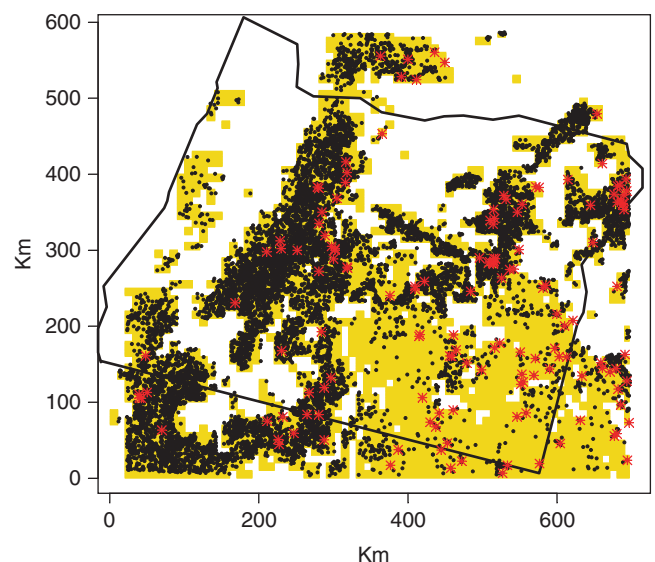


Fig. 2. Locations of all Federal fires between 1989 and 1996 (black dots). The yellow regions correspond to Federal lands. Red stars indicate locations of fires larger than 405 ha (1000 acres).

Estimated probabilities may also be used for estimating the number of expected fires in a given region or forest district. The expected number of fires in a given region on a given day may be estimated by the sum of the daily probabilities in each 1 km² pixel of the region. The procedure assumes one fire occurrence per pixel.

Statistical methods

Estimating probability of fire occurrence

Define a response variable $N_{x,y,t}$ as follows:

$$N_{x,y,t} = 1 \quad \text{if there is a fire at location } (x,y) \text{ and time } t \\ = 0 \quad \text{otherwise.}$$

Next let the probability, p_1 , of a fire occurring at (x,y) and time t be defined by the logistic model

$$p_1 = \text{Prob}\{N_{x,y,t} = 1 | U_{x,y,t}\} = \frac{\exp(\theta_{x,y,t})}{1 + \exp(\theta_{x,y,t})} \quad (1)$$

where $U_{x,y,t}$ is the collection of the values of the explanatory variables (i.e. weather variables and danger indices) at location (x,y) and time t ; $\theta_{x,y,t}$ is a spatially and temporally explicit parameter to be estimated from the data on fire occurrences and the values of the explanatory variables. Ordinary logistic regression procedures are not directly applicable for estimation of the parameters in model (1) because of the following points:

- (1) *Temporal dependence*: Usual logistic regression requires the assumption of independence between observations. This is not the case here because, if a fire occurs at a given location and burns all the available fuel, then the chances of another fire occurring at the same location in the next few weeks might be considerably less than if there were no fire at that location in the near past. We propose to deal with this problem, for now, by assuming that fire occurrences are conditionally independent given values of the explanatory variables. Specifically, by including temporally explicit variables in the model, e.g. fire potential index and fuel moisture, it is possible within the model to have a different expected probability of a fire occurrence a week after a fire because the value of the fuel moisture is changed after the fire.
- (2) *Spatial dependence*: Because of similarities in topography, fuel conditions, or weather at nearby locations fires might often occur in clusters invalidating the assumption of independence of observations. In addition to the spatially explicit explanatory variable (elevation) employed and the spatially and temporally explicit variables (indices and weather variables), we include a vector-valued spatial location variable in the model. The vector valued location variable is included to account for dependencies amongst nearby points and any spatially

explicit topographic or vegetation characteristic not captured by the fuel model or any other variable already in the model. One such variable may be differences in suppression strategies in different regions.

- (3) *Non-linear relationships*: Relationships between the explanatory variables and the probability of fire on the logit-scale could be non-linear. For example, a variable that will most surely have non-linear effects is the spatial location. There is no reason to assume that fire occurrences increase linearly, even on logit scale, as one travels from north to south or from east to west for example. We propose to use non-parametric smooth but otherwise arbitrary functions within generalized additive models to estimate potentially non-linear relationships.
- (4) *Large number of voxels*: Parameter estimation in logistic regression models is done by maximizing the Bernoulli likelihood function

$$\prod_{k=1}^K p_{1k}^{N_k} (1 - p_{1k})^{1-N_k} \quad (2)$$

where, to simplify notation, k refers to the (x,y,t) voxel and K is the total number of voxels. In the present case K is very large (e.g. for the Oregon example and the period of study, $K = 578\,192\,400$, due to the large number of 1 km² Federal land locations and days (voxels) without fires). Brillinger *et al.* (2003) suggest dealing with this problem by randomly selecting a proportion, π , of the voxels with no fire. They develop formulas for estimating p_1 using all voxels with fire and the randomly sampled voxels with no fire. Using the identity

$$\text{Prob}\{A_k | B_k\} = \text{Prob}\{B_k | A_k\} \text{Prob}\{A_k\} / \text{Prob}\{B_k\}$$

where A_k stands for ‘fire in voxel k ’ and B_k stands for ‘voxel k was selected to be in the sample’, one has the conditional probability

$$\text{Prob}\{A_k | B_k\} = \gamma_k = \frac{p_{1k}}{p_{1k} + (1 - p_{1k})\pi}.$$

By elementary algebra, one can show that

$$\text{logit}(\gamma_k) = \text{logit}(p_{1k}) - \text{log}(\pi) \quad (3)$$

where $\text{logit}(p) = \log\{p/(1-p)\}$ and \log is the natural logarithmic function. Consequently, we may use available generalized additive logistic regression programs with the sampled dataset to estimate the parameters γ_k and then use the identity in equation (3) to calculate estimates for p_{1k} , the probability of a fire occurring in voxel k .

The particular model we used to estimate p_{1k} was a logistic regression model with the logit parameter given by

$$\begin{aligned} \text{logit}(p_{1k}) = & \text{fuel}_k + g_1(x_k, y_k) + g_2(\text{day}_k) + g_3(\text{elev}_k) \\ & + g_4(BI_k) + g_5(FP_k) + g_6(KB_k) + g_7(TH_k) \\ & + g_8(W\text{speed}_k) + g_9(RH_k) + g_{10}(DBT_k) \\ & + \text{Weather} \end{aligned} \quad (4)$$

where (x_k, y_k) are location coordinates of the k th response; $fuel_k$ is the fuel model category at location (x_k, y_k) ; day_k is the day in year; $elev_k$ is the elevation (in metres) at location (x_k, y_k) ; $BI_k, FP_k, KB_k, TH_k, RH_k, DBT_k$ are the interpolated fire indices and weather variables at location (x_k, y_k) and day_k ; $Wspeed_k, Weather$ are the wind speed and state of weather at the nearest weather station on that day. The functions $g_i()$ are (non-parametric) smooth functions describing the non-linear relationships between the explanatory variables and the logit of probability of response. The bivariate function $g_1(x_k, y_k)$ is a surface describing the effect of spatial location. It will include the effects of known or unknown spatial explanatory variables not included in the model. This location term is also meant to handle the similarities/dependencies between nearby points. The functions $g_i()$ were estimated simultaneously using *loess* non-parametric smoothers (Cleveland *et al.* 1992) within the generalized additive model, GAM (Hastie 1992) using the S-PLUS statistical package (S-PLUS 2000). We used the jackknife procedure to estimate the uncertainties in the estimated parameters. This involved splitting the 8 years of data randomly into eight subsets and then obtaining eight different estimates of the parameters by dropping one subset at a time from the dataset. Uncertainties are then estimated by calculating the variability in the eight different parameter estimates. This jackknife procedure was also used in Brillinger *et al.* (2003) where a smaller set of explanatory variables was studied.

Estimating the conditional probability of large fires

We have defined large fires as any fire that burned an area greater than 40.5 ha by the time the fire was contained. Now we model the conditional probability, p_2 , of a fire started at time t at location (x,y) becoming a large fire. A logistic regression with the same explanatory variables as those included in equation (4) is employed. Here the application of logistic regression was straightforward because the data used were only from locations where a fire did occur (in this case K in equation (2) is 15 786 voxels). There was no need to use the sampling scheme described in the previous section.

Estimating the unconditional probability of large fires

Once the probability of fire occurrence, p_1 , and the probability of a fire becoming large, p_2 , have been estimated one can estimate the unconditional probability of a fire occurring at a given location and time and then becoming a large fire, p_3 . The unconditional probability of a large fire is given by the identity

$$\text{Prob}\{\text{a large fire occurring}\} = \text{Prob}\{\text{a fire occurring}\} \\ \times \text{Prob}\{\text{a fire becoming large} \mid \text{a fire occurred}\}.$$

Namely, $p_3 = p_1 \times p_2$.

Standard errors for \hat{p}_3 may be estimated using the jackknife method referred to above. Specifically, the data are

split randomly into M subsets and then M different estimates of \hat{p}_3 are obtained by dropping one subset at a time from the dataset. Uncertainties are estimated by calculating the variability between the M different parameter estimates.

Estimating the expected number of fires for a given region and time

Managers are often interested in knowing the expected number of fires in a given region for a given time period. When the number of fires in each voxel is assumed to be a Bernoulli (zero, one-valued) random variable with a different probability in each voxel (i.e. the assumption made in the model above), then the distribution of the total number of fires over multiple voxels is the Poisson-Binomial distribution. When the probabilities in each voxel are small, then the Poisson-Binomial distribution is well approximated by the Poisson distribution (Hodges and LeCam 1960). In our case, the probabilities of fire for a voxel of 1 km² and one day are very small. For example, the maximum estimated probability, over all 1 km² grids in the Oregon region, of a large fire occurring in the month of July 1996, was 0.0004 (or 4 in 10 000). We therefore estimate the expected number of fires for a particular region and time period by summing the estimated probability values of individual voxels and then use the Poisson distribution to obtain approximate confidence intervals.

Application to Oregon data

Of the 11 explanatory variables used to model the probability, p_1 , of fire occurrence, 6 were found to have highly significant effect. The surface and line plots in Fig. 3 are the estimated partial effects of the significant variables: spatial location; day in year; elevation; 1000-h fuel moisture; dry bulb temperature; and state of weather. In other words, the plots in Fig. 3 are estimates of the functions $g_1(x,y)$, $g_2(day)$, $g_3(elev)$, $g_7(TH_k)$, $g_{10}(DBT)$ and the categorical effects of *Weather* in equation (4). They are called the partial effects because they describe the effect of each variable in the presence of all other variables in the model. The horizontal line at level zero provides the overall average response level. The plots for all explanatory variables except spatial location also include approximate point-wise 95% confidence bounds of the estimates produced by the jackknife procedure. A variable is considered to have no apparent significant effect on response at a given level if the corresponding point on the horizontal zero line is within the 95% bounds.

It is to be realized that the confidence bands on the figures are marginal 95% rather than simultaneous. The implication is that ~5% of the inferences made using them can be expected to be incorrect. Wide confidence bounds are due in part to limited numbers of observations near some levels of concern. The estimates in Fig. 3a suggest that probabilities of fire occurrence are lowest in the south-eastern region of Oregon, consistent with the data in Fig. 2. Estimated probability of occurrence seems to be highest on days with lightning

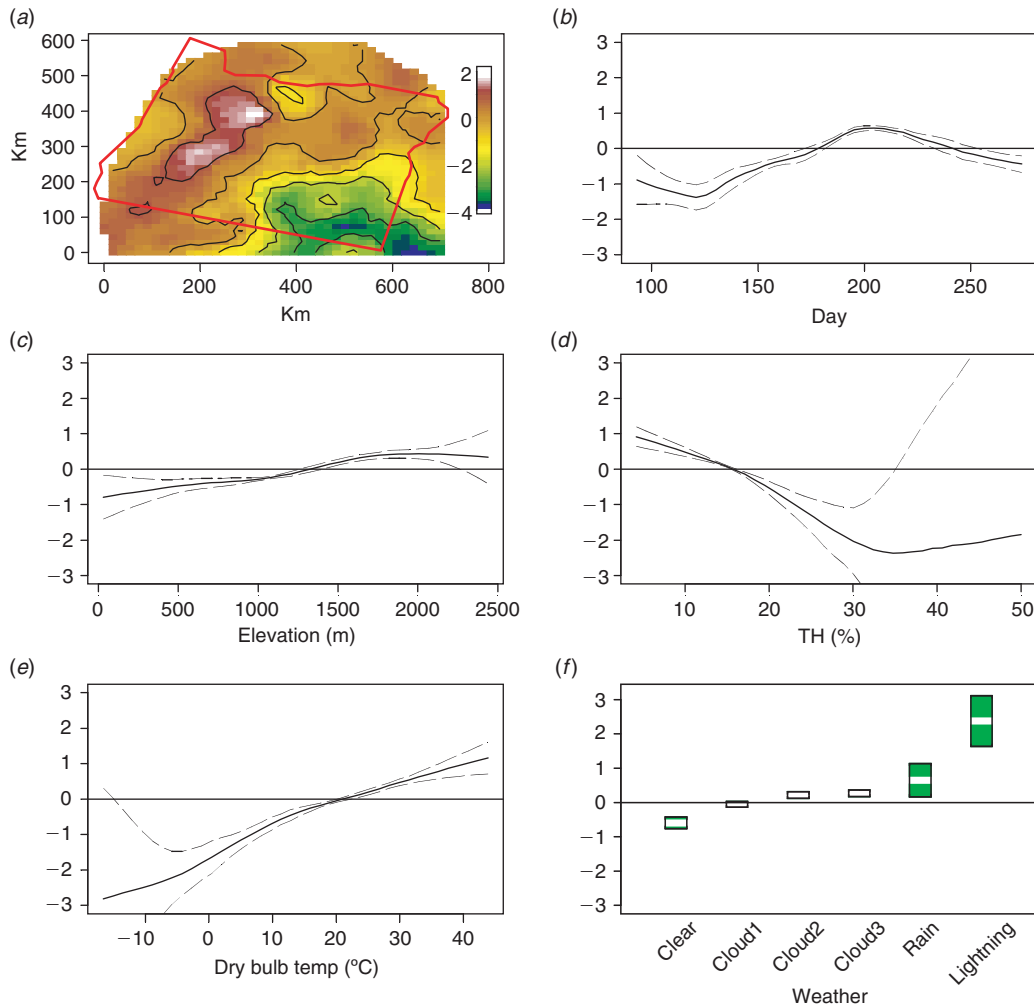


Fig. 3. Estimated partial effects of explanatory variables on the probability of fire occurrence. Probabilities of occurrence appear to be lowest in the south-eastern region. Probabilities of occurrence appear to increase with increasing values of elevation and Dry Bulb Temperatures and to decrease with values of TH. Wide 95% confidence bounds in some places are partly due to the limited number of observations at that level of a variable.

(Fig. 3*f*). The odds of fire occurrence were estimated to be 19.5 times larger for days with lightning activity than days with clear skies. The 95% confidence limits for the estimated odds ratio was 2.97 to 26.18. Estimated probabilities of occurrence appeared to increase with increasing values of elevation and dry bulb temperatures and to decrease with increasing values of 1000-h fuel moisture (*TH*). Fires appear to be more likely to occur when values of *TH* are less than 10 and less likely to occur when *TH* values are greater than 20. The estimated seasonal effect (day in year) indicates higher probabilities of fire occurrences in the summer season. However, it is noteworthy that the seasonal effect is smaller here than in a previous model (Brillinger *et al.* 2003), where no indices or weather variables were used. The smaller seasonal effects seen in the present model may be due to the fact that some of those effects were included through the seasonally changing variables, *TH* and *DBT*. The explanatory variables

BI, *FP*, *KB*, *Wspeed*, and *RH* did not appear to have significant effects on fire occurrence when included simultaneously with the rest of the variables. The location variable was the most significant variable among the ones in the model even when fuel model category was included in the model. This suggests that, in addition to the variables already in the model, there are other topographic and/or vegetation cover variables affecting fire occurrence.

The plots in Fig. 3 may also be used to predict days with ‘higher than normal’ probabilities of fire occurrence. According to the plots in Fig. 3, if the reported dry bulb temperatures are greater than 30°C and the *TH* values are less than 10% then the model predicts higher than average numbers of fires for that location and time of year.

Figure 4 is a plot of the estimated effects of the four significant variables on the conditional probability of an occurrence becoming a large fire (>40.5 ha). Important variables were

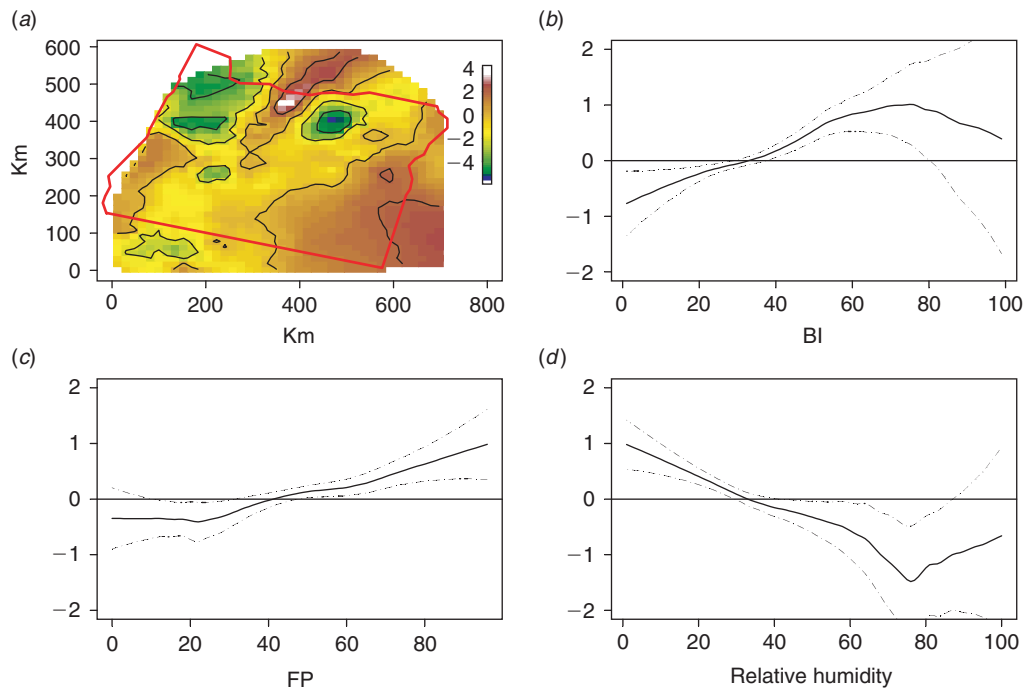


Fig. 4. Estimated partial effects of explanatory variables on the conditional probability of a fire becoming a large fire (>40.5 ha). The region of Oregon with smallest probabilities of fire occurrence (south-eastern region) appears to have some of the highest probabilities of an ignition becoming a large fire.

arrived at by forward and backward stepwise variable selection. It is interesting to note that, except for the spatial location, the variables entering the model for the conditional occurrence of large fires, p_2 , were different from those in the fire occurrence model, p_1 . It is also apparent from Fig. 4 that the south-eastern region of Oregon with relatively low fire occurrences has some of the highest estimates for the conditional probability of large fires.

The probability of an ignition turning into a large fire seems to increase with values of FP and BI and decrease with increasing values of RH (Fig. 4). It is important to note that these effects are the partial effects of each explanatory variable in a model that includes the effects of several variables. For example, the effect of FP , especially for values less than 40, was more pronounced when the variable RH was not in the model. The day-in-year variable (day) was not highly significant (P -value = 0.07), suggesting that the seasonal variables BI , FP , and RH appear to have accounted for most of the reasons for increased large fires in the summer. The location variable, however, was still highly significant, suggesting the need for other spatially explicit variables. Different suppression strategies might be one such variable.

The plots in Fig. 4 may be used to predict days with higher than average conditional probabilities of fires becoming large fires and as such may be particularly useful in practice. According to these plots, if the reported BI values are greater than 40, together with FP values greater than 60 and relative

humidity values less than 30, then the model predicts higher than average numbers of fires once begun turning into large fires. These departures from norm are spatially and temporally explicit. In other words they are departures from the usual for the given location and time of year.

The variable wind speed was found to have no significant effect on either the probability of occurrence (P -value > 0.05) or the probability of a large fire (P -value = 0.33). This does not mean that wind is not an important factor for predicting fire risk but rather that wind speeds measured at weather stations do not appear to be good indicators of risk at surrounding locations. Another variable, state of weather as seen at nearby weather stations, was one of the most important explanatory variables for predicting fire occurrence even though it appears to be a very poor indicator of what the weather actually is at a location some distance from the station. For example, in our data 70% of fire occurrences were reported to be lightning or naturally caused but only 4% of voxels with fire show lightning activity at nearby weather stations. Because of the poor quality of the estimate of state-of-weather variable, the rest of the paper uses a model without the *Weather* variable to prepare maps of estimated risk probabilities and expected numbers of fires.

Figure 5 presents maps of three fire risk probability estimates for two dates for the region including the State of Oregon. The two left panels seem to suggest that the probability of ignition was higher on 29 July 1996 (white and

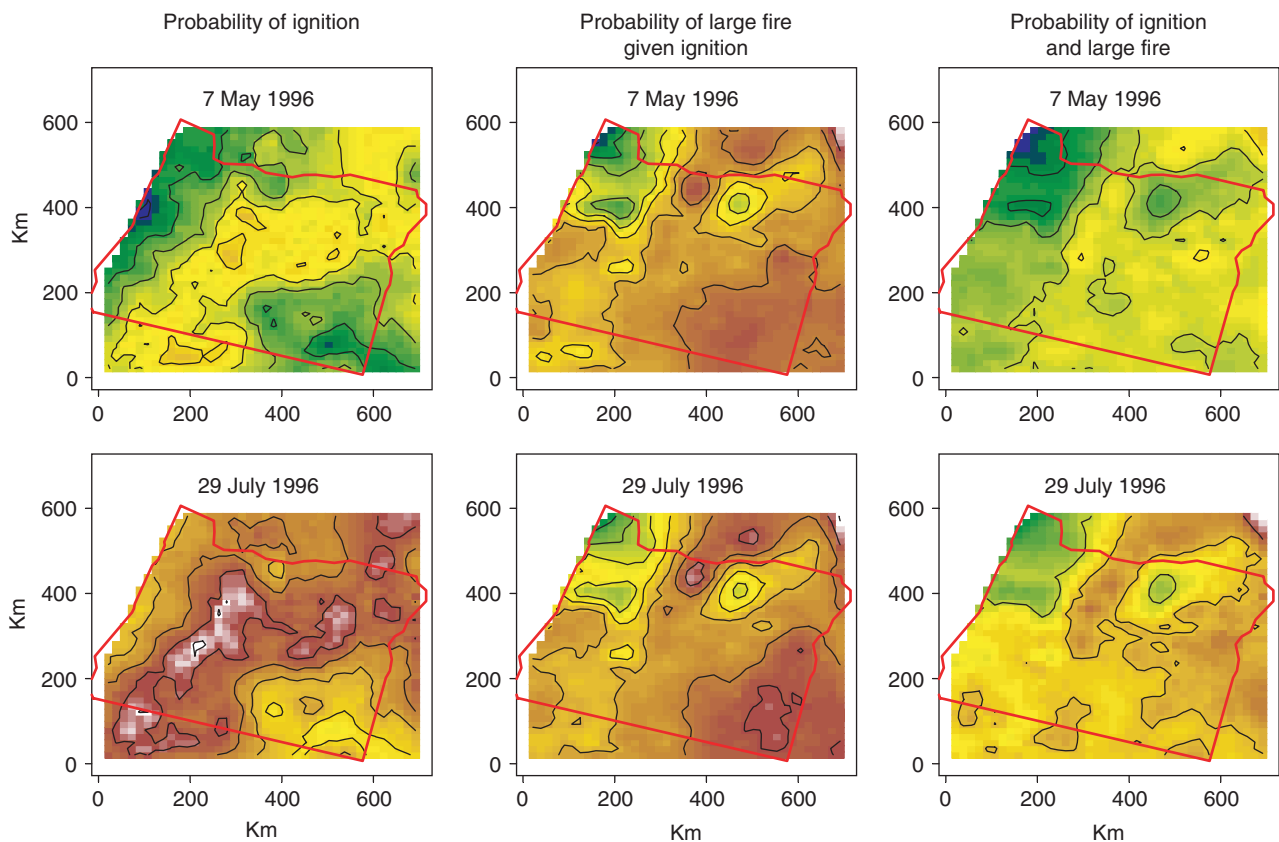


Fig. 5. Estimated probabilities of ignition are relatively lower in the south-eastern region of Oregon than in the Cascades, consistent with the historic fire occurrence map (Fig. 1). The same south-eastern region has relatively high estimates of conditional probabilities of large fires. Products of the first two probabilities are given in the two maps on the right. The colors in increasing order of probability are blue, green, yellow, red/brown, pink/white.

pink colors) than on 7 May 1996 (blues, greens and yellows). Additionally, the probability of fire occurrence on both dates appears to be relatively lower in the south-eastern region of Oregon than in the Cascades. This is consistent with the historic fire locations in Figs 1 and 2. The maps of conditional probability of a large fire given ignition seem to be relatively higher in the south-eastern region, again consistent with the historic locations of large fires in Fig. 2. The two right-most maps indicate that the estimated unconditional probabilities of large fires are small at all locations on 7 May and are relatively larger on 29 July. For example, the estimated probability of a large fire occurring in a 1 km^2 pixel on 29 July in the Oregon region was 2 in 1000. The same probability for 7 May was 4 in 10 000 even though in some locations, given the conditions on 7 May, the chances of an ignition becoming a large fire were as large as 80%.

The plots in Fig. 6 compare actual and predicted numbers of fires per month in the Heppner Ranger District of the Umatilla Forest in Oregon (the Heppner District was used as an example of a region (of size 1100 km^2) that might be of interest to a forest manager). The plots also provide the estimated approximate 99th percentile bounds for the actual

number of fires. The 99th percentile is presented by a shaded region to reflect the uncertainties (95% limits) in the estimation of the percentiles. According to the estimated model we expect 1% of the cases to fall above the 99th percentile if the model is appropriate. In the present example one out of the 48 calculated cases (2.1%) fell above the 99th bound. Our model does not account for the large numbers of fires in the Heppner District in June 1992. The model-predicted probability of at least one large fire in the Heppner District for the 8 years under study to be 4% with a corresponding 95% confidence interval between 0.4 and 30%. There were no reported large fires in this District during the 8 years of the study.

Discussion

We have found the probability-based model presented here useful for examining different aspects of fire risk and for assessing the usefulness of various fire danger indices and weather variables. The particular set of explanatory variables that were found useful depended on the question asked—e.g. what is the expected number of fire occurrences? or,

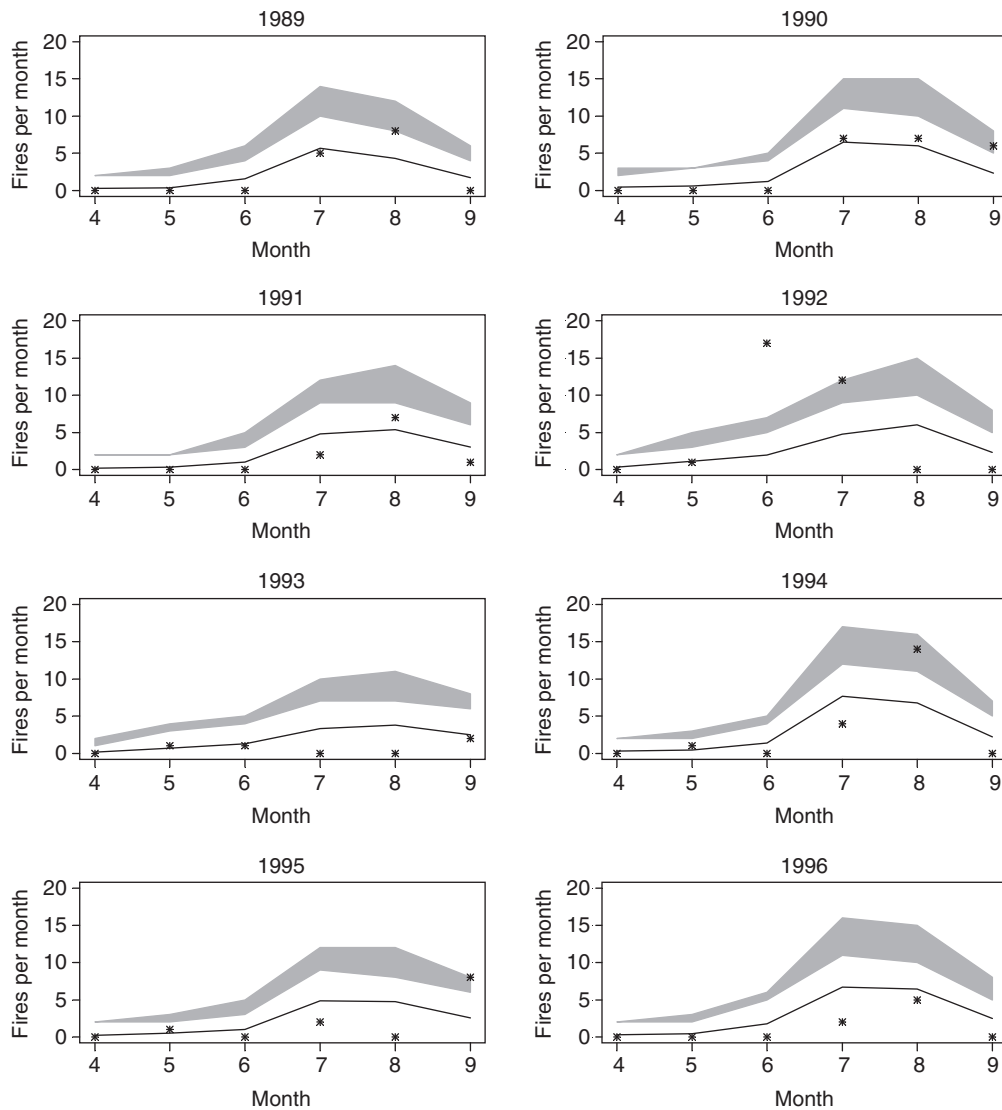


Fig. 6. Numbers of observed (*) and predicted (curves) fire occurrences per month in the Heppner District. The shaded area is the model predicted 99% upper bound for the number of fires per month in this district.

what is the expected number of large fires?—and on the list of variables included in the model. We have demonstrated the use of the model by producing fire risk probability maps for a region including the State of Oregon and estimates of the distribution of numbers of fires for the Heppner Ranger District of the Umatilla Forest. The data from Oregon were limited to Federal lands. This might affect the maps produced for the whole State. Although the maps cover both Federal and non-Federal lands, only the parts in Federal lands are appropriate.

It still remains to be seen if the relationships between explanatory variables and probabilities of fire found while analysing the present data remain the same when our methods are applied to other States, other time periods or with other or ‘better’ danger indices. Location and day in year

were significant variables in our model for predicting fire occurrence. It is hoped that, as better weather or danger variables become available, the significance of the location and day variables in the model will diminish. Until then, these two variables are necessary to account for differences in location and time of day not captured by any of the variables already in the model.

We did not address the importance of triggering mechanisms of fire occurrence (namely, people-caused or lightning-caused) in this paper. Our focus was developing the spatially and temporally explicit probability model and estimation procedure. Cause of fire may be included in the model as an additive or interactive effect. Including this variable in future models may improve the performance of the model.

A further issue is the problem of interpolating explanatory variables to arrive at estimates at the 1 km² pixel size employed in the modeling. In at least two of the variables, wind speed and state of weather, the nearest neighbor value was not adequate. We are presently studying the use of more efficient methods to evaluate weather variables at a 1 km² grid. It might be that there are no interpolation methods that will adequately predict weather conditions at specified locations. A denser grid of weather stations might be required. Another possibility is to use satellite weather data or data projected from climate simulation models (Roads *et al.* 2001).

One of the ultimate purposes of fire risk studies is to be able to predict expected numbers of fires per week during a fire season. The probability methods described here may be used to predict numbers of fires assuming one has available projected values of temporally explicit weather variables. The use of projected fire weather indices introduces another source of variation that must be incorporated in the standard error estimations. The use of scenarios might prove helpful in this connection. We are presently investigating methods for incorporating errors in covariates into the model.

Acknowledgements

The authors appreciate the help of Carolyn Chase, Intermountain Fire Sciences Laboratory, in transferring historical weather, fire danger index and fire history data to the Riverside Fire Laboratory.

References

- Anderson K, Martell DL, Flannigan MD, Wang D (2000) Modeling of fire occurrence in the boreal forest region of Canada. In 'Fire, climate change, and carbon cycling in the boreal forest'. (Eds E Kasischke, BJ Stocks) pp. 357–367. (Springer-Verlag: New York)
- Andrews PL, Bradshaw LS (1997) 'Fires: Fire information retrieval and evaluation system—a program for fire danger rating analysis.' USDA Forest Service, Intermountain Research Station General Technical Report INT-367. 64 pp. (Ogden, UT)
- Brillinger DR, Preisler HK, Benoit JW (2003) Risk assessment: a forest fire example. In 'Science and statistics: A Festschrift for Terry Speed'. (Ed. DR Goldstein) pp. 177–196. (Institute of Mathematical Statistics: Beachwood, OH)
- Burgan RE (1988) '1988 Revisions to the 1978 National Fire-Danger Rating System.' USDA Forest Service, Southeastern Forest Experiment Station Research paper SE-273. 39 pp. (Asheville, NC)
- Burgan RE, Klaver RW, Klaver JM (1998) Fuel models and fire potential from satellite and surface observations. *International Journal of Wildland Fire* **8**, 159–170.
- Chou YH, Minnich RA, Salazar LA, Power JD, Dezzani RJ (1990) Spatial autocorrelation of wildfire distribution in the Idyllwild Quadrangle, San Jacinto mountains, California, USA. *Photogrammetric Engineering and Remote Sensing* **56**(11), 1507–1513.
- Chou YH, Minnich RA, Chase RA (1993) Mapping probability of fire occurrence in San Jacinto mountains, California, USA. *Environmental Management* **17**(1), 129–140.
- Chuvieco E, Salas J (1996) Mapping the spatial distribution of forest fire danger using GIS. *International Journal of Geographical Information Systems* **10**(3), 333–345. doi:10.1080/026937996138070
- Cleveland WS, Grosse E, Shyu WM (1992) Local regression models. In 'Statistical models in S'. (Eds JM Chambers, TJ Hastie) pp. 309–376. (Wadsworth & Brooks/Cole: Pacific Grove, CA)
- Cunningham AA, Martell DL (1972) A stochastic model for the occurrence of man-caused forest fires. *Canadian Journal of Forest Research* **3**, 282–287.
- Dayananda PWA (1977) Stochastic models for forest fires. *Ecological Modelling* **3**, 309–313. doi:10.1016/0304-3800(77)90011-4
- Deeming JE, Lancaster JW, Fosberg MA, Furman RW, Schroeder MJ (1972) 'The National Fire-Danger Rating System.' USDA Forest Service Research Paper RM-84.
- Grissino-Mayer HD (1999) Modeling fire interval data from the American Southwest with the Weibull distribution. *International Journal of Wildland Fire* **9**(1), 37–50. doi:10.1071/WF99004
- Hastie TJ (1992) Generalized additive models. In 'Statistical models in S'. (Eds JM Chambers, TJ Hastie) pp. 249–307. (Wadsworth & Brooks/Cole: Pacific Grove, CA)
- Hastie TJ, Tibshirani R (1991) 'Generalized additive models.' (Chapman and Hall: New York)
- Hodges JL, Le Cam LM (1960) The Poisson approximation to the Poisson Binomial distribution. *Annals of Mathematical Statistics* **37**, 737–740.
- Johnson EA, Gutsell SL (1994) Fire frequency models, methods and interpretations. *Advances in Ecological Research* **25**, 239–287.
- Keetch JJ, Byram GM (1968) (Revised November 1988) 'A drought index for forest fire control.' USDA Forest Service Research Paper SE-38.
- Mandallaz D, Ye R (1997) Prediction of forest fires with Poisson models. *Canadian Journal of Forest Research* **27**, 1685–1694. doi:10.1139/CJFR-27-10-1685
- Martell DL, Bevilacqua E, Stocks BJ (1989) Modelling seasonal variation in daily people-caused forest fire occurrence. *Canadian Journal of Forest Research* **19**, 1555–1563.
- Martell DL, Otukol S, Stocks BJ (1987) A logistic model for predicting daily people-caused forest fire occurrence in Ontario. *Canadian Journal of Forest Research* **17**, 394–401.
- Mosteller F, Tukey JW (1977) 'Data analysis and regression.' (Addison-Wesley: Reading, MA)
- Peng R, Schoenberg F (2001) 'Estimation of wildfire hazard using spatial-temporal fire history data.' Technical report. (Statistics Department, University of California: Los Angeles)
- Reed WJ (1998) Determining changes in historical forest fire frequency from a time-since-fire map. *Journal of Agricultural, Biological and Environmental Statistics* **3**, 430–450.
- Reed WJ, Larsen CPS, Johnson EA, MacDonald GM (1998) Estimation of temporal variations in fire frequency from time-since-fire map data. *Forest Science* **44**, 465–475.
- Roads JO, Chen SC, Fujioka F (2001) ECPC's weekly to seasonal global forecasts. *Bulletin of the American Meteorological Society* **82**, 639–658. doi:10.1175/1520-0477(2001)082<0639:EWTSGF>2.3.CO;2
- S-PLUS (2000) 'User's guide.' (Data Analysis Products Division, MathSoft: Seattle)
- Van Wagner CE (1987) 'Development and structure of the Canadian Forest Fire Weather Index System.' Canadian Forest Service, Ottawa, Ontario Forestry Technical Report No. 35.