A State Space Model for predicting Wildland Fire Risk

Haiganoush K. Preisler[1], and John W. Benoit[1]

[1]USDA Forest Service, Pacific Southwest Research Station, Albany and Riverside, California

Abstract. Wildland fire managers have long desired to know the risks of severe fire events well in advance of its happening. A number of actions are available to address severe fire seasons. However, contingency actions have associated costs and timeliness issues. These issues require information about the likelihood of fire occurrence. In this talk we describe a probability model for predicting wildland fire risks based on non-linear state-space models. By the choice of variables included in the state space, we can handle many situations, for example, the history of the process (locations and times of fires) up to the present in addition to characteristics of the environment that might serve as explanatories. We found the model to be useful for assessing the importance of commonly used fire danger indices and for predicting expected numbers of fires in a region.

Keywords: *Cubic splines; fire danger indices; logistic regression; mutual information; thin-plate spline; uniform residuals.*

Introduction

Fire managers rely on a variety of factors to help determine fire severity and make decisions about resource allocation and fire-fighting tactics. The United States National Fire Danger Rating System (NFDRS) use current and historic weather and fuel conditions to produces a set of indices that are then used to generate fire danger maps[1]. Fire experts use NFDRS indices to determine times for prescribed burning, assess the need for fire suppression resources, and make tactical firefighting decisions. The Canadian Wildland Fire Information System[2] generate daily maps of a relative index that describes how easily vegetation will ignite and how much danger a fire may be. Though fire managers rely heavily on the outputs of fire danger rating systems, their actual relationship to fire occurrence has not been thoroughly examined. To this end we have been developing statistical models for predicting fire risk (Brillinger et al., 2003; Preisler et al., 2004).

In what follows we describe a probability model for predicting wildland fire risks based on non-linear state-space models. By the choice of variables included in the state space, we can handle many situations, for example, the history of the process (locations and times of fires) up to the present in addition to characteristics of the environment that might serve as explanatories. We use the model to assess the usefulness of various fire weather and danger indices on occurrence and spread of fires on federal lands in California and to demonstrate how to generate probability based fire danger maps.

Probability Model

Two probabilities will be used for estimating fire danger at a given location and a given time; 1) the probability of ignition (or fire occurrence) and 2) the conditional probability of spread, defined by the probability of an ignition becoming a large fire (> 40.5 hectare). We will start by defining a random variable

$$Y_{x,y,t} = 1 \quad \text{if} \quad U_{x,y,t} > \theta_{x,y,t}$$
$$Y_{x,y,t} = 0 \quad \text{otherwise}$$

where $Y_{x,y,t}$ indicates a wildland fire at location (x,y), time t, and $U_{x,y,t}$ is a state space variable describing conditions (weather, fuel, $Y_{x,y,t-1}$, etc.) at location (x,y) up to time t.

We consider two cases for the threshold parameter θ.

1) ***Deterministic threshold model*** where

$$\theta_i = \text{logit}\{\Pr[Y_i = 1]\} = \alpha_i + g_1(x_i, y_i) + g_2(day_i) + \sum_k g_k(X_{ki}) \quad [1]$$

X_k = explanatory variables (e.g., temp, fuel moisture, spread index, drought index, etc.)

$g(\ )$ = spline functions (e.g., cubic spline, periodic spline, or thin plate spline.)

$\alpha_i = \log(1/\pi_i)$

π_i = sampling proportion for non fire locations on day_i

$i = (x, y, t) = km^2 \times day$

Further details regarding the fitting of the model are found in Preisler et al. (2004) and Maddala (1992). A similar model is used to estimate the conditional probability of a large fire.

2) ***Random threshold model*** where

---

[1] http://www.fs.fed.us/land/wfas

[2] http://cwfis.cfs.nrcan.gc.ca/en/cwfis_intro_e.php

$$\theta_{x,y,t} = \text{logit}\{\Pr[Y_{x,y,t} = 1]\} = \mu_{x,y,t} + \varepsilon_t \qquad [2]$$

μ is given by equation [1] and ε_t is a latent variable describing the processes up to time t. Brillinger et al. (2004) describe a procedure for fitting a random year effect. Here, we will assume that ε_t are serially correlated day effects with

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t,\ |\rho| < 1 \quad \text{and} \quad u_t \sim IN(0,\ \sigma_u^2) \qquad [3]$$

We estimated the parameter ρ in two stages. First, we estimated μ assuming the fixed effect model in equation [1]. Next we estimated ε_t by

$$\hat{\varepsilon}_t = \frac{1}{n_t} \sum_{x,y} res_{x,y,t}$$

where *res* are the residuals from the fixed effect model in [1]. Replacing the errors in equation [3] by the estimated residuals we get

$$\varepsilon_t = \rho(\hat{\varepsilon}_{t-1} + \nu_{t-1}) + u_{t-1} \text{ where } \nu_t \sim IN(0,\ \sigma_\nu^2).$$

Finally, one may estimate ρ using a maximum likelihood routine for an ARIMA model. In the example below we used the arima function in R (R Development Core Team 2004). We found the usual residuals, i.e., Pearson or deviance residuals, not practical for the purpose of estimating the serial correlation because of the discreteness of the response variable. Figure 1 is a plot of deviance residuals - averaged over locations- from a simulation study using 12,000 independent Bernoulli trails with 30 trials per day. The estimated serial correlation for these residuals was $\hat{\rho} = 0.9 \pm 0.03$ while the true ρ used in the simulation was zero.

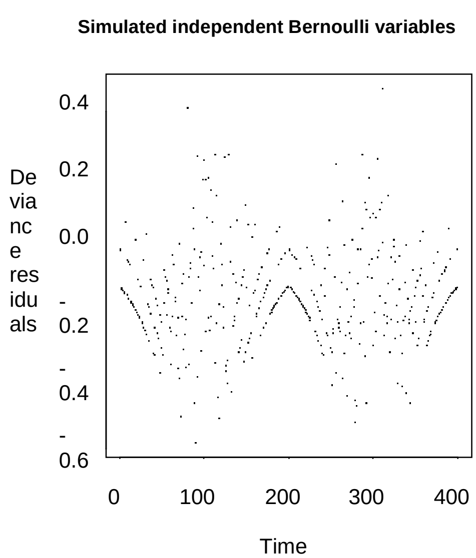

Figure 1: Deviance residuals generated from independent Bernoulli trails with a median response probability of 0.076.

In our analysis we used 'uniform residuals' (Brillinger and Preisler, 1983) which we define as follows. Let $\pi_t = \Pr[\ Y_i = 1 \mid \mathbf{X}]$. For each observation generate two independent uniform random variables, U_{1i}, U_{2i} on the intervals (0, 1- π_t) and (1 - π_t, 1) respectively. Next, calculate the 'uniform residual' $U_i = U_{1i}(1 - Y) + U_{2i}Y$ and their corresponding 'normal residual' $Z_i = \Phi^{-1}(\ U_i)$. Normal residuals, as defined here, have the advantage of being spread out and, consequently, are more useful for estimating the correlation between consecutive observations. Further properties of uniform and normal residuals can be found in Brillinger et al. (2004).

There are many weather and fire danger indices developed as tools for predicting fire danger. Here we propose to study the usefulness of the various indices on estimating the probability of fire occurrence and fire spread. We used the Mutual Information (MI) statistic to study the strength of the statistical dependencies. In particular if Y indicates the occurrence of a fire (or, alternatively, the spread of a fire) and X is the linear predictor as described in [1] then the MI statistic is given by

$$I_{X,Y} = E\left\{\log \frac{p_{X,Y}(X,Y)}{p_X(X)\,p_Y(Y)}\right\}$$

Note that for the bivariate normal case 1-exp(-2 $I_{X,Y}$) is the coefficient of determination. In general, $I_{X,Y} = 0$ when X and Y are independent and $I_{X,Y} \leq I_{Z,Y}$ if Y is independent of X given Z (Brillinger, in press).

The Data

We acquired observed data for all fires that occurred on national forests in California from 2000 through 2003.

The fire data originated from the National Interagency Fire Management Integrated Database (NIFMID).[3] It included the date of the fire, the location, the suspected cause, and the fire size.

We obtained observed weather data from stations in the Weather Information Management System (WIMS) catalog available online since January 2000[4]. Included in each station's daily record were temperature, relative humidity, wind speed, energy release component, spread component, and thousand-hour fuel moisture values (Benoit et al. 2004).

For each day we picked a random sample of locations within California federal lands that did not

[3] http://famweb.nwcg.gov/kcfast/html/ocmenu.htm
[4] http://www.fs.fed.us/land/wfas/wfas29.html

have a fire on that day. The number of non-fire locations sampled each day was proportional to the historic frequency of fires for that day of the year.

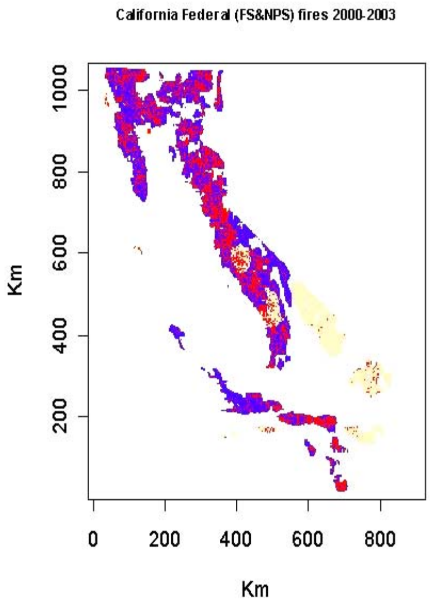

Figure 2: Locations of fires (red dots) on California National Forests (purple) and National Parks (peach) between 2000-2003.

Observed weather variables and NFDRS indices for each day were interpolated from the WFAS station locations to the individual fire or non-fire locations in the sample. The interpolation was done using a generalized additive model with latitude, longitude and elevation as covariates.

Results and Discussion

We studied the effects of seven fire danger indices – thousand-hour fuel moisture (TH), wind speed (WS), spread component (SC), energy release component (ER), dry bulb temperature (DBT), relative humidity (RH) and burning index (BI)- on the probabilities of lighting-caused fires, probabilities of human-caused fires and the conditional probability of large fires. In each case we calculated the MI statistics with Y being the binary response variable; $X = \alpha + g_1(x, y) + g_2(day) + X_k$, and X_k one of the indices.

According to the MI statistics the covariate that was most strongly associated with the occurrence of lighting-caused fires in California was the day-in-year variable (Figure 3). Most fires in California occur in the summer and early autumn months. Two indices, DBT and RH, show a marginal increase in the MI values when added to the model with spatial location and day-in-year.

The covariate that was most strongly associated with the occurrence of human-caused fires was spatial location (Figure 4). The next two most important covariates were day-in-year and elevation. Human-caused fires tend to be at lower elevation. None of the indices appeared to be helpful for predicting human caused fires in California.

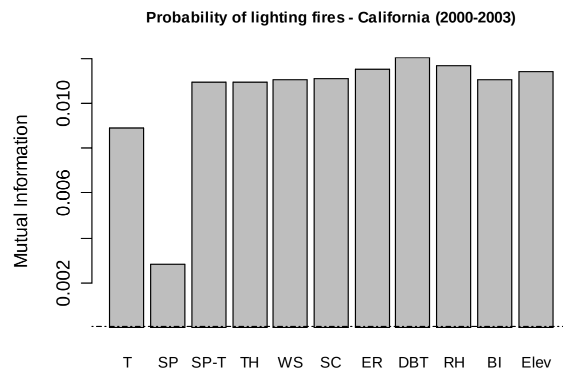

Figure 3: Estimated MI values describing the strength of the association between lighting-caused fires and linear predictors with various combinations of covariates. T is the linear predictor with day-in-year; SP is spatial location; SP-T includes both spatial location and day-in-year, the rest are SP-T model with one additional index. Elev is elevation. Dashed horizontal line at the bottom is the 5% significance level.

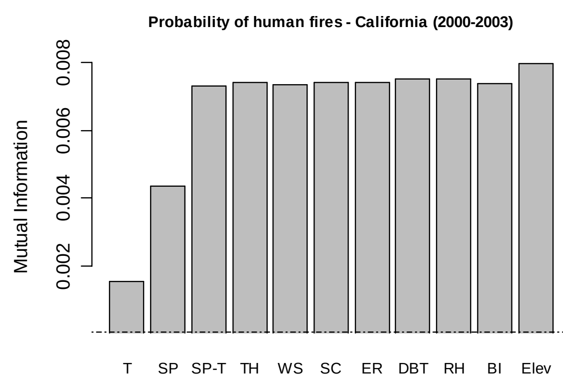

Figure 4: Estimated MI values describing the strength of the association between human-caused fires and linear predictors with various combinations of covariates.

Estimated correlation between days for lighting caused fire was $\hat{\rho} = 0.59 \pm 0.13$. This might be useful for predicting fire occurrence on consecutive days. The corresponding value for human caused fires was $\hat{\rho} = 0.14 \pm 0.11$. Apparently human caused fires on consecutive days are not correlated.

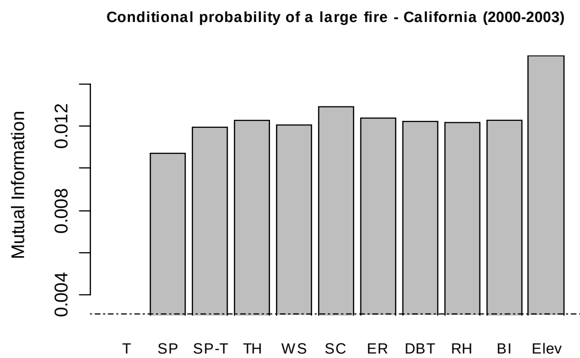

Figure 5: Estimated MI values describing the strength of the association between fire spread and linear predictors with various combinations of covariates. The horizontal dashed line at the bottom is the 95% significant level.

Spatial location and elevation were the two variables most strongly associated with a fire becoming large (Figure 5). The probability of fire becoming large is apparently greater at lower elevations (Figure 6).

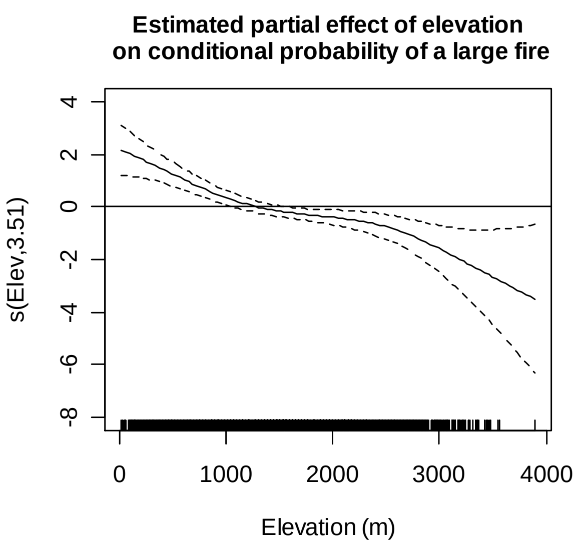

Figure 6: Estimated partial effect of elevation suggesting that the conditional probability of a large fire in California is greater at lower elevations.

The only index that seemed to increase the association with fire spread was the Spread Component! It is important to note that our results were obtained using interpolated weather data from scattered weather stations. It is possible indices generated from other sources of weather data (e.g., data from dynamic weather models) may result in different conclusions regarding the usefulness of fire danger indices in predicting fire occurrence and spread.

The variable day-in-year alone did not have a significant effect on fire spread. However, day-in-year was significant when added to a model with spatial location. According to the estimated partial day effect (Figure 7) fires in California tended to get larger during late fall (November) and none of the indices tested here seem to account for this phenomena.

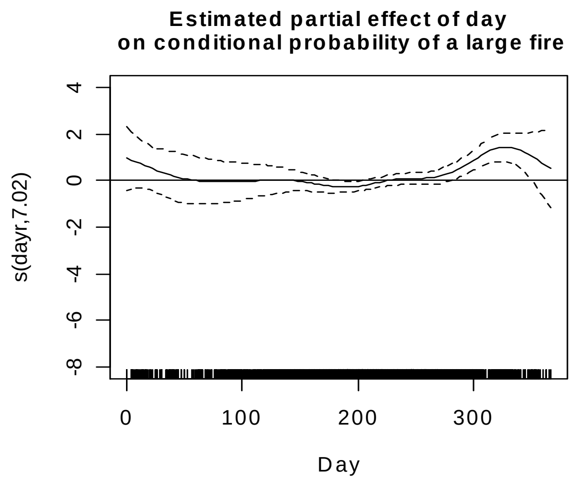

Figure 7: Estimated partial effect of day-in-year. Probability of spread appears to be larger in late autumn than expected from a model that includes an index of spread.

Using the estimated probabilities for human- and lighting-caused fires and the conditional probability of an ignition becoming a large fire we were able to generate ‘fire danger maps’ based on probability of a large fire occurring in a given voxel (km^2x day) given values of the explanatory variables at that location and time. Fire danger maps for two dates in 2003 are given in Figures 8 and 9.

Another useful product are plots of expected total number of fires in a particular region and time interval. As an example of the predictive ability of the model we generated plots of observed and predicted numbers of fires for two-week periods in a region surrounding Yosemite National Park (Figure 10). The model gave reasonable results on average. It is noted that the only fire danger index included in the model (SC) did not account for the greater than expected large fires in the summer of 2003 in Yosemite.

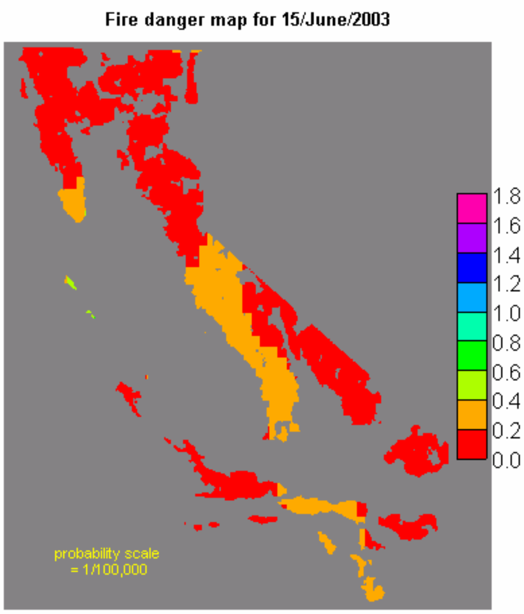

Figure 8: Estimated unconditional probability of a large fire for June 15, 2003. The probability scale is per 100,000 voxels (km^{2}x day).

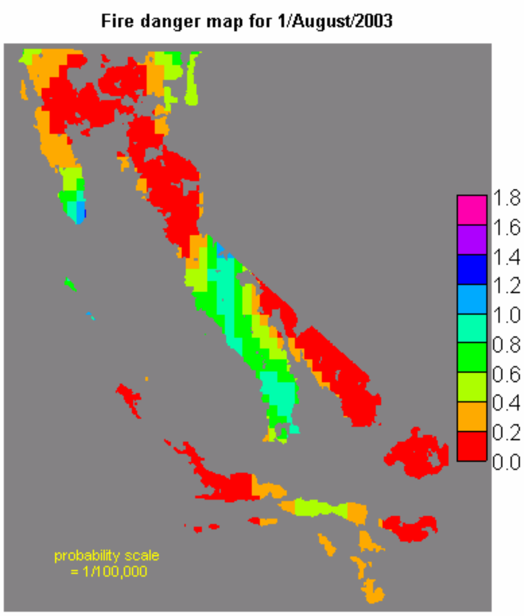

Figure 9: Estimated unconditional probability of a large fire for August 1, 2003. The probability scale is per 100,000 voxels (km^{2}x day).

In conclusion, we were able to produce fire danger maps for California federal lands based on probabilities of fire occurrence and spread. However, we found only one of the fire danger indices marginally useful for predicting fire risk. Part of the reason for this may be because the interpreted values of fire weather and fire danger indices are not good estimates of the actual weather conditions at a given location. We are now working on assessing the utility of weather and fuel data from other sources (e.g. from dynamic weather models and satellite imagery) on predicting probabilities of fire risk.

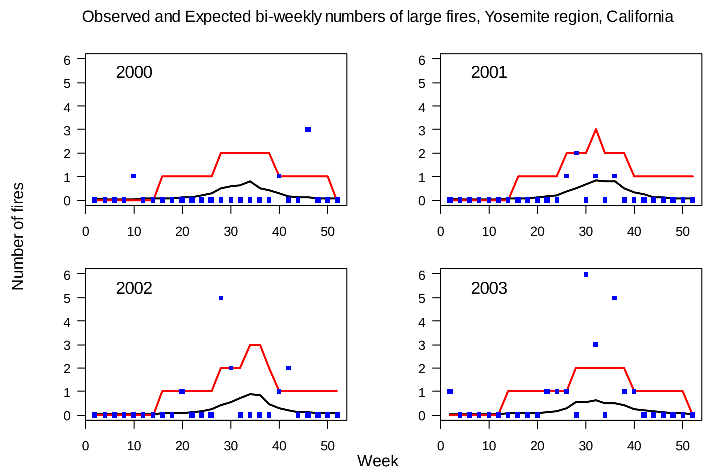

Figure 10: Observed (blue) and predicted (black) numbers of biweekly large fires in the Yosemite region for 2000-2003. The red lines are the point-wise 95% upper bounds.

References

Benoit, J.W., Preisler, H.K. (2004) “Preliminary development of a probability-based fire severity modeling system for California forests” Proceedings of the Second Symposium on Fire Economics, Planning and Policy: A Global View, Córdoba, Spain, April 19-22, 2004.

Brillinger, D. R. and Preisler, H.K. (1983) “Maximum likelihood estimation in a latent variable problem”, Studies in Econometrics, Time Series and Multivariate Statistics. Academic Press, New York 31-65.

Brillinger, D. R. Preisler, H.K. and Benoit , J.W. (2003), "Risk assessment: a forest fire example." In Science and Statistics, Institute of Mathematical Statistics Lecture Notes. Monograph Series.
http://www.fs.fed.us/psw/programs/statistics/staff/hpreisler/WF02061.pdf

Brillinger, D.R., (in press). "Some data analyses using mutual information". To appear in Brazilian J. of Probability and Statistics
www.stat.berkeley.edu/~brill/Papers/bjps1.pdf

Brillinger, D.R., Preisler, H.K. and H. M. Naderi, (2004) "Wildfire chances and probabilistic assessment". TIES Conference proceedings on Spatial Accuracy 2004.
http://stat.berkeley.edu/users/brill/Papers/portland8.ps

Maddala, G.S. (1992). "Introduction to Econometrics" Second Edition, Mac Millan, New York.

Preisler, H.K., Brillinger, D.R., Burgan, R.E. and Benoit, J.W. (2004), "Probability based models for estimating wildfire risk". International Journal of Wildland Fire. Vol 13 (2).
http://www.fs.fed.us/psw/programs/statistics/staff/hpreisler/IMS_Vol40.pdf

R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL
http://www.r-project.org/.