Preliminary Development of a Probability-Based Fire Severity Modeling System for California Forests[1]

John W. Benoit,[2] Haiganoush K. Preisler[3]

Abstract

In this currently ongoing study, we examine historical fire data, weather, and fuel conditions in the national forests in California. We build logistic regression models to determine the probabilities of fire ignitions and of ignitions becoming large fires. We use the same methodology that was developed in a previous study using data for the state of Oregon. Our objective is to further develop this process as a risk assessment tool for fire managers and planners.

Introduction

Fire managers in California rely on a variety of factors to help determine fire severity and make decisions about resource allocation and fire-fighting tactics. California's diverse environmental regions necessitate the use of a range of variables which describe weather conditions, fuel moisture, and fire behavior. An indicator of fire potential considered suitable at one location may be deemed ineffective at another. Each local fire expert might therefore choose a set of indicators to rely on. This can cause difficulty when one wants to compare fire severity across different regions.

In the United States, the National Fire Danger Rating System (NFDRS) is a process that produces a set of indices for any particular area of the United States, given current weather and fuel conditions. These indices give an idea of the fire potential and fire behavior in a given area. Fire experts use NFDRS indices to determine times for prescribed burning, assess the need for fire suppression resources, and make tactical firefighting decisions. Though these indices are heavily relied upon, their actual relationship to fire occurrence has not been thoroughly examined.

In a previous study, we gathered fire history, observed weather station data, and NFDRS index values, along with other geographic layers, for the state of Oregon (Preisler and others 2004). Logistic regression equations were then constructed to estimate, for any given day and location in Oregon, 1) the probability of a fire ignition, and 2) the probability of an ignition becoming a large fire (greater than 100 acres, or approximately 45 hectares). The probability of having a large fire could be computed from these two probabilities.

---

[1] An abbreviated version of this paper was presented at the second international symposium of fire economics, policy, and planning: a global vision, 19–22 April, 2004, Córdoba, Spain.

[2] Computer specialist, USDA Forest Service, Riverside, California.

[3] Statistician, USDA Forest Service, Albany, California.

In the current study, we want to apply and adapt this methodology for use in California. There are a few differences between these two studies – mainly to do with the collected observed data. For example, we have access to more current and higher resolution GIS layers for California than we did for the Oregon study. Also, California is more geographically and climatologically diverse than Oregon. However, these factors should not affect our methodology, as the models we build take such differences into account through explanatory variables; they are robust for any given location.

Data

The Wildland Fire Assessment System (WFAS) provides a collection of observed data and forecast products related to fire severity within the United States. WFAS has made raw (tabular) observed data from stations in the Weather Information Management System (WIMS) catalog available online since January 2000[4]. We obtained this station data for 2000 through 2002 for every day for all stations reporting on each day. Included in each station's daily record were temperature, relative humidity, wind speed, precipitation, energy release component (NFDRS index), spread component (NFDRS index), Keetch-Byram drought index, and ten-, hundred-, and thousand-hour fuel moisture values. It was of benefit that this data included NFDRS information along with weather and fuels conditions, since generating NFDRS indices from historical data is a troublesome process. Prior to 2000, historic NFDRS indices were not archived, nor were some of the necessary input variables into NFDRS easily accessible. For example, observed fuel moisture – a key input into NFDRS – was sparsely recorded at many WIMS stations.[5]

We acquired observed data for all fires that occurred on national forests in California from 2000 through 2002. The fire data originated from the National Interagency Fire Management Integrated Database (NIFMID).[6] It included the date of the fire, the location, the suspected cause, and the fire size. Since this set of fire data contained not only fires originating in, but also fires *that threatened* the forests (e.g. fires started on lands adjacent to the forests), the California national forest boundaries were used as a mask to ensure that we had a list of fires whose ignitions truly occurred within the forest. Also, only fires that consumed more than 0.1 acres and classified as wildfires were included. There were 4746 fires that met this criterion during the study period (Figure 1).

Elevation data for the national forests was obtained at kilometer resolution. Because the intended application of the procedure we are developing is over fairly large regions, the one-kilometer spacing of values was adequate.

Finally, a fuel model data layer for California was taken from the California Department of Forestry and Fire Protection's Fire and Resource Assessment Program (FRAP).[7] This was a map of vegetation grouped according to similar burning

---

[4] See http://www.fs.fed.us/land/wfas/wfas29.html. Before 2000, only maps (graphics) of spatially interpolated values are available from this site.

[5] In our Oregon study, we used NFDRS indices for 1989 through 1996 generated by a process different than that used to produce indices at the WIMS stations in WFAS. While developing our procedure for California, we did not want to combine these two different data sets until we can verify that their general characteristics are similar.

[6] See http://famweb.nwcg.gov/kcfast/html/ocmenu.htm.

[7] See http://frap.cdf.ca.gov/.

characteristics. The fuel model data was resampled to have a cell size of one kilometer.

All data layers were georegistered (spatially aligned) to one another. We converted all coordinates to the Lambert Azimuthal geographic projection, which places locations in a Cartesian plane. This becomes convenient in our methodology, when we want to compute the expected number of ignitions or large fires for a given day.

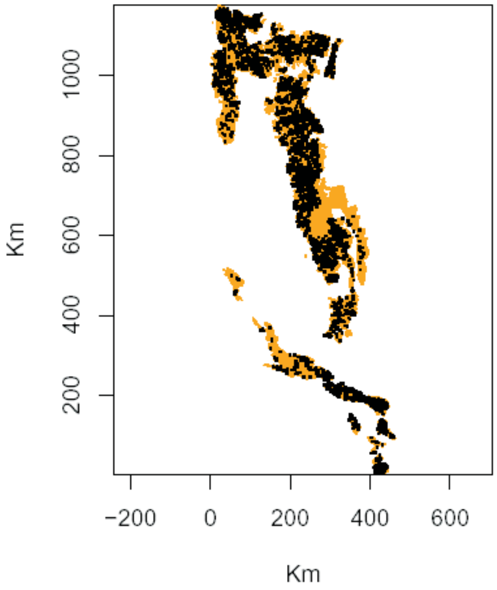

Figure 1—California federal fires from 2000 to 2003. The shaded area represents national forest land within the state of California. The black dots are the locations of fire ignitions.

Methodology

The observed study data was spatially arranged on a Cartesian grid with a one kilometer unit size and temporally by day. Our basic unit for sampling was a single square kilometer for a particular day. This unit was referred to as a *voxel*. In our California national forests data for 2000 through 2002, there were 117,438,750 voxels, with 5,317 of these having at least one fire ignition. Since the proportion of voxels with fires to those voxels without fires is extremely low, a two-stage sampling technique was implemented (Brillinger 2003). This involved defining a set that included all voxels that had fires and a small sample of the voxels with no fire[8]. Parameter estimation can then be performed using this smaller set of voxels rather than on the entire population of the study data.

For each day, observed weather, fuel moisture, and NFDRS values at specific stations are interpolated from the WFAS station location values to the individual voxels. Inverse distance weighted interpolation can be used to arrive at the one-kilometer cell values for any particular day. Alternatively, other interpolation

---

[8] The size of the sample of voxels with no fire is determined by choosing a small proportion (e.g. ~0.15%). The entire collection no-fire voxels are then randomly sampled at this rate.

methods, such as loess or thin-plate splines can be used, and most likely will give more accurate values. The interpolated grids of values are stored for later use.

We now consider an indicator variable such that:

$N_{x,y,t} = 1$ if there was at least one ignition in the voxel at location (x, y) and day t, and

$N_{x,y,t} = 0$ if there were no ignitions in the voxel

Next we can define the probability p_1 that a fire ignition occurs in voxel (x,y,t) with the logistic model:

$$p_{1,x,y,t} = \text{Prob}(N_{x,y,t} = 1 \mid U_{x,y,t}) = \frac{e^{\theta_{x,y,t}}}{1+e^{\theta_{x,y,t}}} \quad (1)$$

where $U_{x,y,t}$ represents the historic values of explanatory variables at voxel (x,y,t) and $\theta_{x,y,t}$ is the set of parameters we wish to estimate. The *logit* of $p_{1,x,y,t}$ is defined as:

$$logit(p_{1,x,y,t}) = \log\left(\frac{p_{1,x,y,t}}{1-p_{1,x,y,t}}\right) \quad (2)$$

From (1) and (2), we have

$$logit(p_{1,x,y,t}) = \theta_{x,y,t} = \beta_{0,x,y,t} + \beta_{1,x,y,t}x_{1,x,y,t} + \ldots + \beta_{n,x,y,t}x_{n,x,y,t} \quad (3)$$

where n is the number of independent variables in the model. The value of $logit(p_{1,x,y,t})$ can range anywhere in $(-\infty,\infty)$. Assuming independence among the voxels, we can use regression techniques to obtain estimates of the coefficients in (3).[9] Using generalized additive model (GAM) would allow for nonlinear relationships among the explanatory variables:

$$\begin{aligned} logit(p_{1,x,y,t}) = {} & g_1(x,y) + g_2(t) + g_3(Elev_{x,y,t}) + g_4(FM_{x,y,t}) + g_4(BI_{x,y,t}) \\ & + g_5(ERC_{x,y,t}) + g_6(KB_{x,y,t}) + g_7(TH_{x,y,t}) + g_8(WS_{x,y,t}) \\ & + g_9(RH_{x,y,t}) + g_{10}(Temp_{x,y,t}) \end{aligned} \quad (4)$$

Here, $Elev_{x,y,t}$ is elevation, $FM_{x,y,t}$ is the fuel model, $BI_{x,y,t}$ is the burning index, $ERC_{x,y,t}$ is the energy release component, $KB_{x,yt,}$ is the Keetch Byram drought index, $TH_{x,y,t}$ is the thousand hour fuel moisture, $WS_{x,y,t}$ is the wind speed, $RH_{x,y,t}$ is the relative humidity, and $Temp_{x,y,t}$ is the temperature at voxel (x,y,t). The functions $g(\bullet)$ are nonparametric smoothers estimated simultaneously with the GAM. $g_1(x,y)$ represents location effect and $g_2(t)$ is the temporal effect in the model.

The conditional probability $p_{2,x,y,t}$ that a large fire occurs (over 100 acres) given that there was an ignition can be modeled using very much the same technique as that for finding $p_{1,x,y,t}$. The same or a similar set of coefficients may be used. In the case of $p_{2,x,y,t}$, however, coefficients are estimated using only the set of the voxels in which fires *did* occur.

---

[9] Naturally, we cannot easily assume that the data is temporally and spatially independent. For example, after a fire has burned an area (a set of voxels), another fire in the same location is not likely to occur for some time. We hope that some of these relationships are accounted for by the explanatory variables with each voxel – a recently burned area may exhibit a change in its fuel moisture, which would affect the chance of another fire occurring in the same location soon afterwards.

Finally, given $p_{1,x,y,t}$ and $p_{2,x,y,t}$, we can easily compute the unconditional probability that a large fire will occur, since

$$p_{3,x,y,t} = \mathrm{P}(\mathbf{A} \cap \mathbf{B}) = \mathrm{P}(\mathbf{A}) \cdot \mathrm{P}(\mathbf{B} \mid \mathbf{A}) = p_{1,x,y,t} \cdot p_{2,x,y,t} \qquad (5)$$

where

A = event of a fire ignition, and

B = event that an ignition becomes a large fire.

A convenient feature of using voxels as our basic unit for computing probabilities is that we can quickly get the expected number of fire events for a particular day. For example, the expected number of fires within the study area or subset of the study area for a specific day is the sum of the values of $p_{1,x,y,t}$ at every kilometer in the desired area for that day. The expected number of ignitions becoming large fires and the expected number of large fires per day can be found in the same way, using $p_{2,x,y,t}$ and $p_{3,x,y,t}$, respectively. Additionally, confidence intervals can be derived for the three probabilities (see Preisler and others 2004).

Further Work

Our study of California is currently in progress. Presented here are the findings as of now, along with the procedure we intend to use (much of which has been performed on our study of Oregon). There are some options that we plan to investigate. We would like to get a larger study period than just the years 2000-2002. The main hindrance has been obtaining historic values of NFDRS indices, which we wish to relate to fire occurrence. We are presently investigating possible methods of generating historic NFDRS values.

Our procedure allows for the testing of different spatial interpolation methods on the observed daily station data. Inverse distance weighted interpolation is one of the simplest schemes to employ (and commonly used in WFAS products), but more accuracy can be found in using locally-weighted methods, such as loess and spline techniques.

The work outlined here has been concerned with observed data. However, the intention is to use the developed logistic regression equations in forecasted data to produce maps of future severity probability (Figure 2). The testing of different weather model outputs will be of interest. We therefore want to eventually integrate different weather models, such as RSM (Roads 2000 & 2002) or MM5, to our procedure.

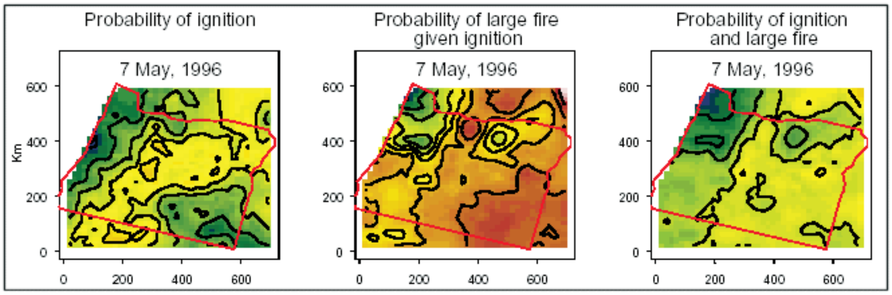

Figure 2—Maps of probabilities of fire ignition, of a large fire given an ignition, and of an ignition and large fire for Oregon. Colors range from blue/dark green (low

probability) to reddish brown (high probability). We will produce similar maps for California.

The intent of this work is to eventually produce a fire severity forecast product that is based on technology and tools being currently used by the fire community. We would hope this product is scaleable, from a national scope down to the forest district level. Finally, it should be easy to use and understand. This probability-based approach shows promise in achieving these tasks.

References

Bradshaw, Larry S.; Deeming, John E.; Burgan, Robert E.; Cohen, Jack D., compilers. 1984. **The 1978 National Fire-Danger Rating System: technical documentation.** Gen. Tech. Rep. INT-169. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 44 pp.

Brillinger, D.R.; Preisler, H.K.; Benoit, J. W. 2003. **Risk assessment: a forest fire example.** In: Goldstein, Darlene R., editor. Institute of Mathematical Statistics Lecture Notes – Monograph Series 40: 177-196.

Preisler, H.K.; Brillinger, David R.; Burgan, Robert E.; Benoit, J.W. 2004. [In press]. **Probability based models for estimation of wildfire.** International Journal of Wildland Fire.

Roads, J.O; Chen, S.-C.; Kanamitsu, M. 2002. **US regional climate simulations and seasonal forecasts.** Journal of Geophysical Research-Atmospheres.

Roads, J.O.; Fujioka, F.; Burgan, R. 2000. **Development of a seasonal fire weather forecast for the contiguous United States.** Preprints. Third symposium on fire and forest meteorology, Long Beach, CA. Amer. Meteor. Soc.; 99-102.

S-PLUS 2000 guide to statistics, vol 1. Data Analysis Products Division, MathSoft, Seattle, WA; 638 p.