Projecting Wildland Fire Severity using RSM Simulations with a Probability Model

Shyh-Chin Chen, Haiganoush K. Preisler, John W. Benoit, and Francis Fujioka
USDA Forest Service USDA, Pacific Southwest Research Station, Riverside, CA.
email: schen@fs.fed.us

Abstract

The National Fire Danger Rating System (NFDRS) indices deduced from RSM 50-km 1-day simulations from January 1998 through December 2003 are used in conjunction with a probability model to determine the expected number of fire occurrences and large fires. The RSM simulation and forecasts are ongoing experimental products from the Experimental Climate Prediction Center (ECPC) at Scripps Institution of Oceanography. The probability model uses non-parametric logistic regression with spline functions for evaluating relationships between covariates and probabilities of fires. The 1-day RSM simulated NFDRS indices and surface meteorological variables are used with the probability model to assess the skill of the system in determining the observed fire severity. The Fire Weather Index (FWI) which was derived from weather variables only, along with NFDRS indices of the Keetch-Byram Drought Index (KBDI) and Energy Release (ER), were found to produce more significant mutual information of the observed big fire events than all other stand-alone weather variables. These selected indices, in addition to historical fire information, were used in the probability model to determine the expected number of fire events. It was shown that the probability model using combined fire danger indices outperformed the one with historic information only. Geographical maps of wildland fire probability were subsequently produced and reasonably well matched the actual fire events. This method paves a feasible way to use climate forecast output from a dynamical meteorological model with a statistical model to predict the probability of wildland fire severity.

I. Introduction

The US Forest Service (USFS) National Fire Danger Rating System (NFDRS) (Deeming *et al.* 1977) reflects the conceived fire behavior and potential from the effects of terrain, weather and fuel condition from fuel models. Although the system was originally developed for the main purpose of firefighter safety, many wildland fire management agencies also use it for quantification of risk element, staffing level, appropriate suppression response, and strategic decision (NWCG Fire Weather Working Team 2005). The reliability and integrity of the system, however, are hampered by difficulties arising from the use of non-stationary, insufficient spatial coverage, and non-standardized data communication and archiving of weather station data. Using global or regional scale weather analysis as weather input is an effective alternative to partially overcome the aforementioned shortcomings of the current NFDRS monitoring system. The weather model can provide a thorough and dynamically consistent data set with ample spatial coverage to the NFDRS. It also has potential value for providing predictive NFDRS with a lead time of a season or longer.

Recently, Roads et al. (2005) evaluated experimental forecasts of NFDRS indices at weekly to seasonal scale using a meteorological model as weather input. They showed that these indices can be well predicted at weekly time scales when validated against the validation indices deduced from the weather model 1-day simulation. How these model-deduced indices are associated with observed fire characteristics such as fire counts and acres burned remains unclear. As demonstrated in Roads et al. (2005), there is only a weak relationship between their validation indices and the observed fire counts/acres burned. In this study, a probability model approach (Brillinger et al., 2003; Preisler and Westerling 2005) is used to assess the skill of the model-deduced monthly fire danger variables in estimating large fire events.

II. Fire danger and fire variables

The fire danger variables in this study were adapted from the meteorological forecasting system developed at the Experimental Climate Prediction Center (ECPC) in La Jolla (Roads *et al.* 2005). The model system uses operational daily 00 UTC analyses from the NCEP Global Data Assimilation (GDAS), which is used for NCEP's global extended range weather forecast, as initial condition for regional forecast up to 16 weeks lead time.

The regional spectral model (RSM) used in this study was originally developed at NCEP (Juang and Kanamitsu 1994; see also Juang *et al.* 1997). The RSM is a regional extension of the global spectral model (GSM; Kalnay et al 1996). The RSM provides an almost seamless transition from the GSM to the higher resolution region of interest (Chen *et al.* 1999) and thus avoids a common regional model problem when using incompatible physics between the driving global model and the nested regional model (Chen 2001). The description of the RSM and the model setup used in this study can be found in Roads et al. (2003).

Global analyses from January 1, 1998 through December 31, 2003 were used to initialize and force at the lateral boundary of the regional spectral model (RSM) for one day integration for each initial day. Horizontal grid space of 60 km was used in RSM. The 1-day forecasted surface weather variables, including temperature, 2-meter relative humidity (R2H) and wind speed from the model, top 10-cm soil moisture contest (SMC1) along with observed precipitation, fuels and slope were the input for the NFDRS indices computation (Burgan 1988). The major difference of our NFDRS calculation from the standard one is the use of model weather output instead of weather station observations. Interested readers can refer to Roads *et al.* (2005) and Burgan (1988) for a more detailed description. Since not all standard NFDRS indices are useful to fire managers, we chose to examine only spread component (SC), energy release component (ER), burning index (BI), ignition component (IC) and Keetch-Byram (KB) drought component. In addition, fire weather index (FWI), 2-meter relative humidity (R2H), and the first 10 cm of soil moisture content (SMC1) from the meteorological model were also included to contrast the skill from NFDRS indices.

FWI (Fosberg 1978; Fujioka and Tsou 1985) is an index derived only from weather variables of temperature, relative humidity and wind speed, by assuming constant grass fuel and equilibrium moisture content as a function of weather variables. This index is not part of the NFDRS and requires only instantaneous value from the weather model. Due to its easy application, FWI has been used for seasonal fire danger forecasting to provide a first look at global wildfire condition (Roads *et al.* 1995). As will be shown, despite its absent use of surface fuel information, FWI offers significant skill in explaining fire occurrence.

This work relied on fire history data sets compiled from Federal land management agency fire reports. Westerling et al. (2003) compiled a gridded one-degree Lat/Lon data set of monthly fire starts and acres burned from approximately 300 thousand fires reported by the USDA Forest Service and the USDI's Bureau's of Land Management and Indian Affairs and the National Park Service for 1980-2003. Since we have meteorological model derived fire danger indices from January 1998 through December 2004, we will only use the fire data for the same period.

III. Statistical Methods

III.a. Probability models

We used logistic regression with piece-wise polynomials (Hastie et al. 2001, Preisler and Westerling 2005) to estimate two probabilities of fire risk. The probability of 'ignition' was defined as the probability of at least one fire occurring in a given one-degree grid cell during a given month. The probability of 'spread' was defined as the conditional probability of a burn area greater than 400 ha given at least one fire occurrence in the one-degree cell during a given month. The product of these two probabilities, namely, the probability that an area greater than some specified value will burn in a given grid cell during a given month of a given year, was used as a metric for fire danger. Hereafter, for ease of understanding, we will use the terms "fire occurrence" and "large fire" for "ignition" and "spread" respectively.

The logistic regression line used to estimate the probabilities of fire occurrence and large fire is specified in the following equation

$$\text{logit}(p_v) = \beta_o + g_1(lon_v, lat_v) + g_2(month_v) + \sum_{m=1} g_{m+2}(X_{mv}) \qquad [1]$$

where the subscript, *v*, indicates the one-degree by one-month voxel; *p* is either the probability of ignition or probability of large fire; *(lon, lat)* are the longitude and latitude of the mid point of the grid cell; X_m are explanatory fire weather and fire danger variables. The terms *g()* are semi-parametric smooth functions (Hastie et al. 2001) such as piecewise polynomials, periodic splines (for estimating month-in-year effect) and thin plate splines (for estimating the spatial surface as a function of *lon* and *lat*). Estimation was done with the R statistical package (R Development Core Team, 2004). Further details of the preceding probability risk model are found in Brillinger et al. (2003), Preisler et al. (2004), Preisler and Benoit (2004) and Preisler and Westerling (2005).

We used the Mutual Information (MI) statistic to study the strength of the statistical dependencies between explanatory variables (e.g. a fire danger index) and the probabilities of fire risk. In particular, we used the MI statistic to select the indices, or combination of indices, with the most 'information' regarding the probability of fire risk. Details of the MI statistic are given in Preisler and Benoit (2004). Three statistical models with logit line were compared using the MI statistic, they are 1) historic model (H), 2) fire danger model (FDI), and 3) combined model (C). H model included all terms except the fire and weather variables in [1]. With this model each cell has a different probability for each location and month of the year but the probabilities do not change from year to year. FDI model used one of the fire and weather variables that was excluded in H. C model applied all term included all weather and fire indices as selected by the MI criterion.

The final models, with the selected set of indices, are next used to estimate the probability of fire occurrence, p_i, and the conditional probability of large fire, p_s. Finally, the probability of a larger fire event is defined by multiplying the two probabilities, namely, $\hat{\pi} = \hat{p}_i \times \hat{p}_s$.

III.b. Accessing model skills

We assessed the overall goodness-of-fit of the final selected model by grouping together all cells with similar estimated probabilities (within 3% of each other) and comparing the observed fraction of responses in each group with the corresponding estimated probability of response. Response here is defined as a voxel with a large fire event. Estimated probabilities for each voxel were produced using cross-validation. Specifically, estimations for a given year were done by using the model parameters from all other years except the year being evaluated.

The skill of the model in estimating the total number of voxels per month with large fire event was assessed by comparing observed numbers of monthly totals for each year with the deduced 50^{th} and 95^{th} percentiles. The 95^{th} percentile includes both natural variation (Poisson) and variation due to the error in the estimated model parameters.

III.c. Maps of Odds Relative to Norm

The methods described here may be used to produce maps showing departures from 'normal' conditions. In this study the 'norm' is the estimated probability of a large fire event produced by using the historic model (H). Since our study was based on six years of data (1998-2003) the 'norm' reflects average conditions during these particular six years. Maps of estimated departure from the norm were produced using the odds ratio statistic. The rules used to produce the maps of odds were as follows: let $\hat{\pi}_C$ be the estimated probability of a large fire in a given voxel using the combined model C described above; and $\hat{\pi}_H$ is the estimated probability from the historic model H. Let $\hat{\theta} = \log(\hat{\gamma})$ be the logarithm of the estimated odds ratio, $\hat{\gamma} = \hat{\pi}_C/(1-\hat{\pi}_C) \div \hat{\pi}_H/(1-\hat{\pi}_H)$, i.e. the logarithm of the odds relative to historic values. Risk maps were produced using the rules:

$$\begin{array}{ll} \text{Lower than historic} & \text{if } \hat{\theta} < -\hat{\sigma} \\ \text{Normal} & \text{if } -\hat{\sigma} \le \hat{\theta} \le \hat{\sigma} \\ \text{Higher than historic} & \text{if } \hat{\sigma} < \hat{\theta} \le \log(3) + \hat{\sigma} \\ \text{Extreme} & \text{if } \hat{\theta} > \log(3) + \hat{\sigma} \end{array} , \qquad [2]$$

where $\hat{\sigma}$ is the standard deviation from the historic odds. A voxel is designated as extreme if the odds ratio is greater than 3, i.e., the odds are at least three times as large as the historic odds.

IV. Results

Mutual information statistics (MI) were estimated for various probability models using fire occurrence and large fire data described above. The indices FWI and 2-meter relative humidity (R2M) indicated the highest relative increase in strength of dependence with fire occurrence when added independently to the historic (H) model. The indices with the highest relative increase in strength of dependence with the conditional probability of large fire were KBDI, FWI, IC and R2H. R2H is one of the input variables for computing FWI, hence the correlation between FWI and R2H was very high (r = -0.92). Because FWI was found to show dependence with both probabilities of fire occurrence and large fire, in the next stage of model development we started with the FDI model that included FWI. Next we developed combined models (C) by adding the rest of the indices, one at a time. The order in which the indices were added to the probability model was such that those with the smallest correlation with FWI were added first. We should point out that correlation reflects only the linear relationship between two variables. Since our probability model as well as the relationship between FWI and other variables is not entirely linear, a significant increase in MI by adding another index, which has high linear correlation with FWI, might be still possible.

The progressively increased MI is shown in Figure 1. Note that the increase of MI after the first few indices was relatively negligible, the final combined model (C) for the probability of fire occurrence included only the indices FWI, ER and KBDI in addition to location and month. The final combined model for the conditional probability of large fire included FWI, ER, KBDI and R2H in addition to location and month. For its relatively non-negligible contribution to the increase of collective MI, the R2H was one of the variables empirically selected for the conditional probability of large fire even thought it was highly correlated with FWI. R2H is basically inversely proportional to FWI when R2H is high. However this linear relationship breaks down during dry and high FWI weather. Obviously wind is playing an even more critical role under the circumstances.

To demonstrate the skill of estimating fire occurrence, Figure 2 shows the observed fraction of large fire events *vs.* the estimated probabilities from historic (H) and combined (C) models. The observed fraction is the ratio of all voxels with the same estimated probability to those with actual large fire observed. The scattered point of observed fractions and estimated probabilities of large fire events were mostly within the expected point-wise 95% confidence bounds, which are represented by the two dashed lines, for both the historic model (H) and the model with combined indices (C). The confidence bounds increase with increasing probabilities due to the small number of voxels in the high probability groupings. Statistically, the overall Chi-square goodness-of-fit statistic dropped from 36.8 (P-value = 0.0008) for model H to 19.2 (P-value = 0.51) for model C. Moreover the increase in the estimated probability within the range of observed conditions ranged between 0 – 0.72 in the combined model as opposed to 0 - 0.56 in the historic model. The latter implies that by using values of weather indices we were able to identify the probabilities of large fire events as high as 72%. Therefore it is shown that the historic model lacks the ability to project high probability for large fire events so that the scattered distribution skews to the upper side of the diagonal line. The combined model, however, gives significant improvement by showing comparable probabilities against those observed within the 95% confidence bounds.

One of the possible outputs of the probability model are maps of departure from the 'norm', as given by estimated odds ratios relative to historic (see [2]). Maps of odds ratio are particularly useful when accompanied by probabilities of large fire events. For example, the estimated odds of a large fire event appeared to be higher than the norm in the southwestern states in May 2002 and in the northwestern states in August 2003 (Figure 3 left panels). Although the overall estimated probabilities for May 2002 in the Southwest (Figure 3 top right panel) were low (<20%), we still expected and observed a few large fire events. On the other hand, in August 2003 the estimated norms for the northwestern states were higher than the norm and because the probabilities were also high (mostly > 50%) we did observe many large fire events.

However, the most useful application of probability model is to estimate the total number of large fire events for a given month over a given area by adding up the probability at all cells. If the period of the analysis data used in the probability model was long enough, this estimate could be applied over a small region for fire management use. Here we show the monthly estimated as well as the observed large fire events in Figure 4 for the northwestern and southwestern states separated by 40N. The estimated numbers are plotted with 50th and 90th percentiles in solid curves. The observed numbers are marked with dots. The 50th percentile estimates from historic model are also given in grey lines. Historically, the Southwest lags the Northwest by one month reaching the peak of large fire occurrence during fire season. Over the northwestern region, a higher than normal number of big fire events were observed and hence well estimated for years 2000 and 2003, except for 2001 summer. In the southwestern region, the inter-annual variations of fire events were not as apparent. However, summer of 2002 shows an observed early peak in June, compared to the historical model, and was well captured in the estimates. The higher and lower odds relative to the historic model over the northwestern and southwestern states for May 2002 and August 2003 respectively can also be found in the time series. Overall, all estimates were distributed around the 50 percentile estimates and definitely under the 90 percentile curves.

V. Summary and Discussion

A statistical method of estimating big wildland fire events has been applied to the monthly mean fire danger indices of the numerical weather prediction products from the Experimental Climate Prediction Center (Roads et al. 2005). Logistic regression with piece-wise polynomials (Hastie et al. 2001, Preisler and Westerling 2005) was used in the statistical model. With the mutual information analysis, it is found that FWI, KBDI, ER, and R2H explained most of the variability in fire occurrences and big fire events. These variables were subsequently chosen to construct the combined probability model for estimating the probability of the big fire events.

The mutual information is not only practical for selecting variables and producing probability maps of fire risk, it is also useful in assessing the adequacy of the fire danger indices in describing the observed wildland fire events at temporal scale longer than a few days. Since the NFDRS is probably designed to support fire fighting tactics, some of the indices, such as SC, BI and IC, are sensitive to short term variation of weather components, especially wind speed. These indices, however, might lose their characteristics when a long term average is taken. Therefore it is not too surprising to see that these indices did not add information to those slow varying indices, such as KB and ER, in describing observed big fire events. What is surprising is that FWI, an index determined by weather variables only, demonstrates a significant contribution to the mutual information statistics. It is not clear whether the simplified relationship of weather variables to the index or the assumed constant grass-type surface fuel preserved the fire characteristic at monthly scale. Further analysis is required.

It is promising that combinations of fire danger indices have good skills in estimating the probability of large fire events at monthly scale. The combined indices model out-performed the historic model, in which neither weather variable nor fire danger index was used in producing reasonable probability, especially at high probability cases. Furthermore, these estimated probabilities at each cell can be developed further into monthly anomaly maps for fire danger, which has been shown to be in good agreement with the observed fire events.

While the traditional anomaly maps are useful to fire managers in identifying high fire risk areas, the most useful application might be the ability to assess the total number of big fire event against historic estimates over a region in a probabilistic manner. Roads et al. (2005) showed that although the meteorological model predicted fire danger indices quite well even at seasonal time scale, the association between the observed fire occurrence/acre-burn and the "observed" fire danger indices were poor, let alone to the predicted one. Part of the reason could be that fire events have such a nondeterministic nature that straightforward point-to-point temporal correlation evaluation is inadequate. Here we propose an alternative to evaluate the association between the derived fire danger indices and the observed fire characteristics through a probabilistic framework over a certain area. The result indicates that the estimated numbers of big fires within 95 and 50 percentiles agree fairly well with those observed.

Similar analysis needs to be done with forecasted monthly fire weather/danger indices to assess the skill of the forecasted variables on predicting large fire events for this method to be truly useful for fire managers. Future work will address the skill of predicting big fire events at different lead times at various geographical regions. Our current work can only be evaluated at a regional scale as big as half the western US due to the limitation of the coarse spatial resolution of fire data and the short duration of the available model derived fire danger indices. If these two shortcomings can be overcome, we may be able to focus over a smaller jurisdictional region so that the prediction can be used for fire management operation.

References

Brillinger, D.R., Preisler, H.K., and Benoit, J.W., (2003), "Risk assessment: a forest fire example." In (D. R. Goldstein Ed.) *Science and Statistics, Institute of Mathematical Statistics Lecture Notes.* Monograph Series, 177-196.

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).

Chen S. -C., Roads JO, Juang H. -M. H., Kanamitsu M. (1999) "Global to regional simulations of California

wintertime precipitation. *J. of Geophys. Research-Atmospheres*, 104(D24), 31517-31532.
Chen S. -C. (2001) "Model mismatch between global and regional simulation". *Geophys. Res. Lett.*, 29 (5) 4.1 – 4.4.
Deeming JE, Burgan RE, Cohen JD (1977) "The National Fire-Danger Rating System-1978". USDA Forest Service, Inter-mountain Forest and Range Experiment Station General Technical Report INT-39. (Ogden, UT)
Fosberg MA (1978) Weather in wildland fire management: the fire weather index. In *Proceedings of the Conference on Sierra Nevada Meteorology*, South Lake Tahoe, NV. Pp. 1-4.
Fujioka FM, Tsou T-H (1985) "Weather in wildland fire mana-gement: the fire weather index". In *Proceedings of the Conference on Sierra Nevada Meteorology*, South Lake Tahoe, NV. pp. 1-4.
Hastie, T.J., Tibshirani, R., Friedman, J. (2001), "*The Elements of Statistical Learning". Data Mining, Inference, and Prediction*. Springer, New York. 533 pp.
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*, 13, 133-142.
Juang H. -M.H. and Kanamitsu M. (1994), "The NMC nested regional spectral model". *Mon. Wea. Rev.*, 122, 3-26.
Juang H. -M. H., Hong S., and Kanamitsu M. (1997), "The NMC regional spectral model. An update". *Bull. Amer. Meteor. Soc.*, 78, 2125-2143.
Kalnay E., M. and colleagues (1996), "The NCEP/NCAR 40-year Reanalysis project". *Bulletin of the American Meteorological Society*, **77**, 437-471.
Preisler, H.K. and Benoit, J.W., (2004), "A State space model for predicting wildland fire risk", Proceedings of the Joint Statistical Meeting, Toronto, August 8-12, 2004.
Preisler, HK and Westerling, AL, (2005) "Estimating risk pro-babilities for wildland fires" Proceedings of the Joint Statistical Meeting, Minneapolis, MN, August 8-12, 2005.
Roads JO, Fujioka F, Chen S-C, Burgan R (2005) "Seasonal fire danger forecasts for the USA". *International Journal of Wildland Fire*, 14, 1-18.
Roads JO, Chen S-C, and Kanamitsu M (2003) "US regional climate simulations and seasonal forecasts. *Journal of Geophysical Research-Atmospheres*, 108 (D16), 8606, doi:10.1029/2002JD002232.
R Development Core Team (2004), "*R: A language and environment for statistical computing*". R Foundation for Statistical Com-puting, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Westerling, A.L., Gershunov, A. and Cayan, D.R., (2003), "Statistical Forecasts of the 2003 Western Wildfire Season Using Canonical Correlation Analysis," *Experimental Long-Lead Forecast Bulletin*, 12(1,2)

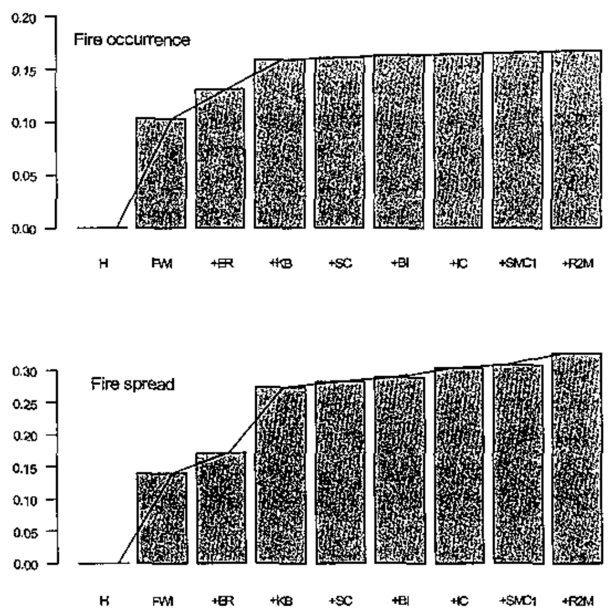

Figure 1: Mutual information statistic for various combined models. All values are relative to the historic (H) value which is set to zero. The models were developed by adding indices consecutively in the order seen in the figures (left to right).

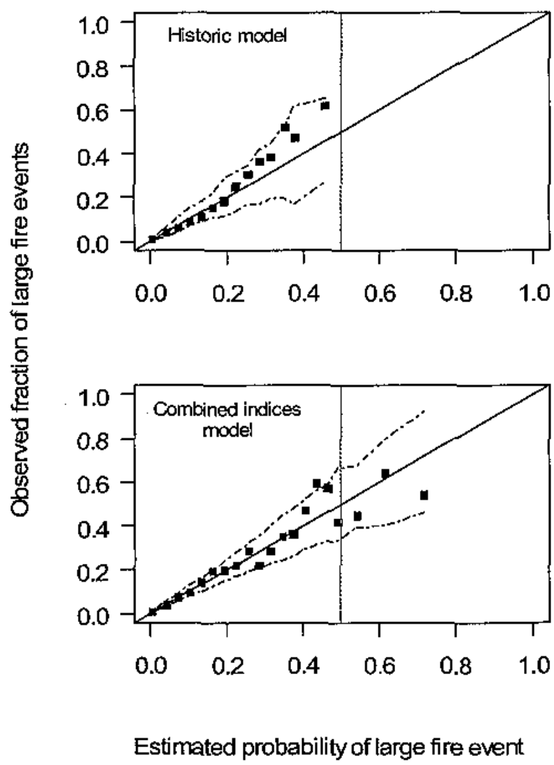

Figure 2: Observed fraction of large events in a given one-degree cell against the estimated probability based on the historic model and the combined model. The dashed lines are the approximate point-wise 95% confidence bounds.

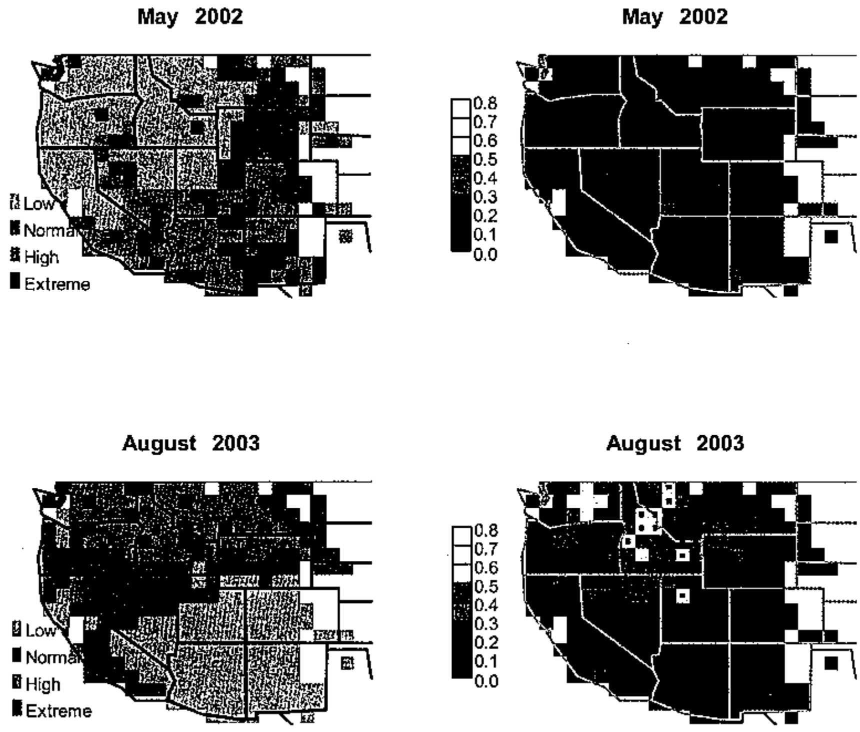

Figure 3: Maps of odds relative to historic (left panels) and estimated probabilities (right panels) of large fire events for two time periods. Note that higher than normal odds resulted in many large fire events in August 2003 and very few events in May. This is consistent with the low estimated probabilities in May and high estimated probabilities in August.

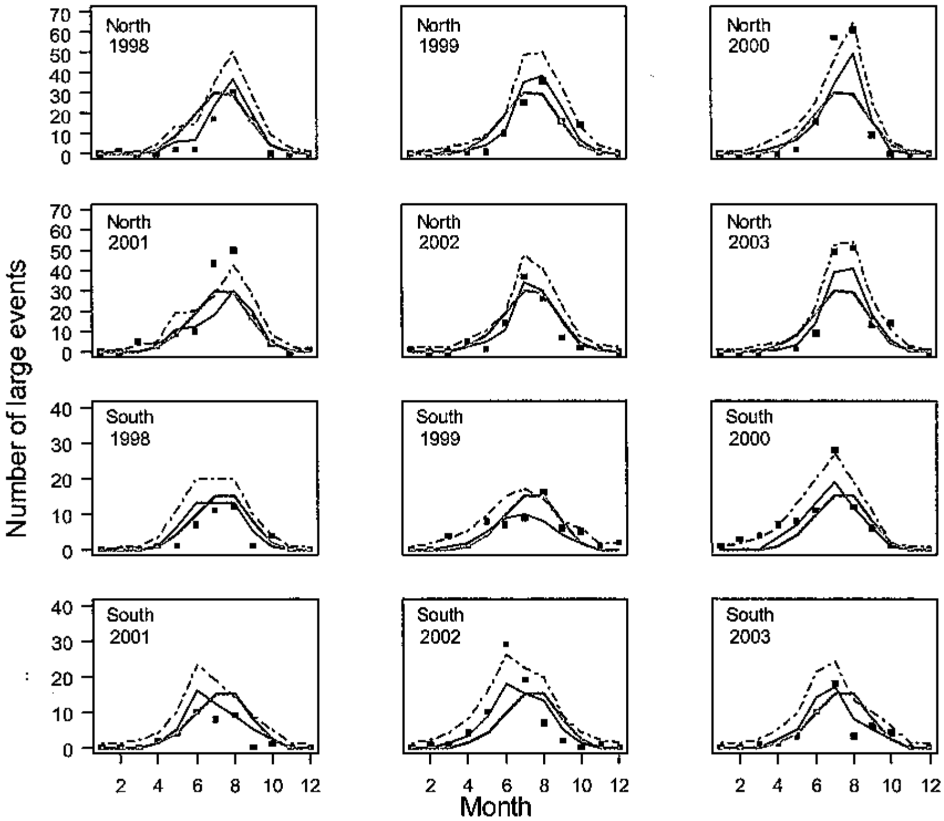

Figure 4: Observed (dots) and estimated (curves) number of one-degree cells with large fire events. Solid curves are the 50th percentile of the fitted distribution. Dashed curves are the upper 95th percentile of the distribution. Grey curves are estimated 50th percentiles of historic model. "North" and "South" are for all cells above and below 40N, respectively.