Estimating Risk Probabilities for Wildland Fires

Haiganoush K. Preisler[1] and Anthony L. Westerling[2]

[1] USDA Forest Service USDA, Pacific Southwest Research Station, 800 Buchanan St., West Annex, Albany, CA 94710 Telephone: +1 510 559-6484; fax: +1 510 559-6440; email: hpreisler@fs.fed.us
[2] Scripps Institution of Oceanography, University of California, 9500 Gilman Drive, San Diego, CA 92093 -0225

Abstract

In 2002. there were 88,458 fires reported on federal lands. These fires burned 6,937,584 acres and 2,381 structures costing taxpayers $1.6 billion for fire suppression. On average, 4,215,089 acres of federal lands burn annually. Forecasting wildland fire risk (occurrence and size) is important to fire managers who desire to know the risks of severe events well in advance of their happening. In this talk, we discuss the estimation of a probability model for forecasting fire risk one month in advance. The model uses 25 years of historic fire data on federal lands in addition to weather and climatological variables.

Keywords: forecasting, Palmer Drought Severity Index, semiparamteric regression, spatial and temporal model.

Introduction

Data

This work relied on two 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 fire reports reported by the USDA Forest Service and the USDI's Bureau's of Land Management and Indian Affairs and the National Park Service for 1980-2000. This fire history was updated through 2004 and used to estimate the risk of ignition in the analysis presented below.

In addition, a large (> 1000 acres) fire history compiled from the same sources by Westerling et al (2005) was used to describe the risk of large fire occurrence. Fire records in the latter data set were cross-checked to eliminate duplicate records and to correct errors in location wherever possible, comparing reported latitude and longitude to coordinates derived from the public land survey system and the boundaries of the reporting land management units.

Average monthly Temperature and Palmer Drought Severity Indices (PDSI) from U.S. Climate Divisions (NCDC 1994) were projected onto a one degree grid to provide a "local" monthly climate record for each grid cell. PDSI is an autoregressive index of combined precipitation, evapotranspiration, and soil moisture that represents cumulative precipitation and temperature anomalies (Alley 1985, Guttman 1991). The index is far from being a perfect proxy for soil moisture (Alley 1984, Karl and Knight 1985), but it has proven to be well correlated with wildfire risks in the western U.S. in particular in numerous studies (see, e.g., Balling et al. 1992; Larsen and MacDonald 1995; Swetnam and Betancourt 1998, Westerling et al 2002 and 2003, Westerling and Swetnam 2003). PDSI is convenient for these purposes due to both its easy availability and due to the fact that it is a normalized index, the values of which provide a moisture index comparable across a diverse landscape.

We also employ two indices describing patterns of sea surface temperatures in the Pacific Ocean known to be associated with multi-year to decadal scale variability in western climate: the El Niño-Southern Oscillation (NINO), and the Pacific Decadal Oscillation (PDO) (Gershunov and Barnett 1998, Dettinger et al 1998). Historical values of these indices were obtained from the NOAA's Climate Prediction Center (http:// www.cpc.ncep.noaa.gov/data/indices/).

Statistical Methods

Probability models

We are interested in estimating wildfire 'danger' at a given location and time. As a metric for fire danger we use the probability, π, that an area greater than some specified value (e.g., 1000 acres) will burn in a given 1x1 degree grid cell during a given month. Let Y_{ijk} be the random size of the area burned in grid i (i = 1, …, I), month j (j =1, …, 12) and year k (k = 1980, …, 2004). Let N_{ijk} be the random number of fires in grid i, month j and year k. Let $\mathbf{X}_{ijk}$ be a matrix of explanatory variable values for grid i, month j and year k. We define probability of fire risk as

$$\pi_{ijk} = \Pr[Y_{ijk} > C \mid \mathbf{X}_{ijk}, \theta] \qquad [1]$$

where C is a critical size of interest (e.g., 1000 acres) and θ is a vector of parameters. The probability in equation [1] may be written as a product of two probabilities,

$$\pi_{ijk} = \Pr[N_{ijk} > 0 \mid \mathbf{X}_{ijk}, \theta_1] \times \Pr[Y_{ijk} > C \mid \mathbf{X}_{ijk} \ \& \ N_{ijk} > 0, \theta_2] \qquad [2]$$

Although the probability in equation [1] may be evaluated directly, we chose to use the two stage evaluation procedure in equation [2] because different relationships may exist between explanatory variables and the probability of 'ignition' (at least one fire)

$$\Pr[N_{ijk} > 0 \mid \mathbf{X}_{ijk}, \theta_1]$$

and the probability of 'spread' given at least one fire

$$\Pr[Y_{ijk} > C \mid \mathbf{X}_{ijk} \ \& \ N_{ijk} > 0, \theta_2] \,.$$

Further details of the preceding probability risk model are found in Brillinger et al. (2003, 2004), Preisler et al. (2004) and Preisler and Benoit (2004).

Estimation

We used logistic regression techniques with piece-wise polynomials to estimate the probabilities in equation [2]. Specifically, we estimated the regression line

$$\operatorname{logit}(p_v) = \beta_o + g_1(lon_v, lat_v) + g_2(month_v) + \sum_{m=1} g_{m+2}(X_{mv}) \qquad [3]$$

where the subscript, v, indicates the 1x1-degreexmonth voxel; p is either the probability of ignition or probability of spread ; *(lon, lat)* are the longitude and latitude of the mid point of the 1x1-degree grid cell; X_m are explanatory variables. In the present study the explanatory variables used were temperature; PDSI value in the previous month; maximum PDSI in the last 12 months; values of NINO and PDO. The terms *g()* are semi-parametric smooth functions such as piecewise polynomials, periodic splines, or thin plate splines (Hastie et al. 2001). These functions are needed to account for non-linear relationships. For example, the relationship between probability of ignition and month is non-linear with more fires (higher probabilities) occurring in the middle of the fire season (summer months) than at the beginning or end of season. Similarly, the relationships between temperature and probability of ignition or spread may not be linear. The basic idea here is to replace the vector of inputs (covariate) with additional variables which are transformations of the original variables. For example, X may be replaced by X and X^2 if the relationship is a second degree polynomial. In the present approach, each (n x1) vector, X, is replaced by an (n x q) matrix of basis functions which are then used in a simple or logistic regression routine. The statistical package R (R Development Core Team, 2004) has a module bs( ) that determines the basis functions of a given vector. Once the basis functions are determined one may use any linear or logistic regression routine because the model is linear in these new expanded variables. For the spatial component, $g(lon\ ,\ lat)$, we utilized the two dimensional version of the basis function, i.e., the thin plate spline function. The required modules for fitting thin plate splines within R were downloaded from the web (Geophysical statistical project, 2002).

The models in equation [3], together with equation [2], when evaluated, give a unique estimate for each grid cell and each month within a year. Historic average probabilities of fire danger for 1980-2004 were estimated in a similar fashion with equation [3] replaced by,

$$\text{logit}(p_v) = \beta_o + g_1(lon_v, lat_v) + g_2(month) \quad [4]$$

In other words, only location and month were included in the model. Historic averages (or climatology) have a unique value for each grid cell and each month, however, the values do not change from year to year.

Forecasting

One-month-ahead forecasts of the probabilities of fire occurrence and fire spread were obtained by first forecasting values of the explanatory variables. We used time series data for a period of 114 years (1890-2004) on the monthly values of the explanatories from NOAA's Climate Predction Center. Next we used autoregressive models (arima module in R) to forecasts one-month-ahead values for each climate division in the US. For temperature we also used month as a regressor in the autoregressive model. The autoregressive model had the best skill in forecasting temperature values (Figure 1). Forecasted PDSI values were reasonable only when the change in PDSI was small. Consequently, in this study we produced only one-month-ahead forecasts of fire risk because it required forecasts of only the temperature variable (PDSI values used in the model were those of the previous month).
It should be noted that further improvements in forecast skill for fire risks may be achieved by, for example, using North Pacific sea surface temperatures as predictors for Temperature and PDSI. However, the emphasis in the present analysis is on demonstrating the feasibility of using logistic probability models for estimating fire risk.

Fire Danger Maps

We produced risk maps using the following rule: risk in a given voxel was defined as

$$\begin{array}{lll} \text{Low} & \text{if} & \hat{\pi} + 2\hat{\sigma} < \alpha_1 \\ \text{Moderate} & \text{if} & \alpha_1 \le \hat{\pi} + 2\hat{\sigma} < \alpha_2 \\ \text{High} & \text{if} & \alpha_2 \le \hat{\pi} + 2\hat{\sigma} < \alpha_3 \\ \text{Extreme} & \text{if} & \hat{\pi} + 2\hat{\sigma} \ge \alpha_3 \end{array}$$

where $\hat{\pi}$ is the estimated probability of > 1000 acres burns in a given voxel as given by equation [1] and [3]; $\hat{\sigma}$ is the jackknife standard error estimate of $\hat{\pi}$, and α_k, k= 1,2,3 are arbitrary probability cutoff points. For the purposes of this study we used the cutoff values 0.1, 0.3 and 0.6. Other cutoff points may be used both for C (defining large fires) and α_k (defining the categories of risk).

Results

The explanatory variables temperature, PDSI in previous month, maximum PDSI in last 12 month, in addition to spatial location and month, had significant effects on both probabilities of fire occurrence and fire spread given ignition. The effects of the two climate change variables, NINO and PDO, were marginally significant with no obvious patterns. On average fire risk (probability of ignition and spread) appeared to increase with decreasing PDSI and increasing maximum PDSI. The latter indicates an increase in fire risk when there are large shifts from high to low PDSI values (Figure 2). Maps of estimated probabilities of large fire incidences (Figure 3) produced by the climatology model (equation [4]) display the areas in the West of the USA with high risks of fire. Areas with highest probabilities of fire risk in August appear to be in California and around Southern Idaho. In October, only a few grid cells in Southern California and Southern Idaho have probabilities greater than 5%.
We assessed the skill of the model with weather variables included (equation [3]) by producing a plot of the observed monthly total number of cells with areas burned > 1000 acres. These were then compared with the predicted numbers, produced by the model, using cross-validation (Figure 4). Specifically, predictions for a given year were done by estimating the model parameters from all other years excepting the year being predicted. Confidence bounds produced included natural (Poisson) variation and variations due to the error in the parameter estimated. The model appears to give reasonable predictions of the total number of grid cells with 'high fire risk'. Seven percent of the observed totals where outside the 95% confidence bounds.
To study the skill of the model for forecasting fire risk at a given grid cell we generated maps that highlight grid cells with the observed values outside the 95% confidence bounds of predicted.

Observed numbers are the number of years (out of 25) with > 1000 acres burned in a given cell. This was compared with the corresponding forecasted values. The spatial pattern of significant departure from forecasted values appeared to be random for the months of August and October (Figure 5). However, it is of interest to note the higher than forecasted observed values in October in some of the cells in Southern California.

We also produced maps of fire risk (Figure 6) based on the probability of > 1000 acres burning in a given voxel. Managers may be interested in studying such maps for a range of past years in order to study the utility of this risk model in their management practices.

Discussion

In this talk we presented methods for estimating, forecasting and mapping fire risk. We found the methods useful for assessing the utility of fire danger indices in forecasting one month ahead large fire events. The estimated statistical models were also useful for producing error rates (precision estimates). Error rates are essential for managers when studying the utility of a given tool (model) for fire risk prediction.

We found the PDSI variables, together with temperature, to be useful indicators for forecasting fire risk one month ahead. A similar study may be done with forecasted fire danger and fire weather indices in order to study the skill of 3 month or 6 month ahead forecasts on fire risk. In addition, recent advances in forecasting temperature and PDSI using observed North Pacific sea surface temperatures (Alfaro et al 2005, Westerling et al 2003) could also be adapted to improve forecast skill for wildfire risks.

References

Alfaro, E.J., Gershunov, A. and Cayan, D.R. (2005), “Prediction of summer maximum and minimum temperature over the western United States: The role of soil moisture and sea surface temperature.” *J of Climate*, submitted.

Alley, W. M., (1984), “The Palmer Drought Severity Index: Limitations and assumptions”. *J. Climate Appl. Meteor.*, 23, 1100–1109.

Alley, W. M., (1985), “The Palmer Drought Severity Index as a measure of hydrologic drought”. *Water Resour. Bull.*, 21, 105–114.

Balling, R. C., Meyer, G. A. and Wells, S. G., (1992), “Relation of Surface Climate and Burned Area in Yellowstone National Park”. *Agric. For. Meteor.*, 60, 285-293.

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.

Brillinger, D.R., Preisler, H.K. and Naderi, H.M. (2004), "Wild fire chances and probabilistic risk assessment". Proceedings of Accuracy 2004 and Ties 2004 Portland, Maine.

Dettinger, M.D., Cayan, D.R., Diaz, H.F., and Meko, D.M.,(1998), “North-south precipitation patterns in western North America on interannual-to-decadal time scales”, *J. of Climate*, 11,3095.

Gershunov, A., and Barnett, T., (1998), “Interdecadal modulation of ENSO teleconnections”, *Bull. Am. Meteorol. Soc.*,79, 2715.

Geophysical Statistics Project. (2002), National Center for Atmospheric Research. http://www.cgd.ucar.edu/stats/Software/Fields

Guttman, N.B., (1991) “Sensitivity of the Palmer Hydrologic Drought Index”. *Water Res. Bull.*, 27, 797-807.

Hastie, T.J., Tibshirani, R., Friedman, J. (2001), “*The Elements of Statistical Learning”. Data Mining, Inference, and Prediction*. Springer, New York. 533 pp.

Karl, T. R. and Knight, R.W., (1985), “Atlas of Monthly Palmer Hydrological Drought Indices (1931-1983) for the Contiguous United States”. *Historical Climatology Series* 3-7, National Climatic Data Center, Asheville, NC.

Larsen, C.P.S., and MacDonald, G.M., (1995), “Relations between tree ring widths, climate, and annual area burned in boreal forests of Alberta”. *Can. J. Forest Res.*, 25, 1746-1755.

NCDC, (1994), “Time Bias Corrected Divisional Temperature-Precipitation-Drought Index”. Documentation for dataset TD-9640. Available from DBMB, NCDC, NOAA, Federal Building, 37 Battery Park Ave. Asheville, NC 28801-2733. 12pp.

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.

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.

R Development Core Team (2004), "*R: A language and environment for statistical computing*". R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Swetnam, T. W. and Betancourt, J. L., (1998), "Mesoscale disturbance and ecological response to decadal climatic variability in the American Southwest". *Journal of Climate,* 11, 3128-3147.

Westerling, A.L., Gershunov, A., Cayan, D.R., and Barnett, T.P., (2002) "Long Lead Statistical Forecasts of Western U.S. Wildfire Area Burned," *International Journal of Wildland Fire*, 11(3,4) 257-266.

Westerling, A.L., Brown, T.J., Gershunov, A., Cayan, D.R., and Dettinger, M.D., (2003), "Climate and Wildfire in the Western United States," *Bulletin of the American Meteorological Society*, 84(5) 595-604.

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).

Westerling, A.L. and Swethnam, T. W. (2003), "Interannual to decadal drought and wildfire in the Western United States". *EOS*, vol 84, 545-560.

Westerling, A. L., Hidalgo, H., Cayan, D.R. and Swetnam, T., (2005), "Trends in large wildfires in western U.S. forests since 1970: Affects of changes in the timing of Spring. In preparation.

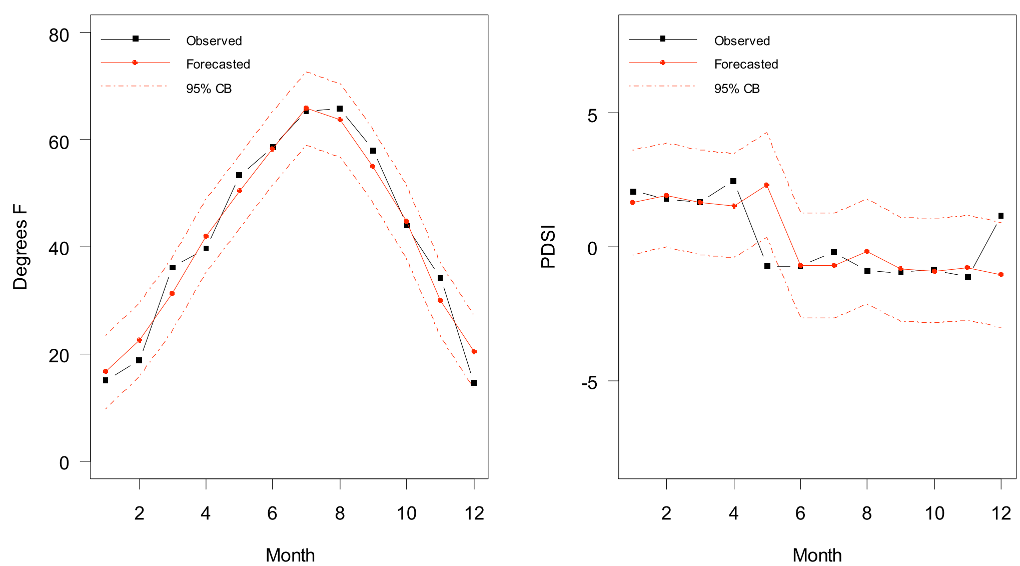

Figure 1: Observed and one-month-ahead forecasts of temperature and PDSI for one climate division in Idaho.

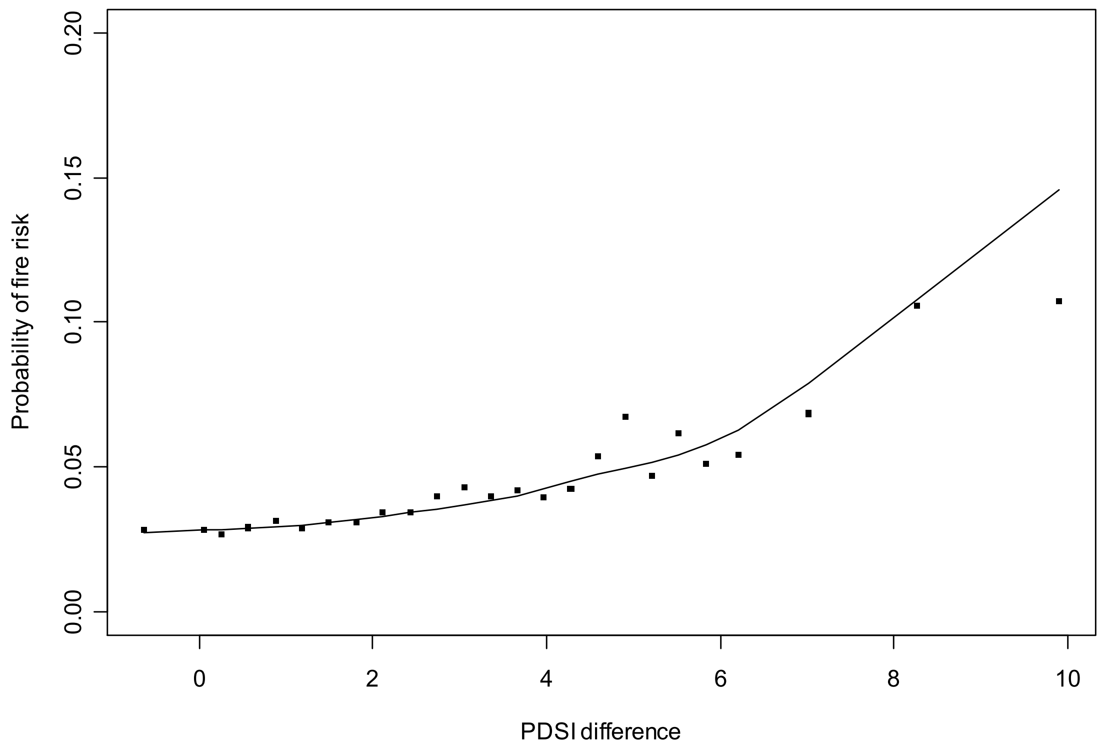

Figure 2: Observed and predicted fraction of voxels with greater than 1000 acres burned plotted versus the difference between maximum PDSI in the past 12 month and PDSI in the previous month. There is a significant increase in probability of large fire risk with an increase in the PDSI difference.

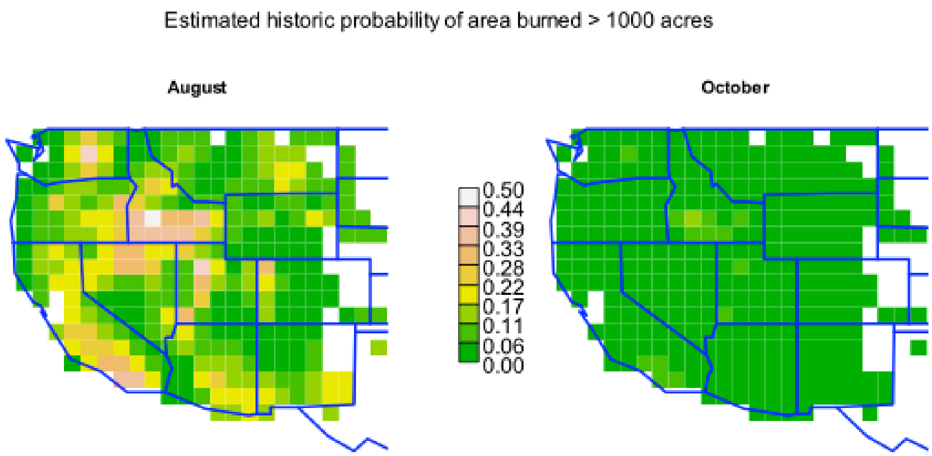

Figure 3: Estimated historic probabilities of greater than 1000 acres burned in a given 1° x 1° grid cell in Western United States. Probabilities were estimated using a model including only spatial and temporal (monthly) effects.

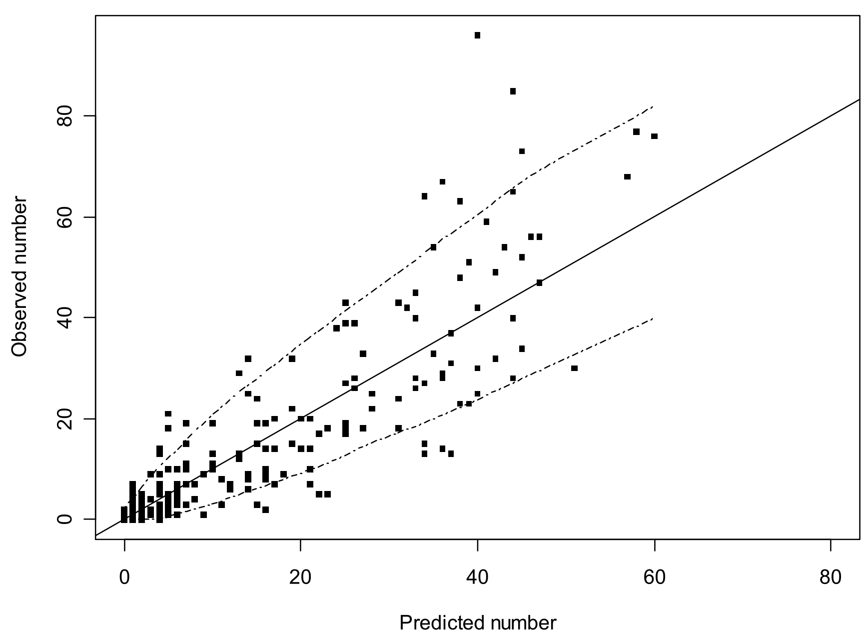

Figure 4 Observed and forecasted monthly total number of grid cells with area burned > 1000 acres. Dashed lines are estimated 95% approximate bounds. 7% of observed totals are outside the bounds.

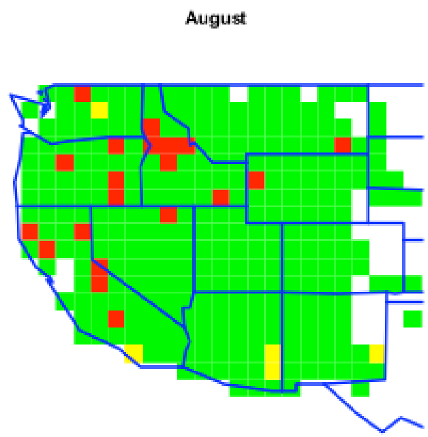

Figure 5 Comparing observed and forecasted number of years (out of 25) with more than 1000 acres burned. Cells marked as Less (Yellow) are those where observed number of years (with > 1000 acres burned) is below the 95% forecasted bound. Cells marked as Greater (Red) are those with observed number larger than the 95% forecasted bound.

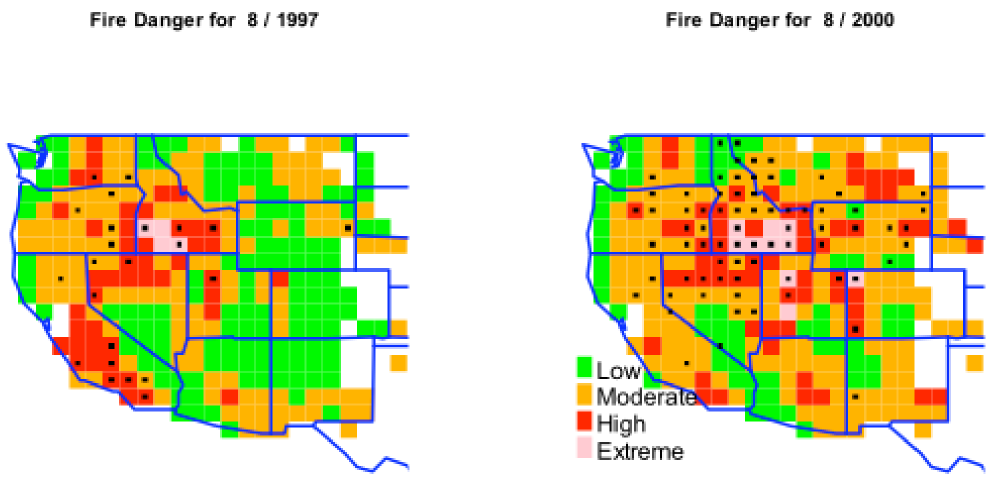

Figure 6: Fire risk maps based on the probability of > 1000 acres burning in a given 1x1 degree grid cell. Black dots are locations of cells with observed areas burned > 1000 acres. Low and high fire danger levels were produced by grouping the estimated probabilities into 4 groups. In the Western US, 1997 was a relatively low fire year while 2000 was a high fire year.