

## Statistical Forecast of the 2001 Western Wildfire Season Using Principal Components Regression

contributed by Anthony L. Westerling<sup>1</sup>, Daniel R. Cayan<sup>1,2</sup>, Alexander Gershunov<sup>1</sup>, Michael D. Dettinger<sup>2</sup> and Timothy Brown<sup>3</sup>

<sup>1</sup>*Scripps Institution of Oceanography, La Jolla, California*

<sup>2</sup>*US Geological Survey, La Jolla, California*

<sup>3</sup>*Desert Research Institute, Reno, Nevada*

The existence of links between seasonal climate anomalies and seasonal fire activity in the Western US (Westerling et al 2001) motivates a forecast of seasonal acres burned (May to October) on a 1 x 1 degree grid in the western contiguous United States using lagged values of the Palmer Drought Severity Index (PDSI). Many areas contain a characteristic pattern that links fuels to climate—high fire activity tends to occur when the preceding year is moist (positive PDSI) and the concurrent year is dry, as illustrated by the relationships shown for the Sierra Nevada and Great Basin in Figure 1. Note that the Great Basin is typically very dry in summer (Osmond et al 1990), so that the important relationship there is with moisture the year before. These relationships motivate our choice of predictor variables below. The forecast model is estimated using a principal components regression (PCR) to calculate linear relationships between principal components of the seasonal acres burned and lagged PDSI data sets.

Acres burned per grid cell were summed for fires starting between May 1 and October 31 and scaled using a  $\log_{10}$  transformation. These data were compiled (Westerling et al 2001) from fire reports from the USDA Forest Service, Bureau of Land Management, and the Pacific Western Region of the National Park Service for the period 1980 – 2000. The 312 grid cells averaging more than one fire per year comprise the predictand data set.

For predictors, 110 western U.S. Climate Division PDSI series are used at seven different lags: January and March immediately preceding, January, March, May, and August one year previous to, and May two years prior to the fire season, for a total of 770 predictor variables.

Since a multivariate regression cannot yield a unique solution if the number of predictor/predictand variables is greater than the number of observations, the dimensions of the predictor and predictand data sets were reduced by substituting the first eight principle components for each of the two data sets. For the predictor data sets, the first eight principal components explain 81% of total variance. Similarly for the predictands, the first eight principal components explain 62% of total variance.

Forecast skill is measured here by the correlation between cross-validated model output and the predictand— $\log_{10}$  of seasonal acres burned—for each grid cell (Figure 2). While this model was not optimized for any particular region, the map in Figure 1 shows the greatest skill in the Rocky Mountains, the Sierra Nevada, central Arizona, and the Great Basin. Note that clear areas can indicate no data or negative correlation (i.e., no skill). The histogram for cross-validated correlation (Figure 3) shows the forecast model does significantly better than random chance. Eighteen percent of the grid cells have cross-validated correlations in the region corresponding to the upper 5% of the t-distribution.

A forecast, developed retrospectively, for the very active fire season of summer 2000, was fairly successful in reproducing the observed acres burned. Cross-validated forecast anomalous acres burned for the 2000 fire season (Figure 4) show a similar spatial pattern in sign and intensity to the actual anomalies (Figure 5). Considering that the 2000 fire season was an extreme year in many locations compared to the previous 20-year record used to estimate the model, this result strongly indicates the utility of this approach to forecasting the western US wildfire season.

The PCR model appears to pick up some of its skill from the spatial structure relating the dominant modes of fire and PDSI variability (i.e., not all skill is derived from local linkages). Figure 1 shows the correlation between seasonal acres burned and 32 lagged monthly PDSI values through August of the fire season for the Sierra Nevada, Great Basin, and Northern Rockies. Note that the correlation between acres burned and lagged PDSI in the Northern Rockies is rather weaker for PDSI scores preceding the fire season than in either the Sierra Nevada or the Great Basin while, in general, skill for the PCR model is higher in the Rockies. The PCR model may be identifying spatial as well as temporal patterns in the drought indices, which provides greater skill in forecasting acres burned in the Rocky Mountains than can be obtained with models using only local lagged PDSI predictors. Alternatively, the apparent skill in the Northern Rockies and elsewhere may be an artifact of the large number of regressors (8) compared to the number of observations (21). The use of cross-validated skill measures may protect against reliance on false skill. Further model validation will be undertaken in future using Forest Service and National Park Service data from 1970-79.

Finally, the prediction of the 2001 fire season was produced using a similar set of lagged PDSI predictors. The 2001 fire season forecast (*Figure 6*) uses persistence in the February 2001 PDSI to model March 2001 PDSI; otherwise variable definitions are the same as for the 2000 forecast. Note that the forecast, while exhibiting positive anomalies in an arc from eastern Washington state through the Rockies and New Mexico, seems to indicate a much less extreme fire season than in 2000.

A wide variety of choices for predictor variables and model specifications remain to be explored. Local regression models also illustrate anomalous fire activity that is likely for specific regions. Figure 7 shows examples of such forecasts for the 2000 and 2001 fire seasons for three regions in the Mojave, Great Basin and Sierra Nevada. The Mojave and Great Basin models each have a single regressor derived from the average interpolated US Climate Divisional PDSI from May twelve months before the fire season. The Sierra Nevada model has two regressors similarly derived from March PDSI of

the previous and contemporaneous years. (March 2001 PDSI is estimated by persisting the February value.) The skill of these models is high—cross-validated R<sup>2</sup> ranging from 0.37 to 0.45, cross-validated correlation from 0.61 to 0.68—and they also indicate a less severe fire season in the Mojave and Great Basin, and only a marginally more intense season in the Sierras compared to last year's prediction.

Further information on climate-fire linkages and seasonal fire forecast procedures, including forecast anomalies in color form using up-to-date PDSI values, is provided at <http://meteora.ucsd.edu/cap>, the California Applications Program (CAP) web page.

## References

- Johnson, R. A. and Wichern D. W., 1998, *Applied Multivariate Statistical Analysis*, Prentice Hall, 816 pp.
- Osmond, C. B., Pitelka, L. F. and Hidy, G. M., (Eds.), 1990, *Plant Biology of the Basin Range*. Ecological Studies 80, Springer-Verlag, 375 pp.
- Westerling, A. L., Brown, T. J., Gershunov, A., Cayan, D. R., Dettinger, M.D., 2001, Climate and Wildfire in the Western United States, submitted to Bulletin of the American Meteorological Society January 2001.

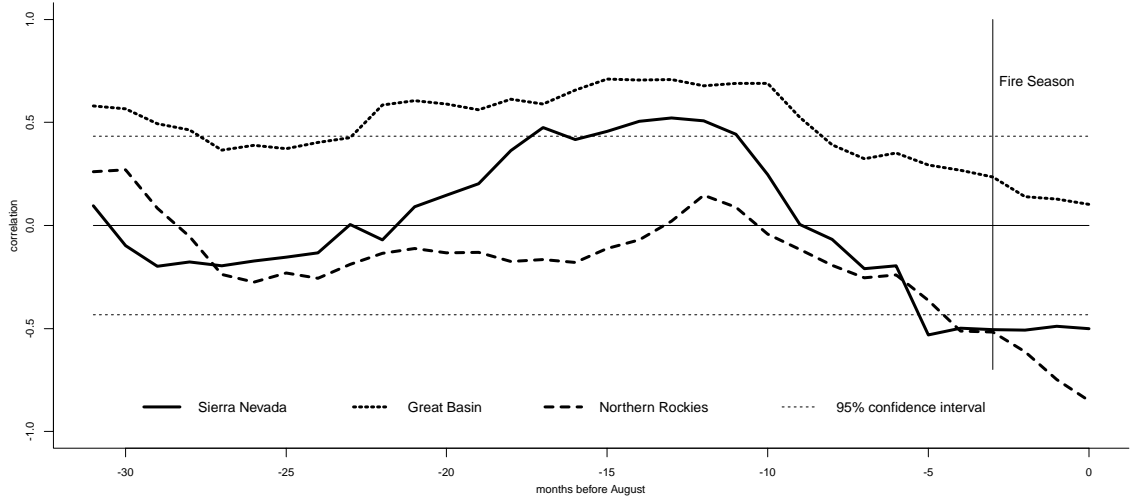


Figure 1. Correlation between  $\log_{10}$  seasonal acres burned and lagged monthly PDSI, with lags expressed as months before the peak of the fire season in August, for the Sierra Nevada, Great Basin and Northern Rocky Mountains.

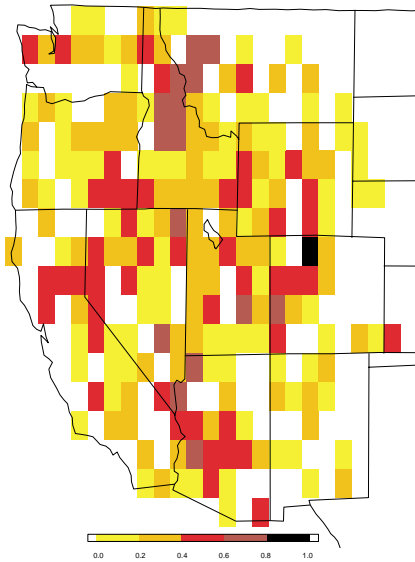


Figure 2. Correlation of cross-validated PCR model output with  $\log_{10}$  transform of seasonal acres burned

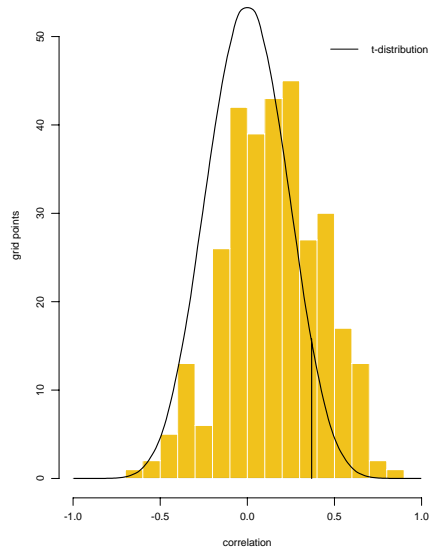


Figure 3. Histogram of cross-validated correlation. Student's  $t$ -distribution is superimposed with 95% confidence interval for right-tail test indicated.

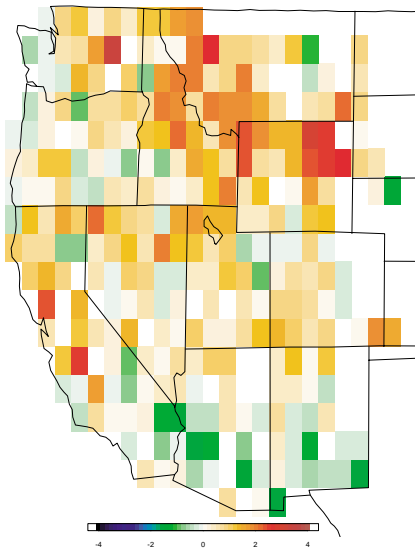


Figure 4.  $\text{Log}_{10}$  transform of forecast anomalous year 2000 seasonal acres burned.

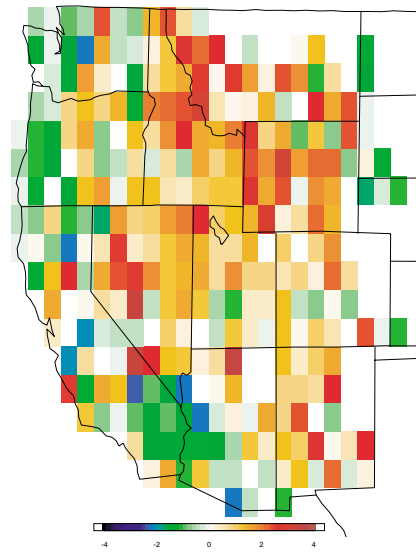


Figure 5.  $\text{Log}_{10}$  transform of anomalous year 2000 seasonal acres burned.

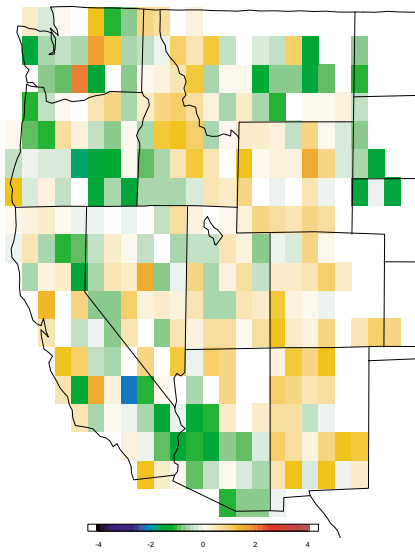


Figure 6.  $\text{Log}_{10}$  transform of anomalous forecast year 2001 seasonal acres burned.

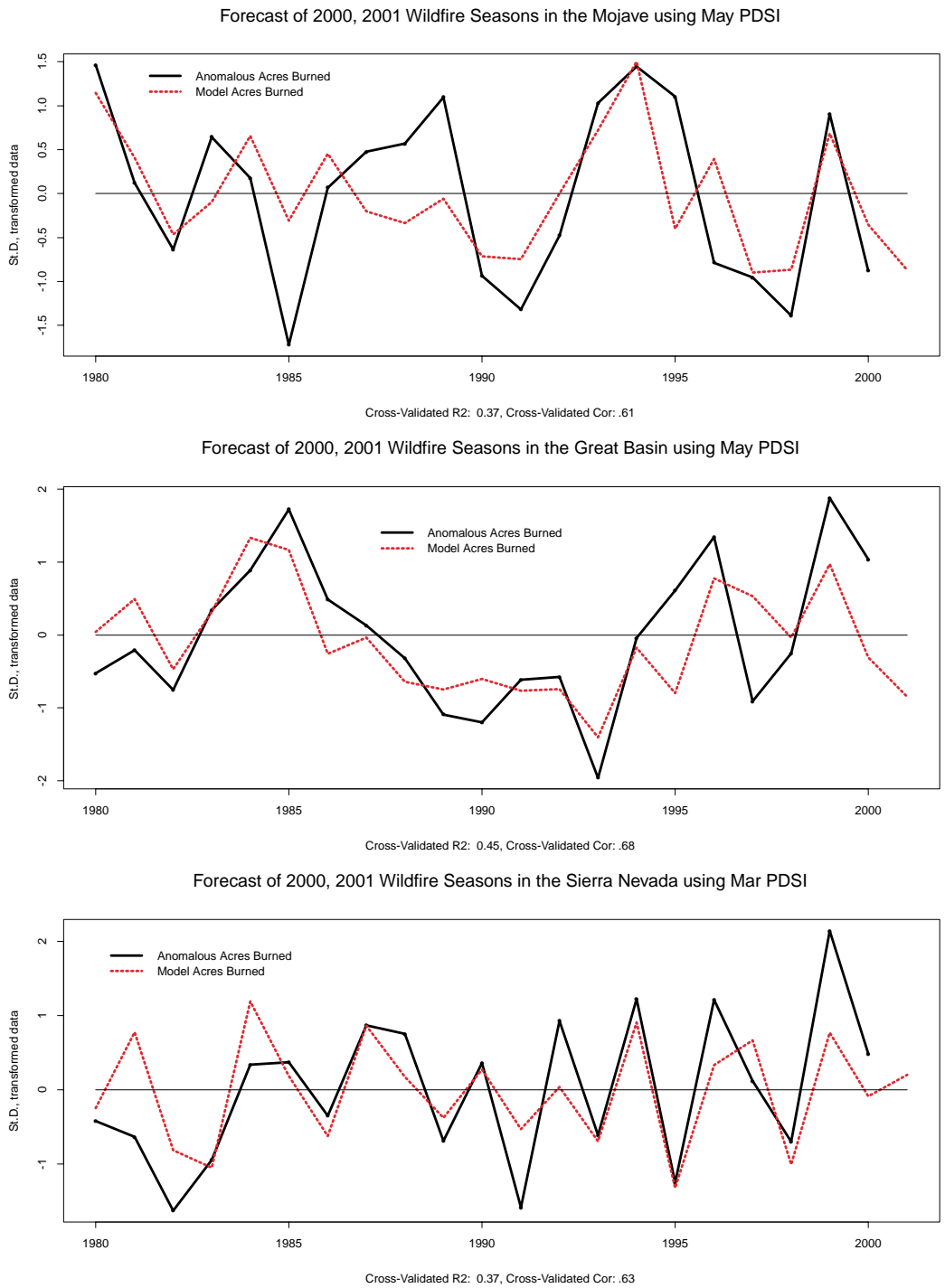


Figure 7. Regional forecasts of normalized  $\log_{10}$  acres burned for the 2000 and 2001 fire seasons using local linear regression models. Mojave (*top*) and Great Basin (*mid*) regions acres burned are modeled as a function of May PDSI from 12 months prior to the fire season. Sierra Nevada (*bot*) acres burned are a function of March PDSI from 2 and 14 months prior to the fire season. (March 2001 PDSI is modeled by persistence in February values.)