CSIRO PUBLISHING

International Journal of Wildland Fire **2009**, *18*, 508–516 www.publish.csiro.au/journals/ijwf

Forecasting distributions of large federal-lands fires utilizing satellite and gridded weather information

Haiganoush K. Preisler[A,E], *Robert E. Burgan*[B], *Jeffery C. Eidenshink*[C], *Jacqueline M. Klaver*[D] *and Robert W. Klaver*[C]

[A]USDA Forest Service, Pacific Southwest Research Station, 800 Buchanan St., West Annex, Albany, CA 94710, USA.

[B]USDA Forest Service, Rocky Mountain Research Station, 5775 West U.S. Hwy 10, Missoula, MT 59808-9361, USA. [Retired]

[C]US Geological Survey (USGS), Earth Resources Observation and Science Center (EROS), 25198 479th Ave, Sioux Falls, SD 57198, USA.

[D]Independent Research Scientist, 2601 S. Holly Ave, Sioux Falls, SD 57105, USA.

[E]Corresponding author. Email: hpreisler@fs.fed.us

Abstract. The current study presents a statistical model for assessing the skill of fire danger indices and for forecasting the distribution of the expected numbers of large fires over a given region and for the upcoming week. The procedure permits development of daily maps that forecast, for the forthcoming week and within federal lands, percentiles of the distributions of (i) number of ignitions; (ii) number of fires above a given size; (iii) conditional probabilities of fires greater than a specified size, given ignition. As an illustration, we used the methods to study the skill of the Fire Potential Index – an index that incorporates satellite and surface observations to map fire potential at a national scale – in forecasting distributions of large fires.

Additional keywords: fire business, fire danger, fire distribution, fire mapping, fire occurrence, semi-parametric logistic regression, spatial mapping, statistical comparisons of fire danger indices.

Introduction

In April 2002, the Bush administration responded to escalating costs associated with large fires by commissioning the Wildland Fire Leadership Council (WFLC) to implement and coordinate the National Fire Plan and the Federal Wildland Fire Management Policy for the United States of America. The WFLC consists of senior-level federal, state, county, and tribal representatives. The WFLC in turn chartered a Strategic Issues Panel on Fire Suppression Costs to examine wildland fire suppression activities. In their report (WFLC 2004), this panel found that 'fire suppression expenditures are overwhelmingly centered in larger fires. From 1980 through 2002 small fires (less than 300 acres (120 ha)) managed by the Forest Service totaled 98.6% of the fires but represented only 6.2% of the total suppression expenditures. Larger fires (greater than 300 acres (120 ha)) represented 1.4% of the fires and a whopping 93.8% of the suppression expenditures'. These statistics strongly support the need for a means to estimate the probability that a fire, once established, will become large.

Systems for evaluating the potential for wildfires have been in existence for some time. Examples include the National Fire Danger Rating System (NFDRS) for the United States (Deeming *et al*. 1977; Burgan 1988) and the Canadian Forest Fire Danger Rating System (CFFDRS, Van Wagner 1987). These systems use weather-station data and spatial interpolation to generate spatially explicit maps of the fire danger and fire weather variables. More recently, some fire danger indices have been based on moderate-resolution remote sensing data (e.g. Relative Greenness, Burgan and Hartford 1993). Additionally, fire danger estimates are being produced at global to regional scales from meteorological climate models (Roads *et al*. 1995). Although these fire danger maps show the relative danger level between locations and dates, they do not provide managers with estimates of expected numbers of large fires. A large fire probability map is an estimate of the likelihood that ignitions will become large fires, given existing levels of fire danger variables. A weighted sum of the probabilities over a given region and time-span can provide an estimate of expected numbers of large fires in a forthcoming day or week, as will be discussed in the following sections.

One procedure for estimating probabilities of fire occurrence, or probabilities of large fires, is the use of logistic regression techniques (Andrews and Bradshaw 1997; Andrews *et al*. 2003; Brillinger *et al*. 2003; Preisler *et al*. 2004; Lozano *et al*. 2007). We discuss how a spatially explicit logistic regression model is used to develop large-fire probability maps using a particular fire potential index. This estimation and mapping technique, however, is not limited to any particular fire danger index. The same methods may be used with any other danger index (or a combination of indices) provided that spatially explicit daily historical values of the index are available for a couple of years. Additionally, we demonstrate how estimates of conditional large-fire

 10.1071/WF08032 1049-8001/09/050508

probabilities, together with information on frequencies of fire occurrence, are used to forecast expected numbers of large fires in the upcoming week for a given region.

The fire potential index (FPI) used in the present study was originally developed at the US Forest Service (USFS) Intermountain Fire Sciences Laboratory and the US Geological Survey Earth Resources Observation and Science Center (Burgan *et al*. 1998). The index has been available to fire managers through the Wildland Fire Assessment System (WFAS) website (http://www.wfas.net and http://gisdata.usgs.gov/website/IVM, accessed 18 June 2009) for several years. It was developed to incorporate both satellite and surface observations in an index that can be used to map fire potential from national to local scales through use of a Geographic Information System (GIS). Recent advances provide the opportunity to use gridded weather forecasts rather than surface weather observations for calculation of the FPI.

The skill of the original FPI model was assessed in 1997 for the state of California for the period 1989 through 1995. Specifically, the actual occurrence of fires was spatially compared with the FPI at 1-km resolution for the same date and geographic location. The FPI was found to have a high statistical correlation (r^2 of 0.72) with 5-year combined fire occurrence (Klaver *et al*. 1997). FPI values at 4.4-km resolution compared with historical fire data from 1995 and 1996 in southern European countries show that FPI values increase as fire occurrence increases (Sebastian-Lopez *et al*. 2002). The FPI was computed from 1981 through 1993 for Kalimantan Island, Indonesia, and assessed by correlating it with derived fire occurrence from the TOMS (Total Ozone Mapping Spectrometer) Aerosol Index. This analysis indicated that FPI trends can be used as an early warning of potential fire problems in a tropical setting (Sudiana *et al*. 2003).

In the following sections, we will describe the procedure for calculating daily FPI values on a 1-km grid followed by some details of the logistic regression technique. We will use the probability model to assess the skill of the FPI in predicting the frequency with which ignitions develop into large fires. Finally, we will demonstrate the methods by producing an example large-fire probability map and 1-week forecasts for expected numbers of large-fire events over a region.

Methods

FPI calculations

Fire potential index maps have been produced for the conterminous United States at 1-km resolution for several years (WFAS site). One kilometre is the base resolution of the vegetation condition information that is derived from satellite observation (Eidenshink 2006). The FPI model operates under the following assumptions: (1) fire potential can be assessed if the proportion of live to dead vegetation is defined; satellite-derived vegetation greenness data (Burgan and Hartford 1993) provide a useful parameterization of this metric; (2) good estimates can be made of how close the dead fuel moisture is to the moisture of extinction; and (3) wind should not be included because it is transitory and difficult to estimate over large geographic areas.

Inputs to the FPI map are (1) calculated dead fuel moistures derived from daily weather observations of temperature, humidity, precipitation, and cloud cover (Fosberg 1971; Fosberg and Deeming 1971; Fosberg *et al*. 1981); (2) a dead fuel extinction moisture map to indicate dead fuel moisture contents at which fires are no longer expected to spread; (3) a maximum live ratio map to indicate the maximum proportion of living vegetation, which is assumed to have a high moisture content; and (4) weekly Relative Greenness (RG) maps (Burgan and Hartford 1993) to adjust the maximum live ratios for current conditions. The FPI algorithm has historically included estimation of only 10-h timelag dead fuel moisture (FPI_{10}), but it used this moisture to represent both 1-h and 10-h fuels, that is, fuel particles that are less than 1 inch (2.5 cm) in diameter. Questions arose concerning whether the FPI could be improved by including moisture contents for larger fuels as well, that is, the 100-h and 1000-h timelag fuels (FPI_{1000}). These fuel particles range in size from 1 to 8 inches (2.5 to 20.3 cm).

The general form of the FPI calculation is:

$$FPI = (1.0 - \mathit{deadfrac}) \times \mathit{dr} \times 100.0 \quad (1)$$

where *deadfrac* = (calculated 10-h dead fuel moisture)/(moisture of extinction) and *dr* = the proportion of the vegetation that is dead.

In the FPI_{10} model, *deadfrac* is calculated as a function of the 10-h timelag fuel moisture, but in the FPI_{1000} model, *deadfrac* is calculated as a function of 10-, 100-, and 1000-h fuel moistures, weighted according to the loading in each dead fuel size class.

For the present study, daily national FPI maps were calculated for the years 2001–03. Meteorological data used to compute dead fuel moisture and subsequent FPI for the conterminous United States were modeled from daily observation data using the National Oceanic and Atmospheric Administration North American Mesoscale (NAM) model. We call this product the observed FPI to differentiate it from the forecasted FPI, which is derived from forecasted weather parameters. The specific process for each pixel was to obtain the static inputs for extinction moisture and maximum live ratio and current inputs for weather and RG, then perform the calculation outlined above. FPI values range from 0 to 100. The FPI will equal 0 when the dead fuel moisture equals the moisture of extinction or when the dead ratio value is 0 (all vegetation is live and fully green). The FPI will attain a value of 100 if all the vegetation is cured and the weighted dead fuel moisture is at its minimum value of 2%. A sample comparison of a fire danger class map and a fire potential map is provided in Fig. 1. Current maps may be obtained at the WFAS site.

Fire occurrence data

We obtained fire occurrence data from the Desert Research Institute (DRI) coarse assessment project (Brown *et al*. 2002). The DRI wildland fire occurrence data include all reported federal-lands fires from the Forest Service, Bureau of Indian Affairs, Bureau of Land Management, Fish and Wildlife Service, and National Park Service. The DRI data are flagged to indicate records with no apparent problems. The fire occurrence data used in the present study were for 1985 through 2005. We used only fires larger than 1 acre (0.4 ha) and removed records that appeared to be duplicates. In the rest of the paper, an ignition will refer to any fire occurrence that was at least 1 acre (0.4 ha) in size when discovered, and a large fire will be defined as a fire

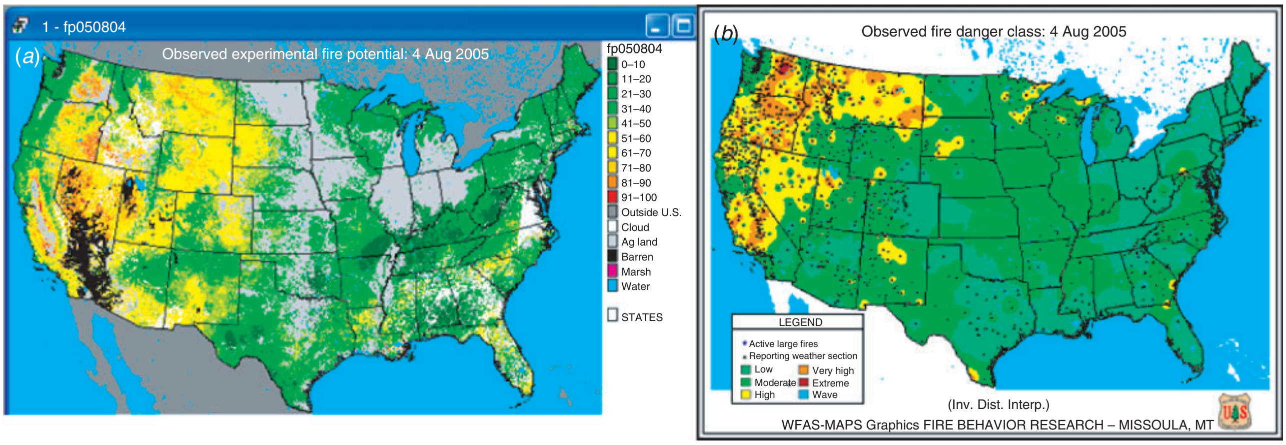

Fig. 1. (*a*) Observed experimental fire potential map, and (*b*) the observed fire danger class map for 4 August 2005.

that burned more than 100 acres (40.47 ha). However, the procedures described below are not limited to any specific definition of large.

Statistical methods

Large-fire probability

The relationship between FPI values and the conditional probability that an ignition within a given 1-km cell in a given day will result in a large fire was estimated using the following logistic regression model:

$$\pi_{ij} = \Pr[Y_{ij} = 1 | m_{ij} = 1, x_{ij}] = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})}$$

$$\text{with } \log\left[\frac{\pi_{ij}}{1 - \pi_{ij}}\right] = \theta_{ij} = A_i + \beta(x_{ij} - \bar{x}) \qquad (2)$$

where π_{ij} is the conditional probability of an ignition at location j and date i becoming a large fire; θ_{ij} is the linear predictor; Y is a binary response variable with $Y_{ij} = 1$ if there was a large fire at location j and date i, and zero otherwise; m_{ij} is the number of ignitions at ij (it is assumed that only one fire can occur in a 1-km^2 area in a given day); x_{ij} is the value of the FPI index at location j and day i; $\bar{x}$ is the mean FPI value over all locations and span of study. A_i and β are the slope and intercepts of the logit regression line to be estimated from the data. We estimated the spatially explicit intercepts by using a two-dimensional smoothing spline function as described in Appendix. In Eqn 2, when the FPI value, x_{ij}, is equal to the mean value, $\bar{x}$, the logit line is equal to the intercept. Consequently, the intercept quantifies the historical average probability of a large fire at location j, given ignition. Using a smooth surface as an estimate of the intercepts has the effect that nearby locations will have similar probabilities; consequently, the spatial correlation that is bound to exist between neighboring points is partially accounted for in the model. Neighboring points are likely to have similar topography and vegetation, among other attributes that are anticipated to cause similar fire behavior (Lozano *et al.* 2007). One may also use a spline function to describe the second term in Eqn 2, i.e. the relationship between the fire danger index and the logit regression line, to account for any possible non-linearities. Nonparametric smooth functions (Hastie *et al.* 2001) such as splines are one way to estimate non-linear relationships without making any *a priori* assumptions about the shape of the relationship. In the present study, however, the relationship between the logit line and the FPI values was not significantly different from linear.

The goodness-of-fit of the estimated probabilities was assessed by producing a reliability diagram where observed values were plotted against estimated probabilities (Hosmer and Lemeshow 1989; Wilks 1995). For the present comparison, we used cross-validation to fit the model. Namely, probability estimates for each of the 3 years (2001–03) were obtained using data only from the other 2 years. Next, estimated probabilities were grouped into cells of width 0.01, and the number of large fires as a fraction of the number of ignitions was evaluated for each group to produce the reliability diagram.

Once the slope and intercepts in Eqn 3 are estimated, maps of conditional large-fire probabilities may be produced for a given date by evaluating the probabilities in Eqn 2 on a 4587 by 2889 grid over the continental United States using the FPI values for that date.

Predicting expected number of large fires

Conditional large-fire probabilities assume that an ignition has occurred; consequently, they are useful for decisions such as ceasing prescribed burn activity, implementing specific public-use restrictions, or making decisions on the number of fire personnel and equipment needed at a given location. In the absence of ignitions, dry fire danger conditions will not result in a fire. A product that may be of use to managers, and that can be evaluated from the large-fire probabilities, is the forecasted number of large fires in a given region of interest (e.g. fire planning units or a Geographic Area Coordination Center (GACC) region – for a map of GACC regions see http://gacc.nifc.gov, accessed 18 June 2009). Expected number of large fires depends not only on the conditional probability of a large fire but also on the frequency of ignitions at a given location and date. For example, using 21 years of historical fire occurrence data (Fig. 2),

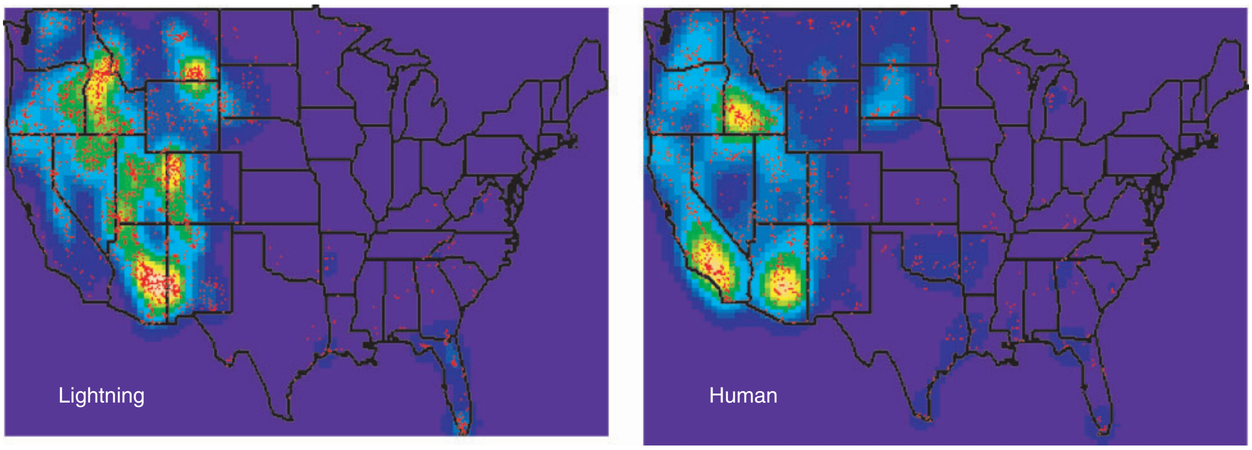

Fig. 2. Locations of historic fire occurrences on federal lands. Red dots are locations of lightning- and human-caused fires during the month of July for the years 1985–2005. The two-dimensional histograms (kernel density estimate) of the fire frequencies are displayed in the background.

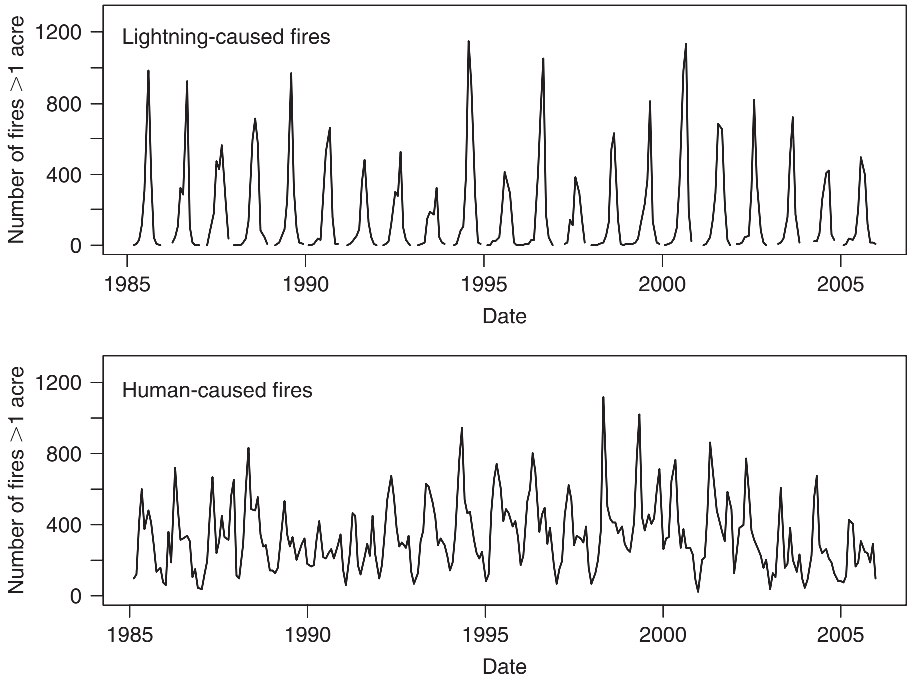

Fig. 3. Observed numbers of lightning- and human-caused fires (size > 1 acre (0.4047 ha)) per month on federal lands.

it is noted that the distribution of ignitions (fires of size >1 acre (0.4 ha)) is not spatially uniform and the pattern differs for lightning-caused fires or human-caused (or non-lightning-caused) fires. Also, frequencies of fire occurrence depend on the day of the year (Fig. 3), with lightning-caused fires occurring mostly in the summer months whereas human-caused fires do not seem to have such a large seasonal effect. These factors have to be taken into account when developing estimates of expected numbers of large fires. One possible estimate for the number of large fires is the weighted sum of the conditional large-fire probabilities over a given area and time-span, with weights given by the probabilities of ignition for that pixel and day. Specifically, the expected value and variance of the number of large fires in a given region, U, may be evaluated by

$$E(N_{kU}) = \sum_{i=k}^{k+6} \sum_{j \in U} w_{ij}\pi_{ij}$$

$$\mathrm{var}(N_{kU}) = \sum_{i=k}^{k+6} \sum_{j \in U} w_{ij}\pi_{ij}(1 - w_{ij}\pi_{ij}) \qquad (3)$$

where N_{kU} is the random number of large fires in region U for the week starting Julian day k; π_{ij} is the conditional probability of a large fire given ignition; and w_{ij} is the probability of ignition for day i and region U.

We obtained estimates of the ignition probabilities, w, by fitting a logistic regression model to the 21 years of fire occurrence data inside federal lands. A separate logistic model was developed for lightning-caused fires and human-caused fires, and the estimate of the ignition probability was set to the sum of the probabilities from the two causes. The explanatory variables we used for the logistic regressions were location and day of the year. The specific regression used was:

$$w_{ij} = \Pr[m_{ij} = 1] = \frac{\exp(c + a_j + \mathbf{b}_U \mathbf{X}_i)}{1 + \exp(c + a_j + \mathbf{b}_U \mathbf{X}_i)} \qquad (4)$$

where m_{ij} is the binary variable set to one if there is a fire (of size >1 acre (0.4 ha)) on day i and location j, and zero otherwise. Only a sample of the locations and days with no observed fires was used because of the large area (all federal lands) in our study. The constant $c = \log(1/\gamma)$ in Eqn 4 was included in the model to adjust for the sampling proportion, γ, of non-fire locations and days (Maddala 1992; Brillinger *et al.* 2003; Preisler *et al.* 2004); $\mathbf{X}_i$ is a non-parametric transformation of the explanatory variable day-in-year (see Appendix). We used a transformation of the day-in-year to account for the seasonality in fire occurrence data. A different day-in-year effect was estimated for each region U. Finally, a_j is the intercept for location j estimated using a smooth surface as described above and $\mathbf{b}_U$ is a vector of the slopes.

Given a 7-day forecast of FPI values, the conditional probability of a large fire, π, for the following 7 days may be estimated by evaluating the probabilities at the forecasted FPI values. For the present study, we did not have FPI values derived from forecasted weather parameters; consequently, we used the observed FPI values for day k as the value for all days in the forthcoming week starting on day k. In an operational setting, forecasted values will be available.

Another statistic that may be of interest to managers is the probability of at least one major fire (e.g. a fire that burns more than 5000 acres (2023 ha)) in region U in the upcoming week. The formula for evaluating this statistic is:

$$\begin{aligned}\Pr[\text{at least one fire} > \text{C}] &= 1 - \Pr[\text{no fires} > \text{C}] \\ &= 1 - \prod_{i=k}^{k+6} \prod_{j \in U} (1 - w_{ij}\pi_{ij}) \qquad (5)\end{aligned}$$

where C = 5000 acres (2023 ha); w_{ij} is the probability of ignition for day i location j, and π_{ij} is the conditional probability of a major fire given by Eqn 2 with slope and intercept evaluated for the new response variable. Namely, the new response variable is set to one if the fire is greater than 5000 acres (2023 ha) and to zero otherwise.

Results and discussion

We used two versions of the FPI (FPI_{10} and FPI_{1000}) as the index for evaluating the conditional probability of a large fire. The two versions of the FPI were similar except in forested areas containing significant loads of larger dead fuels (100-h and 1000-h timelag fuels) (Fig. 4). Although FPI_{1000} values for western forests (fuel models G and H) (Deeming *et al.* 1977) were lower than the rest of the fuel models, the two FPI models had similar

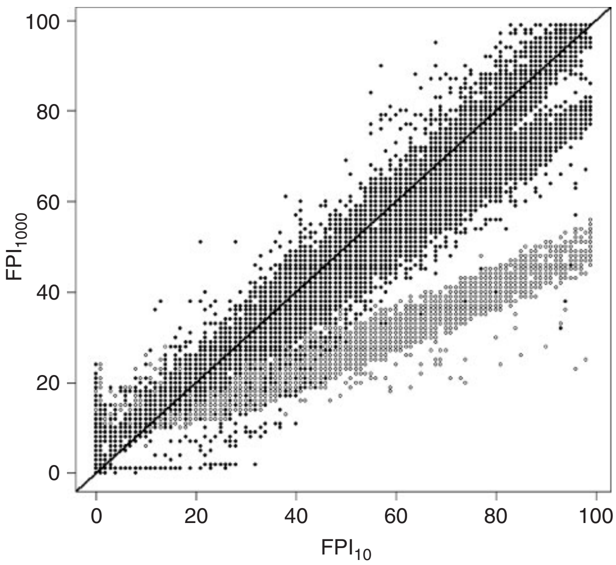

Fig. 4. Ten-hour (FPI_{10}) *v.* 1000-hour (FPI_{1000}) fire potential index for fires during 2001–03. FPI_{10} and FPI_{1000} are similar except for western forests (fuel models G and H) where the FPI_{1000} values are smaller (open dots).

skills in predicting conditional large-fire probabilities. The skill of an index was demonstrated by the significance and magnitude of the estimated slope of the logistic regression in Eqn 2. An index is assumed to have no skill if the estimated slope is not significantly different from zero or if the estimate is significant but its effect on the probabilities is negligible. For the two indices tested here, both had significant slopes; however, the slope for FPI_{10} ($\hat{\beta} = 0.011 \pm 0.0012$) was marginally larger than that for FPI_{1000} ($\hat{\beta} = 0.008 \pm 0.0008$). For this reason, and the fact that FPI_{1000} is more difficult to evaluate, we chose to use FPI_{10} as the index for evaluating the probability maps. The slope in the present logistic regression (Eqn 2) may also be used to calculate the increase in the odds of a large fire as the FPI value increases. For example, according to our results, the odds of an ignition becoming a large fire (>100 acres (40.47 ha)) is between 1.5 and 2.0 times larger when FPI = 80 compared with the odds when FPI = 30. For the entire range of FPI values (0–100), there appears to be a three-fold increase in the odds of a large fire (95% confidence interval is 2.3–3.8).

The logistic regression model appears to be a reasonable approximation for the distribution of large-fire occurrences (given ignition). This is seen in the plot of observed and predicted values when cross-validation was used over the 3 years in our study (Fig. 5).

As an example of the output of our procedures, we produced four maps for 4 August 2005 (Fig. 6). The maps were for (1) estimated conditional probability (%) of a large fire given ignition; (2) approximate 95% upper bound for forecasted number of ignitions (per 10^5 km^2) for the forthcoming week inside federal lands; (3) approximate 95% upper bound for forecasted number of large fires (per 10^6 km^2) for the forthcoming week inside federal lands; and (4) forecasted probability (per 10^6 km^2) of at least one major fire (a fire greater than 5000 acres (2023 ha)) for the forthcoming week inside federal lands. A comparison

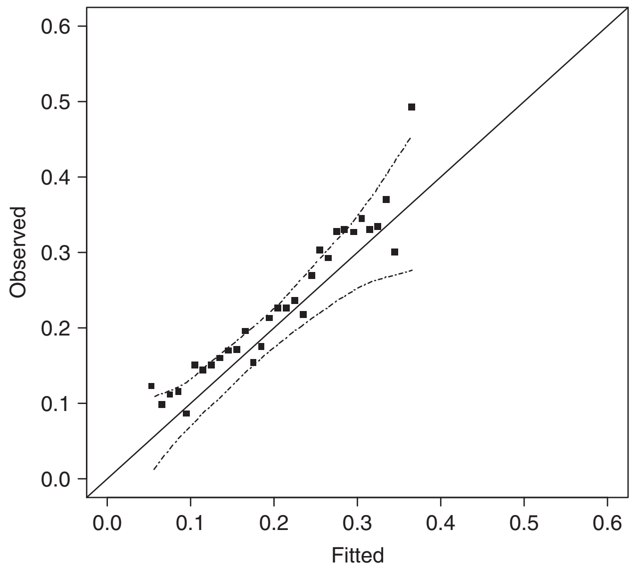

Fig. 5. Observed *v.* fitted values of fraction of large fires as a proportion of number of ignitions. Dotted lines are the approximate point-wise 95% confidence bounds.

between the map of the conditional probabilities for this day and the corresponding map of FPI values (Fig. 1) demonstrates the need for the probability maps. Although the FPI value may be similar in two regions (e.g. southern Arizona and Wyoming), the probabilities are not necessarily the same. For the example noted, the conditional probabilities in Wyoming range between 16 and 30% whereas those in southern Arizona are between 11 and 15%, even though the FPI values in both places are approximately 61–70. Historically, an ignition appears to be more likely to become a large fire in Wyoming than in southern Arizona under the same FPI values. This may result from differences in topography, vegetation, and suppression difficulties, among other reasons that are not reflected in the FPI map but are included in the probability model by the spatially explicit intercept of the logistic regression line. Also noted are the relatively high estimated conditional probabilities for Nevada compared with those in Arizona. For the 21 years under study, there were 1512 observed ignitions during July in Nevada federal lands, with 467 (31%) of them becoming a large fire. During the same period, there were many more ignitions (2504) in Arizona federal lands; however, a smaller percentage (12%) developed into a large fire. This phenomenon is reflected in the third and fourth maps of forecasted numbers of large fires and forecasted probabilities of

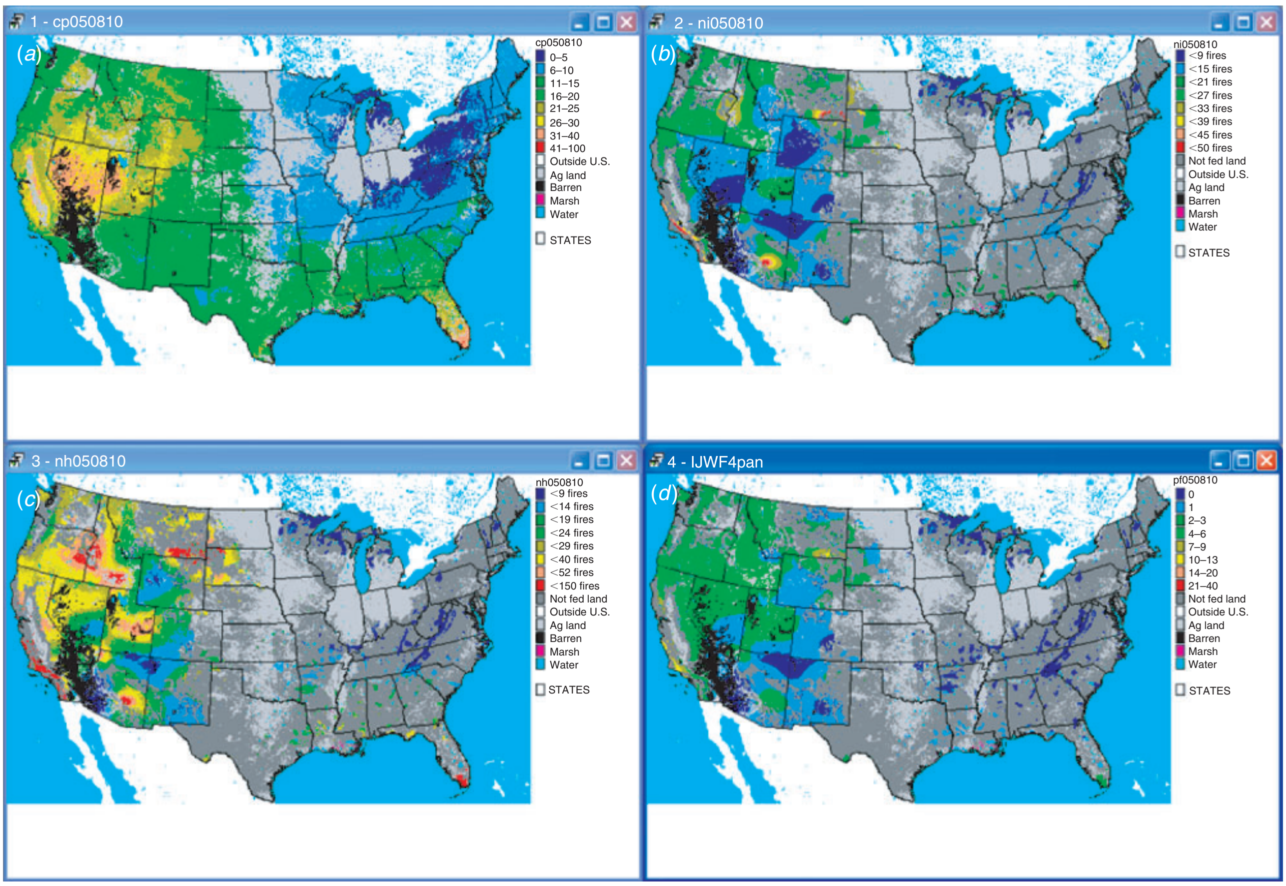

Fig. 6. Four maps generated for 4 August 2005. (*a*) Estimated conditional probabilities (%) of a large fire; (*b*) approximate 95% upper bound for forecasted number of ignitions (per 100 000) for the forthcoming week; (*c*) approximate 95% upper bound for forecasted number of large fires (per million) for the forthcoming week; (*d*) forecasted probabilities (per million) of at least one major fire (>5000 acres (2023 ha)) for the forthcoming week.

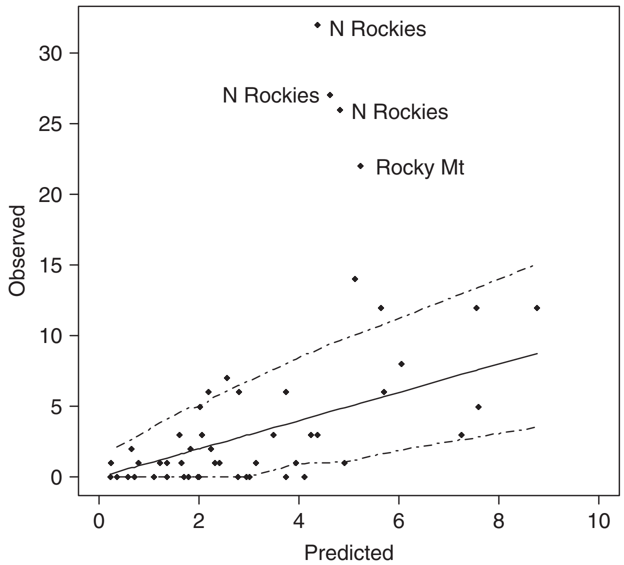

Fig. 7. Observed *v.* forecasted total number of large fires (>100 acres (40.47 ha)) per Geographic Area Coordination Center (GACC) region for the forthcoming week evaluated at five dates during the 2003 and 2005 fire seasons. The dotted lines are the approximate 95% bounds.

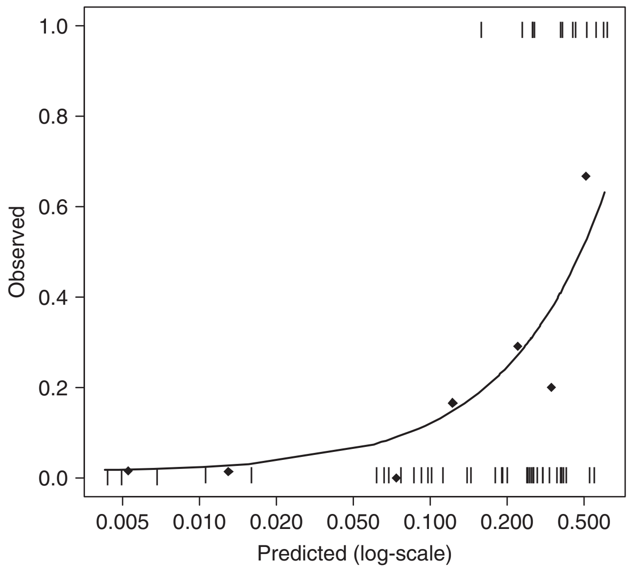

Fig. 8. Observed *v.* predicted proportion of forecasts with at least one very large fire (>5000 acres (2023 ha)) per Geographic Area Coordination Center (GACC) region in a forthcoming week. The hatch marks are the observations (0 if no fire of size >5000 acres and 1 otherwise). Dots are the proportion of the observations when grouped into seven cells.

at least one major fire, respectively. Because of the large number of ignitions in Arizona (see also Fig. 2), more large fires are forecasted there than in Nevada, despite the FPI values and the conditional probabilities being larger in Nevada.

As a goodness-of-fit test for the forecasts, we produced 1-week-ahead predictions for each GACC region and for five dates during the 2003 and 2005 fire seasons. A comparison of the forecasted values with the observed values was used as a check of the skill of the methods (Fig. 7). The forecasts appear reasonable – in most cases, the observed total numbers of large fires per GACC were within the 95% confidence bounds of the forecasts. The four largest observed values are from the Rocky Mountain and the Northern Rocky GACC regions. The skill of the present procedure for predicting large fires seems to be poor for the Rocky Mountain and other regions where many of the ignitions may be due to lightning. Although the present model may be able to predict above-normal numbers of large fires due to dry conditions (as reflected by the FPI), it will not be able to predict these events if the above-average numbers of large fires is due to an unusually large number of fire starts. The fire occurrence probabilities used in the present work are the historical frequencies of fires for a given day and region, and thus do not change from year to year (Eqn 3).

The model appears to have better skill in predicting the probability of at least one major fire (>5000 acres (2023 ha)) per GACC for a forthcoming week. The reliability diagram (Fig. 8) was produced by assigning a value of one when at least one major fire was observed in a given GACC during a forthcoming week and zero otherwise (these are the hatch marks in the figure). Next, the observations were grouped into seven cells (according to the predicted values) and the observed proportion of ones in each cell was plotted against the predicted probabilities. In this sample of cases, there were 18 instances where the predicted probability was between 0.3 and 0.5. In six out of the 18 cases (0.33), we observed at least one fire >5000 acres (2023 ha). In other words, the predicted probability was close to the observed proportion of cases.

Conclusion

In summary, we have presented a statistical model for assessing the skill of fire danger indices and forecasting the distribution of the expected numbers of large fires over a given region and for the upcoming week. As an example, we studied the skill of the FPI. For the 3 years under study, the FPI appeared to have significant skill in predicting large fire occurrence. Although the maps and the goodness-of-fit graphs seem promising, there is room for improvement. For example, we do not have any explanatory variable (index) in the model for the probability of ignition. A fire weather index, such as probability of a lightning storm, may improve the skill of the model in predicting above-average levels of ignitions. At the moment, the ignition probabilities are based on historical averages (albeit the estimates are spatially and seasonally explicit). Using forecasted FPI values may also improve the model output. Finally, it may be interesting to include other fire danger indices (e.g. those in the NFDRS) together with the FPI studied here to see if any other index, or a combination of indices, could improve the skill of the model. The advantage of the probability model presented here is that it could be used with any index or a combination of indices and the skill of the final model may be tested by comparing estimated distributions of fire locations and sizes with those observed.

Acknowledgements

We thank the Desert Research Institute Program for Climate, Ecosystem and Fire Applications for the historical fire occurrence and size data.

References

Andrews PL, Bradshaw LS (1997) FIRES: fire information retrieval and evaluation system – a program for fire danger rating analysis. USDA Forest Service, Intermountain Research Station, General Technical Report INT-GTR-367. (Ogden, UT)

Andrews PL, Loftsgaarden DO, Bradshaw LS (2003) Evaluation of fire danger rating indexes using logistic regression and percentile analysis. *International Journal of Wildland Fire* **12**(2), 213–226. doi:10.1071/WF02059

Brillinger DR, Preisler HK, Benoit JW (2003) Risk assessment: a forest fire example. In 'Science and Statistics, Institute of Mathematical Statistics Lecture Notes'. (Ed. DR Goldstein) Monograph Series, pp. 177–196. (IMS: Bethesda, MD)

Brown TJ, Hall BL, Mohrle CR, Reinbold HJ (2002) Coarse assessment of Federal wildland fire occurrence data. Report for the National Wildfire Coordinating Group. Program for Climate, Ecosystem and Fire Applications Report 02-04. (Desert Research Institute) Available at http://www.cefa.dri.edu/Publications/fireoccurrencereport.pdf [Verified 18 June 2009]

Burgan RE (1988) 1988 revisions to the 1978 National Fire Danger Rating system. USDA Forest Service, Southeastern Forest Experiment Station, Research Paper SE-273. (Asheville, NC)

Burgan RE, Hartford RA (1993) Monitoring vegetation greenness with satellite data. USDA Forest Service, Intermountain Research Station, General Technical Report INT-GTR-297. (Ogden, UT)

Burgan RE, Klaver RW, Klaver JM (1998) Fuel models and fire potential from satellite and surface observations. *International Journal of Wildland Fire* **8**(3), 159–170. doi:10.1071/WF9980159

Deeming JE, Burgan RE, Cohen JD (1977) The National Fire Danger Rating System – 1978. USDA Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-GTR-39. (Ogden, UT)

Development Core Team (2004) 'R: a Language and Environment for Statistical Computing.' (R Foundation for Statistical Computing: Vienna, Austria) Available at http://www.R-project.org [Verified 18 June 2009]

Eidenshink JC (2006) A 16-year time series of 1-km AVHRR satellite data of the conterminous United States and Alaska. *Photogrammetric Engineering and Remote Sensing* **72**(9), 1027–1035.

Fosberg MA (1971) Moisture content calculations for the 100-hour timelag fuel in fire danger rating. USDA Forest Service, Rocky Mountain Forest and Range Experimental Station, Research Note RM-199. (Fort Collins, CO)

Fosberg MA, Deeming JE (1971) Derivation of the 1- and 10-hour timelag fuel moisture calculations for fire-danger rating. USDA Forest Service, Rocky Mountain Research Station, Research Note RM-207. (Fort Collins, CO)

Fosberg MA, Rothermel RC, Andrews PL (1981) Moisture content calculations for 1000-hour timelag fuels. *Forest Science* **27**, 19–26.

Hastie TJ, Tibshirani R, Friedman J (2001) 'The Elements of Statistical Learning. Data Mining, Inference, and Prediction.' p. 140. (Springer: New York)

Hosmer DW, Lemeshow S (1989) 'Applied Logistic Regression.' (Wiley: New York)

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314. doi:10.2307/1390807

Klaver JM, Klaver RW, Burgan RE (1997) Using GIS to assess forest fire hazard in the Mediterranean region of the United States. In '17th Annual ESRI Users Conference', 8–11 July, San Diego, CA. (ESRI: Redlands, CA) Available at http://proceedings.esri.com/library/userconf/proc97/proc97/to300/pap286/p286.htm [Verified 22 June 2009]

Lozano FJ, Suarez-Seoane S, de Luis E (2007) Assessment of several spectral indices derived from multi-temporal Landsat data for fire occurrence probability modeling. *Remote Sensing of Environment* **107**(4), 533–544. doi:10.1016/J.RSE.2006.10.001

Maddala GS (1992) 'Introduction to Econometrics.' 2nd edn. (Macmillan: New York)

Preisler HK, Brillinger DR, Burgan RE, Benoit JW (2004) Probability-based models for estimating wildfire risk. *International Journal of Wildland Fire* **13**, 133–142. doi:10.1071/WF02061

Roads JO, Chen S-C, Fujioka F, Kanamitsu M, Juang H-MH (1995) Global to regional fire weather forecasts. *International Forest Fire News* **17**, 33–37.

Sebastian-Lopez A, San-Miguel-Ayanz J, Burgan RE (2002) Integration of satellite sensor data, fuel type maps and meteorological observations for evaluation of forest fire risk at the pan-European scale. *International Journal of Remote Sensing* **23**(13), 2713–2719. doi:10.1080/01431160110107761

Sudiana D, Kuze H, Takeuchi N, Burgan RE (2003) Assessing forest fire potential in Kalimantan Island, Indonesia, using satellite and surface weather data. *International Journal of Wildland Fire* **12**(2), 175–184. doi:10.1071/WF02035

Van Wagner CE (1987) Development and structure of the Canadian Forest Fire Weather Index System. Canadian Forest Service, Forest Technical Report 35. (Ottawa, ON)

WFLC (2004) Large fire suppression costs – strategies for cost management. Report to Wildland Fire Leadership Council from the Strategic Issues Panel on Fire Suppression Costs. Available at http://www.forestsandrangelands.gov/reports/documents/2004/costmanagement.pdf [Verified 18 June 2009]

Wilks DS (1995) 'Statistical Methods in Atmospheric Sciences.' (Academic Press: London)

Manuscript received 23 February 2008, accepted 26 September 2008

Appendix

We used the R statistical package (Ihaka and Gentleman 1996; Development Core Team 2004) to estimate the coefficients in the logit regression lines given in Eqns 2 and 4. In order to estimate the smooth two-dimensional function of the intercepts, we first used a thin plate spline function that transforms the spatial data (x-coordinate, y-coordinate) for each fire to a matrix of the corresponding radial bases functions (Hastie *et al*. 2001). The required modules for fitting thin plate splines within R were downloaded from the web (Geophysical Statistics Project, National Center for Atmospheric Research, see http://www.cgd.ucar.edu/stats/Software/Fields, accessed 18 June 2009). Once the data are transformed using spline functions, standard logistic regression routine may be used to estimate the coefficients with the bases matrices as the explanatory variables.

The coefficients in Eqn 2 may be estimated simultaneously. However, because we only had FPI values for 3 years, whereas data on fire occurrence and size was available for over 20 years, we chose to do the estimation in two steps. First we estimated the spatial intercepts using 21 years of fire occurrence data using a logistic regression model with spatial location as the only explanatory. Next we used the 3 years of data on fire occurrence and FPI to fit the model in Eqn 2 with the values of the intercepts, A_i, set to their estimates obtained from the first step. It is anticipated that the 21-year dataset would give a better estimate of the historical probabilities than would the 3 years for which FPI is available.

The transformation suggested for day-in-year in Eqn 4 was also obtained using a spline function to account for the nonlinear seasonal effect in fire occurrence data. However, in this case, we used a periodic spline function that produces similar estimates for days at the beginning and end of the year.