CSIRO PUBLISHING
International Journal of Wildland Fire **2011**, *20*, 508–517 www.publish.csiro.au/journals/ijwf

Spatially explicit forecasts of large wildland fire probability and suppression costs for California

Haiganoush K. Preisler[A,G], *Anthony L. Westerling*[B], *Krista M. Gebert*[C,F], *Francisco Munoz-Arriola*[D] and *Thomas P. Holmes*[E]

[A]USDA Forest Service, Pacific Southwest Research Station, 800 Buchanan Street, West Annex Building, Albany, CA 94710, USA.
[B]University of California – Merced, 5200 N. Lake Road, Merced, CA 95343, USA.
[C]USDA Forest Service, Rocky Mountain Research Station, PO Box 7669, Missoula, MT 59807, USA.
[D]University of California – San Diego, 9500 Gilman Drive, San Diego, CA 92093-5004, USA.
[E]USDA Forest Service, Southern Research Station, Forestry Sciences Lab, PO Box 12254, Research Triangle Park, NC 27709, USA.
[F]Present address: USDA Forest Service, Northern Region, 200 East Broadway, PO Box 7669, Missoula, MT 59807, USA.
[G]Corresponding author. Email: hpreisler@fs.fed.us

Abstract. In the last decade, increases in fire activity and suppression expenditures have caused budgetary problems for federal land management agencies. Spatial forecasts of upcoming fire activity and costs have the potential to help reduce expenditures, and increase the efficiency of suppression efforts, by enabling them to focus resources where they have the greatest effect. In this paper, we present statistical models for estimating 1–6 months ahead spatially explicit forecasts of expected numbers, locations and costs of large fires on a 0.125° grid with vegetation, topography and hydroclimate data used as predictors. As an example, forecasts for California Federal and State protection responsibility are produced for historic dates and compared with recorded fire occurrence and cost data. The results seem promising in that the spatially explicit forecasts of large fire probabilities seem to match the actual occurrence of large fires, with the exception of years with widespread lightning events, which remain elusive. Forecasts of suppression expenditures did seem to differentiate between low- and high-cost fire years. Maps of forecast levels of expenditures provide managers with a spatial representation of where costly fires are most likely to occur. Additionally, the statistical models provide scientists with a tool for evaluating the skill of spatially explicit fire risk products.

Additional keywords: fire simulations, generalised Pareto distribution, hydroclimate, logistic regression, moisture deficit, spline functions.

Introduction

For land management agencies such as the US Forest Service (FS), wildland fire management has always been an integral part of the job of caring for the land and protecting lives and valuable resources. Fire management includes a mix of activities that can be planned for, such as hazardous fuel reduction treatments and wildfire prevention and detection, and activities that are more subject to the whims of Mother Nature, such as wildfire suppression. However, the entire wildfire management program, including suppression, is part of the annual budget for the federal land management agencies and, as such, is subject to federal regulations governing the use of funds. In 1870, the legislative appropriations bill included language, later known as the Anti-Deficiency Act, which prohibits departments or agencies from spending more in a fiscal year than they have been provided in their budget (United States Senate 1998). Given that suppression is part of an overall budget they cannot exceed, the FS and other federal land management agencies need estimates of future suppression expenditures both during the budgetary planning process, which occurs 2 to 3 years out, and during the current fiscal year in order to monitor spending.

Over the past decade, the need for such information has grown. Both the magnitude and variability of expenditures have increased substantially over the past two decades. Budgets formulated 2 to 3 years in advance using a 10-year moving average of expenditures often deviate substantially from the amount actually expended. To further complicate matters, agency trust funds, such as the Knudson–Vandenberg fund, were often available to draw from in active fire years, and the funds were repaid in subsequent years. However, these funds have been largely depleted owing to continual borrowing during one active fire year after another. To meet antideficiency regulations, in recent years the FS has either had to request highly uncertain emergency supplemental funding from

10.1071/WF09087 1049-8001/11/040508

Congress or transfer funds from other programs within the FS to pay for suppression.[A]

Owing to these issues, it is important for the FS and other federal land management agencies to have advance warning of the likelihood that actual fire suppression expenditures will exceed the amount appropriated for that fiscal year. It is also important that the agencies have an indication of the magnitude of likely suppression expenditures in order to plan for shortfalls in spending. To that end, researchers have been working on providing forecasts of both upcoming fire activity and likely suppression expenditures (see Gebert and Schuster 1999; Bachelet *et al*. 2000; Westerling *et al*. 2002; Gebert *et al*. 2007; Preisler and Westerling 2007; Abt *et al*. 2008, 2009; Prestemon *et al*. 2008).

There are several ongoing research projects aimed at forecasting suppression expenditures at various lead times. The Rocky Mountain Research Station has developed within-season forecast models, which are currently being used by both the FS and the Department of Interior to monitor spending during the fire season (Gebert and Schuster 1999). These forecasts use a 'best-professional judgment' approach, where forecasts of upcoming fire activity are produced by personnel in the Predictive Services group at the National Interagency Fire Center in Boise, Idaho. These predictions of fire activity are then used to produce forecasts of monthly suppression expenditures that are added to actual year-to-date expenditures to arrive at a forecast for annual suppression expenditures.

Prestemon *et al*. (2008) have developed models, evaluated in autumn and in spring, which use climate and trend variables to estimate suppression expenditures for October–September fiscal year. These models have the advantage of being more scientifically based than the fire-season forecasts but thus far cannot be updated, and the forecasts are provided only twice per year. Also, none of these projects use spatially explicit fire history, land surface and climate data. The advantage of using spatial data to produce the forecasts is the possibility of being able to use the forecasts to inform managers not only of how much might be spent to suppress fires but also where the expenditures might actually occur. There also exists the possibility to reduce expenditures, or to at least increase the efficiency of suppression and prevention efforts, by using spatially explicit forecasts to focus resources where they will have the greatest effect.

The work by Bachelet *et al*. (2000) describes a spatially explicit dynamic vegetation model (MAPSS: Mapped Atmosphere–Plant–Soil System) that includes a fire module (MC1). Currently, MAPSS is being used to produce seasonal forecasts of fire occurrence probability and expected area burned, known as fire risks. The fire occurrence probabilities from the MC1 modules, however, are consensus probabilities 'defined as the percentage of climate scenarios (out of a total of five) that predict a fire in the timeframe mentioned' (http://www.fs.fed.us/pnw/mdr/mapss/fireforecasts/index.shtml, accessed 5 April 2011). As such it is not easy to assess the skill of these forecasts because consensus probabilities cannot be compared directly with historic fire frequency records. In contrast, the skill of large fire probability forecasts estimated from fire occurrence data may be compared directly with observed fire occurrences and sizes, as will be discussed below.

In this work, we propose a statistical model that is used to provide spatially explicit forecasts of suppression costs. As an intermediary step, climate data up to present are used to predict the number of large fires (≥200 ha) on a 0.125° grid for 1 to 6 months ahead, referred to herein as the 'upcoming season'. The estimation is done in two steps. First, we estimate a statistical model relating fire suppression costs – per fire – to fire size, vegetation and topography. Next, we develop and estimate a probability model for forecasting fire occurrence and size. The model estimates probability of occurrence of large fires per 0.125° grid-cell per month, using vegetation, topography and climate variables up to present as explanatory variables. We also estimate the distribution of fire sizes for all fires ≥200 ha. Finally, the two models above are combined to produce spatially explicit forecasts of suppression costs for the upcoming fire season. As an example, our methods are applied to develop a wildfire-forecasting model for California Federal and State protection responsibility areas.

Data and statistical methods

Spatial domain

The spatial domain for this analysis covers the current combined fire protection responsibility areas within the State of California of the California Department of Forestry and Fire Protection and contract counties (combined here as 'CDF'), the US Department of Agriculture's Forest Service (USFS) and the US Department of Interior's National Park Service (NPS), Bureau of Land Management (BLM) and Bureau of Indian Affairs (BIA). The spatial resolution is a 0.125° latitude by longitude grid (~12 km resolution).

Fire history

A history of large wildfires (≥200 ha) for California for 1985–2003 was assembled from digital fire records obtained from CDF (see http://frap.cdf.ca.gov/, accessed 5 April 2011), and FS, NPS, BLM and BIA (see http://fam.nwcg.gov/famweb/weatherfirecd/index.htm, accessed 18 April 2011). The methods used in compiling a fire history from these data are described in Westerling *et al*. (2006, 2009), and Westerling and Bryant (2008). Westerling *et al*. (2003) describe the federal fire histories. The result is a 0.125° gridded monthly dataset of frequencies of fires ≥200 ha in size and of the total area burned in these large wildfires. Federal fires were allocated to the grid cell in which they were reported to have ignited. CDF fires were reported as polygon perimeters and were allocated to the grid cell corresponding to their centroid. Fires were assigned to the month in which they were discovered. In many cases, fires continued to burn for additional months, but the means to apportion area burned by month were unavailable.

Whereas wildfires managed by the Fish and Wildlife Service, the Department of Defence and the Bureau of Reclamation were

[A]In the fiscal year 2010, a 'FLAME FUND' (Federal Land Assistance, Management, and Enhancement Fund) was established as part of the Interior Appropriations Bill (77-21). This fund, which is separate from the regular appropriations, is intended to reduce the likelihood of these transfers from other programs. The bill also requires forecasts of expected suppression spending several times a year.

not included, the fire history assembled here is sufficiently comprehensive to allow estimation of fire risks in a diverse array of California fire regimes. Our prediction models could easily be extended to cover additional parts of the state where fire histories of comparable quality and duration are not available, using the model coefficients derived for the areas described above.

Vegetation characteristics

Coarse vegetation characteristics such as forested land area and the vegetated fraction of each grid cell were compiled from the Land Data Assimilation System (LDAS) for North America's 0.125° gridded vegetation layers that use the University of Maryland vegetation classification scheme with fractional vegetation adjustment (UMDvf) (Hansen *et al*. 2000; Mitchell *et al*. 2004). The UMDvf scheme has 14 coarse surface categories derived from 1-km Advanced Very High Resolution Radiometer (AVHRR) satellite data collected from April 1992 to March 1993. We combined these to obtain the vegetation categories analysed here: Forest (the Evergreen Needleleaf and Broadleaf Forest categories, the Deciduous Needleleaf and Broadleaf Forest categories, and the Mixed Cover category), Woodland (the Woodland and Wooded grasslands categories), Grassland (the Grassland category) and Shrubland (the Closed and Open Shrubland categories), Crop, Bare, Open Water, and the fraction of each grid cell in the Forest, Woodland, Shrubland and Grassland categories above. We were unable to distinguish between evergreen and deciduous forest in this analysis because too little area of the latter was included in the study area to support a statistical analysis at a 12-km resolution.

Topography

Topographic data on a 0.125° grid were also obtained from LDAS. The LDAS topographic layers are derived from the GTOPO30 Global 30 Arc Second (~1 km) Elevation Data Set (Gesch and Larson 1996; Verdin and Greenlee 1996; Mitchell *et al*. 2004). We tested mean and standard deviation of elevation, slope and aspect as explanatory variables in our statistical model.

Hydroclimate

We used a 'nowcast' from the University of Washington and Princeton University Westwide Seasonal Hydrologic Forecast System to get up-to-date gridded hydroclimate data throughout the fire season (http://www.hydro.washington.edu/forecast/westwide/spatial/ncast/index.shtml, accessed 5 April 2011). Based on the index station method (Wood and Lettenmaier 2006), the data describing the preceding month are available at the beginning of every month, allowing us to issue timely seasonal forecasts with monthly forecast updates based on recent climate observations. This system uses historical (1960–2009) climate data obtained from a sample of National Cooperative Development Corporation stations, including maximum and minimum temperature, precipitation, and wind speed regridded from Global Reanalysis data, together with LDAS vegetation and topography, to drive the Variable Infiltration Capacity (VIC) hydrologic model at a daily time step in full energy mode (Liang *et al*. 1994; Maurer *et al*. 2002; Hamlet and Lettenmaier 2005; Wood and Lettenmaier 2006). The output gridded hydroclimatic variables include actual evapotranspiration (AET), soil moisture, relative humidity (RH), surface temperature (TMP) and snow-water equivalent (SWE).

We used average monthly temperatures calculated from the VIC input data and, as indicators of drought stress, cumulative moisture deficits. We calculated the cumulative water-year moisture deficit for the preceding 2 years, for the current water year through March, and for each month afterwards through the fire season. Moisture deficit (D) was calculated from Potential Evapotranspiration (PET) and AET (D = PET − AET). PET was estimated by using the Penman–Monteith equation (Penman 1948; Monteith 1965).

Population

We included population as a potential explanatory variable, given that human-caused ignitions are important in many parts of California. In addition, population may be a proxy for other variables such as infrastructure, accessibility and suppression resource availability. Gridded population estimates were obtained from the Center for International Earth Science Information Network's Socioeconomic Data and Applications Center at Columbia University. We used the Gridded Population of the World Version 3 (http://sedac.ciesin.columbia.edu/gpw/, 5 April 2011) at 2.5-arc-minutes resolution, adjusted to match United Nations population totals. We aggregated these data to produce population counts on the LDAS 0.125° grid.

Estimating suppression cost per fire

We obtained fire suppression cost data per fire for a sample of fires of sizes greater than 40 ha (or 100 acres) for the years 1995 through 2007. These were obtained from a database created and maintained by the Rocky Mountain Research Station, which includes fire-specific suppression expenditures and fire characteristic information for a large set of federal wildland fires (see Gebert *et al*. 2007 for a full description of the database). Although our fire occurrence data include both Federal and State protection responsibility fires, the expenditure costs are only from federal wildland fires. Consequently, if forecast costs are to be extended to estimate both federal and CDF fires, we will need to make the assumption that suppression costs for federal and CDF fires are similar.

We developed a statistical regression model relating cost per fire to various explanatory variables, including fire size. The specific explanatory variables tested were elevation, slope, aspect, standard deviation of the elevation, percentage forest, vegetation fraction and population. The variables were evaluated for the 0.125° grid cells containing the fire.

Following is the final model with only the significant variables included:

$$y = \beta_{\mathrm{ok}} + \beta_{1k} \cdot esd + \beta_{2k} \cdot vegf + \beta_3 \log(hec) + \varepsilon \qquad (1)$$

where y is the square root of the suppression cost, hec is the size of the fire in hectares, esd is the standard deviation of elevation, $vegf$ is vegetation fraction, β values are parameters to be estimated and ε is white noise. The square root of cost was used because the residuals from this fit were best approximated by the

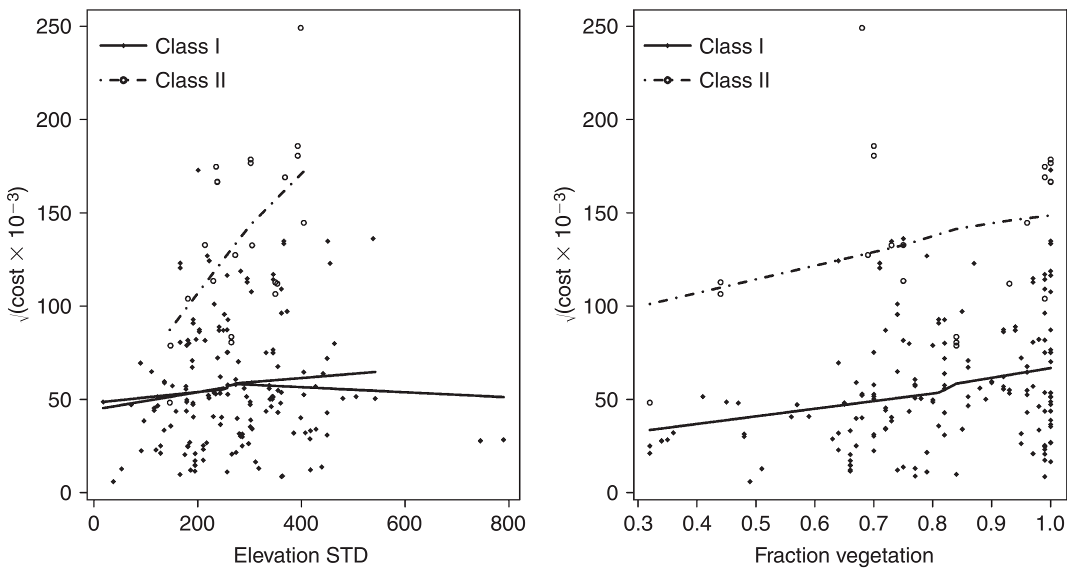

Fig. 1. Suppression costs per fire *v.* elevation standard deviation and fraction vegetation, for two fire size classes, Class I: 200–8500 ha, and Class II: >8500 ha. The two solid lines through the points on the left panel are the fitted curves with and without the two outliers at elevation standard deviations >700.

normal distribution. The subscript k in Eqn 1 stands for one of two fire size classes: Class I, fires between 200 and 8500 ha, and Class II, fires ≥8500 ha. In our preliminary exploratory analysis, we noted that the slopes and intercepts of the relationships between the *vegf* and *esd* variables appeared to be affected by the fire size (Fig. 1). In particular, for fires greater than 8500 ha, the standard deviation of elevation seemed to have a larger effect on area burned than for fires <8500 ha. Consequently, in our final model in Eqn 1, different slopes were assigned to two fire size classes. The standard deviation of elevation (*esd*) is an index of surface roughness that may be indicating how easy it is for a fire to spread given the terrain, as well as how accessible the terrain is for firefighters. Fraction vegetation (*vegf*) describes how much vegetated area there is in the 0.125° grid that can carry a fire in that location. In our sample, the correlation between *vegf* and *esd* was −0.2.

Estimating probability of large fire occurrence

Using land surface (topography and vegetation), population and hydroclimate, expected numbers of large fires for the upcoming season were predicted by fitting spatially explicit logistic regression models. The statement for the probability of a large fire occurrence was as follows: let $r_{ij} = 1$, if there is a fire of size ≥200 ha at location i in month j, and zero otherwise. Then r_{ij} is a Bernoulli random variable with probability of response given by

$$p_{ij} = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})}$$

and with the linear predictor

$$\theta_{ij} = \beta_j + g(long_i, lat_i) + \sum_m g_m(X_{mij}) \qquad (2)$$

The spatial covariate ($long_i$, lat_i) is the longitude, latitude pair of each 0.125° grid cell in California State and Federal lands; the covariate X_{mij} is the mth explanatory variable from the list of variables, including topography, vegetation and lagged climate variables for location i and on date j. The parameters β_j, one for each month, and the non-parametric functions g and g_m are estimated from the data. Note that the complement of the response probability (i.e. $1 - P$) is the probability of 'no fire' or a fire of size less than 200 ha.

We used spline functions for evaluating g_m and thin-plate spline for evaluating the two-dimensional spatial function g (Hastie *et al.* 2001). We used the generalised additive modules of R Development Core Team (2008) in the R statistical package to carry out the estimation and assess the significance of the various explanatory variables. Similar models were used in Preisler and Westerling (2007) and Preisler *et al.* (2008) for studying relationships between various fire danger indices and probability of large fire occurrence in the western United States.

Probability estimates were evaluated for 1–6 months ahead. For example, using the previous 2 years of monthly climate data up to the end of March, we evaluated response probabilities for the months of April to September for that year (Table 1). At the end of March, the response probabilities for April are 1-month-ahead forecasts, whereas that for September is a 6-months-ahead forecast (Table 1; Model 2 for April and Model 3 for the rest). At the end of April, we updated the climate variables to include values up to the end of April, then estimated response probabilities for May–September (Table 1: Model 2 for May and Model 4 for rest) and so on. We also evaluated response variables with only spatial location and month as explanatory variables (Table 1, Model 1). The latter response probabilities were used as the historic estimates for a given location and month. The historic probabilities are invariant from year to year, and they are used to describe the 'norm' for the years in the study (1985–2003) for a given location and month.

Forecast probabilities may also be used to produce maps of significant departures from normal condition. Here, the 'norm' is considered to be the historic average probabilities evaluated from a model with no climate variables. Departure from normal conditions may be displayed by mapping the odds of a large fire in the present year relative to historic odds. If the odds of a large fire, $P/(1-P)$, on a given 0.125° grid-cell and for a given year, are significantly greater than the historic odds, then that cell is designated as having higher than normal odds. A cell is designated to be significantly higher than normal if the odds for that year were larger than one standard deviation above the historic odds. The standard deviation was estimated using the jack-knife procedure (Efron and Tibshirani 1993) where 19 different sets of coefficient estimates were evaluated, each using historic data from all 19 years but one and then calculating the jack-knife standard errors of the 19 values.

In the next section, we fit a generalised Pareto distribution (GPD) to the empirical fire size distribution to estimate the expected size of a fire given the occurrence of a fire of at least 200 ha.

Estimating conditional distribution of large fire size

Histograms of observed large fire sizes are often best characterised by heavy-tailed distributions, such as the log-normal or the Pareto distributions. These distributions have often been used successfully to characterise catastrophic events such as earthquakes (Brillinger 1993) and fires (Moritz 1997). Other distributions used to model fire sizes include the truncated exponential distribution (Cumming 2001). Lately, Ramesh (2005), Holmes *et al*. (2008) and Schoenberg *et al*. (2003) have demonstrated that the GPD is a useful model for characterising large fire sizes in particular when the data are truncated at the lower end. A more comprehensive list of citations on the use of GPD for modelling fire sizes can be found in Holmes *et al*. (2008). In our case, only fires greater than 200 ha are included in the data. The GPD scale and shape parameters are estimated from the data and a threshold level, which will be set to 200 ha. The scale and shape parameters for our data were estimated within the R statistical package using modules from the 'ismev' library (R Development Core Team 2008). One may also include explanatory variables (Holmes *et al*. 2008); however, none of the variables in our list seemed to have a significantly important effect on the fire size given that the fire has already exceeded 200 ha. This might not be surprising given the fact that suppression efforts (a variable not studied here) may be one of the most important explanatory variables for the eventual size of a fire that is already greater than 200 ha.

The goodness-of-fit of the fitted distribution was assessed by simulating 5000 observations from the GPD with a scale and shape parameter set at the values estimated from the data and then comparing the quantiles of the simulated data with those of the observed fire sizes. Simulated values (r) from the GPD were generated by the formula, $r = \log(200) + \hat{\sigma} \cdot (U^{-\hat{\alpha}} - 1)/\hat{\alpha}$, where U is a random variable from a uniform (0,1) distribution (Hastings and Peacock 1975; Davison 2003) and where $\hat{\alpha}$ and $\hat{\sigma}$ are values of the shape and scale parameters estimated from the observed large fire sizes.

Forecasting spatially explicit fire suppression costs

In the 'Estimating suppression cost per fire' section above, we developed a regression model for estimating suppression costs for a given fire, given fire size and some site characteristics. As fire locations and sizes are not known for an upcoming season, we decided to simulate them given the estimated response probabilities and the estimated distribution of fire sizes developed above. By generating multiple simulations of fires and then fire sizes, given a large fire, we can produce a distribution of expected suppression costs at the end of March for the rest of the fire season. Fire occurrence for each month and each pixel was simulated by drawing a random sample from the Bernoulli distribution with probability of success set to the forecast probability of a large fire as given by Eqn 2. Next, for all pixels and each month where the response was one (i.e. a large fire occurrence was forecast), we generated a realisation from the GPD using the method described in the section on 'Estimating conditional distribution of large fire size'. The projected cost of fire expenditures at each pixel was next estimated by Eqn 1 with fire size and size class given by the simulated values. Averages over 1000 simulations per pixel were then mapped to produce spatially explicit cost estimates for the upcoming fire season, covering the March–September period.

Table 1. Models and significant variables used for predicting 1–6 months ahead spatially explicit probabilities of large fires in California

Variable codes are: Spatial, two-dimensional smooth function of latitude and longitude; Month, categorical variable for each month between April and September; Elevation, average elevation (m) over the 0.125° pixel; %Forest, percentage forested land in the 0.125° pixel; pCM, previous years cumulative moisture deficit for October–May; CM, present year cumulative moisture deficit for March–September; pMD, previous month's average moisture deficit; pT, previous month's average temperature; T3–T6, present year average temperature for the months of March to June respectively; MD6, present year average moisture deficit for the month of June

Model	Variables included
(1) Historic	Spatial, Month, Elevation, %Forest
(2) 1-month ahead	Spatial, Month, Elevation, %Forest, pCM, CM, pMD, pT
(3) End of March	Spatial, Month, Elevation, %Forest, pCM, CM,T3
(4) End of April	Spatial, Month, Elevation, %Forest, pCM, CM,T4
(5) End of May	Spatial, Month, Elevation, %Forest, pCM, CM,T5
(6) End of June	Spatial, Month, Elevation, %Forest, pCM, CM,T6
(7) End of July	Spatial, Elevation, MD6
(8) End of August	1-month ahead model (as there is only 1 month, September, being forecast)

Results

Suppression costs per fire

We observed a significant relationship between the standard deviation of elevation and suppression cost per fire. There appeared to be an increase in suppression costs when the elevation around the fire was more variable. However, this increase in cost was only apparent in the largest size class (Fig. 1). The effect of fraction vegetation on suppression cost was also found to be significant; however, there was no significant difference in the slope for the two fire size classes. The overall multiple correlation coefficient for Eqn 1 was 70%. Comparison of observed costs with predicted costs (Fig. 2) demonstrates how, even with a multiple correlation of 70%, there still remains a large degree of unexplained variability in costs per fire. Correlation alone is not a sufficient statistic when describing the skill of a model.

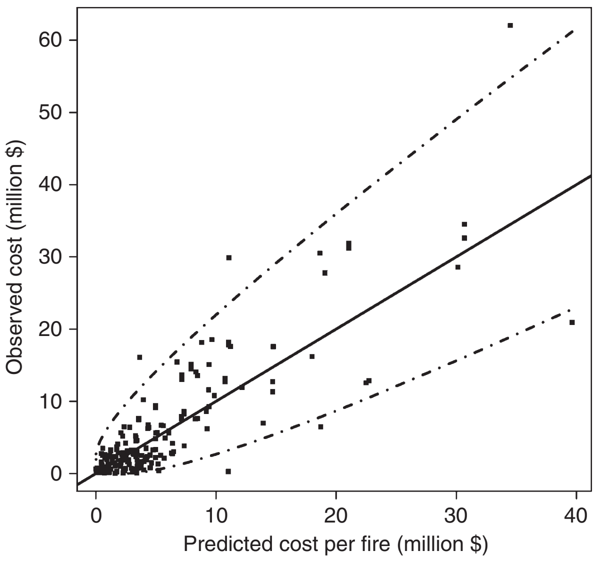

Fig. 2. Observed *v.* predicted costs (US$) per fire based on fire size, elevation standard deviation and vegetation fraction. The dashed lines are the approximate 95% confidence bounds plotted about the 45° line (solid line).

Forecasts of large fire occurrence and size

The following variables were found to have statistically significant effects on the historic probabilities of large fires: spatial location, month-in-year, elevation and percentage forested land (Table 1, Model 1). Note that no climate variables are used to evaluate the historic fire occurrence probabilities because historic probabilities are supposed to estimate overall average monthly levels for a given location. The following variables were found to have significant effects on the monthly probabilities of large fire occurrence; spatial location, month-in-year, elevation, percentage forested land in addition to the previous year's cumulative moisture deficit for October–May, the present year cumulative moisture deficit for October–March, the previous month's average moisture deficit, and temperature (Table 1, Model 2). These results are consistent with those of a previous study over the western United States (Preisler and Westerling 2007).

As a measure of the overall fit of the models, we produced reliability plots for Models 1, 3, 4 and 5 of Table 1. We produced

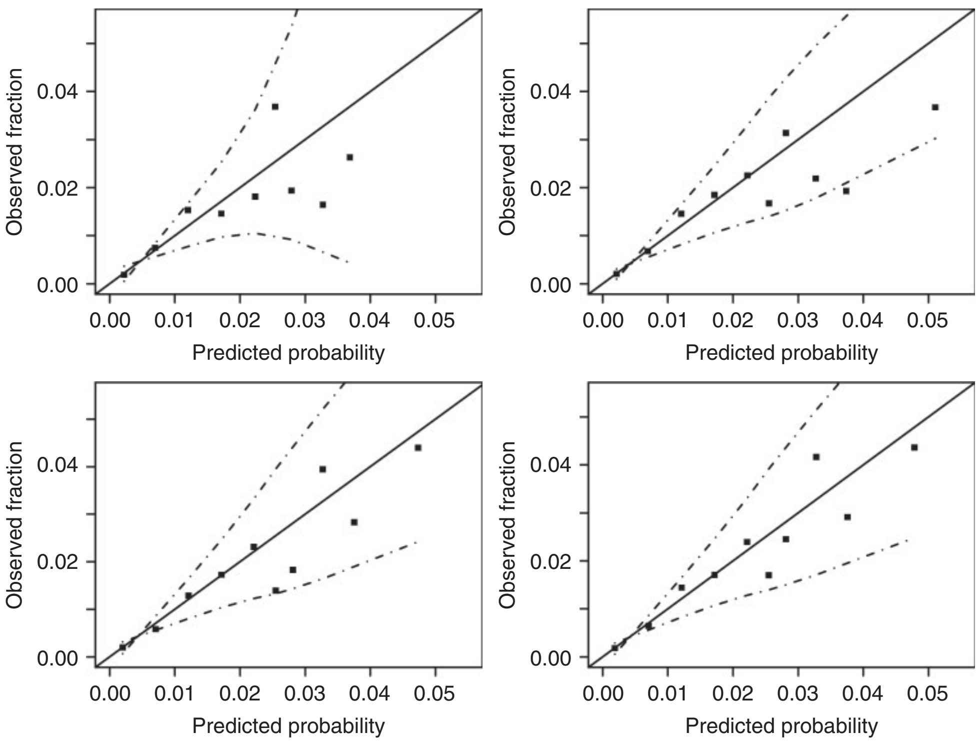

Fig. 3. Observed *v.* predicted probability of ≥200 ha fires for four different forecasts. The dashed lines are approximate 95% confidence bounds plotted about the 45° line (solid line).

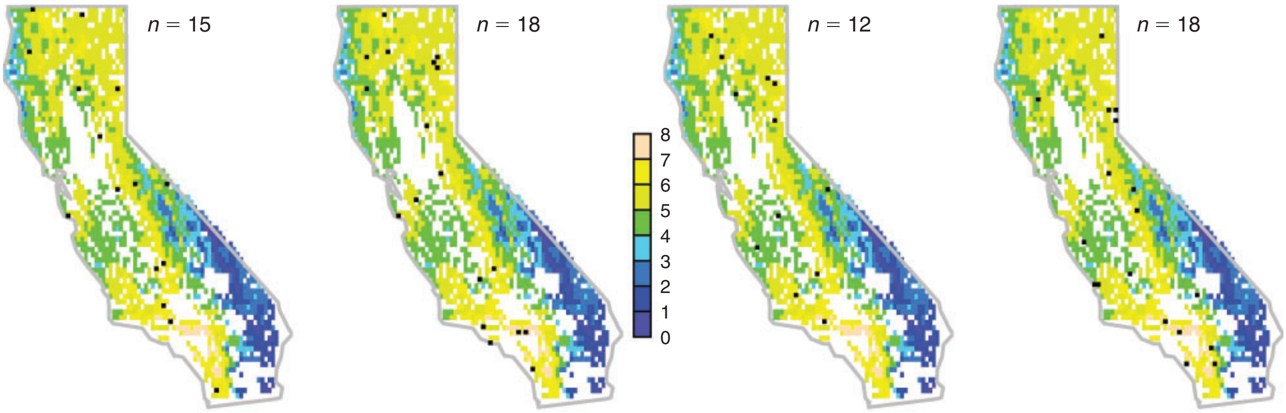

Fig. 4. Maps of forecast large fire probabilities (%) per 0.125° pixel, with potential locations of fires (•) from three simulations (three left panels) compared with observed locations of fires (•) for August 1994 (right panel). *n* is the total number of large fires in each panel. The year 1994 was a relatively high fire year.

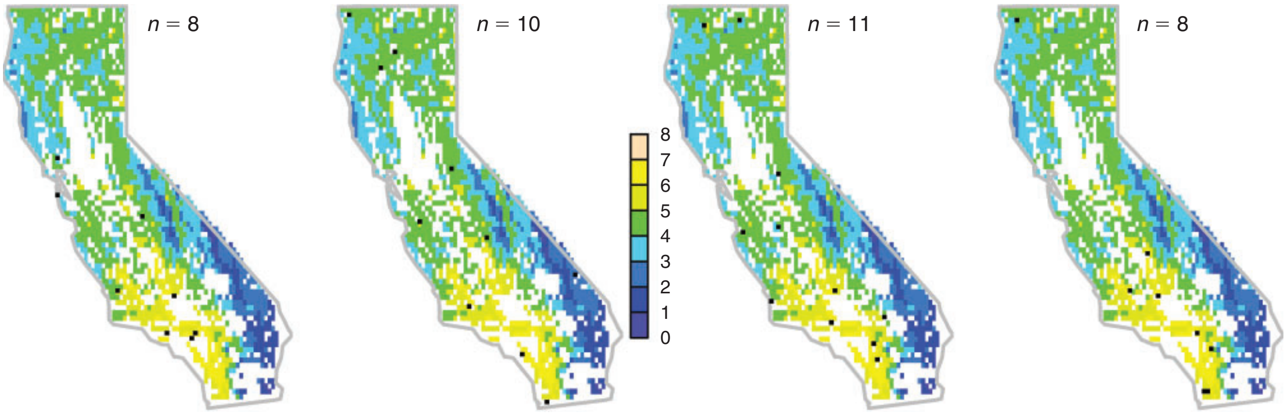

Fig. 5. Maps of forecast large fire probabilities (%) per 0.125° pixel, with potential locations of fires (•) from three simulations (three left panels) compared with observed locations of fires (•) for August 1995 (right panel). *n* is the total number of large fires in each panel. The year 1995 was a relatively low fire year.

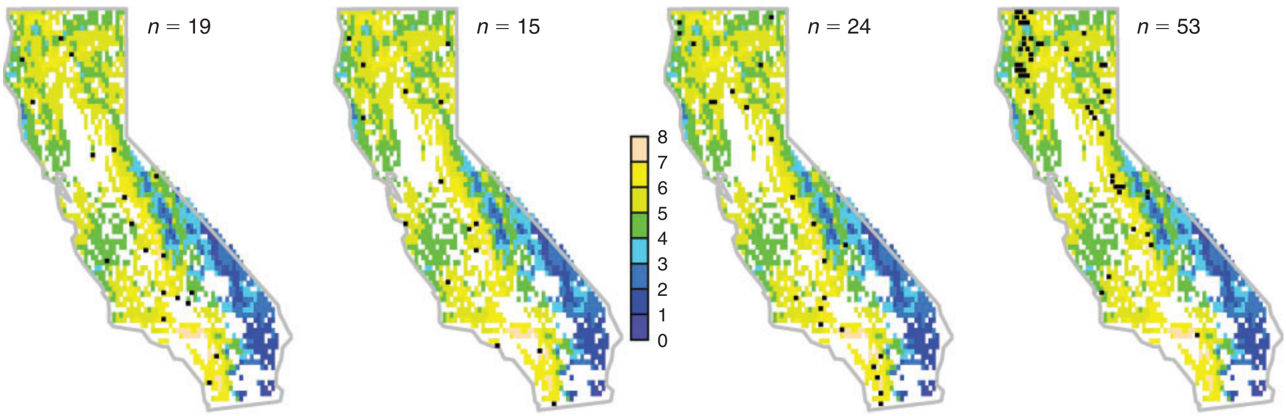

Fig. 6. Maps of forecast large fire probabilities (%) per 0.125° pixel, with potential locations of fires (•) from three simulations (three left panels) compared with observed locations of fires (•) for August 1987 (right panel). *n* is the total number of large fires in each panel. There were many lightning-caused fires in northern California in 1987.

goodness-of-fit plots by dividing the estimated probabilities for June through September into 10 classes, and plotting the fraction of cases with observed large fire, in each class, against the midpoint of the class (Fig. 3). For a good fit, the points will be scattered close to the 45° line within the 95% confidence bounds. The forecasts done in May appear to be an improvement over the historic model and over the forecasts made at the beginning of the fire season (March). In order to assess the goodness-of-fit

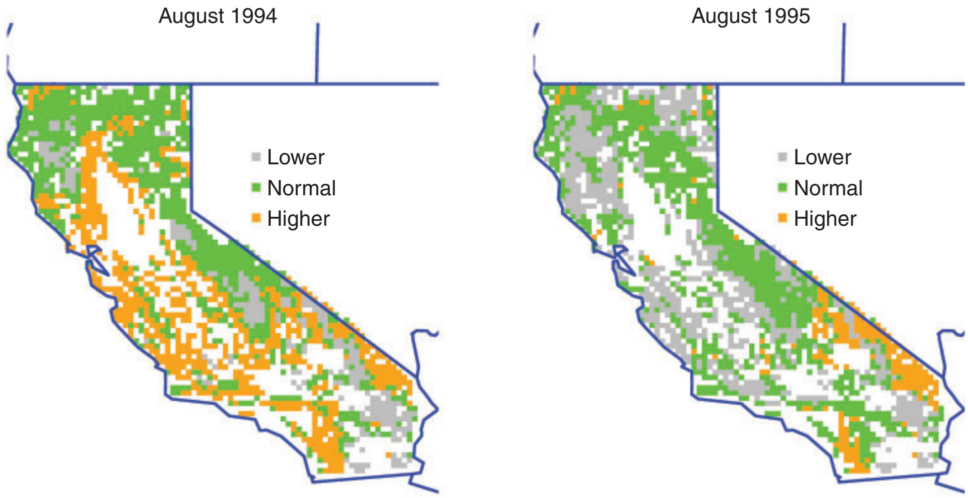

Fig. 7. Maps of forecast odds of large fire occurrences in August 1994 (left) and 1995 (right) relative to historic odds developed with data up to end of March.

of the model spatially, we produced maps of observed locations of fires for a given month and compared them with a set of maps with simulated fire locations (Figs 4–6). The simulations were done using the estimated probabilities to generate a response (fire or no fire of size ≥200 ha) for each 0.125° grid cell and for a given month. The year 1994 (Fig. 4) was a high fire year. The spatial pattern and numbers of observed large fires are similar to the simulated outcomes. The latter is an indication that the observed fires can reasonably be looked at as a realisation from our estimated distribution of large fires for that month and year. The year 1995 (Fig. 5) was a low fire year. Again, the simulations are seen to be similar to the observed numbers and pattern of fires. A similar series of plots were developed for 1987 (Fig. 6). Here, we see how the simulation plots may be used to study the limitations of our model. The pattern in the map with observed fires seems to be different from the three simulations. In particular, there is a cluster of fires that occurred in the northwestern region of California. Our model does not take into account lightning events. The cluster of large fires in northern California during the summer of 1987 is due to a larger than average number of lightning events – hence more ignitions than the historic average – resulting in a greater number than expected large fires. Although our model takes into account 'clustering' of fires due to similarities in the topography or vegetation of nearby points by including a spatial term in the logistic model, it does not take into account causes of clustering of ignitions due to lightning events.

One product of our probability modelling is forecast odds maps relative to historic averages. The relative odds maps for August 1994 and August 1995 (Fig. 7) demonstrate the utility of these maps. Our forecasts made at the end of March for the upcoming month of August seem to capture the high and low fire seasons in these 2 years. Similar maps for other historic years (together with the latest forecasts) are posted on the web at https://wildfire.ucmerced.edu/forecast.

The GPD appeared to be a good fit to the distribution of observed large fire sizes when the observed distribution of large fire sizes was compared with the simulated GPD (Fig. 8).

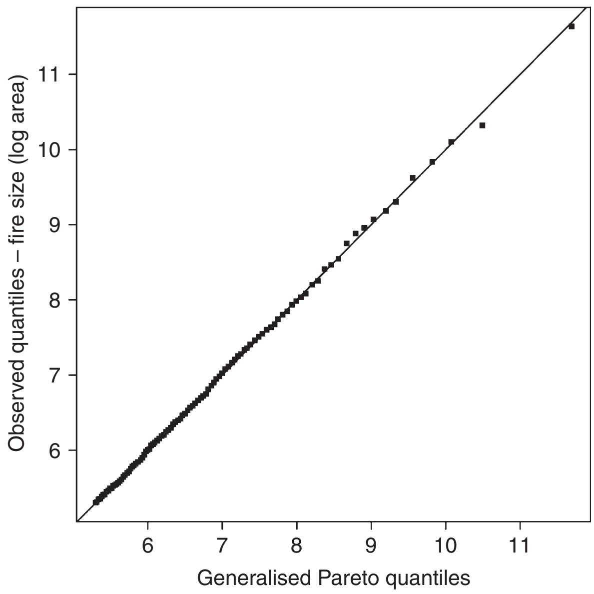

Fig. 8. Quantiles of observed fire sizes against those generated from the generalised Pareto distribution with parameters estimated from the historic fire size data.

Spatially explicit fire suppression costs

As an example of the final output of our modelling procedure, we produced a map of forecast costs for the 1994 fire season and compared them with estimated costs (Fig. 9). The estimated costs came from the National Interagency Fire Management Integrated Database (NIFMID), which includes a field containing estimated suppression expenditures on each fire by the reporting agency. Because the NIFMID cost data do not include CDF fires, we multiplied the forecast costs per pixel by the fraction of the 0.125° pixel within California with federal responsibility of fire suppression.

The general spatial pattern of higher forecast costs for the Los Angeles area, southern California, and the northernmost region of California seem to match well with the pattern of large fire

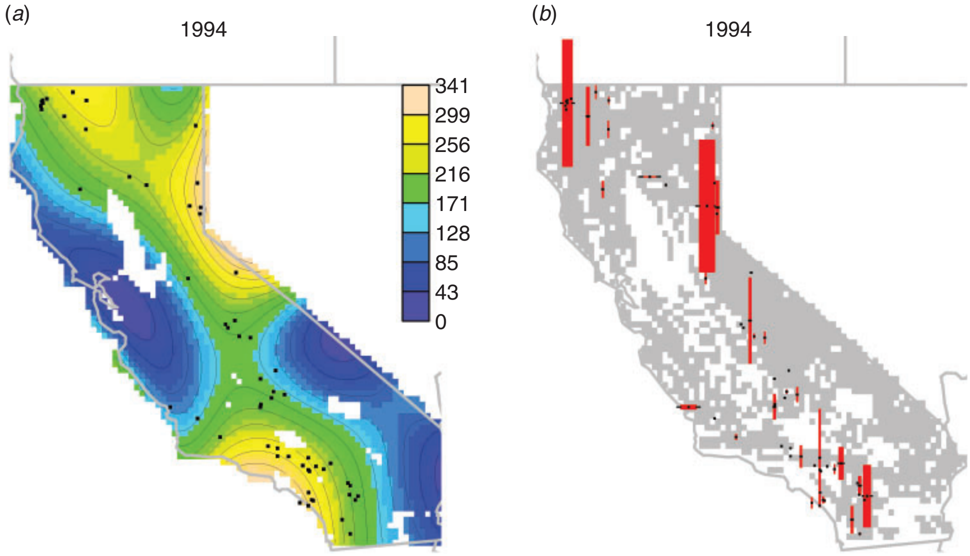

Fig. 9. (*a*) Map of forecast average expenditure costs (background colours) per pixel – in US$1000 – for the 1994 fire season generated using data up to end of March, superimposed with locations of observed wildland fires (black dots). (*b*) Locations of observed wildland fires during 1994 fire season (black dots) with a rectangle centred at each fire giving the individual fire size (width of rectangle) and suppression cost (height or rectangle).

occurrences and costs (Fig. 9). This general spatial pattern did not change for the 1995 forecast (figure not included), although the overall level of expenditure was lower owing to the smaller number of fires in 1995 (Fig. 5). We also produced a map of the standard deviations of forecast cost (Fig. A1 of the Accessory publication, see http://www.publish.csiro.au/?act=view_file&file_id=WF09087_AC.pdf). Note that the forecast costs are per given 0.125° pixel, and they reflect the chance of a large fire in that pixel together with the cost of suppression given a fire. Consequently, the cost of suppressing a large fire in a particular pixel may be large. However, if the probability of a large fire occurrence is small, then the forecast cost will not be as high. Inversely, if the cost of fire suppression is low in a given pixel, but the probability of occurrence is high, then the forecast cost will be higher accordingly.

Summary and conclusions

In the last decade, increases in fire activity and, subsequently, suppression expenditures by federal land management agencies have caused budgetary problems for the involved agencies and increased scrutiny of spending by government oversight agencies. As federal agencies increase their efforts to contain the costs of suppressing wildfires, spatial forecasts of upcoming fire activity and likely expenditures may help the agencies to reduce expenditures, or to at least increase the efficiency of suppression and prevention efforts, by enabling them to focus resources where they will have the greatest effect.

The methodology outlined in this article shows promise for helping with this effort. The spatially explicit forecasts of large fire probabilities seem to match the actual occurrence of large fires well, with the exception of years with widespread lightning events, which remain elusive to prediction efforts. Suppression costs, as previous researchers have found (Gebert *et al*. 2007; Prestemon *et al*. 2008), are difficult to predict, and the models tested in the present effort also left a large degree of unexplained variability. This is not unexpected, however, as suppression expenditures are influenced by a wide array of non-biophysical factors that are not readily captured in a statistical model, especially a spatially explicit model (see Canton-Thompson *et al*. 2006, 2008 for some of the other factors influencing suppression costs). However, even in light of this, our forecasts of suppression expenditures did seem to differentiate between low- and high-cost fire years and regions and, consequently, can provide managers with a spatial representation of where costly fires are most likely to occur.

Additionally, the information provided by these models may prove useful as independent variables in models designed to forecast annual suppression expenditures, such as those produced by Prestemon *et al*. (2008) or Gebert and Schuster (1999). Thus far, however, this methodology has only been tested for the Pacific Southwest Region of the USDA Forest Service in California. In order to be useful for predicting nationwide suppression expenditures, the methodology will have to be tested for the rest of the United States.

Acknowledgements

This work was supported by the USDA Forest Service's Rocky Mountain, Pacific Southwest and Southern Research Stations, and by National Oceanic and Atmospheric Administration's Regional Integrated Science and Assessment Program for California at Scripps Institution of Oceanography. We thank two reviewers for their thorough review and helpful comments on the first draft of the paper.

References

Abt KL, Prestemon JP, Gebert KM (2008) Forecasting wildfire suppression expenditures for the United States Forest Service. In 'The Economics of Forest Disturbances: Wildfires, Storms and Invasive Species'. (Eds TP Holmes, JP Prestemon, KL Abt) pp. 341–360. (Springer Publishing: New York)

Abt KL, Prestemon JP, Gebert KM (2009) Wildfire suppression cost forecasts for the US Forest Service. *Journal of Forestry* **107**(4), 173–178.

Bachelet D, Daly C, Lenihan JM, Neilson R, Parton W, Ojima D (2000) Interactions between fire, grazing, and climate change at Wind Cave National Park, SD. *Ecological Modelling* **134**, 229–244. doi:10.1016/S0304-3800(00)00343-4

Brillinger DR (1993) Earthquake risk and insurance. *Environmetrics* **4**, 1–21. doi:10.1002/ENV.3170040102

Canton-Thompson J, Thompson B, Gebert KM, Calkin DE, Donovan GH, Jones G (2006) Factors affecting fire suppression costs as identified by incident management teams. USDA Forest Service, Rocky Mountain Research Station, Research Note RMRS-RN-30. (Fort Collins, CO)

Canton-Thompson J, Gebert KM, Thompson B, Jones JG, Calkin DE, Donovan G (2008) External human factors in incident management team decision-making and their effect on large fire suppression expenditures. *Journal of Forestry* **106**(8), 416–424.

Cumming SG (2001) A parametric model of the fire-size distribution. *Canadian Journal of Forest Research* **31**, 1297–1303. doi:10.1139/CJFR-31-8-1297

Davison AC (2003) 'Statistical Models.' (Cambridge University Press: Cambridge, UK)

Efron B, Tibshirani R (1993) 'An Introduction to the Bootstrap.' (Chapman & Hall: New York)

Gebert KM, Schuster EG (1999) Predicting national fire suppression expenditures. In 'Proceedings of Symposium on Fire Economics, Planning, and Policy: Bottom Lines'. (Eds A González-Cabán, PN Omi) USDA Forest Service, Pacific Southwest Research Station, General Technical Report PSW-GTR-173, pp. 21–30. (Albany, CA)

Gebert KM, Calkin DE, Yoder J (2007) Estimating suppression expenditures for individual large wildland fires. *Western Journal of Applied Forestry* **22**(3), 188–196.

Gesch DB, Larson DS (1996) Techniques for development of global 1-km digital elevation models. In 'Pecora Thirteen, Human Interactions with the Environment–Perspectives from Space', 20–22 August 1996, Sioux Falls, SD. (CD-ROM) (American Society for Photogrammetry & Remote Sensing: Bethesda, MD)

Hamlet AF, Lettenmaier DP (2005) Production of temporally consistent gridded precipitation and temperature fields for the continental US. *Journal of Hydrometeorology* **6**(3), 330–336. doi:10.1175/JHM420.1

Hansen MC, DeFries RS, Townshend JRG, Sohlberg R (2000) Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* **21**, 1331–1364.

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

Hastings NAJ, Peacock JB (1975) 'Statistical Distributions.' (Butterworth & Co: London)

Holmes TP, Hugget RJ, Westerling AL (2008) Statistical Analysis of Large Wildfires. In 'Economics of Forest Disturbance: Wildfires, Storms, and Invasive Species', Forestry Sciences series, Vol. 79. (Eds TP Holmes, JP Prestemon, KL Abt) pp. 59–77. (Springer: Dordrecht, the Netherlands)

Liang X, Lettenmaier DP, Wood EF, Burges SJ (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research* **99**, 14 415–14 428. doi:10.1029/94JD00483

Maurer EP, Wood AW, Adam JC, Lettenmaier DP, Nijssen B (2002) A long-term hydrologically based data set of land surface fluxes and states for the conterminous United States. *Journal of Climate* **15**, 3237–3251. doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

Mitchell KE, Lohmann D, Houser PR, Wood EF, Schaake JC, Robock A, Cosgrove BA, Sheffield J, Duan Q, Luo L, Higgins RW, Pinker RT, Tarpley JD, Lettenmaier DP, Marshall CH, Entin JK, Pan M, Shi W, Koren V, Meng J, Ramsay BH, Bailey AA (2004) The multiinstitution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research* **109**, D07S90. doi:10.1029/2003JD003823

Monteith JL (1965) Evaporation and environment. In 'Symposium of the Society for Experimental Biology, The State and Movement of Water in Living Organisms', Vol. 19. (Ed. GE Fogg) pp. 205–234. (Academic Press, Inc.: New York)

Moritz MA (1997) Analyzing extreme disturbance events: fire in Los Padres National Forest. *Ecological Applications* **7**, 1252–1262. doi:10.1890/1051-0761(1997)007[1252:AEDEFI]2.0.CO;2

Penman HL (1948) Natural evaporation from open water, bare soil, and grass. *Proceedings of the Royal Society of London. Series A* **193**, 120–146.

Preisler HK, Westerling AL (2007) Statistical model for forecasting monthly large wildfire events in western United States. *Journal of Applied Meteorology and Climatology* **46**(7), 1020–1030. doi:10.1175/JAM2513.1

Preisler HK, Chen SC, Fujioka F, Benoit JW, Westerling AL (2008) Wildland fire probabilities estimated from weather model-deduced monthly mean fire danger indices. *International Journal of Wildland Fire* **17**, 305–316. doi:10.1071/WF06162

Prestemon JP, Abt KL, Gebert KM (2008) Suppression cost forecasts in advance of wildfire seasons. *Forest Science* **54**(4), 381–396.

R Development Core Team (2008) R: a language and environment for statistical computing. (R Foundation for Statistical Computing: Vienna, Austria) Available at http://www.R-project.org [Verified 5 April 2011]

Ramesh NI (2005) Semi-parametric analysis of extreme forest fires. *Forest Biometry, Modeling and Information Sciences* **1**, 1–10. Available at http://cms1.gre.ac.uk/conferences/iufro/fbmis/A/5_1_RameshNI_1.pdf [Verified 5 April 2011]

Schoenberg FP, Peng R, Woods J (2003) On the distribution of wildfire sizes. *Environmetrics* **14**, 583–592. doi:10.1002/env.605

United States Senate (1998) The Congressional Budget Process: An Explanation. Committee Print. S. Prt. 105–67. (Government Printing Office: Washington, DC)

Verdin KL, Greenlee SK (1996) Development of continental scale digital elevation models and extraction of hydrographic features. In 'Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling', 21–26 January 1996, Santa Fe, NM. (CD-ROM) (NCGIA Publications: Santa Barbara, CA)

Westerling AL, Bryant BP (2008) Climate change and wildfire in California. *Climatic Change* **87**(Suppl. 1), 231–249. doi:10.1007/S10584-007-9363-Z

Westerling AL, Gershunov A, Cayan DR, Barnett TP (2002) Long lead statistical forecasts of western US wildfire area burned. *International Journal of Wildland Fire* **11**, 257–266. doi:10.1071/WF02009

Westerling AL, Brown TJ, Gershunov A, Cayan DR, Dettinger MD (2003) Climate and wildfire in the western United States. *Bulletin of the American Meteorological Society* **84**(5), 595–604. doi:10.1175/BAMS-84-5-595

Westerling AL, Hidalgo HG, Cayan DR, Swetnam TW (2006) Warming and earlier spring increases western US forest wildfire activity. *Science* **313**, 940–943. doi:10.1126/SCIENCE.1128834

Westerling AL, Bryant BP, Preisler HK, Hidalgo HG, Das T (2009) Climate change, growth and California wildfire. Public Interest Energy Research CEC-500-2009-046-F. (California Energy Commission: Sacramento, CA) Available at http://www.energy.ca.gov/2009publications/CEC-500-2009-046/CEC-500-2009-046-F.PDF [Verified 19 April 2011]

Wood AW, Lettenmaier DP (2006) A testbed for new seasonal hydrologic forecasting approaches in the western US. *Bulletin of the American Meteorological Society* **87**(12), 1699–1712. doi:10.1175/BAMS-87-12-1699

Manuscript received 8 August 2009, accepted 13 August 2010