Spatially Explicit Forecasts of Large Wildland Fire Probability and Suppression Costs for California

Haiganoush K. Preisler¹,⁶, Anthony L. Westerling², Krista M. Gebert³ and Francisco Munoz-Arriola⁴, Thomas P. Holmes⁵

¹ USDA Forest Service, Pacific Southwest Research Station, Albany, CA, USA
² University of California, Merced, CA, USA
³ USDA Forest Service, Rocky Mountain Research Station, Missoula, MT, USA
⁴ University of Washington, Seattle, WA, USA
⁵ USDA Forest Service, Southern Research Station, Research Triangle Park, NC, USA
⁶ Corresponding author. Email: hpreisler@fs.fed.us

Running Head: Estimation of large wildland-fire probabilities and suppression costs
1 ABSTRACT

2 In the last decade increases in fire activity and suppression expenditures have caused
3 budgetary problems for the involved federal agencies. Spatial forecasts of upcoming fire
4 activity and costs have the potential to help reduce expenditures, and increase the
5 efficiency of suppression efforts, by enabling them to focus resources where they have
6 the greatest impact. In this paper we present statistical models for estimating spatially
7 explicit forecasts (one to-six months ahead) of expected numbers and costs of large fires
8 on a 1/8-degree grid with vegetation, topography and hydro-climate data used as
9 predictors. As an example, forecasts for California Federal and State lands are produced
10 for historic dates and compared with recorded fire occurrence and cost data. The results
11 seem promising in that the spatially-explicit forecasts of large fire probabilities seem to
12 match the actual occurrence of large fires, with the exception of years with wide-spread
13 lightning events, which remain elusive. Forecasts of suppression expenditures did seem
14 to differentiate between low and high cost fire years. Maps of forecasted levels of
15 expenditures provide managers with a spatial representation of where costly fires are
16 most likely to occur. Additionally, the statistical models provide scientists with a tool for
17 evaluating the skill of spatially explicit fire risk products.

18 Additional Keywords: Generalized Pareto distribution, Hydro-climate, Logistic
19 regression, Moisture Deficit, Fire simulations, Spline functions.

20 Brief summary: This paper presents a statistical model for producing maps of California
21 federal and state lands showing expected numbers and expected suppression costs of
large fires (>200ha) for the upcoming summer months using vegetation, topography and
hydro-climate data used as predictors.

1. Introduction

For land management agencies such as the U.S. 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 some 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
these factors (suppression is part of the their overall budget and they can’t go over their
budget), the FS and other federal land management agencies need estimates of future
suppression expenditures both during the budgetary planning process (which occurs two
to three years out) and during the current fiscal year in order to monitor spending.

Over the past decade, the need for such information has grown. With both the
magnitude and variability of expenditures increasing substantially over the past two
decades, budgets formulated two to three years in advance (based upon the 10-year
moving average of expenditures) are often much different than 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 due to continual borrowing due to one active fire year after another. To meet anti-deficiency regulations, in recent years the FS has either had to request emergency supplemental funding from Congress (which does not always occur) or transfer funds from other programs within the FS to pay for suppression.¹

Due 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 Prestemon et al 2008, Gebert 2007, Preisler and Westerling 2007, Westerling et al 2002, Bachelet et al 2000, Gebert and Schuster 1999).

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

¹ In 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.
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 for use in the Fall and Spring of the current fiscal year (which runs from October to September), which use climate, drought and trend variables to estimate suppression expenditures. These models have the advantage of being more scientifically based than the fire season forecasts but thus far are not updateable, with the forecasts being provided only twice per year. None of these projects, however, use spatially-explicit fire history, landsurface, 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 impact.

The work by Bachelet et al (2000) describes a spatially explicit dynamic vegetation model (MAPSS) that includes a fire module (MC1). Currently, MAPSS is being used to produce seasonal forecasts of fire risks (fire occurrence probability and expected area burned, but not suppression costs) at a somewhat coarser resolution than that used here (a 1/2-degree versus 1/8-degree grid). The fire occurrence probabilities from the MC1 modules, however, are consensus probabilities ‘defined as the percentage
of climate scenarios (out of a total of 5) that predict a fire in the timeframe mentioned\(^2\). 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 may be used to provide spatially-explicit forecasts of suppression cost for an upcoming fire season. As an intermediary step, climate data up to present are used to predict the number of large fire (≥ 200ha) on a 1/8 degree grid for the upcoming season (one to six months ahead). 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 (≥ 200ha) per 1/8 degree grid-cell per month, using vegetation, topography and climate variables (e.g., temperature, moisture deficit, etc.) up to present, as explanatory variables. We also estimate the distribution of fire sizes for all fires ≥ 200ha. 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.

2. Data and Statistical Methods

2.2 Spatial Domain

\(^2\) http://www.fs.fed.us/pnw/mdr/mapss/fireforecasts/index.shtml
The spatial domain for this analysis is a 1/8-degree lat/long grid (~12 kilometer [km] resolution) covering 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 U.S. Department of Agriculture’s Forest Service (USFS), and the U.S. Department of Interior’s National Park Service (NPS), the Bureau of Land Management (BLM), and the Bureau of Indian Affairs (BIA).

2.2 Fire History

A history of large (≥ 200 ha) wildfires for California for 1985–2003 was assembled from digital fire records obtained from CDF (obtained online at http://frap.cdf.ca.gov/), and FS, NPS, BLM, and BIA (see http://fam.nwcg.gov/fam-web/weatherfired/). The methods used in compiling a fire history from these data are described in Westerling et al. (2006, online supplement, and 2009), and Westerling and Bryant (2008). Westerling et al. (2002) describe the federal fire histories. The result is a 1/8-degree gridded monthly data set 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 (reported as polygon perimeters) 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 we did not have the means to apportion area burned by month.

While wildfires managed by the Fish and Wildlife Service (FWS), the Department of Defense (DOD), and the Bureau of Reclamation (BOR) were 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.

2.3 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 1/8-degree gridded vegetation layers that use the University of Maryland
vegetation classification scheme with fractional vegetation adjustment (UMDvf)
(Mitchell et al. 2004; Hansen et al. 2000). The UMDvf scheme has 14 coarse surface
categories derived from 1 km AVHRR satellite data collected from April 1992 to March
1993. We combined these to obtain the vegetation categories analyzed 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 Vegetation
Fraction (the fraction of each grid cell in the Forest, Woodland, Shrubland and Grassland
categories above). We did not 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.

2.4 Topography

Topographic data on a 1/8-degree grid were also obtained from LDAS. The LDAS
topographic layers are derived from the GTOPO30 Global 30 Arc Second (~1km)
Elevation Data Set (Mitchell et al. 2004; Gesch and Larson 1996; Verdin and Greenlee
1996). We tested mean and standard deviation of elevation, slope, and aspect as explanatory variables in our statistical model.

2.5 Hydroclimate

We use 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). 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 NCDC COOP 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 use 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 two years, for the current water year through March, and for each month afterwards through the fire season. Moisture deficit (D) was calculated from PET and AET (D = PET - AET). Because Potential
Evapotranspiration (PET) was not easily extracted from the version of the VIC model used here, we used the Penman-Monteith equation to estimate PET directly (Penman 1948; Monteith 1965).

2.6 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, suppression resource availability, etc. Gridded population estimates were obtained from the Center for International Earth Science Information Network’s (CIESIN) Socioeconomic Data and Applications Center at Columbia University. We used the Gridded Population of the World Version 3 at 2.5 arc-minutes resolution, adjusted to match UN population totals. We aggregated these data to produce population counts on the LDAS 1/8-degree grid.

2.7 Estimating suppression cost per fire

We obtained fire suppression cost data per fire for a sample of fires of sizes greater than 40 ha (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 forecasted 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; percent forest; fraction vegetation and population. The variables were evaluated for the 1/8 degree grid cells containing the fire.

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

\[ y = \beta_{ok} + \beta_{1k} \cdot esd + \beta_{2} \cdot vegf + \beta_{3} \log(hec) + \epsilon \]  

[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 fraction vegetation. The square root of cost was used because the residuals from this fit were best approximated by the normal distribution.

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 (Figure 1). In particular, for fires greater than 8500ha, the standard deviation of elevation seemed to have a larger effect on area burned than for fires \(< 8500ha.\)

Consequently, in our final model in equation [1] different slopes were assigned to two fire size classes. The subscript \( k \) in equation [1] stands for one of two fire size classes:

Class I, fires between [200-8500) ha and Class II, \( \geq 8500\)ha. 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 fire-fighters. Fraction vegetation (\( vegf \)) describes how much vegetated area there is in the 1/8 degree grid that can carry a fire in that location. In our sample the correlation between \( vegf \) and \( esd \) was -0.2.
2.8 Estimating probability of large fire occurrence

Using land surface (topography and vegetation), population, and hydroclimate, expected numbers of large fires (≥ 200ha) 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 ≥ 200ha at location \( i \) 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

\[
\theta_{ij} = \beta_j + g(\text{long}_i, \text{lat}_i) + \sum_m g_m(X_{mij})
\]  

The spatial covariate \((\text{long}_i, \text{lat}_i)\) is the longitude, latitude pair of each 1/8 degree grid cell in California State and Federal lands; the covariate \(X_{mij}\) is the \(m^{th}\) explanatory variable from the list of variables, including topography, vegetation and lagged climate variables for location \( i \) and on date \( j \). The parameters \(\beta_j\), one for each month, and the nonparametric 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 200ha’.

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 generalized additive modules in the R statistical package (R Development 2008) 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 one-month to six-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 one-month-ahead forecasts, while that for September is a six-month 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 do not change 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.

Forecasted 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 1/8 degree grid-cells and for a given year, is significantly greater than the historic odds, then that cell is designated as being higher than normal. 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 Jackknife procedure (Efron and Tibshirani, 1993) where 21 different sets of coefficient estimates were evaluated,
each using historic data from all the years but one, then calculating the jackknife standard
errors of the 21 values.

In the next section we fit a Generalized 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 200ha.

2.9 Estimating conditional distribution of large fire size

Histograms of observed large fire sizes are often best characterized by heavy-
tailed distributions, such as the log-normal or the Pareto distributions. These distributions
have often been used successfully to characterize catastrophic events such as earthquakes
and others have demonstrated that the generalized Pareto distribution (GPD) is a useful
model for characterizing 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 modeling fire
sizes can be found in Holmes et al. (2008). In our case only fires greater than 200ha are
included in the data. The GPD has two parameters (scale and shape) that are estimated
from the data and a threshold level, which will be set to 200ha. The scale and shape
parameters for our data were estimated within the R statistical package using modules
from the ismev library (R Development 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 a fire has already exceeded 200ha.
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 5,000 observations from the generalized Pareto distribution with a scale and shape parameter set at the values estimated from the data and then comparing the quantiles of the simulated data to those of the observed fire sizes. Simulated values from the generalized Pareto distribution were generated by the formula, \( r = \log(200) + \sigma \cdot \left( U^{(-\alpha)} - 1 \right) / \alpha \), where \( U \) is a random variable from a Uniform(0,1] distribution (Hastings and Peacock 1975, Davison, 2003) and where \( \alpha, \sigma \), are values of the shape and scale parameters estimated from the observed large fire sizes.

2.10 Forecasting spatially explicit fire suppression costs

In section 2.7 above, we developed a regression model for estimating suppression costs for a given fire, given fire size and some site characteristics. Since 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 forecasted probability of a large fire (≥ 200ha) as given by equation [2]. Next, for all pixels and each month where the response was one (i.e., a large fire occurrence was forecasted) we generated a realization from the GPD using the method described in section 2.9. The projected cost of fire expenditures at each pixel was next estimated by equation [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 (March- September).

3. Results

3.1 Suppression costs per fire
We observed a significant relationship between the standard deviation of elevation and
suppression cost per fire. This effect appeared to be mostly due to the fires in class II
(Figure 1). 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. 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 equation [1] was 70%. The
comparison of observed costs to predicted costs (Figure 2) demonstrates how, even with a
multiple correlation of 70%, there still remains a large degree of unexplained variability
in costs per fire. That is why correlation alone is not a sufficient statistic when describing
the skill of a model.

3.2 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 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 mid-point of the class (Figure 3). For a good fit the points will be scattered close to the 45 degrees line within the 95% confidence bounds. The forecasts done in mid season (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 of the model spatially we producing maps of observed locations of fires for a given month and compared them with a set of maps with simulated fire locations (Figure 4-6). The simulations were done using the estimated probabilities to generate a response (fire/no fire of size ≥ 200ha) for each 1/8th degree grid cell and for a given month. The year 1994 (Figure 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 realization from our estimated distribution of large fires for that month and year. The year 1995 (Figure 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 (Figure 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 nine simulations. In particular, there is a cluster of fires that occurred in the 
northwestern region of California. Our model does not take into account lighting events. 
The cluster of large fires in northern California during the summer of 1987 is due to a 
larger than average numbers of lightning events - hence more ignitions than the historic 
average - resulting in greater than expected large fires. While our model takes into 
account ‘clustering’ of fires due to similarities in the topography or vegetation of near-by 
points (by including a spatial term in the logistic model) it does not take into account 
causes of clustering of ignitions due to lighting events.

One of the products of our probability modeling is forecasted odds maps relative 
to historic averages. The relative odds maps for August 1994 and August 1995 (Figure 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 two 
years. Similar maps for other historic years (together with the latest forecasts) are posted 
on the web at https://wildfire.ucmerced.edu/forecast.

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

3.3 Spatially explicit fire suppression costs
As an example of the final output of our modeling procedure we produced a map of forecasted costs for the 1994 fire season and compared them with estimated costs (Figure 9). The estimated costs came from the National Interagency Fire Management Integrated Database (NIFMID), which includes a field containing estimated suppression expenditures on that fire by the reporting agency. Because the NIFMID cost data do not include CDF fires, we multiplied the forecasted costs per pixel by the fraction of the 1/8-degree pixel within California with federal responsibility of fire suppression.

The general spatial pattern of higher forecasted costs for the Los Angeles area, Southern California, and the northern most region of California seem to match well with the pattern of large fire occurrences and costs (Figure 9). This general spatial pattern did not change for the 1995 forecast (figure not included), although the overall level of expenditure was lower due to the smaller number of fires in 1995 (Figure 5). We also produced a map of the standard deviations of forecasted cost (Supplemental Figure 1). Note the forecasted costs are per given 1/8-degree pixel, and they reflect both 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 forecasted 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 forecasted cost will be higher accordingly.

4. 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 impact.

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 wide-spread lightning events which remain elusive to prediction efforts. Suppression cost forecasts, as previous researchers have found (Prestemon et al. 2008, Gebert 2007), are difficult to predict, and the models tested in this effort also left a large degree of unexplained variability. This is not unexpected, however, since 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 and 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 (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 NOAA’s Regional Integrated Science
and Assessment Program for California at Scripps Institution of Oceanography. We
would like to thank two reviewers for their thorough review and helpful comments on the
first draft of the paper.
References


   UK.
   NY, NY, USA.
   individual large wildland fires. Western Journal of Applied Forestry 22(3):188- 
   196.
   expenditures. P. 21-30 in Proc. of Symposium on fire economics, planning, and 
   precipitation and temperature fields for the continental U.S. Journal of 
7 Hastie, T.J., Tibshirani, R., Friedman, J. (2001). The Elements of Statistical Learning:
   London. 130pp.
   Wildfires," Chapter 4 Economics of Forest Disturbance: Wildfires, Storms, and


1 R Development Core Team (2008). R: A language and environment for statistical
2 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-
4 Schuster, E.G., D.A. Cleaves, E.F. Bell. (1997). Analysis of USDA Forest Service fire-
7 Environmetrics 14: 583-592.
9 Committee Print. S. Prt. 105-67. 58 pgs.
11 Change, Growth and California Wildfire. Public Interest Energy Research,
12 California Energy Commission. Sacramento, CA. in press (draft available online
13 at California Climate Portal).
17 Earlier Spring Increases Western U.S. Forest Wildfire Activity. Science 313: 940–
18 943. DOI:10.1126/science.1128834. Online supplement.
20 Climate and Wildfire in the Western United States. Bulletin of the American
    Statistical Forecasts of Western U.S. Wildfire Area Burned. *International Journal
    of Wildland Fire* 11(3,4): 257–266. DOI: 10.1071/WF02009.

    forecasting approaches in the western U.S. *Bulletin of the American
    Meteorological Society*, 87(12), 1699-1712, doi:10.1175/BAMS-87-12-1699.
Table 1: Models and significant variables used for predicting one-month to six-month ahead spatially explicit probabilities of large fires in California.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables included</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Historic</td>
<td>Spatial, Month, Elevation, %Forest</td>
</tr>
<tr>
<td>2) 1-Month ahead</td>
<td>Spatial, Month, Elevation, %Forest, pCM, CM, pMD, pT</td>
</tr>
<tr>
<td>3) End of March</td>
<td>Spatial, Month, Elevation, %Forest, pCM, CM, T3</td>
</tr>
<tr>
<td>4) End of April</td>
<td>Spatial, Month, Elevation, %Forest, pCM, CM, T4</td>
</tr>
<tr>
<td>5) End of May</td>
<td>Spatial, Month, Elevation, %Forest, pCM, CM, T5</td>
</tr>
<tr>
<td>6) End of June</td>
<td>Spatial, Month, Elevation, %Forest, pCM, CM, T6</td>
</tr>
<tr>
<td>7) End of July</td>
<td>Spatial, Elevation, MD6</td>
</tr>
<tr>
<td>8) End of August</td>
<td>1-month ahead model (since there is only one month, September, being forecasted)</td>
</tr>
</tbody>
</table>

Spatial = two dimensional smooth function of latitude and longitude
Month = categorical variable for each month between April – September.
Elevation = average elevation (m) over the 1/8 degree pixel.
%Forest = percent forested land in the 1/8 degree 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.
Figure Legends

Figure 1: Suppression costs per fire versus elevation standard deviation and Fraction vegetation, for two fire size classes, Class I: (200-8500) ha and Class II: >8500 ha.

Figure 2: Observed versus predicted costs per fire based on fire size, elevation STD and vegetation (% forest). The dashed lines are the approximate 95% confidence bounds plotted about the 45º line (solid line).

Figure 3: Observed versus predicted probability of ≥200 ha fires for 4 different forecasts. The dashed lines are approximate 95% confidence bounds plotted about the 45º line (solid line).

Figure 4: Maps of forecasted large fire probabilities (%) per 1/8 degree pixel, with potential locations of fires (•) from 3 simulations (3 left panels) compared with observed locations of fires (•) for August 1994 (right panel). N is the total number of large fires in each panel. 1994 was a relatively high fire year.

Figure 5: Maps of forecasted large fire probabilities (%) per 1/8 degree pixel, with potential locations of fires (•) from 3 simulations (3 left panels) compared with observed locations of fires (•) for August 1995 (right panel). N is the total number of large fires in each panel. 1995 was a relatively low fire year.

Figure 6: Maps of forecasted large fire probabilities (%) per 1/8 degree pixel, with potential locations of fires (•) from 3 simulations (3 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 lighting caused fires in Northern California in 1987.
Figure 7: Maps of forecasted odds of large fire occurrences in August 1994 (left) and 1995 (right) relative to historic odds developed with data up to end of March.

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

Figure 9: (a) Map of forecasted average expenditure costs (background colors) per pixel – in $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 centered at each fire giving the individual fire size (width of rectangle) and suppression cost (height or rectangle).

Supplemental Figure 1: Map of standard errors of forecasted costs generated from 1000 simulations of the model.