SFB 649 Discussion Paper 2009-032

Weather-based estimation of wildfire risk

Joanne Ho*
Martin Odening**

* University of Washington, USA
** Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Weather-based estimation of wildfire risk

Joanne Ho[1]
College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100 and Department of Agricultural Economics, Humboldt-Universität zu Berlin, D-10099 Berlin, Germany

Martin Odening
Department of Agricultural Economics and Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin, Philippstraße 13, D-10115 Berlin, Germany

Abstract

Catastrophic wildfires in California have become more frequent in past decades, while insured losses per event have been rising substantially. On average, California ranks the highest among states in the U.S. in the number of fires as well as the number of acres burned each year. The study of catastrophic wildfire models plays an important role in the prevention and mitigation of such disasters. Accurate forecasts of potential large fires assist fire managers in preparing resources and strategic planning for fire suppression. Furthermore, fire forecasting can *a priori* inform insurers on potential financial losses due to large fires. This paper describes a probabilistic model for predicting wildland fire risks using the two-stage Heckman procedure. Using 37 years of spatial and temporal information on weather and fire records in Southern California, this model measures the probability of a fire occurring and estimates the expected size of the fire on a given day and location, offering a technique to predict and forecast wildfire occurrences based on weather information that is readily available at low cost.

Keywords: biased sampling, forest fires, fire occurrence probabilities, fire weather

JEL Classification: C24; C25; Q23; Q54

---

[1] Address of correspondence to Joanne Ho: *University of Washington, College of Forest Resources, Box 352100, Seattle, WA, 98195-2100, Email: jjho@u.washington.edu* This work was supported by Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk," the Deutscher Akademischer Austausch Dienst, and the National Science Foundation through the Integrative Graduate Education and Research Traineeship. We wish to thank members of the SFB and the U.S. Forest Service Pacific Southwest Research Station for helpful comments.

1 Introduction

Catastrophic wildfires in California have become more frequent in past decades, while insured losses per event have been rising substantially. On average, California ranks the highest among states in the U.S. in the number of fires as well as the number of acres burned each year. Lending to economic and popluation growth in the wildland-urban interface, the financial risk of wildfires have increased drastically throughout the years. Annual insured losses from catastrophic fires in California increased 1725 percent in real terms from 1964 to 2007. In Southern California alone, the number of housing units increased from roughly 750,000 homes in 1990 to over 850,000 in 2005, while the number of acres burned were consistently around 31 million acres each year. The study of catastrophic wildfire models plays an important role in the prevention and mitigation of such disasters. The ability to accurately predict the occurrence and severity of a forest fire is of major importance to land managers and fire managers. In addition to the accuracy of the predictions, the ability to forecast fire events in advance is also valuable to fire managers, buying them more time to organize personnel and the necessary equipment. Fire managers rely on these predictions to allocate resources for fire suppression operations.

The physical components of fire are based on three factors: fuel, heat and oxygen. Generally, heat determines the ability of fuel to ignite and stay ignited. Any combustible material can be considered fuel, while oxygen allows the fire to burn. For wildfires in forested areas, fire intensity and severity are based on the following components: fuel load, topography, and weather conditions (Agee, 1993). For this reason, wildfires are often closely related to the weather of the present and previous days, while the weather of forthcoming days may dictate the duration of the fire. Hotter temperatures often leads to lower moisture content in the air and soil, which leads to a higher combustibility of fuels. High wind speeds provide the oxygen for the flames. Wildfires are usually ignited by either lightning or human action. Trees, branches, needles, or any other combustible materials such as homes and other structures, are considered fuel. Topography as well as wind speed and direction play a role in the spread of fire, since wind provides oxygen and assists the fire in extending to neighboring unburned areas. For example, wind can carry embers from one location to another, where they serve as a source of ignition in an unburned area.

There are a number of tools to help fire managers determine the severity of fire and make decisions with regard to resource allocation and strategic fire management. First of all, so called fire indices have been defined for that purpose. Some indices are simply based on weather information, while others involve additional information on the forest such as vegetative cover and topography. These indices are often based on modeling of the physical components of fire. One example is the U.S. National Fire danger Rating System (NFDRS), which uses current and historic weather and fuel information to create indices, which are then expressed in maps. The Canadian Wildland Fire Information System generates maps describing the fuel conditions and the likelihood of ignition. Relating specifically to weather conditions, the Fosberg fire weather index, which uses temperature, relative humidity and wind speed to evaluate the potential influence of weather on a wildland fire. The Keetch-Byram drought index utilizes precipitation to estimate the dryness of the soil and duff layers. To date, the accuracy of these tools has not been thoroughly tested, leaving no consensus on the overall superiority of any one tool over the others.

The first probability-based fire risk model began with Bratten (1982). Bratten constructs a model to predict the probability for initial attack of a forest fire and probability for its escape. The initial attack module calculates the probability of the fire escaping , which is defined as „any result from the initial attack or fire behavior calculations in which fire size or intensity exceed values beyond which the initial attack model is invalid." Secondly, the large fire module estimates probabilities for the final fire size and the associated cost of damage.

Preisler et al. (2004), Preisler and Benoit (2004), and Brillinger (2006) present models for estimating probabilities of fire on a particular day on a 1 km grid on Federal land. Preisler et al. (2004) defines and estimates the probability of a small fire occuring and the probability of a large fire occuring, defining large fire as fires equal to or larger than 40.5 hectares. By calculating the joint probability of these two, they find the probability of a small fire becoming large. Preisler et al. (2004) use a logit function to estimate the log odds of each of the probabilities mentioned above, using the following as explanatory variables: fuel category, location, day in year, elevation, Burning Index, Fire Potential Index, Keetch-Byram Drough Index, Thousand-hour fuel moisture, wind speed and direction, relative humidity, and dry bulb temperature.

Recently, Brillinger (2006) estimates the probability of fire using a penalized quasi-likelihood function to compute three logit models, differing in their explanatory variables. The first model includes only location and day of year, while the second model includes the year in addition to variables in the first model. The third model includes a vector of independent normals in addition to the first model. The authors use 1 km by 1 km spatial pixels for each day to construct space-time cells (voxels), where voxels with fire occurrence were included, while voxels with no fire were sampled proportional to the total number of fires on that day. For each day, the authors randomly sample locations within California federal lands that did not experience fire on the given day. The authors conclude that elevation is an unnecessary explanatory variable in such a space-time model, and that random effects are necessary in order to estimate future probabilities. Oddly enough, the authors only select data from locations where fire have occurred, and only selection a sample of locations where fire have not occurred. Although the sample selection is proportional to the total number of fires on that day, this sampling practice renders a spatial bias to the estimations.

This paper offers an explorative statistics approach to predict wildfires. A probability model based on the two-stage Heckman procedure is employed for predicting wildland fire risks. Using spatial and time-series information on weather and fire history, the model measures the probability of a fire occurring on a given day and estimates the expected size of the fire. The model is applied to 37 years of data on fire occurrence and daily historical weather records in southern California.

2 Data

Daily fire occurrence records from 1970 to 2007 on U.S. federal wildland are provided by the Desert Research Institute (DRI) (Brown et al. 2002). These records are collected from the USDA Forest Service (USFS), and the Department of Interior (DOI) agencies Bureau of Indian Affairs (BIA), Bureau of Land Management (BLM), U.S. Fish and Wildlife Service (FWS) and National Park Service (NPS). The DRI data includes the following: Discovery date of event, latitude and longitude coordinates, number of acres burned, the cause of ignition, the agency reporting the event, and the state in which the event occurred. Between 1970 and 2000, 43.4 percent of fires were reported as natural caused fires, while 56.6 percent were reported as human caused. Even though these federal fire records are marked by frequent errors, there lacks alternative forms of centralized collection of fire information, while most other sources are advised not to be used for statistical purposes. Cross-referencing fire observations across agency records, Brown et al. identifies and corrects for faulty records, which are mainly due to incorrect or the lack of date and/or location. Depending on the agency, between 50 and 83 percent of the records are found to be of acceptable accuracy. Due to technological improvements to data management, fire records are likely to be more accurate through time. In the present study, a binary random variable “ignition” is created, where ignition is equal to 1 if at least one fire appears on the given day, and 0 otherwise. Out of 138,580 observations, 80.98 percent do not observe fire, while 19.02 percent show observations of at least one fire.

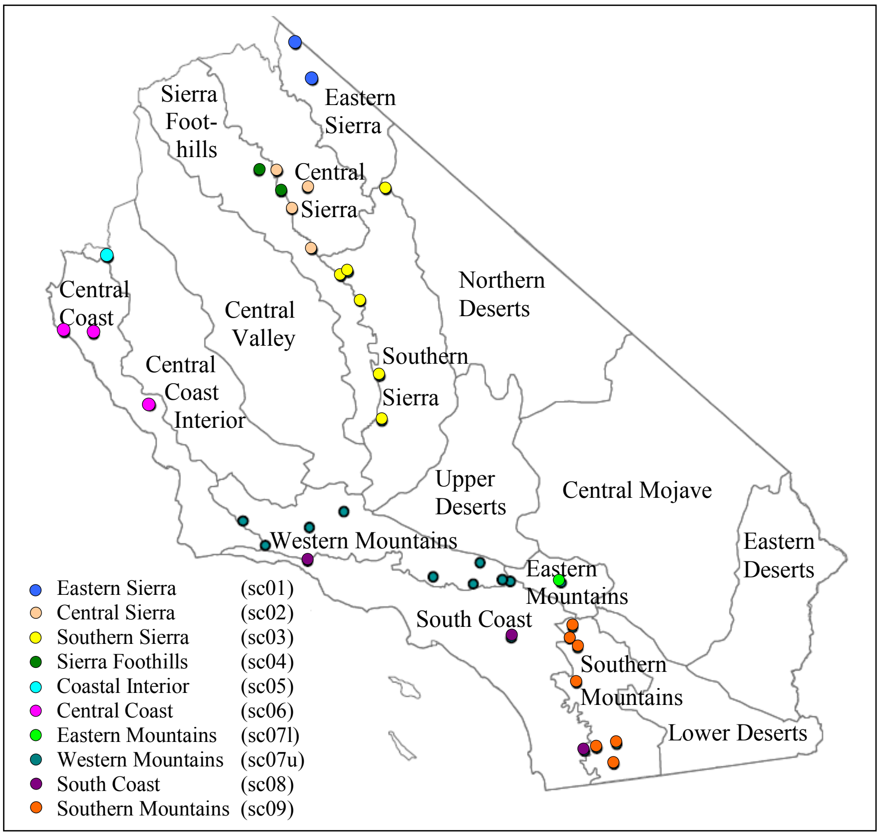
Table 1: Frequency of observation by fire size classes

Size class	A	B	C	D	E	F	G	
Acres	0-0.24	0.25-9.99	10-99.99	100-299.99	300-999.99	1000-4999.99	≥5000	Total
Frequency	125,763	8,798	2,153	688	538	396	244	138,580
Percent	90.75	6.35	1.55	0.50	0.39	0.29	0.18	100.00

Daily weather data are recorded by Remote Automatic Weather Stations (RAWS). Because weather stations are not established simultaneously and require periodic maintenance, weather records do not have the same starting year, and are subject to temporary as well as permanent discontinuation. For this reason, only the weather stations with the longest and most continuous set of records between 1970 and 2007 are selected for this study. Furthermore, weather stations are not spatially distributed uniformly, as they are only established on federal forested lands, where weather records are likely to be needed for research purposes. Weather variables are constructed using the data collected from a total of 48 RAWS stations.

In order to minimize spatial deviation between the fire observations and weather records, the data are spatially nested into ten groups. The boundaries for these groups are defined by Predictive Service Areas (PSAs) provided by the Southern California Geographic Area Coordination Center (OSCC). Based on the Fire Danger Rating system, PSAs are regions of Southern California distinguished by their geographical, topographical, ecological, and climatic attributes that lead to similar fire behaviors. Shapefiles of PSAs are directly obtained from OSCC. Figure 1 shows the location of all RAWS stations and the boundaries of PSAs.

Figure 1: RAWS locations in Southern California



The following weather data acquired directly from RAWS records are included in this study: weather station number, date of record, dry bulb temperature *(temp)* measured in degrees Fahrenheit, relative humidity *(rh)* measured in percent, wind direction from the north *(wind_n)* and northeast *(wind_ne)* as dummy variables indicating the presence of Santa Ana winds, wind speed *(windspd)* measured in miles per hour, precipitation amount *(pptamt)* measured in inches, and precipitation duration *(pptdur)* measured in hours. Deriving from the temperature and relative humidity records, vapour pressure deficit *(vpd)* is specified as

$$vpd = E_s - E_a \quad (1)$$

where E_s represents the saturated vapor pressure (how much water vapor the air can hold at a given temperature) and E_a represents the actual vapor pressure (how much water vapor there is in at the air).

With the exception of *temp,* most weather variables exhibit strong skewness to the left with extremely long right tails. While *temp* is mostly distributed normally, it exhibits slight skewness to the right. The summary statistics for each of the variables are provided in Table 2. Table 3 presents the correlation between weather variables.

Table 2: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
temp	10677	76.50579	12.24744	17	112.5
rh	10677	38.65022	18.67222	1	100
vpd	10677	21.9366	12.56183	0	80.3
pptamt	10652	5.440257	71.10028	0	4950
pptdur	10677	0.6938513	2.710366	0	24
windspd	10677	6.788143	3.470349	0	50

Table 3: Correlation between Weather Variables

Variable	temp	rh	vpd	pptamt	pptdur	windspd
temp	1					
rh	-0.6384	1				
vpd	0.9188	-0.7598	1			
pptamt	-0.1289	0.1827	-0.123	1		
pptdur	-0.3932	0.5262	-0.3683	0.3337	1	
windspd	0.0718	-0.1749	0.1078	-0.0351	-0.0964	1

3 Statistical model

The model we suggest answers two questions: First, given the weather conditions for any given day, what is the probability that fire will occur? Second, if a fire has occurred, how big can we expect the fire to be? In similar spirit as Preisler et al. (2004) and Preisler and Benoit (2004), this paper uses probabilistic estimations to address different risks associated with small verses large fires. What differs in this paper is the statistical method employed. Rather than using the logit model and the Poisson-binomial distribution as in Preisler et al. (2004) and Preisler and Benoit (2004), this paper demonstrates probabilistic estimations using the Heckman two-stage procedure.

Since over 80 percent of observations do not exhibit fire occurrence while weather conditions are recorded every day, this high volume censorship of the dependent variable may lead to biased estimates of the impact of weather variables. In order to address this bias, this paper employs the Heckman two-stage procedure to differentiate between non-fire observations and fire observations. The first stage is the selection model, which answers the first objective question above on the likelihood of fire. The second stage is a linear regression model, which answers the second objective question on the expected size of fire.

Consider a response variable Z_{it} where

$$Z_{it} = \begin{cases} 1 & \text{if there is a fire at location } i \text{ on day } t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The selection mechanism is as follows:

$$Z_{it}^{*} = \gamma' \mathbf{w}_i + u_{it} \tag{3}$$

where

$$Z_{it} = \begin{cases} 1 & \text{if} \quad Z_{it}^{*} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\mathbf{w}_i$ is a vector of observed weather variables, γ is the estimated parameter for $\mathbf{w}_i$, and u is a vector of random errors. Z_{it}^{*} is the unobserved, selection variable. Using the probit model standard in the Heckman selection model,

$$\begin{aligned} \text{Prob}(Z_{it} = 1) &= \Phi(\gamma' \mathbf{w}_i) \\ \text{Prob}(Z_{it} = 0) &= 1 - \Phi(\gamma' \mathbf{w}_i) \end{aligned} \tag{5}$$

where $\Phi(\cdot)$ is the cumulative distribution function.

To estimate the expected size of fire, consider the estimated size of fire to be

$$\begin{aligned} Y_{it} \quad &= \beta' \mathbf{x}_{it} + \varepsilon_{it}, \text{observed only if } Z_{it} = 1 \\ &(u_{it}, \varepsilon_{it}) \text{ bivariate normal with } E\begin{pmatrix} u_{it} \\ \varepsilon_{it} \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{and } E\begin{pmatrix} u_{it} & \varepsilon_{it} \end{pmatrix}\begin{pmatrix} u_{it} \\ \varepsilon_{it} \end{pmatrix} = \begin{bmatrix} \sigma_u^2 & \rho_{u\varepsilon} \\ \rho_{u\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix} \end{aligned} \tag{6}$$

The expected size of fire, adjusted for selection bias, is therefore given by

$$E[y_{it} \mid z_{it} = 1] = \beta' \mathbf{x}_{it} + \rho \sigma_\varepsilon \lambda(\gamma' \mathbf{w}) \tag{7}$$

where λ is the inverse mills ratio, computed by

$$\hat{\lambda} = \frac{\phi(\gamma' \mathbf{w}_{it})}{\Phi(\gamma' \mathbf{w}_{it})} \tag{8}$$

All weather variables *(temp, rh, vpd, pptamt, pptdur, windspd)* are included as regressor candidates for the selection (probit) model, along with the squared of these weather terms and their interaction terms.

As no comprehensive theoretical model is available that supports the specification of the empirical model we pursue an exploratory approach. For each location, roughly 220 thousand regressions for every combination of inclusion and exclusion of weather variables are run. Using the Bayesian information criteria (BIC) as an indicator for goodness of fit, the best fit regression is then selected as the final model specification. In order to reduce the level of computational intensity, only models with coefficients that have logical interpretations are included, while models with illogical coefficient interpretations are discarded. Namely, positive coefficients are expected for *temp* and *windspd* or their squared terms, while negative coefficients for *rh*, *vpd*, *pptamt*, and *pptdur* or their squared terms are expected. The BIC is constructed by

$$BIC = -2ln(\ell)] + [ln(N)]k \tag{9}$$

where $ln(\ell)$ denotes the log likelihood, k denotes the degrees of freedom, and N denotes the number of observations. The BIC is used as a measure of fit and complexity of the model. The lower the score, the better the fit. While the log likelihood provides an indicator for goodness of fit, the second term of equation (9) penalizes the BIC score for the inclusion of terms that only negligibly improve the goodness of fit. Although the Akaike information criteria (AIC) performs a similar function, the BIC measure is chosen in favor of the AIC because it assigns a stronger penalty for including more terms.

The exclusion of regressions yielding coefficients with unexpected signs (negative or positive) is based on theoretical intuition. For example, a positive coefficient for temperature is expected, because higher temperatures should most often lead to a higher potential for fire. An apparent illustration of this intuition is that more fires are expected in the summer compared to winter. Measurements of moisture, such as relative humidity and precipitation are expected to have negative coefficients, because damp fuels are more difficult to ignite and stay ignited. Wind speed is expected to have a positive coefficient, because wind assists the spread of fire once a fire has occurred. Furthermore, hot, dry, and fast winds such as the Santa Ana winds tend to dry out the fuels quicker than non-windy conditions. Hence, fast winds in combination with hot weather before ignition preconditions the fuels to become more easily igniteable.

4 Results

4.1 Selection Model

Model specifications of probit estimations using time-series modelling for each of the ten locations differ in the inclusion of variables, as well as the significance of variables. Only temperature is consistently significant at the 5 percent level in all of the locations, while relative humidity is significant in six of ten locations. The interaction term $temp \times rh$ is significant at the 5 percent level in seven of ten locations, while $temp \times windspd$ and $rh \times windspd$ are significant in half of the locations. The rest of the variables are either never significant or are significant in less than half of the locations. Time-series probit parameter estimates for each location are listed in Table 4.

The McFadden's Pseudo-R^2 measures the percent improvement of fit of the model compared to the null model. Table 4 shows Pseudo-R^2 ranging from 0.0569 in the Coastal Interior (sc05) to 0.2262 in Southern Sierra (sc03). This indicates that using weather variables to estimate the probability of fire occurrence improves the skill of prediction by 5 percent for the Coastal Interior over the skill of a probit model with a constant only, while the prediction in the Southern Sierra is improved by 22 percent. The average Pseudo-R^2 across all estimations is 0.13063.

The model is now used for an *ex post* prediction of fire probabilities in each PSA. Predicted probabilities of ignition for each day of the year from time-series estimates are averaged across the sample time period (1970-2007) and are presented in Figure 4. Obviously, there are higher chances of ignition from May until end of October. Furthermore, it can been seen that the locations are distinctly different from one another. More specifically, one can compare the relationships between mountainous and coastal regions, as well as across the northern and southern regions. In the north, there is generally a positive relationship between elevation and peak ignition probability. Mountainous regions have higher probabilities of ignition than coastal regions throughout the year. With the Central Sierras (sc02) exhibiting the highest peak probability of ignition, Eastern Sierra (sc01) exhibit similar trends as the Foothills (sc04). Similarly, the Western Mountains (sc07u) also exhibit much higher probabilities of ignition compared to the Eastern Mountains (sc07l). Contrary to the elevation-ignition probability relationship in the north, ignition probabilities in the South Coast (sc08) region generally exceeds that in the Southern Mountains (sc09) year-round. Predictions of ignition probabilities in the Southern Coast are comparable to those in the Eastern Mountains.

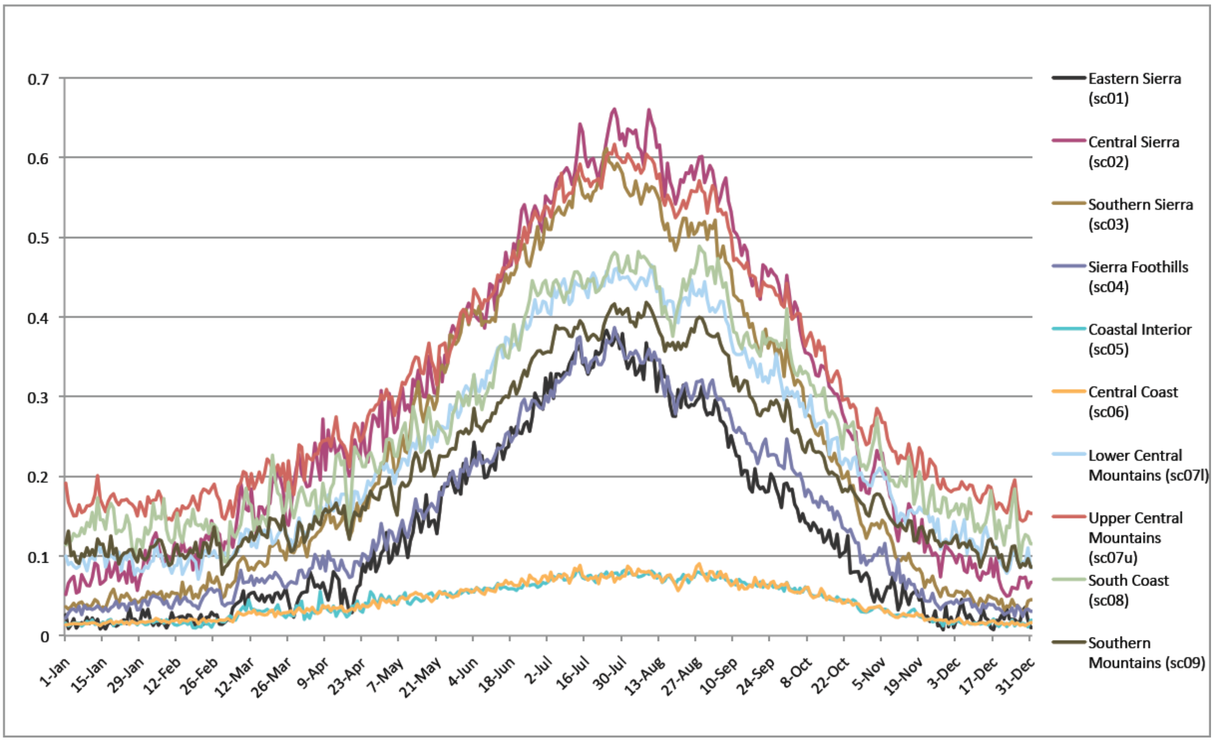
Table 4: Estimated coefficients of regional selection models

	Region									
Independent Variable	(sc01)	(sc02)	(sc03)	(sc04)	(sc05)	(sc06)	(sc07l)	(sc07u)	(sc08)	(sc09)
constant	-3.475403**	-3.750916**	-6.952771**	-7.320773**	-5.367578**	-4.060536**	-2.035498**	-5.112145**	-3.966031**	-6.10278**
	(.1393287)	(.1036097)	(.2635167)	(.493342)	(.4047125)	(.1865491)	(.1170619)	(.2034703)	(.1185068)	(.3094253)
pptamt × *pptdur*			.0001024	-.0000568		-.0001076	-.0002059*	-.0000934		
			(.0001324)	(.0000543)		(.0000932)	(.0001017)	(.0000512)		
pptamt^2		-.0000306**								-.0000231*
		(.0000113)								(.00001)
pptamt × *windspd*									-.0001006	
									(.0007568)	
pptdur^2	-.0085478**	-.0022124**	-.0102233**						-.0014872	
	(.0023606)	(.0006066)	(.0023206)						(.0018313)	
pptdur × *windspd*									-.0023291	
									(.0048052)	
rh	.0446853**		-.0206827**				-.0399217**	-.0195486**		
	(.0030402)		(.003236)				(.0036752)	(.002276)		
rh × *pptamt*					-.0002029				.000048	
					(.0001098)				(.0001143)	
rh × *pptdur*									.0000323	
									(.0007057)	
rh^2	-.0003865**						.000225**			.0001378**
	(.0000713)						(.0000362)			(.0000325)
rh × *windspd*				-.0028253**	.0009861**	.0007339**	-.0007009**		.0001253	-.0020333**
				(.0005363)	(.0002352)	(.0001878)	(.0001686)		(.0001312)	(.0002235)
temp		.0368823**	.1206258**	.100309**	.0557762**	.0294839**		.0943783**	.039916**	.087498**
		(.0017148)	(.0075095)	(.0087774)	(.0065413)	(.0021863)		(.0056548)	(.0014448)	(.0054885)
temp × *pptamt*	9.71e-06	.0001279**							-.000131	
	(6.60e-06)	(.000032)							(.0001552)	
temp × *pptdur*	.0023483**		.0021307**						.0001577	
	(.0004142)		(.0003502)						(.0008552)	
temp × *rh*							.0008871**			
							(.00005)			
temp^2	-.0011698**		.0014472**	.0012112**				.0014253**		.000814**
	(.000142)		(.0001804)	(.0001836)				(.0001516)		(.0001142)
temp × *windspd*		.0017754**	.0008246**	-.0023406**	-.0008044**		-.0008241**			-.0018238**
		(.0001147)	(.0000893)	(.0007248)	(.0001721)		(.0002433)			(.0003657)
vpd	.0740557**		-.1192159**	-.0780073**	-.0226451**		.0280493**	-.0918048**		-.0617207**
	(.0034086)		(.0117054)	(.0098166)	(.0054166)		(.0037569)	(.0091699)		(.005963)
windspd				.3432121**			.1143883**	.0227521**	.0380511**	.2030891**
				(.0702366)			(.0191691)	(.0046128)	(.0060724)	(.0286643)
windspd^2		-.0128995**	-.0029773**	-.0105045**					-.0007943	
		(.0019032)	(.0007254)	(.0025841)					(.0004764)	
No. Observations	6047	10661	11822	10761	8616	9791	11490	12839	10652	13023
McFadden's Pseudo R^2	.1262	.2103	.2262	.1469	.0569	.061	.1259	.1309	.1184	.105
Log Likelihood	-2730.7702	-5369.8254	-5259.516	-4183.4332	-1619.2542	-1722.7158	-5545.442	-7074.0181	-5625.8936	-6070.4264

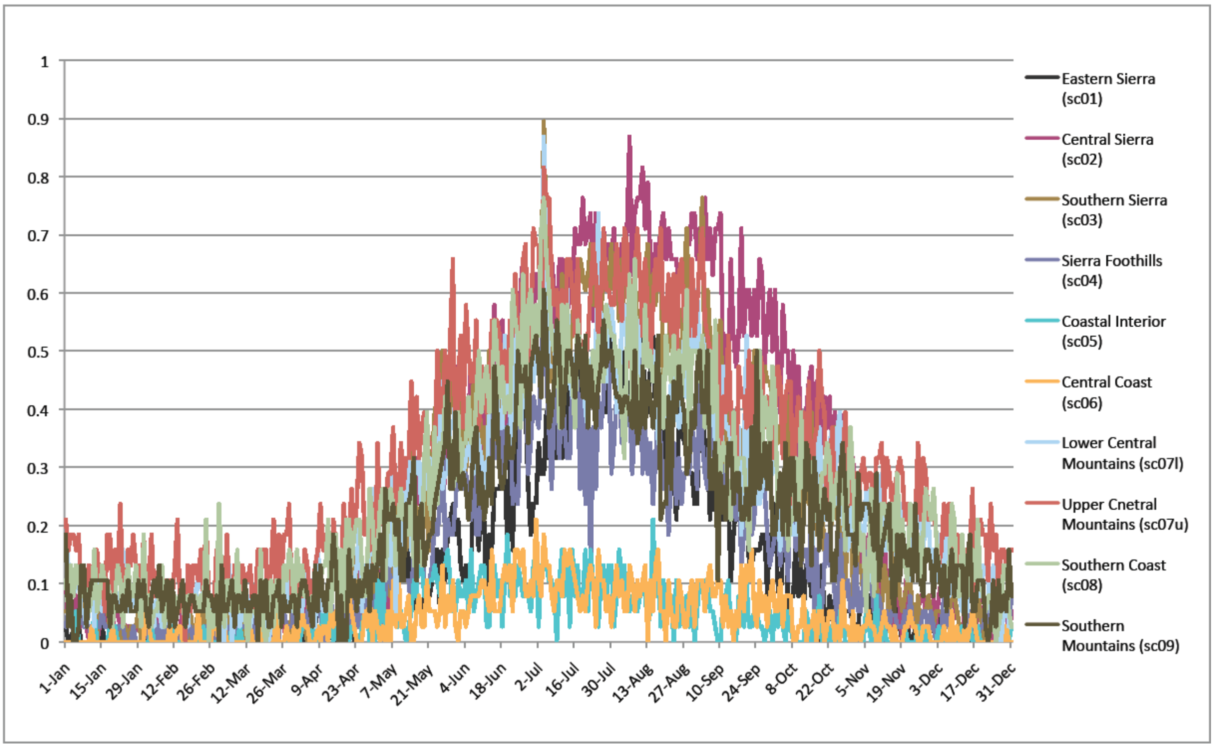
NOTE: standard errors in parentheses. ** denotes significance at 1 percent level. * denotes significance at 5 percent level.

In order to evaluate the estimated selection model we contrast the predicted probabilities with the empirical probabilities based on historical frequencies of ignition (Figure 4).

Figure 4: Average probabilities (1970-2007)



(A) Predicted probabilities



(B) Empirical probailities

As a way of evaluating the skill of prediction of the selection model, we use the Fosberg Fire Weather Index (FFWI) as a benchmark of comparison. The FFWI is defined by a quantitative model that provides a nonlinear filter of meteorological data which results in a linear relationship between the combined meteorological variables of relative humidity and wind speed, and the behavior of wildfires. Like the selection model, the FFWI uses only weather data that are easily attainable from RAWS stations, thus serving as a comparable benchmark. The calculation for the FFWI is taken from Goodrick (2002), and is then scaled to a percentage so that it can be compared to probabilities.

Illustrations in Figure A1 in the appendix compare the results from the selection model with the FFWI and empirical probabilities. From these illustrations, several characteristics are noteworthy. Firstly, it is apparent that the FFWI fits poorly to the empirical observations, as it fails to report seasonal changes in fire occurrence, likely because the FFWI does not take temporal and spatial information into account. The probabilities resulting from the time-series selection model estimation follow the annual trend of fire occurrence, although it fails to predict the idiosyncratic characteristics for each day of the year. The estimations generally overestimate for the spring and fall, while the peak fire season in the summer is underestimated.

Inducing from the illustrations in Figure A1, one may suggest that the empirical annual trend may serve as a better predictor than either the selection model results or the FFWI. However, this suggests that fire managers would simply consult a calendar to inform oneself of the fire occurrence likelihood for that day of the year, disregarding the idiosyncratic weather conditions for a particular day. The selection model described in this paper improves the skill of estimating the probability upon simply knowing the day of the year.

4.2 Outcome Model

The outcome model is an ordinary least squares (OLS) regression estimating the expected acres burned on a given day, given the weather conditions, and conditional on the occurrence of ignition. The outcome model estimation is obtained using the maximum likelihood estimation using STATA, where the outcome estimation is influenced by the probability of occurrence of the event as estimated in the probit model. The conditionality on the occurrence of ignition is signified in the inverse mills ratio expressed in equation (8).

Again, using the all subset selection, the outcome model was tested with the inclusion of different combinations of weather variables, along with seven cumulative days of leads and lag for each weather variable. The final selection of the model specification is chosen by the highest adjusted -R^2. The results are presented in Table 5. The regressions were poorly fitted, not to exceed an R^2 of 0.1379. Figure A2 in the appendix illustrates the difference between actual and predicted acres burned of four locations that are representative of the predictions from other regions. Most of the predictions cannot successfully predict extreme cases, when larger areas are burned during fire season. With the exception of the Upper Central Mountains, most predictions are relatively flat, exhibiting only a variation around an average throughout the year. While most predictions do not exceed 400 acres, many of the predictions peak around 100 acres. Among the locations that are shaped similarly to Central Sierra (sc02) are Eastern Sierra (sc01), Lower Central Mountains (sc07l), the Coastal Interior (sc05), and the Sierra Foothills (sc04), which reaches it's annual average maximum around 200 acres. Predictions with more variability such as that seen in the South Coast (sc08) are in the Southern Sierra (sc03) and South Mountains (sc09). The Upper Central Mountains (sc07u) has the best seasonal predictions, while the Central Coast (sc06) has the poorest fit, with predictions burned area ranging from -5562 acres to 4050 acres.

Although in practice, it is impossible to achieve negative acres of burned area, negative acres are predicted in this outcomes model. Under circumstances of normality of the data, the Heckman procedure corrects for the truncation of the predictions at zero by controlling for selection bias. However, because of the inclusion of a few events of extremely large fires, the OLS procedure finds a best-fit equation that compromises the fit of smaller fires in order to account for the large fires.

Table 5: Estimated coefficients of regional outcome model

	Region									
Independent Variable	(sc01)	(sc02)	(sc03)	(sc04)	(sc05)	(sc06)	(sc07l)	(sc07u)	(sc08)	(sc09)
constant	2546.998	-26.35809	-6800.868**	-6839.844**	3875.202	21641.4	1519.627	-11529.84**	18769.26**	5854.417*
	(2,273.5900)	(546.9324)	(.)	(219.5497)	(4,742.8860)	(24,121.5500)	(1,721.2330)	(505.4421)	(3,932.4910)	(2,422.0990)
temp	6.416162	-5.811591	12.24316	57.85088**	-59.38608	62.35709	-30.61875	111.4909**	-26.49368	-38.92001
	(17.5536)	(5.9489)	(6.6409)	(2.3778)	(34.8078)	(158.4735)	(18.6218)	(4.8005)	(42.4322)	(31.2104)
rh	-7.299908	-4.613356	-59.41647**	0.0005913	-11.46897	66.39592	-17.92743**	-4.715055	-55.28176**	-28.22586
	(8.8450)	(3.3842)	(4.8963)	(0.0283)	(20.2881)	(67.1565)	(6.8032)	(3.6468)	(21.1395)	(14.7799)
windspd	-4.791106	-61.10788**	137.7925**	69.86059**	-176.4859**	89.69558	-18.7026	72.91417**	144.6202	300.0966**
	(12.2879)	(19.3988)	(32.0695)	(12.3825)	(64.8514)	(344.3456)	(27.6456)	(18.5594)	(77.6529)	(72.3838)
wind_n	455.8588	-114.7714	-965.6732	-0.0036018	-637.1033	-1924.486	159.4209	0.0152677	-215.741	-130.1421
	(552.4108)	(162.0382)	(507.2704)	(0.4988)	(865.0675)	(2,308.3260)	(352.1876)	(2.6912)	(781.7325)	(846.9514)
wind_ne	-162.4813	-66.8629	46.78901	-0.0055038	-691.6636	-1383.934	-17.27912	-0.0062053	-1949.436	-4113.006
	(543.2946)	(565.7310)	(260.0724)	(2.1944)	(671.7582)	(2,791.3600)	(365.5638)	(5.6309)	(2,671.0380)	(3,375.3110)
lag temp	-12.22076*		0.2313016	0.0000151	6.211002	-83.12575*		-3.20E-06		
	(5.2576)		(0.8918)	(0.0071)	(3.4164)	(38.1445)		(0.0170)		
lag rh	7.915418**	1.452998		0.0000302			4.13467*		16.86772**	6.749418*
	(2.7478)	(0.8608)		(0.0034)			(1.8150)		(5.3321)	(3.3823)
lag vpd	15.04349*	2.915588*		7.60E-06		99.42489**	13.5749**	0.0001121	32.11088**	17.98654**
	(6.1301)	(1.1715)		(0.0072)		(31.6343)	(3.6001)	(0.0171)	(6.5638)	(5.0233)
lag pptdur	-38.76346*			-0.0001375						
	(18.1287)			(0.0217)						
lag windspd		7.4287*			-18.62138	81.94066	-11.33707*	-0.0001514		-42.76649**
		(3.5294)			(15.7879)	(73.4504)	(5.6723)	(0.0303)		(13.9056)
lead temp				0.0000162			-2.126987		-49.60973**	-23.04856**
				(0.0039)			(1.7919)		(11.1600)	(7.7287)
lead rh				0.0000101	-6.118779*					10.39181*
				(0.0027)	(2.6491)					(4.1815)
lead vpd			10.79325**						30.66974**	25.41203**
			(0.8707)						(8.5717)	(8.8657)
lead pptamt	-0.0221802									
	(0.2478)									
lead pptdur					54.31532					-72.88774*
					(44.7067)					(31.5256)
lead windspd			11.19614**		44.91106**		8.044163		-43.27482**	
			(4.0887)		(17.0393)		(5.6690)		(13.7378)	
No. Observations	5716	10428	11504	10348	8460	9669	10208	12618	9941	12728
Log Likelihood	-10110.08	-32762.01	-29284.76	-13348.06	-3370.808	-4766.256	-16867.95	-41691.78	-28213.94	-30413.98

NOTE: standard errors in parentheses. ** denotes significance at 1 percent level. * denotes significance at 5 percent level.

5 Summary and Conclusions

This paper develops a weather based statistical model for the prediction of the occurrence and the size of wildfires. The model has been applied to wildfires in Southern California. The results of this statistical exercise are mixed. On the one hand, the probability of the occurrence of wildfires is estimated fairly well. Our model outperforms simpler *ad hoc* approaches such as the Fosberg Fire Index. On the other hand, the prediction of the size of wildfires is rather disappointing. The inability to accurately estimate the size of the fire may be the result of several statistical attributes. Firstly, the poorness of fit is not surprising, given that the variability in weather occurrences is not as large as the size of fires. The size of fire is small on most days, while extremely large fires do not necessarily occur every year. When large fires occur, they only occur over the course of a few days. On the contrary, variability of temperature and rainfall cannot be so drastic as to explain such large changes in fire events. Given this characteristic of distribution of the data, it is not surprising that the variability of weather does poorly in explaining and predicting the size of fires. Secondly, this study neglects non-weather variables that are likely to explain the size of fire. While weather variables may have considerable information to predict the probability of fire occurrence, the size of the fire may be greatly influenced by other characteristics of the land, such as the fuel type, the connectivity of fuels, the slope and aspect of the location, proximity to road and other human activities, whether the land is urban, and so on. In practice, the urban landscape can be considered as a fuel type, since the fuel type of wooden homes and asphalt-paved roads conduct fire differently from forested land or chaparral. One way to improve the estimates of the second stage of the Heckman procedure would be to include some variables that measure the land characteristics mentioned above. Alternatively, other statistical approaches could be investigated, such as parametric calibrations using a homogeneous Poisson process (cf. Haerdle and Lopez-Cabrera 2007) focusing only on the occurrence of large fires.

As an overall conclusion we find that a weather-based statistical model is able to support the informational basis for a precautionary fire management. From an actuarial viewpoint the contribution of this model is limited as it fails to predict the size of fires correctly. Large fires, however, are in general most expensive and thus insurance companies are particularly concerned about them. This, in turn, means that it is rather difficult to reinsure against wildfire losses with the help of weather derivatives or other risk transfer products that are solely based on weather indices.

References

Agee, J.K. (1993): Fire Ecology of Pacific Northwest Forests. Island Press, Washington DC.

Bratten, F.W. (1982): Probability Model for Analyzing Fire Management Alternatives: theory and structure. United States Department of Agriculture, Forest Service. Pacific Southwest Forest and Ranger Experiment Station. Berkeley, California.

Mills, T.J., and Bratten, F.W. (1982): FEES: Design of a Fire Economics Evaluation System. United States Department of Agriculture, Forest Service. Pacific Southwest Forest and Ranger Experiment Station. Berkeley, California.

Brillinger, D.R., Preisler, H.K., and Benoit, J.W. (2006): Probabilistic Risk Assessment for Wildfires. Environmetrics. 17: 623-633.

Brown, T.J., Hall, B.L., Mohrle, C.R., and Reinbold, H.J. (2002): Coarse Assessment of Federal Wildland Fire Occurrence Data. Desert Research Institute (DRI), CEFA Report 02-04

Goff, J.A. (1957): Saturation Pressure of Water on the new Kelvin Temperature Scale. Transactions of the American Society of Heating and Ventilating Engineers. 347-354, American Society of Heating and Ventilating Engineers, Murray Bay, Que. Canada.

Goodrick, S.L. (2002): Modification of the Fosberg Fire Weather Index to Include Drought. International Journal of Wildland Fire. 11, 205-211.

Greene, W.H. (2000): Econometric Analysis. 4th Ed., Prentice-Hall, 929.

Haerdle, W. and Lopez-Cabrera, B. (2007). Calibrating CAT Bonds for Mexican Earthquakes. SFB 649 Discussion Paper 2007-037.

Heckman, J.J. (1979): Sample Selection Bias as a Specification Error. Econometrica. January, 47: 153-161.

Preisler, H.K., and Benoit, J.W. (2004): A State Space Model for Predicting Wildland Fire Risk. Proceedings of the Joint Statistics Meetings, Toronto.

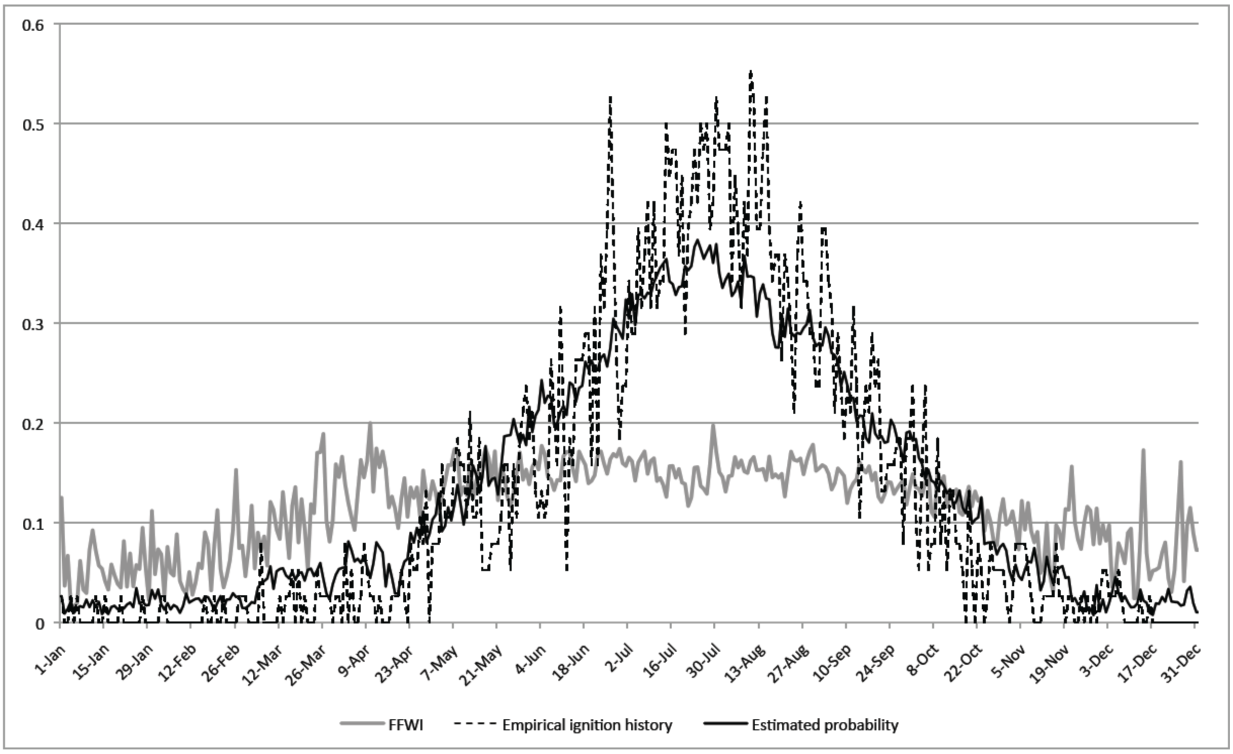
Preisler, H.K., Brillinger, D.R., Burgan, R.E. and Benoit, J.W. (2004): Probability Based Models for Estimation of Wildfire Risk. International Journal of Wildland Fire, 13: 133-142.

Preisler, H.K., Burgan, R.E., Eidenshink, J.C., Klaver, J.M., and Klaver, R.W. (unpublished): Forecasting Distributions of Large Federal-lands Fires Utilizing Satellite and Gridded Weather Information.

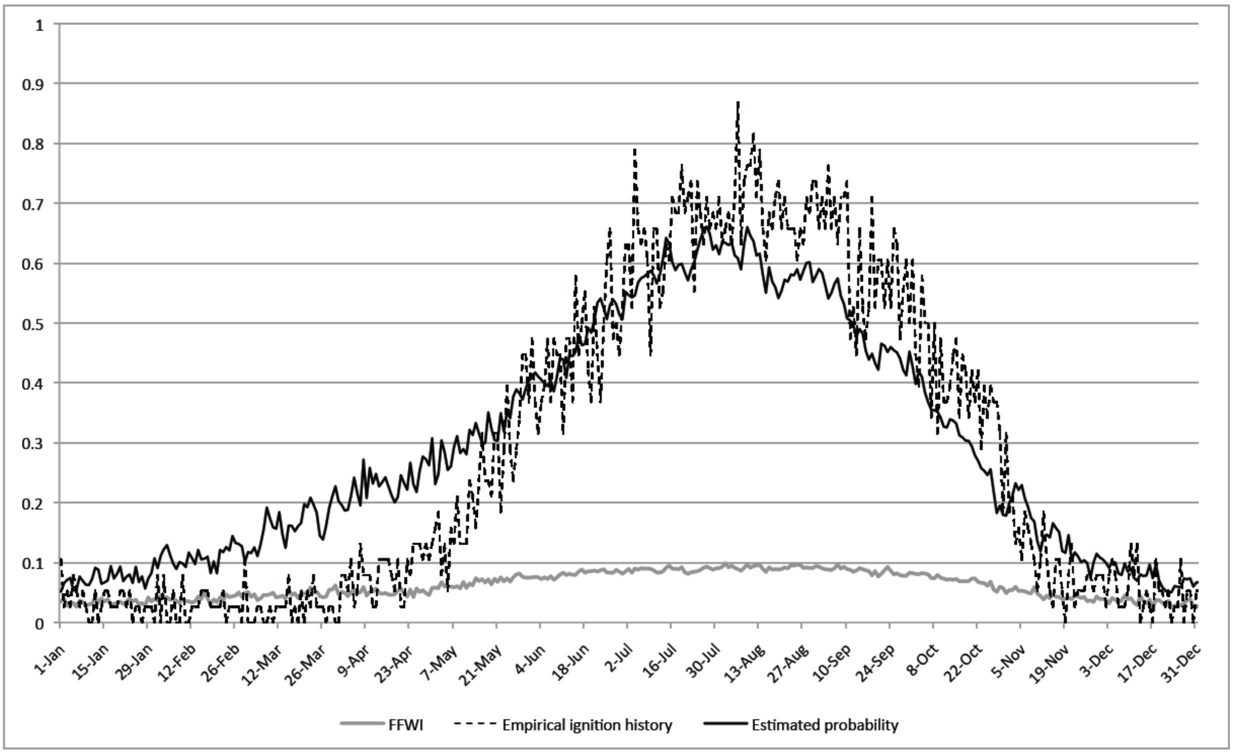
RAWS (2008): National Fire and Aviation Management Web Applications. Available at: http://fam.nwcg.gov/fam-web/weatherfirecd.

Appendix

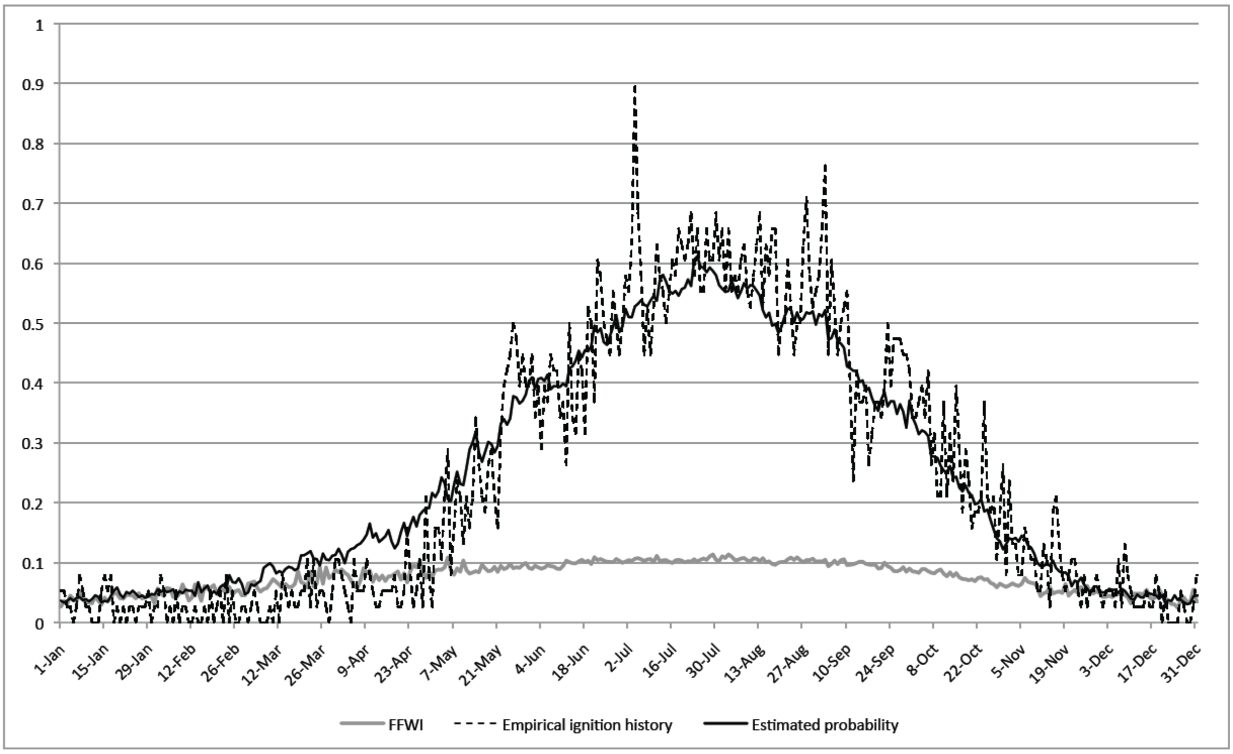
Figure A1: Annual probabilities estimated by selection model and FFWI



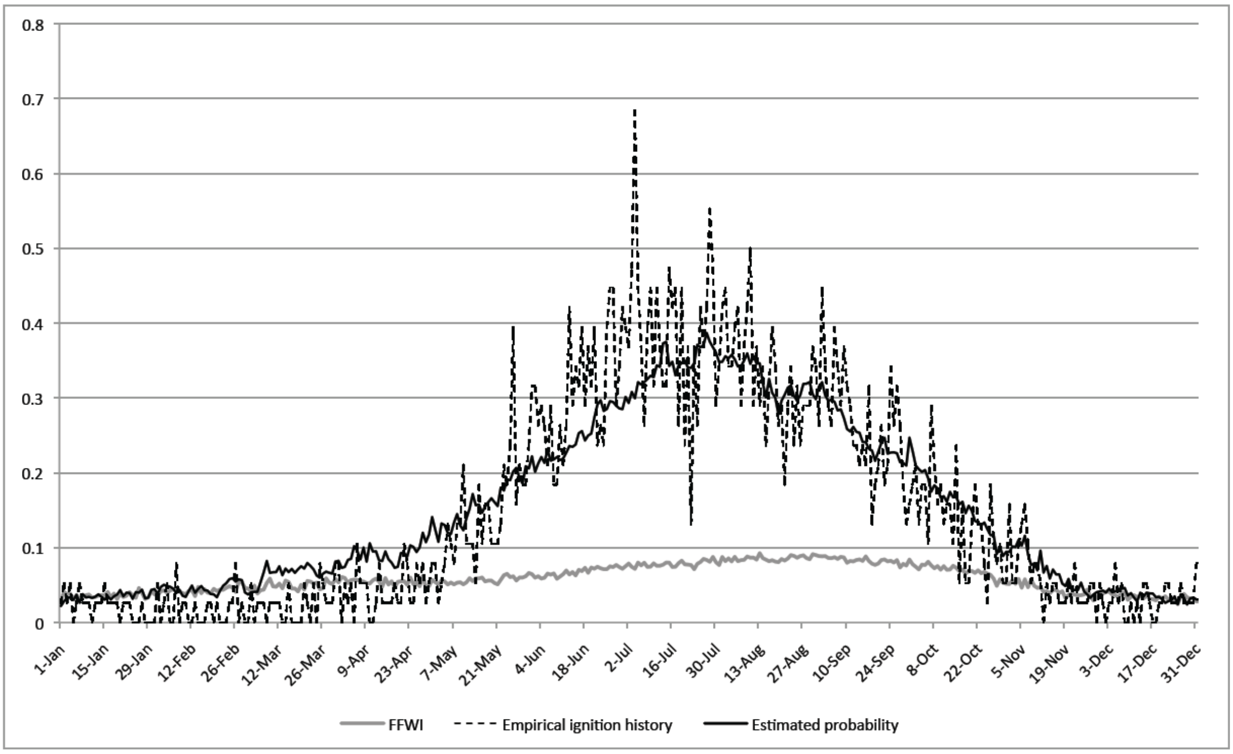
[Eastern Sierra (sc01)]



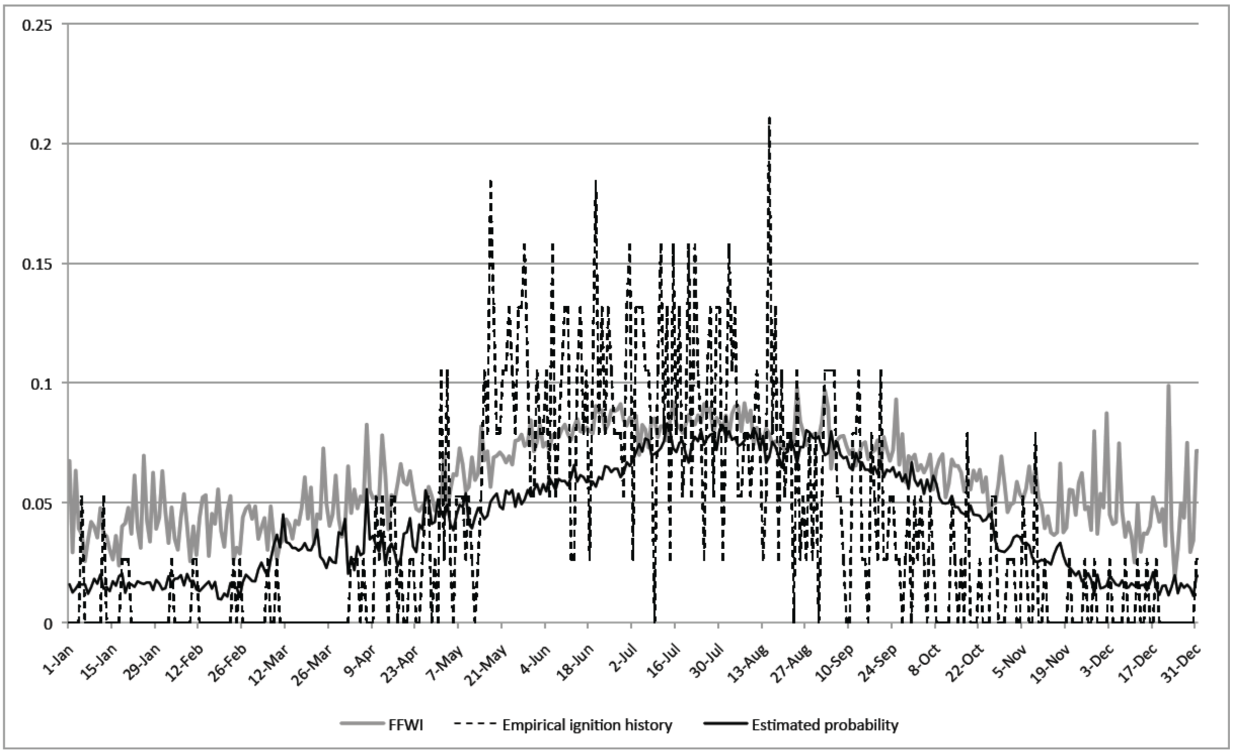
[Central Sierra (sc02)]



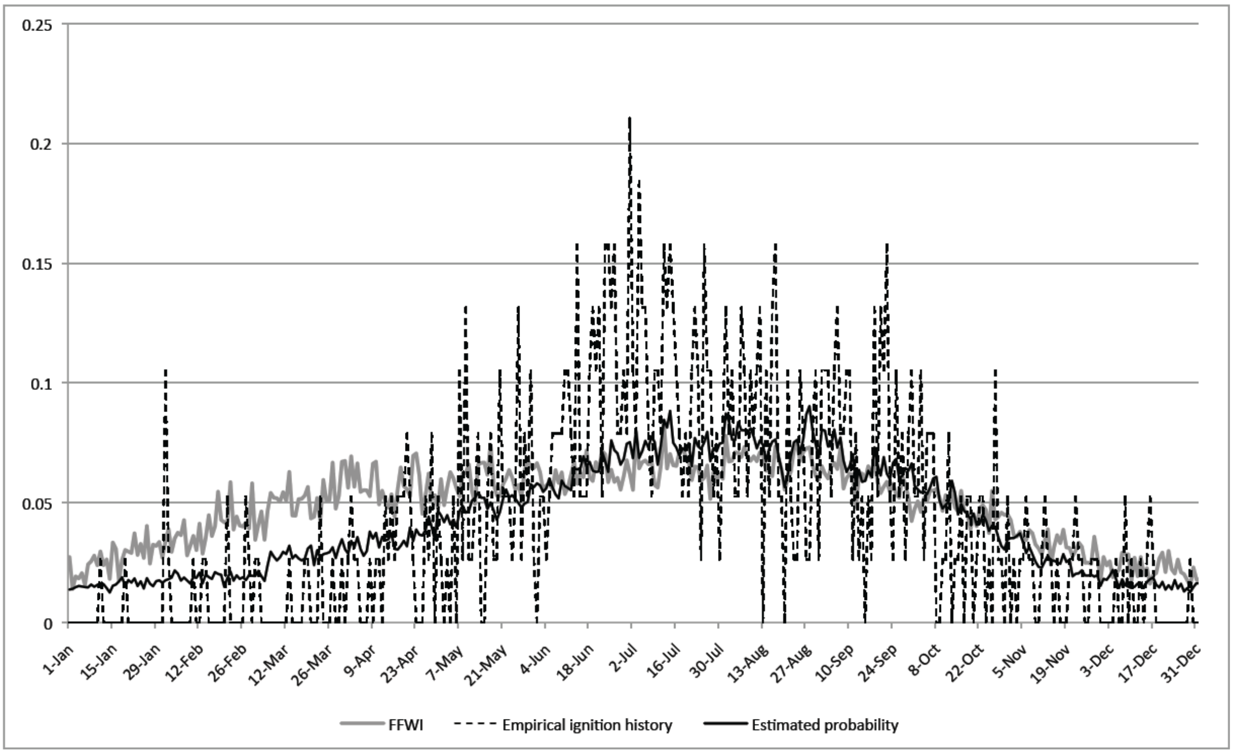
[Southern Sierra (sc03)]



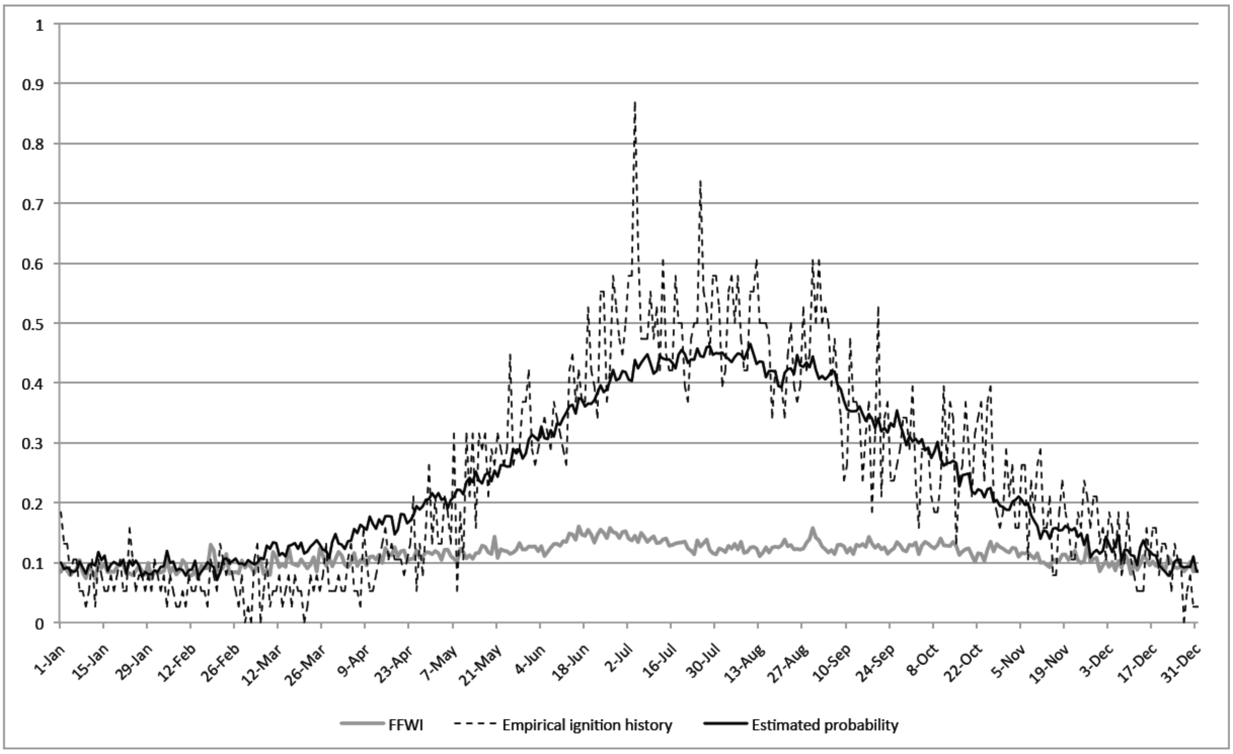
[Sierra Foothills (sc04)]



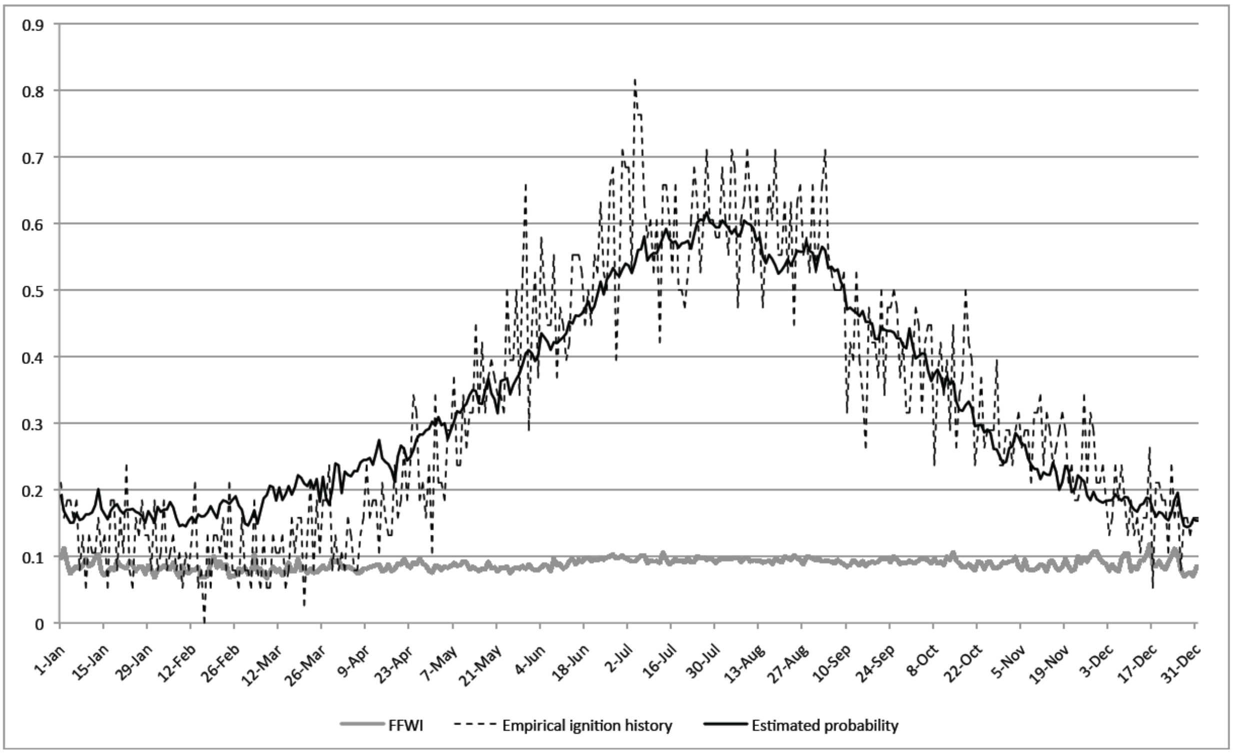
[Coastal Interior (sc05)]



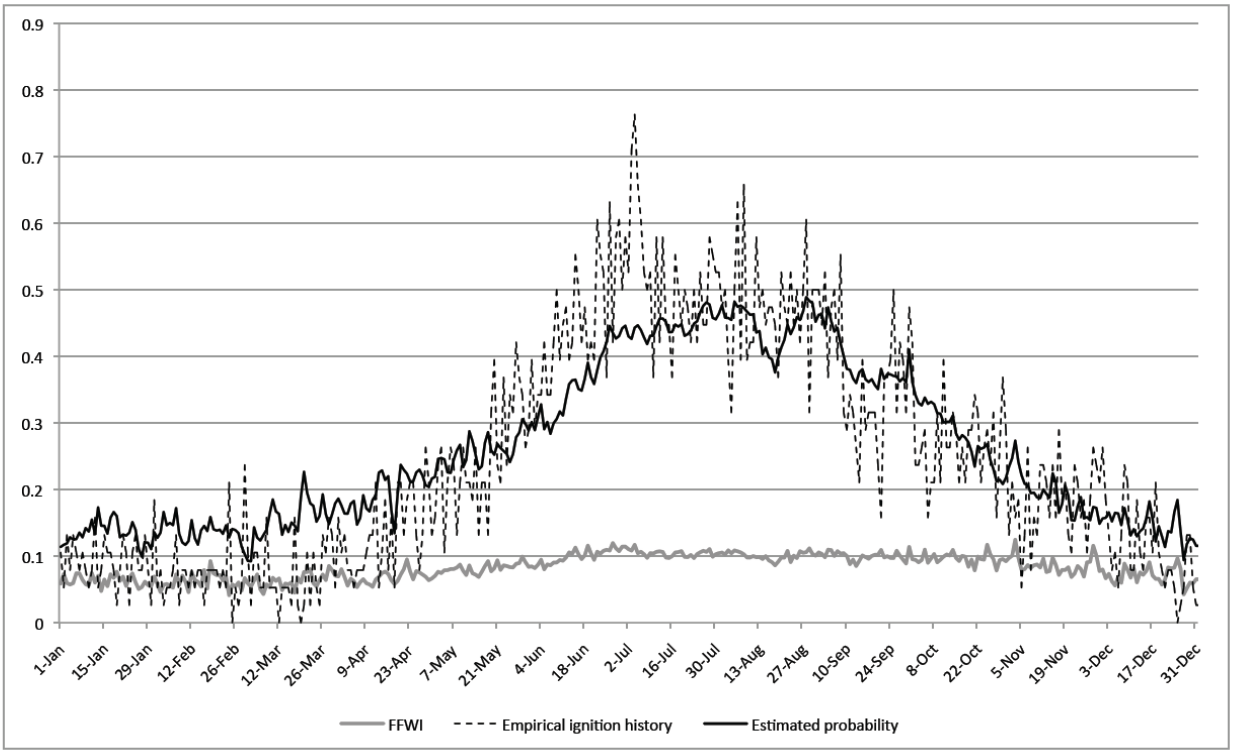
[Central Coast (sc06)]



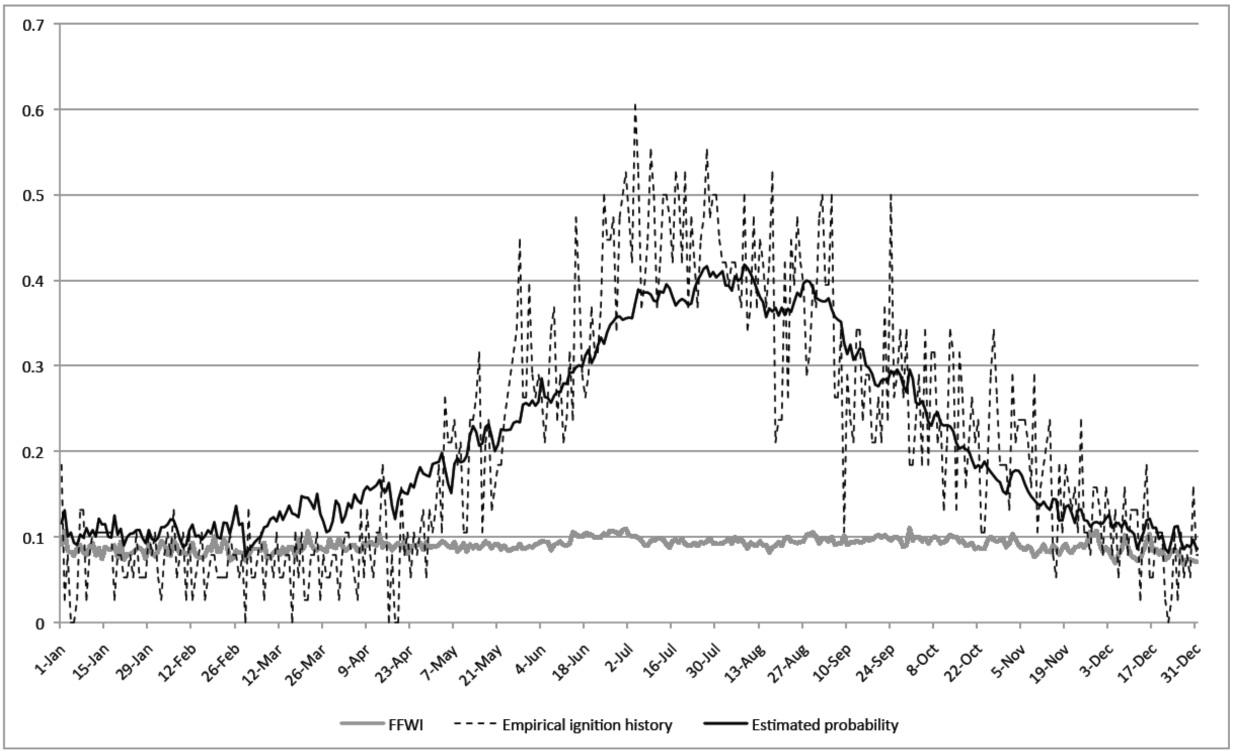
[Eastern Mountains (sc07l)]



[Western Mountains (sc07u)]

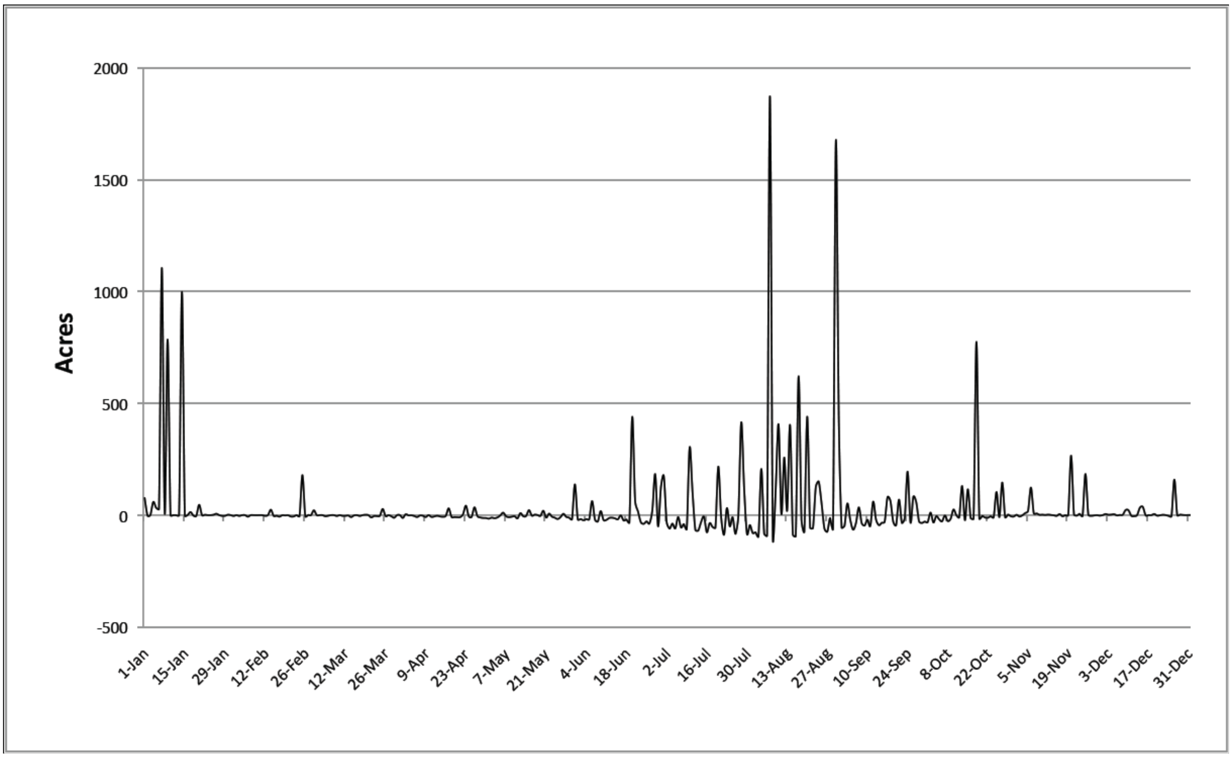


[South Coast (sc08)]

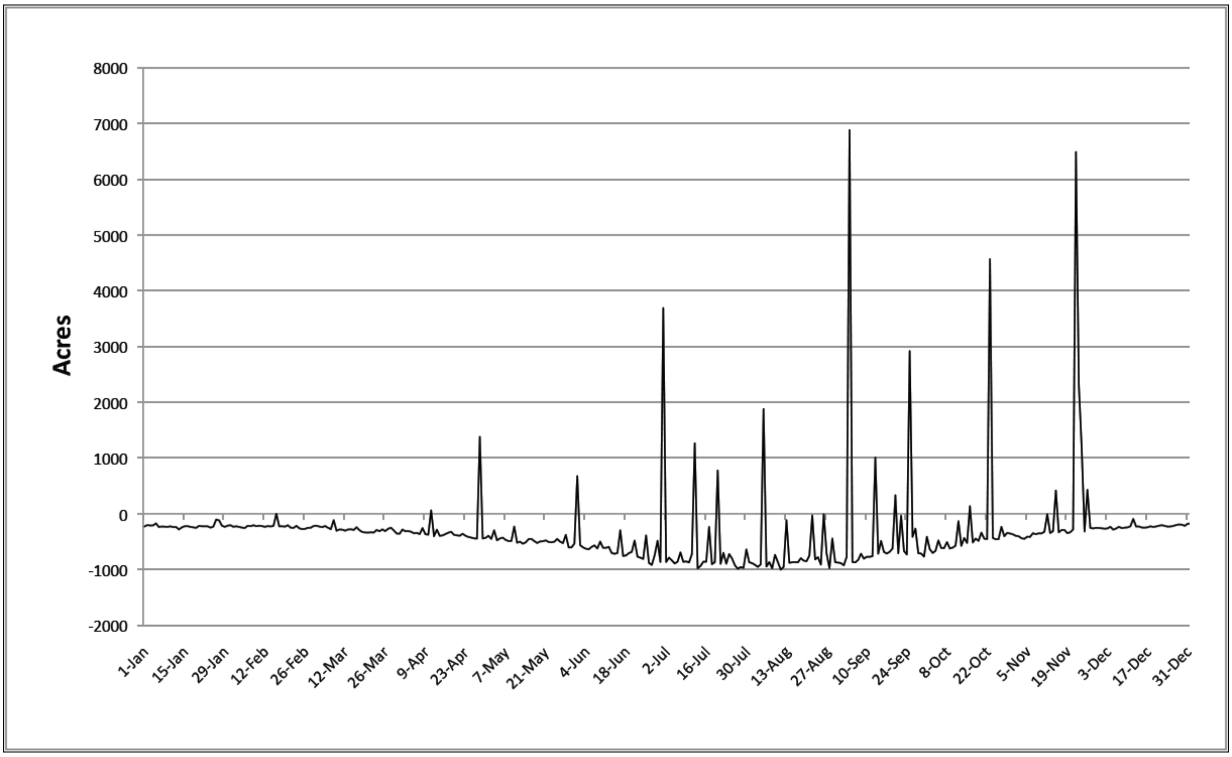


[Southern Mountains (sc09)]

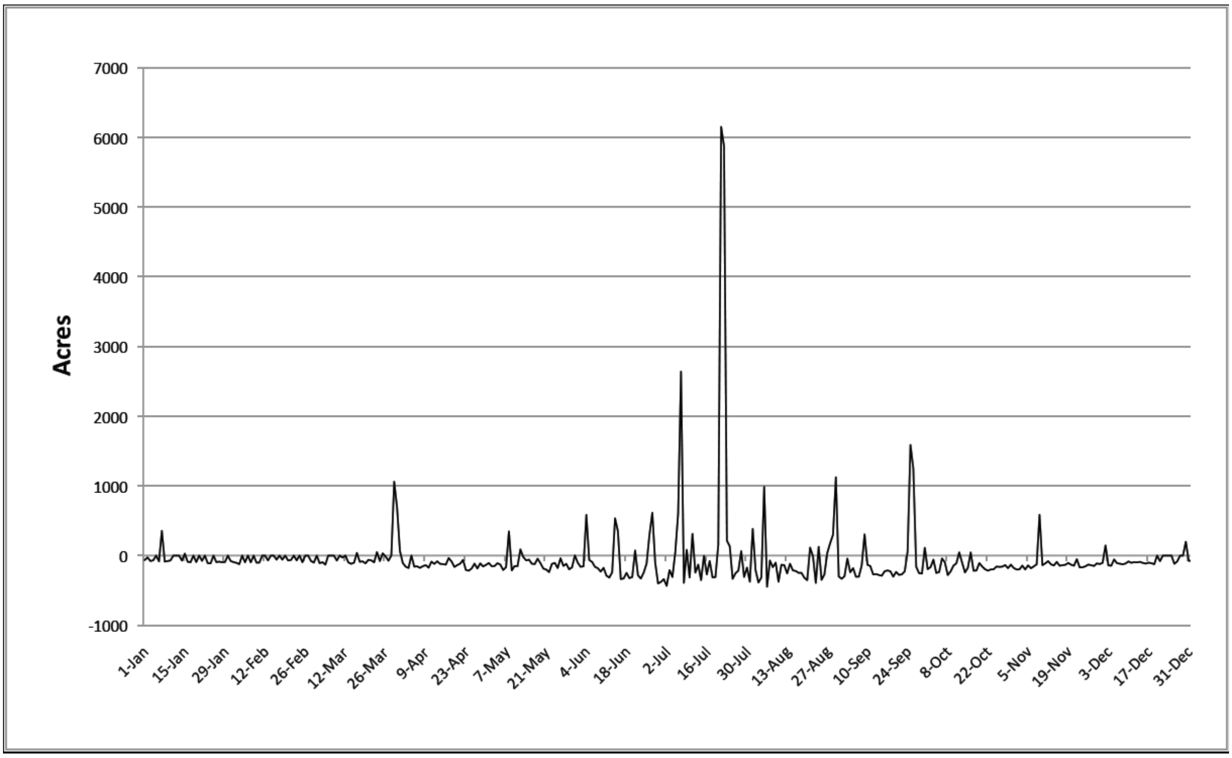
Figure A2: Difference between actual and predicted acres burned (1970-2007)



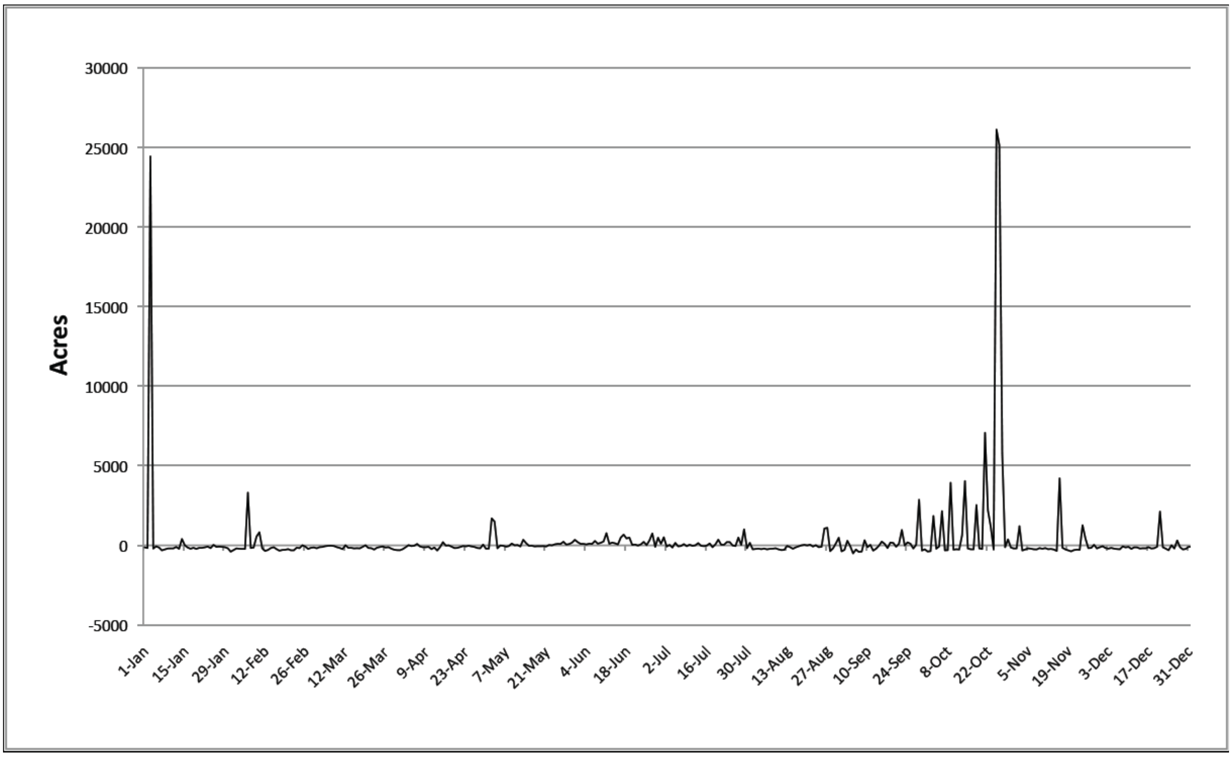
[Central Sierra (sc02)]



[Upper Central Mountains (sc07u)]



[Southern Sierra (sc03)]



[South Coast (sc08)]

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Implied Market Price of Weather Risk" by Wolfgang Härdle and Brenda López Cabrera, January 2009.
002 "On the Systemic Nature of Weather Risk" by Guenther Filler, Martin Odening, Ostap Okhrin and Wei Xu, January 2009.
003 "Localized Realized Volatility Modelling" by Ying Chen, Wolfgang Karl Härdle and Uta Pigorsch, January 2009.
004 "New recipes for estimating default intensities" by Alexander Baranovski, Carsten von Lieres and André Wilch, January 2009.
005 "Panel Cointegration Testing in the Presence of a Time Trend" by Bernd Droge and Deniz Dilan Karaman Örsal, January 2009.
006 "Regulatory Risk under Optimal Incentive Regulation" by Roland Strausz, January 2009.
007 "Combination of multivariate volatility forecasts" by Alessandra Amendola and Giuseppe Storti, January 2009.
008 "Mortality modeling: Lee-Carter and the macroeconomy" by Katja Hanewald, January 2009.
009 "Stochastic Population Forecast for Germany and its Consequence for the German Pension System" by Wolfgang Härdle and Alena Mysickova, February 2009.
010 "A Microeconomic Explanation of the EPK Paradox" by Wolfgang Härdle, Volker Krätschmer and Rouslan Moro, February 2009.
011 "Defending Against Speculative Attacks" by Tijmen Daniëls, Henk Jager and Franc Klaassen, February 2009.
012 "On the Existence of the Moments of the Asymptotic Trace Statistic" by Deniz Dilan Karaman Örsal and Bernd Droge, February 2009.
013 "CDO Pricing with Copulae" by Barbara Choros, Wolfgang Härdle and Ostap Okhrin, March 2009.
014 "Properties of Hierarchical Archimedean Copulas" by Ostap Okhrin, Yarema Okhrin and Wolfgang Schmid, March 2009.
015 "Stochastic Mortality, Macroeconomic Risks, and Life Insurer Solvency" by Katja Hanewald, Thomas Post and Helmut Gründl, March 2009.
016 "Men, Women, and the Ballot Woman Suffrage in the United States" by Sebastian Braun and Michael Kvasnicka, March 2009.
017 "The Importance of Two-Sided Heterogeneity for the Cyclicality of Labour Market Dynamics" by Ronald Bachmann and Peggy David, March 2009.
018 "Transparency through Financial Claims with Fingerprints – A Free Market Mechanism for Preventing Mortgage Securitization Induced Financial Crises" by Helmut Gründl and Thomas Post, March 2009.
019 "A Joint Analysis of the KOSPI 200 Option and ODAX Option Markets Dynamics" by Ji Cao, Wolfgang Härdle and Julius Mungo, March 2009.
020 "Putting Up a Good Fight: The Galí-Monacelli Model versus 'The Six Major Puzzles in International Macroeconomics'", by Stefan Ried, April 2009.
021 "Spectral estimation of the fractional order of a Lévy process" by Denis Belomestny, April 2009.
022 "Individual Welfare Gains from Deferred Life-Annuities under Stochastic Lee-Carter Mortality" by Thomas Post, April 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Pricing Bermudan options using regression: optimal rates of convergence for lower estimates" by Denis Belomestny, April 2009.
024 "Incorporating the Dynamics of Leverage into Default Prediction" by Gunter Löffler and Alina Maurer, April 2009.
025 "Measuring the effects of geographical distance on stock market correlation" by Stefanie Eckel, Gunter Löffler, Alina Maurer and Volker Schmidt, April 2009.
026 "Regression methods for stochastic control problems and their convergence analysis" by Denis Belomestny, Anastasia Kolodko and John Schoenmakers, May 2009.
027 "Unionisation Structures, Productivity, and Firm Performance" by Sebastian Braun, May 2009.
028 "Optimal Smoothing for a Computationally and Statistically Efficient Single Index Estimator" by Yingcun Xia, Wolfgang Härdle and Oliver Linton, May 2009.
029 "Controllability and Persistence of Money Market Rates along the Yield Curve: Evidence from the Euro Area" by Ulrike Busch and Dieter Nautz, May 2009.
030 "Non-constant Hazard Function and Inflation Dynamics" by Fang Yao, May 2009.
031 "De copulis non est disputandum - Copulae: An Overview" by Wolfgang Härdle and Ostap Okhrin, May 2009.
032 "Weather-based estimation of wildfire risk" by Joanne Ho and Martin Odening, June 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".