UC Irvine
UC Irvine Previously Published Works

**Title**
Controls on the spatial pattern of wildfire ignitions in Southern California

**Permalink**
https://escholarship.org/uc/item/1mx8j8mf

**Journal**
International Journal of Wildland Fire, 23(6)

**ISSN**
1049-8001

**Authors**
Faivre, N
Jin, Y
Goulden, ML
*et al.*

**Publication Date**
2014

**DOI**
10.1071/WF13136

**License**
CC BY 4.0

Peer reviewed
Controls on the spatial pattern of wildfire ignitions in Southern California

Nicolas FaivreA,C, Yufang JinB, Michael L. GouldenA and James T. RandersonA

A Department of Earth System Science, University of California, 2101 E Croul Hall, Irvine, CA 92697-3100, USA.
B Department of Land, Air and Water Resources, University of California, Davis, CA 95616-8627, USA.
C Corresponding author. Email: nfaivre@uci.edu

Abstract. Wildfire ignition requires a combination of an open spark, and suitable weather and fuel conditions. Models of fire occurrence and burned area provide a good understanding of the physical and climatic factors that constrain and promote fire spread and recurrence, but information on how humans influence ignition patterns is still lacking at a scale compatible with integrated fire management. We investigated the relative importance of the physical, climatic and human factors regulating ignition probability across Southern California’s National Forests. A 30-year exploratory analysis of one-way relationships indicated that distance to a road, distance to housing and topographic slope were the major determinants of ignition frequency. We used logistic and Poisson regression analyses to model ignition occurrence and frequency as a function of the dominant covariates. The resulting models explained ~70% of the spatial variability in ignition likelihood and 45% of the variability in ignition frequency. In turn, predicted ignition probability contributed to some of the spatial variability in burned area, particularly for summer fires. These models may enable estimates of fire ignition risk for the broader domain of Southern California and how this risk may change with future population and housing development. Our spatially explicit predictions may also be useful for strategic fire management in the region.

Additional keywords: biophysical drivers, fire frequency, fire ignition, human influence, Mediterranean ecosystems, spatial regression model, wildland fire risk.

Received 21 August 2013, accepted 7 May 2014, published online 28 July 2014

Introduction

Wildland fire regimes in Southern California are influenced by climate, ecosystem properties, the rate of human-caused ignition and fire suppression. More than 90% of the fires in Southern California are human ignited and rapid response often extinguishes ignitions that could otherwise become large wildfires (Keeley 1982). Previous studies have reported that the frequency of small fires in coastal Southern California increased during the late 20th century (Keeley et al. 1999). Increasing population size and an expansion of housing into fire-prone wildland areas (Hammet et al. 2007) has increased ignition risk (Keeley and Fotheringham 2001), and when coupled with severe fire weather (e.g. Santa Ana winds), has resulted in several recent catastrophic fire episodes (Keeley et al. 2009). Most ignitions lead to small fires with relatively insignificant effects (Strauss et al. 1989). Approximately 90% of fire ignitions accounted for only 1% of the total area burned between 1980 and 2010 in Southern California, whereas only 3% of ignitions led to fires larger than 400 ha and accounted for 96% of area burned during this period (USDA Forest Service 2010). A combination of altered ignition frequencies, changing climate and expansion of housing near wildland areas has modified fire risk in Southern California and presents a serious threat to human lives and property (Syphard et al. 2012, 2013).

Landscape-scale disturbances such as fire contribute to the distribution of vegetation and habitats across California (Callaway and Davis 1993). Fire is a natural process whose occurrence and magnitude are regulated by environmental and ignition agents. A location’s fire environment, as determined by fuel, weather and topography, affects the occurrence and spread of fire (Countryman 1972). Topographic factors such as elevation, slope and aspect influence fuel characteristics including moisture content, and thus indirectly control fire occurrence and behaviour (Agée 1993; Pyne et al. 1996). Under normal weather conditions, the propensity to burn is fuel dependent and is controlled by the amount, arrangement and physical characteristics of vegetation (Whelan 1995). Extreme weather, including strong winds or extended periods with low humidity, over-ride this fuel dependency and dramatically increase fire risk in Southern California (Keeley and Zedler 2009; Moritz et al. 2010; Jin et al. 2014). Weather and climate, especially precipitation and temperature, affect moisture content and thus the flammability of both live and dead plant material (Verdú et al. 2012). Wind speed, relative humidity and air temperature control fire spread rate and direction, and thus possible future changes in weather and climate have the potential to modify wildfire risk (Cayan et al. 2008). Climate models predict a hotter and drier climate throughout California by the mid to late 21st
Data and methods

Study area

Our main study domain covers 23,500 km² of wildland areas within USA National Forests in Santa Barbara, Ventura, Los Angeles, San Bernardino, Orange, Riverside and San Diego counties. The United States Forest Service (USFS) has used the administrative boundaries of National Forests as the spatial template for recording fire ignition locations. We therefore used this layout to develop our model, focusing on the Los Padres, Angeles, San Bernardino and Cleveland National Forests (Fig. 1). Southern California experiences a Mediterranean climate with a long, dry summer and a relatively short and mild rainy season (Bailey 1966). Contrasting patterns of temperature and rainfall lead to a diverse range of vegetation associations (Franklin 1998). Particularly widespread vegetation types include chaparral, open oak woodland, coastal sage scrub, valley grassland, oak woodland and coniferous forest (Di Castri et al. 1981; Arroyo et al. 1995; Davis and Richardson 1995). The region experiences intense human pressure: over 22 million people lived in Southern California in 2010 and an extensive road network connects numerous communities (source: US Census Bureau 2012). California has ~8500 miles (~13 700 km) of state and federal highways and the average road density within the National Forests in 2000 was ~1.3 km km⁻² (US Census Bureau 2000).

Datasets: fire ignitions and fire perimeters

We extracted the ignition records for 1980–2009 from the USFS FIRESTAT database of individual fire incident reports (USDA Forest Service 2010). This period overlapped with the availability of information on human and biophysical factors and was chosen because earlier records were less reliable for ignition location and date. The location of fire origin was only specified to within ~0.8 km for some fires. This resulted in an artificial clustering of ignitions at ~1.6-km intervals in some areas. Given this uncertainty, we opted to carry out our analysis at a 3 × 3-km resolution. This resolution was chosen as a trade-off between higher resolution grids where ignition location error had a larger effect and lower resolution grids where the loss of spatial information reduced the usefulness of model predictions for management applications. Grids of varying sizes, from 1 to 5 km, were tested in preliminary sensitivity analyses: the 3-km resolution yielded a large continuous range of ignition frequencies, which aided model development. A Mantel test confirmed there was no evidence of spatial autocorrelation at 3-km resolution ($r = -0.065, P > 0.05$). We added a 5-km external buffer to account for ignitions near the National Forest boundaries. We further reduced the noise in the data by excluding ignitions that initiated fires less than 0.1 acres ($\sim 400$ m²). ArcGIS desktop computer software (ArcGIS Desktop, Environmental Systems Research Institute, Redlands, CA) was used for all digital map analyses.

In a final step we compared our ignition estimates with the observed spatial patterns of fire frequency to quantify the role of ignition frequency in explaining the local fire return interval. We computed a fire frequency map using historical perimeter data for fires larger than 100 acres ($\sim 0.4$ km²) compiled by the California Department of Forestry and Fire Protection’s Fire
Resource Assessment Program (FRAP 2010). We calculated fire frequency in each 3 x 3-km grid cell (i.e. number of fires during the 1980–2009 period per square kilometre), weighting each fire within a grid cell by its fractional burned area. We also classified these burns into Santa Ana and non-Santa Ana fires based on the fire start date reported in the FRAP (2010) database and the time series of Santa Ana days, following the approach described by Jin et al. (2014). Santa Ana days were identified when the north-easterly component of the daily mean wind speed was greater than 6 m s \(^{-1}\) at the exit of the largest gap across the Santa Monica Mountains (Hughes and Hall 2010).

Datasets: human factors
Recent studies identified numerous predictors of the occurrence of ignitions in densely populated areas (Syphard et al. 2008; Catry et al. 2009; Martinez et al. 2009). Fire ignitions recorded over the last three decades tend to be clustered around transportation networks and near urban areas. We therefore considered five variables to describe the human footprint: (1) distance to a major road, (2) distance to a minor road, (3) road density, (4) distance to low-density housing and (5) population density.

We used the US Census Bureau’s TIGER road data (Topologically Integrated Geographic Encoding and Referencing; US Census Bureau 2000) to estimate the road density per grid cell and the distance from cell centroid to nearest road. Highways and state roads were classified as major roads (Fig. 2a), whereas streets and vehicle trails were classified as minor roads. This differentiation helped account for traffic volume as a controller of ignition.

We computed average housing density for 1980–2009 based on 1990 and 2000 US Census data (Hammer et al. 2004, 2007) (Fig. 2b). The WUI is often defined as areas with less than 50% vegetation and at least one house per 40 acres (6.2 houses km \(^{-2}\)) that are located within 1.5 miles (~2.4 km) of an area over 500 ha that is more than 75% vegetated (Stewart et al. 2007). Assuming that ignitions most likely occur close to or within interface and intermix communities (Syphard et al. 2007, 2008), we used the distance to the nearest housing area with a density greater than 6.2 housing units km \(^{-2}\) as an indicator of the proximity to the WUI (Fig. 2b). In addition, we used the WUI vector maps created by the USFS (Radeloff et al. 2005) to calculate the number of ignitions within and outside the WUI areas.

Finally, we used the best available, fine-grained spatial database on population demographics from the US Census Bureau’s block group data for 2000 (US Census Bureau 2001) to compute average population density per grid cell (Fig. 2c). This variable captured the direct influence of human presence within and outside the National Forest boundaries. All human variables were summarised at 3-km resolution.

Datasets: biophysical factors
We considered ten variables that were related to topography, land cover and climate: elevation, slope, south-westness, forest...
cover, shrubland cover, grassland cover, cover of other land cover types, annual average daily maximum temperature, annual average daily minimum temperature, and cumulative winter precipitation. We used the 3-arc-second digital elevation model from the US Geological Survey National Elevation Dataset (NED) to calculate the slope and aspect for each 3 × 3-km grid cell using ArcGIS software (Gesch et al. 2002) (Fig. 2d). Aspect was transformed trigonometrically to a south-facing index referred to as ‘south-westness’ (cos(aspect – 225°)) following Beers et al. (1966). This index provided a measure of sun exposure and dryness within each grid cell. Flat terrain with a slope of less than 5° was excluded from the aspect analyses.

We assessed vegetation characteristics using the most recent and comprehensive land cover dataset at 100-m resolution from FRAP (2002). We classified the vegetation in the National Forests into three major types: ‘shrubland’ (60% of the area), ‘forest/woodland’ (22%) and ‘grassland’ (6%). The remaining non-vegetated land cover types were grouped as ‘other’,
included agricultural land, urban, desert, wetland, water and barren (Fig. 2e). We calculated the fraction of each class within each 3 × 3-km grid cell. The monthly averages of precipitation and daily maximum and minimum temperature were taken from the gridded Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset at 800-m resolution (Daly et al. 2002). In our analysis we summarised winter precipitation (September–March; Fig. 2f), and annual mean maximum and minimum temperatures for each 3 × 3-km cell over the 1981–2009 period.

Model building: logistic and Poisson regressions

Our input dataset consisted of the 15 explanatory variables described above, which were spatially averaged within each 3 × 3-km grid cell (Fig. 3). We sought to determine the influence of those predictors on two dependent variables: (i) the occurrence of ignition (presence or absence within a cell) and (ii) the frequency of ignition (number of ignitions within a cell).

We used a logistic regression approach to model the presence or absence of ignitions (Kleinbaum et al. 2002; Hosmer and Lemeshow 2005). Logistic regression has been used to successfully model the probability of fire occurrence at a range of geographic scales (Chou et al. 1993; Chuvieco et al. 1999; Vasconcelos et al. 2001). Logistic regression is expressed as:

\[
\logit(p_i) = \ln \left( \frac{p_i}{1 - p_i} \right) = a_0 + b_1 x_{i1} + \ldots + b_j x_{ij}
\]  

where \( p_i \) is the probability of an ignition in the cell \( i \) and \( x_{ij} \) is the value of the \( j \)th predictor in the cell \( i \). The underlying distribution is binomial and the logit function is defined as the natural logarithm (ln) of the probability of ignition occurrence.

We found exponential relationships between the explanatory variables and the number of ignitions (Fig. 4) and thus used a Poisson regression model (Agresti 2002) for ignition frequency expressed as:

\[
\log(y) = a_0 + b_1 x_{i1} + \ldots + b_j x_{ij}
\]  

Poisson models provide several advantages including the ability to represent a skewed distribution and the restriction of predicted values to non-negative numbers (Gardner et al. 1995). The Poisson probability distribution of observing any specific count \( y \) for an outcome \( Y \) and where \( i \) describes the average rate of ignitions is given by:

\[
\text{Pr}(Y = y) = \frac{y^i e^{-i}}{y!}
\]

Model building: variable selection and model validation

We sought to identify which variables were most important for controlling the spatial distribution of ignitions, independent of their interaction with other explanatory variables. We first examined the correlation matrix among explanatory variables...
variables. We quantified the relative importance of retained explanatory variables by estimating their percentage contribution to the model goodness of fit (i.e. maximised log-likelihood).

Our goal was to develop simplified logistic and Poisson models, with a reduced set of explanatory variables as a compromise between model fit and model complexity. Logistic nested models were compared and examined using inferential and descriptive statistics. We used the likelihood ratio test and the Wald statistic to assess overall model fit and the respective contribution of individual predictors to fitted models. Receiver operating characteristic (ROC) analysis was performed to quantify the area under the curve (AUC) – a measure of the predictive capability of the logistic model to identify candidates that had an ignition event. We also used the ROC curve to select the optimal threshold probability or cut-off value for the probability that an ignition would occur in a given cell. Using the validation samples we cross-validated the best multiple regression model and built contingency tables of observed and expected responses to evaluate model accuracy, precision, sensitivity and specificity (Hilbe 2009). As with the multiple logistic regression, we tested the significance of nested Poisson models and the significance of individual parameters using the likelihood ratio test and the Wald statistic. We also calculated Pearson’s correlation coefficients for validation samples between observed and predicted values to assess model goodness of fit.

Model building: performance improvement

Poisson regression is a form of generalised linear models that assumes the conditional variance to equal the conditional mean. Therefore, a Poisson model is usually too restrictive when predicting count data, which manifests as data over-dispersion (i.e. the variance exceeds the mean) or as estimates of considerably fewer zero counts than are actually observed in the sample (Long 1997). As an alternative approach to Poisson regression, we tested a negative binomial (NB) regression (Agresti 2002; Gelman and Hill 2007; Hardin and Hilbe 2007), which uses a dispersion parameter $\phi$ to handle the variance independent of the mean.

Based on their respective contributions to model fit, we restricted the number of predictors to a subset of four explanatory variables while ensuring model performance was not significantly altered. We compared the performance of Poisson and NB models to ordinary linear regression, using Pearson’s correlation coefficients, the root mean square error (RMSE), and the percentage of bias between simulated and observed data. We estimated how the four main predictors influenced model performance by comparing the nested models to the 4-parameters final model.

Predictive ignition frequency mapping

We applied the resulting NB model to estimate the ignition frequency for all $3 \times 3$ km grids within the National Forest boundaries, and analysed the spatial distribution of the residual ignition frequency. We used the Pearson correlation coefficient to estimate how much of the spatial variability in fire frequency derived from FRAP fire perimeter data was explained by the predicted variability of ignitions within National Forest boundaries. We also separately considered Santa Ana and non-Santa Ana fires following Jin et al. (2014) to compare the controls of ignition patterns on fire frequency for both fire types.
Table 1. Correlations among the 15 explanatory variables used in the analysis
The table indicates the Pearson’s correlations between all independent variables. Please refer to the main text for more information on individual variables. Interactions with significant correlations ($P < 0.001$) superior to 0.5 (inferior to −0.5 for negative correlations) are indicated in bold

<table>
<thead>
<tr>
<th></th>
<th>d.MajR</th>
<th>d.minR</th>
<th>d.Housing</th>
<th>Droad</th>
<th>Dpopulation</th>
<th>elevation</th>
<th>slope</th>
<th>south-westness</th>
<th>tree</th>
<th>shrub</th>
<th>grass</th>
<th>other</th>
<th>Tmin</th>
<th>Tmax</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.54</td>
<td>0.38</td>
<td>0.36</td>
<td></td>
<td>−0.42</td>
<td>−0.43</td>
<td>−0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>−0.42</td>
<td>−0.43</td>
<td>−0.36</td>
<td>0.18</td>
<td>0.26</td>
<td>0.14</td>
<td>−0.33</td>
<td>−0.28</td>
<td>0.33</td>
<td>0.43</td>
<td>0.1</td>
<td>−0.49</td>
<td>−0.28</td>
<td>0.41</td>
</tr>
<tr>
<td>0.03</td>
<td>0.07</td>
<td>0.15</td>
<td>−0.09</td>
<td>−0.15</td>
<td>0.16</td>
<td>0.01</td>
<td></td>
<td></td>
<td>0.11</td>
<td>0.11</td>
<td>0.1</td>
<td>−0.19</td>
<td>−0.16</td>
<td>0.66</td>
</tr>
<tr>
<td>0.11</td>
<td>0.11</td>
<td>0.1</td>
<td>−0.19</td>
<td>−0.16</td>
<td>0.66</td>
<td>0.27</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.50</td>
<td>−0.15</td>
<td></td>
</tr>
<tr>
<td>−0.18</td>
<td>−0.18</td>
<td>0.13</td>
<td>0.06</td>
<td>−0.03</td>
<td>−0.24</td>
<td>−0.36</td>
<td>−0.02</td>
<td>−0.12</td>
<td>−0.12</td>
<td>−0.16</td>
<td>0.26</td>
<td>0.63</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>−0.32</td>
<td>−0.25</td>
<td>−0.31</td>
<td>0.66</td>
<td>0.52</td>
<td>−0.32</td>
<td>−0.46</td>
<td>0.13</td>
<td>−0.29</td>
<td>−0.29</td>
<td>−0.60</td>
<td>−0.02</td>
<td>−0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>−0.16</td>
<td>−0.16</td>
<td>−0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>−0.81</td>
<td>−0.11</td>
<td>−0.16</td>
<td>−0.63</td>
<td>−0.63</td>
<td>0.26</td>
<td>0.05</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>−0.26</td>
<td>0.34</td>
<td>0.1</td>
<td>−0.23</td>
<td>−0.09</td>
<td>0.34</td>
<td>0.58</td>
<td>−0.12</td>
<td>0.28</td>
<td>0.28</td>
<td>−0.18</td>
<td>−0.45</td>
<td>−0.16</td>
<td>−0.51</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Univariate logistic and Poisson regression results for all variables influencing the occurrence and frequency of fire ignitions in Southern California National Forests
Values and direction (i.e. positive or negative) of the coefficients indicate the influence of covariates (the driver variables) towards the response variables (ignition occurrence or frequency)

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Fire occurrence (Logistic model)</th>
<th>Fire frequency (Poisson model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Binary response variable</td>
<td>Continuous response variable</td>
</tr>
<tr>
<td></td>
<td>Coefficient</td>
<td>$P$-value</td>
</tr>
<tr>
<td>Human accessibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to major roads (km)</td>
<td>−0.11 ± 0.010</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Distance to minor roads (km)</td>
<td>−0.32 ± 0.040</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Distance to housing (km)</td>
<td>−0.05 ± 0.005</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Urban development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population density (1000 persons km$^{-2}$)</td>
<td>−0.007 ± 0.004</td>
<td>0.09</td>
</tr>
<tr>
<td>Road density (km roads km$^{-2}$)</td>
<td>0.006 ± 0.002</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Topography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>0.44 ± 0.090</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>0.007 ± 0.004</td>
<td>0.08</td>
</tr>
<tr>
<td>South-westness (0–1)</td>
<td>−0.14 ± 0.120</td>
<td>0.225</td>
</tr>
<tr>
<td>Land cover types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tree (%)</td>
<td>0.67 ± 0.19</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Shrub (%)</td>
<td>0.11 ± 0.14</td>
<td>0.436</td>
</tr>
<tr>
<td>Grass (%)</td>
<td>−1.97 ± 0.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Others (%)</td>
<td>−0.42 ± 0.17</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Climate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature maximum (°C)</td>
<td>−0.03 ± 0.01</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Temperature minimum (°C)</td>
<td>−0.04 ± 0.02</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Winter precipitation (mm year$^{-1}$)</td>
<td>0.57 ± 0.41</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Results
Spatial distribution of fire ignitions
Most of the wildland fire ignitions in the National Forests during 1980–2009 occurred near major roads and close to urban housing (Figs 1–3). Ignition points were clustered around populated areas, major infrastructure and highways, implying a strong influence by human factors (Fig. 3). The WUI, defined as areas where housing meets or intermingles with undeveloped wildlands (Stewart et al. 2007), had particularly high ignition densities (i.e. near the border of National Forests in Los Angeles, San Bernardino and Orange counties; Figs 1–3). Ignitions were much less frequent in sparsely populated areas such as Santa Barbara and Ventura counties (Fig. 1). The WUI covered only 5% of National Forest area but accounted for 40% of ignitions. Ignition density was considerably higher within the WUI (0.6 ignitions km$^{-2}$) than in more remote areas (0.03 ignitions km$^{-2}$).

A quantitative analysis of the relationship between ignition density and human variables confirmed that ignitions were most
common near roadways and housing (Fig. 4). Approximately 60% of all ignitions occurred within 1 km of a major road, and ignition density declined with distance more rapidly from minor roads than from major roads (Fig. 4). Approximately 75% of ignitions occurred within 5 km of areas with a density of housing units greater than 6.2 km\(^{-2}\). Ignition density peaked ~2 km and decreased exponentially in areas further away from housing (Fig. 4). Ignition density was highest in areas with intermediate levels of topographic complexity, with slopes between 20 and 40% (Fig. 4).

**Influence of human and biophysical variables on ignition occurrence**

Univariate logistic regressions showed that all human-related variables except population density were significant in explaining ignition occurrence ($P < 0.05$) (Table 2). Ignition occurrence was positively correlated with road density and negatively correlated with distance from major roads, minor roads and low-density housing. The presence or absence of ignitions also was related to vegetation type, with a significantly higher likelihood of ignition in forest and lower probability in grassland and non-vegetated areas (Table 2). We estimated that 53% of 1980–2009 ignitions occurred in shrublands and 22% in forests (Fig. 2e). Elevation was the only topographic variable significantly correlated with ignition occurrence: the likelihood of ignition increased with elevation. Precipitation was significantly correlated with ignition presence (Table 2).

The final model based on stepwise backward selection had 10 significant variables ($P < 0.0001$):

\[
\logit(p_i) = -4.65 - 0.12 \times d_{MajR} + 0.001 \times elev - 0.03 \\
\times d_{Housing} + 2.04 \times shrub - 0.22d_{Hou} + 1.88 \times tree + 0.12 \\
\times T_{max} + 0.02 \times slope + 0.02 \times D_{roads} - 0.002 \times D_{population}
\]

where $d_{MajR}$ is the distance to major roads (km), $elev$ is elevation (m), $d_{Housing}$ is distance to nearest housing area with density greater than 6.2 units km\(^{-2}\) (km), $shrub$ is the percentage cover of shrubland, $d_{Hou}$ is the distance to minor roads (km), $tree$ is the percentage cover of tree, $T_{max}$ is the annual average daily maximum temperature ($^\circ$C) from 1980 to 2009, $slope$ is the percentage slope, $D_{roads}$ is road density (kilometres of roads per square kilometre) and $D_{population}$ is population density (number of persons per square kilometre). The analysis of modelled variance indicated that not all variables contributed equally to the model fit: $d_{MajR}$, $elev$, $d_{Hou}$, shrub and $d_{Hou}$ together explained over 87% of the model variance (Fig. 5a). The variables $tree$, $slope$, $T_{max}$ and $T_{min}$ were highly correlated with $elev$ (Table 1), and their contribution to the model may have been masked by the apparent strong influence of elevation (Fig. 5a). Likewise, the contributions of $D_{roads}$ and $D_{population}$ may have been masked by multicollinearities with $d_{MajR}$, $d_{Housing}$, shrub and $d_{Hou}$ (Table 1; Fig. 5a).

The ROC analysis using only the five strongest parameters resulted in an AUC of 0.72, which indicated that the reduced model was reasonably able to distinguish where ignitions were most likely to occur. Our cross-validation demonstrated that the model correctly predicted 67% of the observed distribution of ignition occurrence.

**Influence of human and biophysical variables on ignition frequency**

We found that all variables related to human presence significantly explained variability of ignition frequency (Table 2). The variables $d_{MajR}$, $d_{Housing}$ and $d_{Hou}$ were the most influential human factors for ignition frequency (Figs 5b, 6). In contrast to the logistic regression results, all climate variables were significant and positively correlated with ignition frequency. Ignitions were more frequent in areas with warmer temperatures and higher precipitation (Table 2). The slope and shrub cover also had significant and positive influences on ignition frequency and contributed substantially to explaining the variability of ignition patterns (Table 2; Fig. 5b). The final Poisson regression following backward selection retained all explanatory variables ($P < 0.001$) except annual average daily minimum temperature:

\[
\log(\text{Ignition}_{FREQ}) = -4.38 - 0.14 \times d_{MajR} - 0.05 \times d_{Housing} \\
+ 0.02 \times slope - 0.23 \times shrub + 0.001 \\
\times elev + 0.15 \times T_{max} + 0.02 \times D_{roads} - 0.001 \times D_{population} \\
+ 0.84 \times prec + 0.23 \times swindex + 0.67 \times tree
\]
The model equation for NB regression was the same as that for Poisson regression.

\[
\log(\text{Ignitions}_{FREQ}) = 0.97 - 0.11 \times d_{\text{majR}} - 0.04 \times d_{\text{Housing}} + 0.03 \times \text{slope} - 0.25 \times d_{\text{minR}}
\]

The most influential predictors of ignition frequency were \(d_{\text{majR}}, d_{\text{Housing}}, \text{slope}, d_{\text{minR}}, \text{shrub}, \text{elev}, T_{\text{max}}, D_{\text{roads}}, D_{\text{population}}\), and \(\text{tree}\) were defined as above, \(\text{prec}\) is the annual average cumulative winter precipitation (September–March) (mm year\(^{-1}\)) and \(\text{swindex}\) is the south-westness index. The most influential predictors of ignition frequency were \(d_{\text{majR}}, d_{\text{Housing}}, \text{slope}\) and \(d_{\text{minR}}\); these variables combined explained \(\sim 85\%\) of model variance (Fig. 5b). The comparison of Poisson and NB models to a linear model showed a clear improvement of model performance, with a reduced AIC value (Table 3). The linear model was inferior to the other models: the estimated coefficients of linear regression showed significantly higher standard errors (Table 3). Our results indicated that over-dispersion was better captured by the NB, as the dispersion estimate was closer to 1. The NB model improved the fit compared to the Poisson model with a significantly reduced bias while showing similar Pearson correlation and RMSE values (\(R^2 = 0.45; \text{RMSE} = 2.79\)) (Fig. 6). The form of the model equation for NB regression was the same as that for Poisson regression:

The distance to major road alone explained \(33\%\) of the observed spatial variance in ignition frequency; distance to housing and slope explained another \(10\%\); and distance to minor roads explained the remainder. We caution that the primary influence of the variables retained in the final Poisson model may be confounded by multicollinearity (Table 1). For example, the importance of slope may have been overestimated due to implicit contributions from \(\text{elev}, \text{prec}, T_{\text{max}}\) or land cover.

**Table 3. Summary of fitted regression models for ignition frequency**

The top part of the table gives coefficient estimates (with standard errors) for each explanatory variable. The second portion of the table compares model performance and reports the number of estimated parameters, maximised log-likelihood, AIC criterion and estimates of dispersion after model fitting.

<table>
<thead>
<tr>
<th>Model predictors</th>
<th>Linear model</th>
<th>Generalised linear models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson</td>
<td>Negative binomial</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>2.10 ± 0.06</td>
<td>1.00 ± 0.03</td>
</tr>
<tr>
<td>distance to major roads</td>
<td>-0.83 ± 0.08</td>
<td>-0.14 ± 0.007</td>
</tr>
<tr>
<td>distance to housing</td>
<td>-0.49 ± 0.07</td>
<td>-0.04 ± 0.003</td>
</tr>
<tr>
<td>slope percentage</td>
<td>0.58 ± 0.07</td>
<td>0.03 ± 0.001</td>
</tr>
<tr>
<td>distance to minor roads</td>
<td>-0.28 ± 0.09</td>
<td>-0.28 ± 0.03</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-5322.3</td>
<td>-4129.7</td>
</tr>
<tr>
<td>AIC criterion</td>
<td>3223.4</td>
<td>3223.4</td>
</tr>
<tr>
<td>Dispersion (\phi)</td>
<td>0.32</td>
<td>0.20</td>
</tr>
</tbody>
</table>

The spatial distribution of ignition frequency predicted using the NB model showed good agreement with the observed ignition pattern (\(R^2 = 0.45\); Figs 6, 7a, b). The proximity to human infrastructure strongly determined ignition frequency (Fig. 7b; Table 3). The model accurately captured the relative lack of ignitions in remote, interior areas (e.g. Los Padres National Forest). Likewise, the model accurately predicted high ignition frequency in many areas near major roads and housing (e.g. Los Angeles County). The model underestimated ignition frequency along high-traffic transportation corridors (e.g. Interstate 5) and in close proximity to some populated urban areas (e.g. San Bernardino) (Fig. 7c). Thus, the variables used to predict ignition frequency (i.e. distance to major roads and distance to housing) were insufficient for discriminating areas where human pressure exceeded a certain threshold.

**Relationship between ignition frequency and fire frequency**

We found a significant positive relationship between ignition frequency and fire frequency. The grided ignition frequency observations within National Forests explained \(3.0\%\) of the spatial variance of observed fire frequency (\(P < 0.001, n = 2625\)). For this same domain, the NB model explained \(5.3\%\) of the observed fire frequency variance (\(P < 0.001\)). We found that many areas with a high potential risk of ignition did not burn between 1980 and 2009 (Fig. 8), indicating that a substantial component of burned area variability was not explained by the drivers of ignition.
We repeated the previously described correlation analysis separately for Santa Ana vs. non-Santa Ana fires (Jin et al. 2014). Santa Ana fires accounted for 45.0% of total burned area across Southern California and 18.0% of the number of fires over the 30-year study period. For non-Santa Ana fires, observed ignitions explained 6.5% ($P < 0.001$) of observed fire frequency and the estimated ignition patterns from the NB model explained 12.2% of fire frequency ($P < 0.001$) within the National Forests.

Our NB approach improved model performance over commonly used linear and Poisson models. The NB model better captured the clustering patterns of ignitions around urban development and transportation corridors with a reduced set of variables describing human accessibility and urban development. Ignition occurrence is most strongly determined by distance to major roads and housing, elevation and the proportion of shrub cover. Elevation may be strongly determined by distance to major roads and housing, elevation and the proportion of shrub cover. Elevation may be

**Discussion**

Our modelling approach allowed us to identify the combination of factors influencing the spatial distribution of ignitions. We found that proximity to roads and housing were the dominant controls for ignition frequency. All variables describing human accessibility and urban development were significantly correlated with ignition frequency. These results provided evidence that human activities are the primary source of ignition in Southern California and are consistent with studies that also found increased ignition frequency near transportation corridors (Stephens 2005) and WUIs (Syphard et al. 2007). Environmental variables resulted in a higher density of ignitions for mid-level slopes and forest land cover types. Past studies that considered generalised linear models to predict the spatial distribution of ignition frequency arrived at similar conclusions (Yang et al. 2007; Syphard et al. 2008). Although topographic features usually influence fire intensity (i.e. spread rate) and the distribution of burns across the landscape (Beaty and Taylor 2001; Alexander et al. 2006), variations in elevation cause variations in fuel type, moisture and phenology, which in turn control the conditions for fire ignition (Swetnam et al. 2011). Our logistic approach confirmed that ignition occurrence is most strongly determined by distance to major roads and housing, elevation and the proportion of shrub cover. Elevation may be capturing the secondary influence of temperature, slope and tree cover because these variables were collinear. The logistic model captured $\sim 70\%$ of ignition likelihood at 3-km resolution, which we considered satisfactory considering the heterogeneity and the large area investigated. Ignition occurrence was strongly conditioned by fuel type, with 75% of ignitions occurring in forest and shrubland.

Ignitions were less important, although significant, in controlling the spatial pattern of Santa Ana fires, with observed and predicted ignitions explaining 3.1 and 4.9% ($P < 0.001$) of observed fire frequency.

**Fig. 7.** Spatial maps of ignition frequency in unit of numbers of ignitions per grid cell across the National Forests at a 3-km resolution during 1980–2009 as (a) recorded in the FIRESTAT database and (b) predicted by the negative binomial model. The residual ignition frequency (observation – prediction) is shown in (c).

**Fig. 8.** Historical area-weighted fire frequency in USFS lands based on 1980–2010 fire perimeters data (FRAP 2010).
and wilderness areas, it is likely that ignition risk will increase. Traffic trends are likely to follow the WUI expansion, which would imply higher traffic volumes along fast-growing corridors such as Ventura, Orange and San Diego Counties (Crane et al. 2002). Our analysis demonstrated that major roads carry higher ignition risk than secondary roads. As a result, rising traffic among highways, such as Interstate 5, which crosses the Los Padres National Forest, and Interstate 15, which contours the San Bernardino National Forest (Fig. 3), will likely increase the ignition rate in these areas. Our approach may prove useful for both fire mitigation and urban planning. For example, it may be used to project ignition risk based on projections of future climatic and human activity across Southern California.

Acknowledgements

This study was supported by NASA Interdisciplinary Science grant NNX10AI4G to the University of California, Irvine. We thank the USFS and the California Department of Forestry and Fire Protection for providing the compiled ignition and fire perimeter data. We thank researchers from UCI’s Earth System Science department for their comments on the earlier version of the manuscript, and two anonymous reviewers for their valuable comments.

References


Predicted ignition frequency explained a small but significant amount (i.e. 12%) of the observed spatial patterns in non-Santa Ana fire frequency within the National Forests. Wildland areas that are likely to experience greater numbers of ignitions coincide with areas characterised by recurrent burning. Nevertheless, in more remote areas than the WUI, such as Los Padres National Forest, where fuel fragmentation is not a limiting factor, fires tend to spread away from ignition sources and burn more frequently (Syphard et al. 2008). For Santa Ana fires, the lower correlation of ignition frequency to fire frequency suggested that the ignition controls on burned area patterns were considerably weaker relative to other factors as compared to non-Santa Ana fires. As burned area per se is not a function of ignition probability only (Archibald et al. 2009), additional variables related to fuel moisture, fuel continuity, fuel load and wind speed (Moritz et al. 2010) need to be considered for modelling burnt area. Similarly, interactions between biophysical factors such as wind speed and precipitation or vegetation type may need consideration given the environmental heterogeneity in the region. Here, we addressed single-term effects of explanatory variables as we sought to build a simplified model of ignition frequency patterns. Although we also investigated quadratic and interactive terms between biophysical variables and between human variables, the results were not conclusive. An important future step is to combine our estimates of ignition frequency with other data sources to model the spatial distribution of burned area using a similar framework.

The predictive maps of ignition frequency generated in this study are synthetic measures of the spatial influence of human and environmental drivers on the current landscape. An important related question is how ignition patterns will evolve during future decades. Increasing human influence through densification and expansion of the WUI is expected to directly affect the wildland ignition regime (Hammer et al. 2007, Radeloff et al. 2010). California’s population is projected to increase to ~49 million in 2025, a 44% increase from 2000. Although most population growth will occur in urban centres, housing density within 10 km of wildlands is projected to increase by ~80% by 2030 in California (Miller et al. 2011). We found distance to nearest housing area with density greater than 6.2 units km–2 was the second most influential control of ignition risk. As a consequence of future housing growth at the periphery of rural

Besides the errors input to the underestimation of ignitions frequency, two contrasting patterns of ignition merit discussion: a peri-urban ignition pattern in counties with dense development and a wildland ignition pattern in counties with sparse housing. Contrasting human settlement density were reported by Badia-Perpinya and Pallares-Barbera (2006). This partly explains why the scattered parts of the WUI, were underestimated. It may be possible to refine model predictions.

Acknowledgements

This study was supported by NASA Interdisciplinary Science grant NNX10AI4G to the University of California, Irvine. We thank the USFS and the California Department of Forestry and Fire Protection for providing the compiled ignition and fire perimeter data. We thank researchers from UCI’s Earth System Science department for their comments on the earlier version of the manuscript, and two anonymous reviewers for their valuable comments.

References


Predicted ignition frequency explained a small but significant amount (i.e. 12%) of the observed spatial patterns in non-Santa Ana fire frequency within the National Forests. Wildland areas that are likely to experience greater numbers of ignitions coincide with areas characterised by recurrent burning. Nevertheless, in more remote areas than the WUI, such as Los Padres National Forest, where fuel fragmentation is not a limiting factor, fires tend to spread away from ignition sources and burn more frequently (Syphard et al. 2008). For Santa Ana fires, the lower correlation of ignition frequency to fire frequency suggested that the ignition controls on burned area patterns were considerably weaker relative to other factors as compared to non-Santa Ana fires. As burned area per se is not a function of ignition probability only (Archibald et al. 2009), additional variables related to fuel moisture, fuel continuity, fuel load and wind speed (Moritz et al. 2010) need to be considered for modelling burnt area. Similarly, interactions between biophysical factors such as wind speed and precipitation or vegetation type may need consideration given the environmental heterogeneity in the region. Here, we addressed single-term effects of explanatory variables as we sought to build a simplified model of ignition frequency patterns. Although we also investigated quadratic and interactive terms between biophysical variables and between human variables, the results were not conclusive. An important future step is to combine our estimates of ignition frequency with other data sources to model the spatial distribution of burned area using a similar framework.

The predictive maps of ignition frequency generated in this study are synthetic measures of the spatial influence of human and environmental drivers on the current landscape. An important related question is how ignition patterns will evolve during future decades. Increasing human influence through densification and expansion of the WUI is expected to directly affect the wildland ignition regime (Hammer et al. 2007, Radeloff et al. 2010). California’s population is projected to increase to ~49 million in 2025, a 44% increase from 2000. Although most population growth will occur in urban centres, housing density within 10 km of wildlands is projected to increase by ~80% by 2030 in California (Miller et al. 2011). We found distance to nearest housing area with density greater than 6.2 units km–2 was the second most influential control of ignition risk. As a consequence of future housing growth at the periphery of rural


Hardin JW, Hilbe J (2007) *Generalized Linear Models and Extensions.* (StataCorp Press: College Station, TX)


www.publish.csiro.au/journals/ijwf