UCLA

UCLA Electronic Theses and Dissertations

Title

County-level Analysis of the Impact of Temperature and Population Increases on California Wildfire Data

Permalink

<https://escholarship.org/uc/item/2753572w>

Author

Baltar, Matthew

Publication Date

2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

**County-level Analysis of the Impact of
Temperature and Population Increases on
California Wildfire Data**

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Matthew Baltar

2015

© Copyright by
Matthew Baltar
2015

ABSTRACT OF THE THESIS

County-level Analysis of the Impact of Temperature and Population Increases on California Wildfire Data

by

Matthew Baltar

Master of Science in Statistics

University of California, Los Angeles, 2015

Professor Frederic R. Paik Schoenberg, Chair

The extent to which the apparent increase in wildfire incidence and burn area in California from 1990 to 2006 is affected by population and temperature increases is examined. Using generalized linear models with random effects, we focus on the estimated impacts of increases in mean daily temperatures and populations in different counties on wildfire in those counties, after essentially controlling for the overall differences between counties in their overall mean temperatures and populations. We find that temperature increase appears to have a significant positive impact on both total burn area and number of observed wildfires. Population growth appears to have a much less pronounced impact on total burn area than do annual temperature increases, and population growth appears to be negatively correlated with the total number of observed wildfires. These effects are especially pronounced in the winter season and in Southern California counties.

The thesis of Matthew Baltar is approved.

Yingnian Wu

Hongquan Xu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2015

To my family.

TABLE OF CONTENTS

1	Introduction	1
1.1	Data from U.S. Geological Survey	2
2	Methods	5
2.1	Poisson GLM Models	7
3	Results	12
3.1	Random Effect of County	12
3.2	Entire Year and All Counties	12
3.3	Temporally Grouped Data	15
3.4	Spatially Grouped Data	17
4	Conclusion	20
4.1	Model Diagnostics	20
4.2	Discussion	22
	References	24

LIST OF FIGURES

1.1	Observed survivor function of the areas burned in California wild-fires from 1/1/1990 - 12/31/2006, along with the survivor function $(1 - F(x))$ corresponding to the tapered Pareto distribution fit by maximum likelihood.	3
2.1	(a) Histogram of number of fires per county-month. An exponential density fit by maximum likelihood is shown. (b) Histogram of total area burned per county-month. An exponential density fit by maximum likelihood is shown.	6
2.2	(a) Change in population from county-mean population versus change in log burn area minus log county-mean burn area. (b) Change in temperature from county-mean temperature versus change in log burn area minus log county-mean burn area. Least squares lines overlaid.	8
2.3	Map of California showing fire locations. Areas of different shades represent possible grouping of counties for spatial analysis.	11
3.1	(a) Estimated random effect of County for area burned. (b) Estimated random effect of County for number of fires. (c) Estimated random effect of month for area burned (d) Estimated random effect of month for number of fires.	13
3.2	(a) Sonoma County's population over time. (b) Sonoma County's annual number of observed wildfires vs. Sonoma County's population, with a least squares line overlaid.	19

4.1	Binned plots comparing observed monthly number of wildfires with fitted monthly number of wildfires (a,b), binned by population and temperature respectively, observed monthly total area burned with fitted monthly total area burned (c,d), binned by population and temperature respectively, and observed monthly number of wildfires with fitted monthly number of wildfires (e,f), binned by changes in population and changes in temperature, respectively. The bin sizes were determined so that each bin has the same number of points, resulting in different ranges for each bin. 15 bins were used for each plot. 95% confidence intervals shown for each point.	21
-----	--	----

LIST OF TABLES

2.1	RMS of within-county differences, between one month and the next, of the covariates used to fit models 2.1 and 2.2, using various subsets of the dataset.	10
3.1	Estimated coefficients in models 2.1 and 2.2, using data from all counties and months. Standard errors in parentheses. Coefficients in bold are significant at the 0.10 level. All models in this paper fit using the "glmer" function in the "lme4" package in R.	14
3.2	Estimated coefficients for population and temperature using temporally split data. Each model was fit using exclusively data from either summer, winter or fall. Standard errors in parentheses. Coefficients in bold are significant at the 0.10 level.	15
3.3	Coefficients for models fit using spatially split data. Divided into coastal, inland, and southern California regions, as shown in Figure 4. Standard errors in parentheses. Coefficients in bold are significant at the 0.10 level.	17

CHAPTER 1

Introduction

The Western United States in general, and California in particular, have exhibited a noticeable increase in both wildfire frequency and total annual area burned over the past several decades (Arno and Allison-Bunnell 2002, Westerling et al. 2003, Stephens 2005, Westerling et al. 2006, Miller et al. 2009). A possible explanation for these increases is climate change, as increased temperatures may cause wildfires to ignite more easily and become more difficult to extinguish, and also cause land to become increasingly arid and its vegetation flammable. In addition, as the vast majority of these wildfires are ignited as a result of human activity (Keeley 1982), the observed increase in wildfire may in part be attributed to human population increases in California over this same time period. Previous studies have looked at the relationship between population and wildfire sizes (Preisler et al., 2011), but not specifically the effect of changes in population within a given county. The goal of this paper is to determine and quantify the extent of the role played by human population and temperature increase in affecting fire frequency and area burned. To this end, we employ spatial-temporal wildfire, population, and mean temperature data in California for the 17 years spanning from January 1, 1990 to December 31, 2006, using California counties as the spatial units of analysis, and examine the estimated coefficients in generalized linear models in order to discern the relative impacts attributable to population and temperature change.

1.1 Data from U.S. Geological Survey

The wildfire data in this analysis were catalogued and compiled by the United States Geological Survey (USGS), Western Ecological Research Center, and were provided to us by CalFire. They consist of wildfire occurrences on CalFire protected areas from Jan. 1, 1990 to Dec. 31, 2006. The dataset consists of several thousand wildfires and represents an accumulation of a wealth of wildfire data collected from an assortment of different sources. The wildfire dataset is perhaps exceptional in its completeness, and is believed to be at least approximately complete for wildfires down to 10ha, since 1990. Keeley (1982) analyzed historical wildfire activity in this region and estimated that 16.2% of wildfires (or 13.1% of the area burned) in this region were caused by lightning, and the remainder of the ignitions were attributable to human activities.

Wildfire burn areas appear often to be roughly Pareto or tapered Pareto distributed (Cumming 2001, Malamud et al. 2005), and hence the completeness of a wildfire dataset may be assessed by observing the extent to which the survivor function of the areas burned appears to be approximated by the tapered Pareto distribution as suggested in Schoenberg et al. (2003). Figure 1.1 shows that the wildfire sizes are well approximated by a tapered Pareto law for fires above approximately 5 ha; such fires account for over 98% of the total area burned in our time frame. The data are organized by county; for each of the 58 counties in California we have records of all observed wildfires on CalFire protected areas occurring during the selected time period, as well as population estimates by year, and daily high temperature. Daily high temperature records for each county were obtained from www.almanac.com/weather/history/CA, and in cases where more than one reporting station was available for a county, the average of records from those stations was used. Population data are publicly available from the California Department of Finance; see www.dof.ca.gov/research/demographic/reports/view.php.

Figure 1.1: Observed survivor function of the areas burned in California wild-fires from 1/1/1990 - 12/31/2006, along with the survivor function $(1 - F(x))$ corresponding to the tapered Pareto distribution fit by maximum likelihood.

Counties with virtually no recorded wildfire activity, or containing Cal Fire lands but providing their own fire protection and thus not included in the data set, were discarded from the analysis; these were the following nine counties: Alpine, Imperial, Los Angeles, Marin, Mono, San Francisco, Sierra, Sutter and Ventura. The focus of this paper is on the remaining 49 counties with 9,996 county-months for the analysis.

CHAPTER 2

Methods

Note that the results of a simple linear regression of area burned or number of fires on temperature and population would say little about the impact of temperature and population changes on wildfire, as most of the variation in the covariates would be due to overall differences in temperature and population between counties. The regression estimates would be dominated by the correlation between characteristics (average population and temperature) of the counties and wildfire activity in those counties, which is not our focus here. Instead, we seek to identify the extent to which changes in population and temperature, within a given county, appear to influence burn area and number of wildfires. Further, we do not wish to combine similar counties, but instead leave each of the 49 counties as a separate entity so that spatial heterogeneity of the results may be investigated. We incorporate a random effects model to account for the fact that different counties will have different population sizes and mean temperatures, and correspondingly quite different numbers of wildfires and area burned between them. We also include a random effect term for the month of the year to emphasize the temporal aspect of the analysis and to account for varying temperatures across seasons during the year.

A histogram of the total number of observed wildfires per county per month (Figure 2.1a) closely resembles the exponential distribution. Note also that observed wildfires per month must be nonnegative, so a random effects regression model assuming roughly Gaussian error terms seems inappropriate. Instead, in

Figure 2.1: (a) Histogram of number of fires per county-month. An exponential density fit by maximum likelihood is shown. (b) Histogram of total area burned per county-month. An exponential density fit by maximum likelihood is shown.

consideration of Figure 2.1a, we propose the use of a random effects generalized linear model (GLM) with a log-link function, to model this response variable as a function of population growth and temperature increase. The fact that the number of observed wildfires in a county-month is count data suggests fitting a Poisson regression model in particular. While the distribution of the areas burned in individual fires appears approximately tapered Pareto (Figure 1.1), the distribution of total area burned in each county-month (Figure 2.1b) appears approximately exponential, like wildfire frequencies, so we fit a Poisson GLM with a log-link when modeling the total area burned as the response variable as well. A log-link function for area burned seems appropriate considering Figure 2.2, where we see a roughly log-linear relationship between change in log-area burned from logmean area burned in a county plotted against each of our two covariates, population and temperature.

2.1 Poisson GLM Models

Overdispersion is frequently a concern when fitting data with Poisson models. The burn area data does exhibit this problem, having a variance considerably higher than the mean. To account for the overdispersion present in monthly burn area, we fit a model containing an observation-level random effect term (Elston et al. 2001). These considerations suggest models of the form:

$$Y_{ijk} \sim \text{Poisson}(\exp(\mu_{1i} + \nu_{1j} + \tau_{1ijk} + \alpha_1 P_{ij} + \beta_1 T_{ij})) \quad (2.1)$$

and

$$Z_{ijk} \sim \text{Poisson}(\exp(\mu_{2i} + \nu_{2j} + \tau_{2ijk} + \alpha_2 P_{ij} + \beta_2 T_{ij})) \quad (2.2)$$

where Y and Z denote the total area burned and number of observed wildfires,

Figure 2.2: (a) Change in population from county-mean population versus change in log burn area minus log county-mean burn area. (b) Change in temperature from county-mean temperature versus change in log burn area minus log county-mean burn area. Least squares lines overlaid.

respectively, index i represents the county, index j represents the month and k the year in consideration, P denotes population and T is temperature. Parameters μ_{1i} and μ_{2i} represent the baseline for the mean number of wildfires and total area burned for county i across all months, respectively, parameters ν_{1j} and ν_{2j} represent the baseline for the mean number of wildfires and total area burned for month j , while α_1 and α_2 denote the estimated coefficients for population increase, and β_1 and β_2 the estimated coefficients corresponding to temperature increases. The observation-level random effect term given to each data point is denoted τ .

Models 2.1 and 2.2 allow us to focus on temperature increases and population growth within each county and the effect of these changes on wildfire incidence and total area burned, while essentially controlling for the differences in overall population and mean temperature between counties and months.

Models 2.1 and 2.2 were initially fit using data from all seasons, for each county-year, leaving out the random effects for months. However, a problem with this is that a particularly cold winter could average out an exceptionally hot summer in a given county and thus the effects of temperature changes on our response variables would be masked. To account for this, we added random effect terms for month, ν_{1j} and ν_{2j} , and fit the models 2.1 and 2.2 to all the county-months as well as separately after splitting the data temporally into fire seasons delineated as suggested in Nichols et al. (2011). The first season, the most active season and covering the summer months, extends from May 25 to September 26. This is followed by an intense fall season characterized by fewer but larger fires. This season begins September 27 and continues until November 7. The remaining months, November 8 to May 24, cover the mostly dormant, winter season. This seasonal splitting allows us to investigate whether the overall trends are present in these shorter time periods which have more consistent weather throughout them.

Each county-month has its own background rate of total area burned or number of fires estimated in the models. The parameters α_1 , α_2 , β_1 , and β_2 , which are

	Entire Year	Winter	Summer	Fall	Coastal	Inland	Southern
RMS of month-to-month temperature changes ($^{\circ}\text{C}$)	4.31	1.18	1.52	2.82	3.74	4.99	3.29
RMS of month-to-month population changes (people)	3813	3813	3813	3813	2205	1821	11644

Table 2.1: *RMS of within-county differences, between one month and the next, of the covariates used to fit models 2.1 and 2.2, using various subsets of the dataset.*

constrained to be the same across all counties, are influenced then by changes in population and temperature within a county. After both population and temperature have been standardized (converted to z-scores), the parameters are also in standard units. Table 2.1 shows the root mean square (RMS) of the county-wide month-to-month variations in the covariates in each subset of the data being analyzed. Thus, for instance, the size of a typical increase in population from one year to the next is 11,644 for the southern counties, and roughly 2,000 for the coastal and inland counties. The population data for each county is annual, and thus the RMS of the population changes from one year to another are identical to 3813 for each of the seasonal subsets. We fit all of our models by Laplace approximation using R software (2013).

A mapping of California wildfires (Figure 2.3) suggests a possible grouping of counties for an improved spatial analysis. We thus also re-fit models 2.1 and 2.2 using only subsets of the data, such as data from all counties but only in fall, or data from the entire year but only for coastal counties, etc.

Figure 2.3: Map of California showing fire locations. Areas of different shades represent possible grouping of counties for spatial analysis.

CHAPTER 3

Results

Tables 3.1, 3.2 and 3.3 present the results of 14 fitted models, all of which follow either equation 2.1 or 2.2, but differ in the data being used to fit the model, either spatially or temporally.

3.1 Random Effect of County

Figure 3.1 displays the random effects for our models. The estimated random effects for county show correlation between fire activity and total burn area, as we might expect. Southern California and Northern California both appear to have low and high activity counties, though the southernmost part of California is considerably active (Riverside and San Diego counties). The estimated monthly random effects reveal, not surprisingly, that the hot summer months tend to have both larger and more frequent fires. The curve for area burned has a spike in October which is caused by the year 2003. In 2003 San Diego county had over 150,000 ha burned in what was called the 2003 Fire Siege.

3.2 Entire Year and All Counties

Table 3.1 shows the results for the models fit with data for all months and for all counties. Temperature has a positive slope in both models, meaning higher temperatures are associated with increasing total area burned and fire frequency. Area burned (Model 2.1) increases along with population and temperature, but typical

Figure 3.1: (a) Estimated random effect of County for area burned. (b) Estimated random effect of County for number of fires. (c) Estimated random effect of month for area burned (d) Estimated random effect of month for number of fires.

Data from Entire Year	$\hat{\alpha}$	$\hat{\beta}$
Model 2.1: Area Burned (ha)	0.024 ha/100,000 increase in population (0.032)	0.209 ha/°C (0.011)
Model 2.2: No. of Wildfires	-0.040 wild-fires/100,000 increase in population(0.013)	0.063 wildfires/°C (0.003)

Table 3.1: Estimated coefficients in models 2.1 and 2.2, using data from all counties and months. Standard errors in parentheses. Coefficients in bold are significant at the 0.10 level. All models in this paper fit using the "glmer" function in the "lme4" package in R.

temperature variations appear to have a much stronger effect on area burned than do typical population increases. Indeed, one may interpret what the fitted model implies for a given county given a typical increase in population or temperature as indicated in Table 2.1: an increase of about 4.31 °C (1 RMS temperature increase) in a typical county would correspond to an expected increase of about 146% (+/- 4.8%) in its monthly number of wildfires (since $\exp\{0.209 \cdot 4.31\} \sim 2.461$) and an expected increase of about 31% (+/- 1.3%) in its monthly number of observed wildfires ($\exp\{0.063 \cdot 4.31\} \sim 1.312$). Meanwhile, a typical jump in population of 3,813 people would correspond to an increase in a typical county's annual area burned of about 0.09% (+/- 0.1%) ($0.024 \cdot 3,813 / 100,000 \sim 1.00093$) and a decrease in its total annual number of observed wildfires of about 0.15% (+/- 0.05%), ($\exp(-0.040 \cdot 3,813 / 100,000) \sim 0.9985$).

The fact that typical monthly temperature variations appear to outweigh population increases in influencing area burned and number of observed wildfires is hardly surprising, since wildfires are well known to be more prevalent and often larger during especially hot years (Pyne et al. 1996, Johnson and Miyanishi 2001, Moritz et al. 2005).

The result of the fitted model 2.2 indicating that population increases tend

Seasonal Models		Summer	Winter	Fall
Model 3.1: Response = Area Burned	$\hat{\alpha}_1$ (ha/100,000 in- crease in pop.)	0.204 (0.075)	-0.018 (0.043)	0.056 (0.050)
	$\hat{\beta}_1$ (ha/°C)	0.210 (0.043)	0.368 (0.064)	0.226 (0.044)
Model 3.2: Response = No. of Fires	$\hat{\alpha}_2$ (wild- fires/100,000 in- crease in pop.)	-0.013 (0.023)	-0.043 (0.026)	-0.038 (0.026)
	$\hat{\beta}_2$ (wildfires/°C)	0.047 (0.011)	0.182 (0.026)	0.065 (0.014)

Table 3.2: *Estimated coefficients for population and temperature using temporally split data. Each model was fit using exclusively data from either summer, winter or fall. Standard errors in parentheses. Coefficients in bold are significant at the 0.10 level.*

on average to correlate with decreased fire frequency may seem curious. It is conceivable that as a county becomes more urbanized and developed, there is less burnable land remaining and thus decreased potential for a wildfire to ignite. Sonoma County is a prime example. From figure 3.2 one sees that population has increased steadily in Sonoma County over the observed time interval studied here, while the number of observed wildfires has decreased sharply. The least-squares fit is shown and confirms the downward trend.

3.3 Temporally Grouped Data

Table 3.2 shows the coefficients for our models fitting the data split according to wildfire seasons. These models do not contain the monthly random effect term seen in models 2.1 and 2.2, and instead are of the form:

$$Y_{ijk} \sim \text{Poisson}(\exp(\mu_{1i} + \tau_{1ijk} + \alpha_1 P_{ij} + \beta_1 T_{ij})) \quad (3.1)$$

and

$$Z_{ijk} \sim \text{Poisson}(\exp(\mu_{2i} + \tau_{2ijk} + \alpha_2 P_{ij} + \beta_2 T_{ij})) \quad (3.2)$$

where parameters μ , τ , α and β are defined as in models 2.1 and 2.2. With this temporally focused analysis we see that, as in the entire-year fits of models 2.1 and 2.2 from Table 2, temperature increases are associated with increases in both total area burned and fire frequency. Population increases, however, are associated with decreases in the number of observed wildfires in a county. Our discussion thus shifts to the seasonal variation in effect size between temperature and population.

The fits for model 3.1 in Table 3.2 show that the increases in total area burned attributable to population increases are most evident in the active summer season, less so in the intense fall season and considerably less substantial in the dormant winter season. Somewhat surprisingly, it is the warm summer season's total area burned and number of wildfires that appears to be least affected by changes in temperature. In the fit for summer, one sees a 37.6% increase in burn area corresponding to a 1 RMS (1.52 °C) increase in temperature, which is considerably lower than the 54.4% and 89.2% increases in burn area corresponding to comparable temperature changes in winter and fall, respectively. In terms of wildfire activity, this same 1 RMS increase in temperature is associated with a 7.5% increase in number of fires, which is again less than is seen in winter and fall, with changes of 23.9% and 20.0%, respectively.

It is worthwhile to note that fall has the highest variability in temperature of the three seasons (1 RMS is 2.82 °C); if instead of a 1 RMS increase we consider a 1 °C increase in temperature for each season, the fitted seasonal model parameters from Table 3.2 suggest that winter would have a 44% increase in monthly burn area, whereas fall and summer would have increases of only about 25%. Note that the fall season includes the most area burned for any single county-year in the dataset; this is partially attributable to the 2003 Fire Siege in San Diego County. The fall season accounts for 30% of all area burned in the data, yet represents only 11% of the total number of fires.

Regional Models		Coastal	Inland	Southern
Model 2.1: Response = Area Burned	$\hat{\alpha}_1$ (ha/100,000 in- crease in pop.)	0.040 (0.053)	0.036 (0.047)	-0.072 (0.064)
	$\hat{\beta}_1$ (ha/°C)	0.290 (0.024)	0.562 (0.020)	0.405 (0.045)
Model 2.2: Response = No. of Fires	$\hat{\alpha}_2$ (wild- fires/100,000 in- crease in pop.)	0.020 (0.025)	-0.017 (0.020)	-0.070 (0.018)
	$\hat{\beta}_2$ (wildfires/°C)	0.116 (0.008)	0.150 (0.010)	0.119 (0.013)

Table 3.3: Coefficients for models fit using spatially split data. Divided into coastal, inland, and southern California regions, as shown in Figure 4. Standard errors in parentheses. Coefficients in bold are significant at the 0.10 level.

Considering now changes in population, total area burned in summer and number of fires in winter are the most significantly affected. The effect of population is not as pronounced as that of temperature, but is present, and appears to follow a pattern qualitatively similar to that for the entire year; population increases are associated with increases in area burned, but a decrease in fires ignited. In summer, a 1 RMS change in population corresponds to a 0.8% increase in total burn area, yet a not statistically significant 0.04% decrease in number of fires. The same population increase (3,813 people) in winter corresponds to a 0.2% decrease in wildfire activity.

3.4 Spatially Grouped Data

Figure 2.3 displays a map of California wildfire locations in the 49 counties used in our analysis. Three very distinct clusters are visible; fires in coastal counties, counties with fires east of the San Joaquin and Sacramento Valleys and west of the Sierra Nevada mountain range, and counties in southern California. We divide the data spatially referring to the darkest shaded cluster as the coastal region, the medium grey shaded area as the inland region, and the lightest shaded as the southern region.

Coefficients for the spatially grouped data can be seen in Table 3.3. The fits

to model 2.1 indicate that population increases have their strongest effect in the southern region, where a typical population change of 11,644 people results in a 0.85% decrease in both total monthly area burned and number of wildfires. Population's effect is only significant in the southern region, and then only for number of fires. As in the entire-state analysis (Table 3.1) and the temporally divided analysis (Table 3.2), population increases appear to be associated with decreases in observed wildfire frequency, while temperature increases are associated with increases in fire activity in all three regions.

Temperature increases, meanwhile, are significant across all regions and for both burn area and wildfire frequency. Temperature's effect is most strongly seen in the inland region, where a 1 RMS increase in temperature is associated with a 111% increase in wildfire activity, whereas a comparable change in the southern and coastal regions produce 48% and 54% increases respectively. A similar trend is seen for total area burned, where comparable temperature changes affect the inland region approximately 4 times as much as either the coastal or southern regions.

Figure 3.2: (a) Sonoma County's population over time. (b) Sonoma County's annual number of observed wildfires vs. Sonoma County's population, with a least squares line overlaid.

CHAPTER 4

Conclusion

4.1 Model Diagnostics

It is important to assess how well these models actually fit the data. The standard deviation of the monthly number of observed wildfires across all county-years in the dataset is 4.84 fires. This may be compared to the standard deviation of the residuals for model 2.2 of 2.92 fires; the model 2.2 explains nearly 40% of the variation observed. Correspondingly, the standard deviation of annual total area burned across all county-months is 1,959 ha, whereas the residual standard deviation for model 2.1 is 1954 ha; model 2.1 accounts for only 0.3% of the observed variation in monthly burn area changes. This percentage is small due to the effect of the outlier San Diego, 2003, having a large residual. When we remove this point from the data, model 2.1 accounts for 14% of the variation in monthly burn area changes.

Figure 4.1 shows binned plots comparing observed responses with fitted responses. From Figs. 4.1a and 4.1b one sees that the model 2.2 predicts number of fires very accurately; we see a clear linear relationship between actual and fitted number of fires, whether we bin by population or temperature. Fig. 4.1c also shows a closely linear relationship between observed and fitted annual burn areas, though from Fig. 4.1d, when binned by temperature, the relationship shows some mild departures from linearity. The outlier seen in Fig. 4.1c contains San Diego county, 2003.

Figure 4.1: Binned plots comparing observed monthly number of wildfires with fitted monthly number of wildfires (a,b), binned by population and temperature respectively, observed monthly total area burned with fitted monthly total area burned (c,d), binned by population and temperature respectively, and observed monthly number of wildfires with fitted monthly number of wildfires (e,f), binned by changes in population and changes in temperature, respectively. The bin sizes were determined so that each bin has the same number of points, resulting in different ranges for each bin. 15 bins were used for each plot. 95% confidence intervals shown for each point.

As discussed earlier, the coefficients in the models are influenced by changes in population and temperature. Thus, we can also check the model with binned plot where we are not simply binning by temperature or population, but by the change in each of those variables from the year 1990 to any given year. Figures 4.1e and 4.1f show these plots and in both cases we do see a linear structure indicating rather satisfactory fit overall.

4.2 Discussion

The random effect Poisson regression models for total area burned and total number of observed wildfires as functions of county-wide temperature increases and population increases suggest that temperature increases are positively associated with both total area burned and number of observed wildfires, which is not surprising. The models suggest that far more of the variation in area burned and number of wildfires may be attributable to changes in temperature than to changes in population, especially in the coastal and inland counties. The variability in these effects between spatial locations is extremely pronounced. This is most clearly seen in differences between the fitted models for the coastal and southern California regions, where population growth and temperature appear to exhibit nearly opposite effects on total burn area and fire frequency.

It is important to note that the observations here are only for the time span 1/1/1990 to 12/31/2006, and are only for wildfires on CalFire protected lands, so further investigation may be necessary to confirm these results more generally. Of course, wildfire ecology is extremely complex (Pyne et al. 1996; Johnson and Miyanishi 2001; Moritz et al. 2005), and several important covariates, such as windspeed, precipitation, drought indices, bark beetle populations, cloud cover, etc., may act as confounding factors in our analysis. Future work may include the use of Burning Index (BI) and/or the variables determining the BI as covariates,

though some previous analyses have shown the BI not to be an accurate predictor of wildfire incidence or area burned in southern California (Schoenberg et al. 2007, Schoenberg et al. 2009, Xu and Schoenberg 2011). In addition, the model with total area burned as the response had highly skewed residuals and several highly influential observations. Further, while the completeness of this dataset is thought to be rather exceptional, some missing observations may have influenced the results here. The results here should be tested and confirmed in the future with other datasets and for other spatial-temporal regions.

REFERENCES

- [1] Arno, S.F., and Allison-Bunnell, S. (2002). *Flames in our Forest: Disaster or Renewal?* Island Press, Washington, D.C.
- [2] Cumming, S. (2001). *A parametric model of the fire-size distribution*. Can. J. For. Res. 31, 1297- 1303.
- [3] Elston, D. A. and R. Moss and T. Boulinier and C. Arrowsmith and X. Lambin. (2001) *Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks*. Parasitology 122(5): 563-569
- [4] Johnson, E.A. and Miyanishi, K. (2001). *Forest Fires: Behavior and Ecological Effects*. Academic Press, San Diego.
- [5] Keeley, J.E. (1982). *Distribution of lightning- and man-caused wildfires in California*. Proceedings of the Symposium on Dynamics and Management of Mediterranean-type Ecosystems, C.E. Conrad and W.C. Oechel, editors, USDA Forest Service, Pacific Southwest Forest and Range Experiment Station, General Technical Report PSW-58, pages 431–437.
- [6] Keeley, J.E. (2009). *Fire intensity, fire severity and burn severity: a brief review and suggested usage*. International Journal of Wildland Fire 18, 116-126.
- [7] Keeley, J.E., Fotheringham, C.J., and Morais, M. (1999). *Reexamining fire suppression impacts on brushland fire regimes*. Science 284, 1829-1832.
- [8] Keeley, J.E., Safford, H., Fotheringham, C.J., Franklin, J., and Moritz, M. (2009). *2007 southern California wildfires: lessons in complexity*. Journal of Forestry Sept., 287-296.
- [9] Malamud, B.D., Millington, J.D.A., and Perry, G.L.W. (2005). *Characterizing wildfire regimes in the United States*. Proceedings of the National Academy of Sciences of the USA 102 (13), 4694-4699.
- [10] Miller, J. D., H.D. Safford, M. A. Crimmins, and A.E. Thode. (2009) *Quantitative evidence for increasing forest fire severity in the Sierra Nevada and southern Cascade Mountains, California and Nevada, USA*. Ecosystems 12, 16-32
- [11] Moritz, M.A., Morais, M.E., summerell, L.A., and Doyle, J. (2005). *Wildfires, complexity, and highly optimized tolerance*. Proceedings of the National Academy of Sciences of the USA 102, 17912-17917.
- [12] Nichols, K., Schoenberg, F.P., Keeley, J., and Diez, D. (2011). *The application of prototype point processes for the summary and description of California wildfires*. Journal of Time Series Analysis 32(4), 420-429.

- [13] Pyne, S., Andrews, P., and Laren, R. (1996). *Introduction to Wildland Fire*. Wiley, New York.
- [14] Preisler, H.K, Westerling, A.L., Gebert, K.N., Munoz-Arriola, F., and Holmes, T.P. (2011). *Spatially explicit forecasts of large wildland fire probability and suppression costs for California*. Int. J. Wildland Fire 20, 508-517.
- [15] Schoenberg, F.P., Peng, R., and Woods, J. (2003). *On the distribution of wildfire sizes*. *Environmetrics* 14(6), 583-592.
- [16] Schoenberg, F.P., Chang, C., Keeley, J., Pompa, J., Woods, J., and Xu, H. (2007). *A critical assessment of the burning index in Los Angeles County, California*. International Journal of Wildland Fire 16(4), 473-483.
- [17] Schoenberg, F.P., Pompa, J.L., and Chang, C. (2009). *A note on non-parametric and semiparametric modeling of wildfire hazard in Los Angeles County, California*. Environmental and Ecological Statistics 16(2), 251-269.
- [18] Stephens, S. L. (2005). *Forest fire causes and extent on United States Forest Service lands*. International Journal of Wildland Fire. 14, 213-222.
- [19] Westerling A.L., Brown, T.J., Gershunov, D.R., Cayan D., Dettinger, M.D. (2003). *Climate and wildfire in the Western United States*. Bulletin of the American Meteorological Society 84, 595-604.
- [20] Westerling, A. L., H. G. Hidalgo, D. R. Craven, and T. W. Swetnam. (2006) *Warming and earlier spring increase Western U.S. forest wildfire activity*. Science 313, 940-943.
- [21] Xu, H., and Schoenberg, F.P. (2011). *Point process modelling of wildfire hazard in Los Angeles County, California*. Annals of Applied Statistics 5(2a), 684 - 704