Title

Advancements in Modeling Forest Fires with the Stoyan-Grabarnik Statistic

Author

Hollister, Brooke

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Advancements in Modeling Forest Fires with the Stoyan-Grabarnik Statistic

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

by

Brooke Maressa Hollister

2024

ABSTRACT OF THE THESIS

Advancements in Modeling
Forest Fires with the
Stoyan-Grabarnik Statistic

by

Brooke Maressa Hollister
Master of Science in Statistics
University of California, Los Angeles, 2024
Professor Frederic R. Paik Schoenberg, Chair

Spatio-temporal point processes are a common method to analyze data that involves event occurrences in space and time, such as wildfires. Model parameters for a point process are typically fitted using maximum likelihood estimation, which finds parameter values that maximize the probability of observing the data according to the specified model. This method, however, often involves finding a complex and non-closed-form integral. The Stoyan-Grabarnik (SG) statistic is a way to find model parameters for a spatial point process that is faster and easier than maximum likelihood estimation and does not require computing or approximating a computationally intensive integral. This work uses the SG statistic methodology to estimate model parameters for forest fire ignitions occurring in National Forest System lands in California between 2008-2012. The models utilize covariates such as precipitation, wind speed, temperature, and evaporation and are evaluated for a variety of subsets of the data, including size and cause over northern and southern California. The results show that modeling accuracy is not compromised while also revealing interest-

ing patterns in the relationship between fire ignitions and weather conditions. Results in this work could help advance modeling efficiency and provide insights pertinent to fire risk management.

The thesis of Brooke Maressa Hollister is approved.

Qing Zhou

Hongquan Xu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2024

To my mom, whose words of
encouragement empowered me
to keep pushing forward.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

(Acknowledgments omitted for brevity.)

CHAPTER 1

Introduction

Point process models are used to describe patterns of event occurrences, or points, that happen spatially, such as earthquakes, fires, or crime [Oga98, NSK13, KS23]. Maximum likelihood estimation is a common method for estimating parameters that govern the conditional intensity in a point process model [Rei05]. The parameter vector of a point process N can be estimated by Equation 1.1 [KS23], where λ represents the conditional intensity (the rate of events per unit area $\times$ unit time) for a realization of the point proccess N $\{(t_i, x_i, y_i,)\}_{i=1}^{n} = \{\tau_i\}_{i=1}^{n}$.

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \left(\sum_i log\lambda\left(\tau_i; \theta\right) - \int_0^T \int \int \lambda\left(\tau; \theta\right) \, dt \, dx \, dy \right) \tag{1.1}$$

More specifically $\lambda\left(\tau\right)$ is defined as:

$$\lambda\left(\tau\right) = \lim_{h \downarrow 0} \frac{\mathbb{E}[N(t, t+h)]|H_t}{t} \tag{1.2}$$

This method, however, can be computationally intensive due to the integral in the likelihood function. It has been shown that the Stoyan-Grabarnik (SG) statistic in Equation 1.3 is a consistent estimator, under general conditions, for parameters of a point process on a spatio-temporal region $\mathcal{I}$ [KS23]. The SG statistic also does not require computation or approximation of an integral term [KS23].

$$S_{\mathcal{I}}\left(\theta\right) = \sum_{i:\tau_i \in \mathcal{I}} \frac{1}{\lambda\left(\tau_i; \theta\right)} \tag{1.3}$$

The SG statistic is the inverse of the conditional intensity, which is the instantaneous rate of events that occur at a given time. One useful property of the SG statistic is that the

expectation of the sum of the SG statistic happens to be the volume of the spatial region [SG91].

$$\mathbb{E}[S_{\mathcal{I}}(\theta)] = \mathbb{E}[\sum_{i:\tau_i \in \mathcal{I}} \frac{1}{\lambda(\tau_i;\theta)}] = |\mathcal{I}_j| \tag{1.4}$$

The SG stat was first proposed in 1991 as a measure for the mean value of marks for a marked Gibbs point process [SG91]. A marked Gibbs point process is a point process that can account for interactions, dependency, and repulsion between points where each point has an associated magnitude, or a mark [GSG96]. For example, a mark for an earthquake point process could be the magnitude of the earthquake. Since its introduction, the SG statistic has been recommended as a goodness-of-fit diagnostic for assessing the goodness-of-fit for a spatial point process [BTM05] and determining the optimal bandwidth for kernel smoothing when estimating the conditional intensity for a spatial point process [CL18]. More recently the SG statistic was proposed by Kresin and Schoenberg [KS23] to be consistent for estimating parameters of a spatial-temporal point processes, which has been the main motivation for this work. They show that when the spatio-temporal region $\mathcal{I}$ is divided into subregions, or cells, model parameters can be fit optimally by minimizing the sum of squares of the integral of the spatial distribution of residuals across all cells [KS23].

$$\tilde{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{j=1}^{p} \left(\sum_{i:\tau_i \in \mathcal{I}_j} \frac{1}{\lambda(\tau_i;\theta)} - \mathbb{E}\left[\sum_{i:\tau_i \in \mathcal{I}_j} \frac{1}{\lambda(\tau_i;\theta^*)} \right] \right)^2 \tag{1.5}$$

In Equation 1.5, the SG statistic is summed over all points in a given cell and the expected value of the SG statistic is summed over all points. Then the sum of all of the differences over all of the cells is minimized to get the parameter vector $\tilde{\theta}$. This allows the parameters to be adjusted according to the best fit between the observed data and the model predictions. Kresin and Schoenberg demonstrated a few examples of applying this technique to model crime and earthquakes [KS23].

This study aims to showcase the effectiveness of the SG stat in the context of hind-casting California fires on National Forest System lands for 2008-2012. Forest fires have become more intense and more frequent in the western US due to climate change, resulting in major

property loss, ecological damage, and threats to human safety [SJA20]. There are a number of ways that fire and land management agencies, like Cal Fire, assess fire risk. Currently, Cal Fire, relies on Fire Hazard Severity Zone maps to assess wildfire risk [Autnd]. Although these maps are comprehensive, they are also highly complex and require substantial resources to create. Another common measure for wildfire is the Burning Index (BI). The predictive power of the BI, however, may be limited [PSW05, XS11]. Point process methods have emerged as an approach for wildfire modeling due to their theoretical foundation and ability to account for inherent dynamic spatial patterns [SJV13, NSK13, JMS12]. Recent approaches offer their own unique strategies, but they seem to be more computationally intense than using the SG statistic, or they have variable results and high uncertainty [QDM21, YWD15, SSM13].

In this work, the conditional intensity (λ) represents the rate of forest fire ignitions per km $\times$ year. Various weather covariates, such as precipitation and wind, are included in the mathematical model that describes λ. In addition to demonstrating the performance of modeling with the SG statistic, hopefully, the results will provide insight into the spatial distribution patterns of forest fire ignitions across California and shed light on the potential driving factors of forest fire ignitions. Furthermore, the insights gained from this research could be informative for future fire risk assessment and land management practices [WHC06]. This project was started by Abigail Coelho, who performed the initial data pre-processing and modeling [Coe23]. I have continued her work by refining some methods to improve model convergence and residuals. Additionally, I have recalculated the precipitation covariate measure to enhance the accuracy of the models.

CHAPTER 2

Data

2.1 Forest Fire Data

Forest fire occurrence data is from the US Department of Agriculture Forest Service [USD24]. There are 329,480 instances of forest fires in this dataset in the US from 1987 to 2021. Only fires in California for 2008-2012 have been selected arbitrarily for manageability, leaving 7179 entries. The relevant variables selected for this project include ignition location in latitude and longitude, ignition date, cause (lightning, campfire, miscellaneous, children, debris burning, smoking, arson), and acres burned. Note that the fires recorded in this dataset are strictly from National Forest System lands. Consequently, wildfires that occurred outside of National Forest System lands are not included in this work. Although this reduces the scope of the work, the central focus on assessing the accuracy of using the SG statistic remains unchanged, prioritizing this goal over forest fire prediction.

After subsetting the California fires for the years 2008-2012, rows that did not have all of the information for ignition date, location, and cause were dropped. Modeling was performed for the following combinations of subsets of the data: unnatural/human-caused (n = 4489), natural/lightning caused fires (n = 2690), northern (n = 4197), southern california fires (n = 2982), as well as fires larger than 1 acre (1195), and fires larger than 10 acres (n = 496). In the analysis, data with "lightning" as the cause is considered a natural-caused fire, while fires with any other cause, were designated as unnatural. Figure 2.1 shows the ignition locations of natural and unnatural caused fires. According to the National Park Service, about 85%

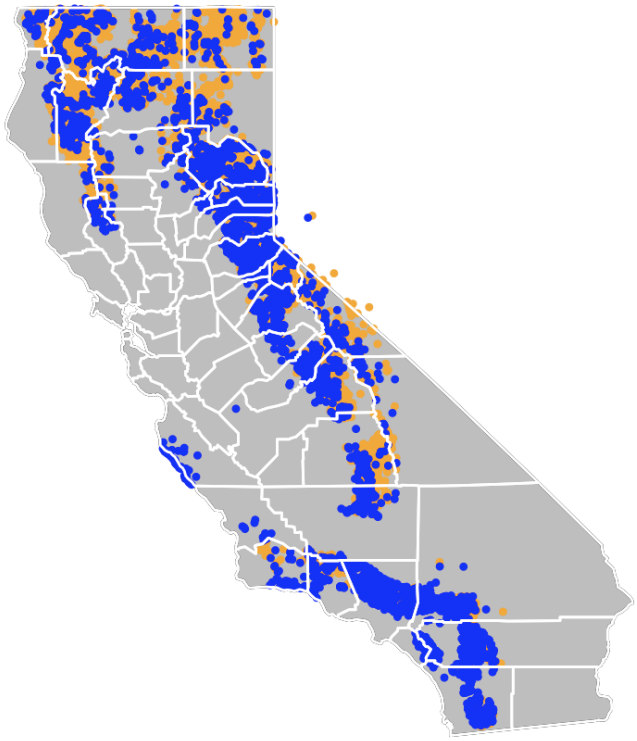


Figure 2.1: Lightning-caused fires in orange (n = 2690) and non-lightning caused fires in blue (n = 4489)

of fires in the US are caused by humans, so fires with a "Miscellaneous" cause were grouped into the "unnatural" category as well (US Forest Service Research Data Archive).

2.2 Weather Data

The weather dataset, from 1,233 National Oceanic and Atmospheric Administration stations in California, was used to determine the weather conditions associated at the location and on the date of each fire ignition [Nat24]. The variables selected from this data set for modeling are monthly precipitation, maximum temperature, minimum temperature, average wind speed, and evaporation. These variables were selected due to their association with fire behavior [al17, BFL14, KSF09]. Fuel composition and topographical features have also

shown to be important factors in wildfire ignition and spread [SRK07]. This study, however, focuses mainly on demonstrating the efficacy of the Stoyan Grabarnik statistic for modeling point processes. Therefore, these factors will not be examined since they are beyond the scope of this paper, but they would be interesting factors to examine in future work.

Entries for temperature and wind speed that were beyond the .999 and .001 quantile were removed, as evaporation and precipitation exhibit more natural variability [Coe23]. Since the lack of accumulated precipitation, leading to dry conditions [Jon10, al17], has a significant effect on wildfire ignitions, a 4 day rolling precipitation calculation was used in Coelho's modeling [Coe23]. In this work, the 4 day rolling average has been changed to a seasonal precipitation calculation that sums rain from November to March, which is typical rainy season in California [LDC18], preceding a fire. For fires that happened in January, February, or March, precipitation is calculated from the preceding November to the previous month. For example if a fire happened in February 2011, then the rainy season was calculated using the period from November 2010 to January 2011.

2.3 Merging Fire and Weather Data

A k-dimensional tree was used to connect each fire with the nearest weather station and its corresponding weather readings on the day of the fire. A k-dimensional tree is a data structure that organizes points along given dimensions. In this case, the tree organizes the fire origins and weather stations according to latitude, longitude, and date. This method is less computationally intensive than calculating distances between every station and every fire.

Not all stations have equipment to record all of the weather variables and not all stations record every day. Therefore, the closest stations to a given fire with the most recent measurements available were selected for missing variables. Therefore some fires may only have an approximation of the weather conditions on that day. Figure 2.2 shows the number

Weather Variable	Number of Stations Recording Variable
Average wind speed (AWND)	83
Maximum Temperature (TMAX)	724
Minimum Temperature (TMIN)	724
Evaporation (EVAP)	17
Seasonal Precipitation (PRCP_SZN)	1549

Figure 2.2: Number of stations that recorded each weather variable.

of stations that recorded each weather variable [Coe23, Nat24].

CHAPTER 3

Methods

In order to model forest fires occurrences as a point process with the SG statistic, a spatial division framework is needed. The region of California was divided into 100 cells on a 10 x 10 grid, each cell, or bin, being 104.607 km by 114.263 km, or 11952.710 km^2. Dividing the region into cells enables parameters to be estimated more locally and allows for the capture of data patterns in that cell. The grid partitioning scheme is arbitrary, a notion supported by Kresin and Schoenberg [KS23]. Out of these 100 cells, only 40 cells contained fires.

The chosen model for the conditional intensity, or the instantaneous rate of wildfire ignitions given a time t and location x, y, is

$$\lambda\left(t, x, y\right) = \gamma m(x, y) + \alpha W\left(t, x, y\right) \tag{3.1}$$

where $m(x, y)$ is a background rate and $W(t, x, y)$ is a linear combination of weather covariates. The background rate serves as an indicator of the typical frequency of fire ignitions within a specific region over time. It can be thought of as a way to encapsulate natural periodicity and seasonality of fire behavior [PSW05]. The background rate was calculated with a min-max normalization of the count of wildfire occurrences in each cell over the years 2005 – 2007. There were a total of 5,036 fire instances across all of the cells during this period. These years are omitted in the model estimation, so that the background rate and covariate effects may be estimated separately, as in [Oga98]. The remaining 7,179 fires during 2008-2012 are used for model estimation.

Parameters are estimated by minimizing the sum of squared differences between the SG statistic and its expected value. As shown below in Equation (3.2), the expected value of the

SG stat is the size of the spatio-temporal region $|I_j|$, or in other words, each 11952.710 km^2 cell over the period 2008-2012.

$$\begin{aligned}\tilde{\theta} &= \underset{\theta\in\Theta}{\text{argmin}} \sum_{j=1}^{p} \left(\sum_{i:\tau_i\in\mathcal{I}_j} \frac{1}{\lambda(\tau_i;\theta)} - \mathbb{E}\left[\sum_{i:\tau_i\in\mathcal{I}_j} \frac{1}{\lambda(\tau_i;\theta^*)} \right] \right)^2 \\ &= \underset{\theta\in\Theta}{\text{argmin}} \sum_{j=1}^{p} \left(\mathcal{S}_{\mathcal{I}_j}(\theta) - \mathbb{E}\left[\mathcal{S}_{\mathcal{I}_j}(\theta^*) \right] \right)^2 \\ &= \underset{\theta\in\Theta}{\text{argmin}} \sum_{j=1}^{p} \left(\mathcal{S}_{\mathcal{I}_j}(\theta) - |\mathcal{I}_j| \right)^2 \end{aligned} \tag{3.2}$$

Once the intensity function $\lambda(\tau_i;\theta)$ is calculated for each point using the weather data and the inverse is summed for all points in the cell, the expected value of the inverse conditional intensity is subtracted off. Since the expected value of $\mathcal{S}_{\mathcal{I}_j}(\theta^*)$, which is $\mathcal{I}_j$, happens to be the size of the cell times number of years, this computation is relatively straightforward. This is repeated for each cell and then the squared differences for each cell are summed. This function is minimized iteratively according to the algorithms in the 'optim' function in R to get the parameter estimates.

Model testing begins by evaluating the performance of a model with just one variable, such as the background rate. The starting parameter estimate for this variable is set as 1 in 'optim'. Then optimization is performed iteratively using parameter estimates from the last run until the parameter estimate converges, or until the SG statistic stops decreasing. This approach ensures the parameter space is adequately explored while utilizing a stopping criterion based on the SG statistic. Then, additional variables, such as maximum temperature, are progressively incorporated one at a time, with each new initial parameter estimate set to 1. Table 3.1 provides a list of the models with variables in the order that they were added. The impact on the model's predictive ability is assessed according to the same criterion mentioned before. Throughout this process, different combinations of variables are tested, sometimes adding new variables and at other times removing some, to determine the most effective set of predictors for the conditional intensity. Most models converged quickly

and only required 2 iterations. Modeling was performed for the following subsets of data: All fires, natural caused, unnatural caused, northern CA fires, northern natural, northern unnatural, southern CA, southern natural, southern unnatural, fires larger than 1 acre, and fires larger than 10 acres.

Variables
Bkgd
Bkgd, awnd
Bkgd, awnd, tmax
Bkgd, awnd, tmax, evap
Bkgd, awnd, tmax, evap, tmin
Bkgd, awnd, tmax, evap, tmin, prcp
Bkgd, awnd, tmax, evap, tmin, prcp, constant
Bkgd, constant
Bkgd, constant, awnd,
Bkgd, constant, awnd, evap
Bkgd, constant, awnd, evap, tmax
Bkgd, constant, awnd, tmax
Bkgd, constant, prcp
Bkgd, constant, tmax
Bkgd, constant, tmax, prcp
constant

Table 3.1: Example of models tested in sequential order.

A test data set of 100 random fire points (with dates and coordinate locations) was generated for each region for the purpose of model evaluation. Like the actual data, every point was matched to weather conditions with a k-dimensional tree. The conditional intensities

for the test data set was computed based on the same models for the actual data set for each fire incident. The intensities were aggregated by region and then average intensity for each region was computed. Each model's performance is evaluated based on the mean absolute residual. The mean absolute residual provides a summary estimate of the discrepancy between the model's predictions and the actual rate of fire ignitions. The mean absolute residual was computed by first subtracting the actual rate of fire ignition (actual intensities) from the predicted rate of fire ignitions (predicted intensities) for each region. Then all of the absolute differences were averaged.

CHAPTER 4

Results

4.1 Rate of Ignited Fires Statewide

Of all the models that were designed to predict forest fires in California (n = 7179), the model that incorporates the background rate, a constant term, and the maximum temperature variable, performs the best in terms of mean absolute residual (Table A.2). The mean absolute residual represents the mean of the differences between the actual conditional intensities of forest fires for each region and the predicted intensities from the testing data set for each region. Recall that the conditional intensity is the rate of fires per unit area per year. The calculations in the code I inherited were in terms of megameters, therefore all of the tables are also in terms of megameters as the measurement for area. For the sake of consistency and ease of building upon Coelho's work [Coe23], I opted to retain the parts of the code using megameters. It is also a little bit easier on the eyes to interpret the residuals when they are expressed in megameters, as they tend to be small decimals when expressed in kilometers. Therefore, the conditional intensity is the rate of fire ignitions per megameter squared per year, or simply divide by 1,000,000 and intensity will be per kilometer squared per year.

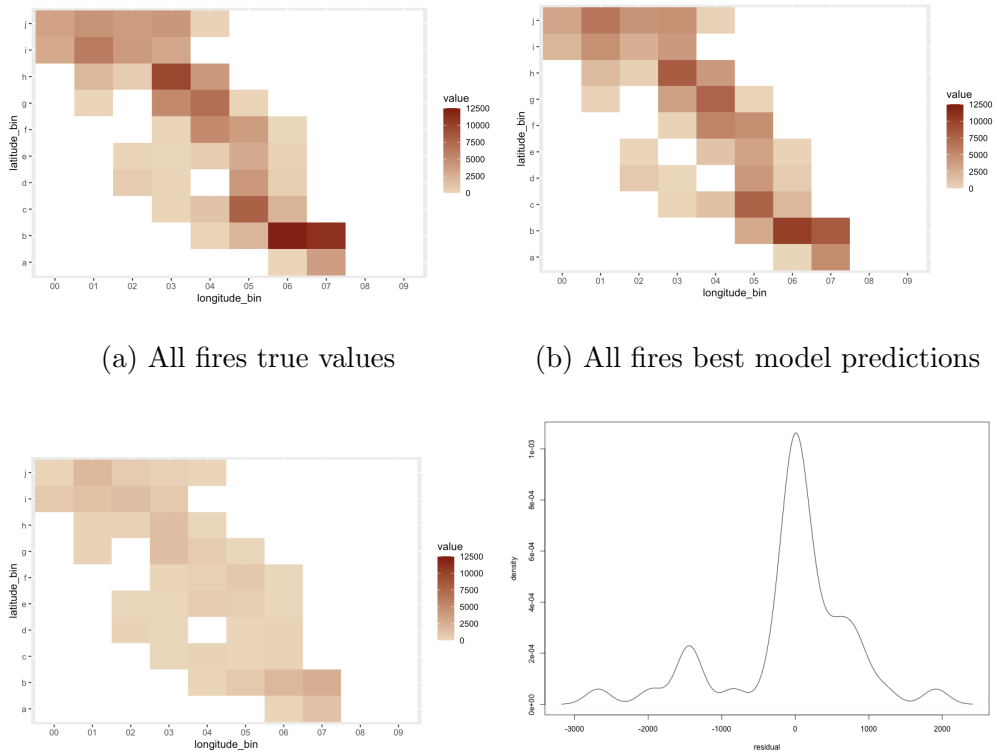
To provide context for interpreting the residuals, there are about 35 fire ignitions per region per year. The actual mean conditional intensity across all 40 regions is about 3000 ignitions per megameter squared per year, translating to about .003 ignitions per kilometer squared per year. The standard deviation among all the conditional intensities across the

all regions is about 3170 per megameter squared per year, which is reasonable considering that some regions in California are more prone to fires than other regions due to differences in terrain and population density.

The constant model assigns one uniform conditional intensity for all regions, making it a fair benchmark for assessing the effectiveness of each model. In terms of mean absolute residual, the best model performs 86% better than the constant model and has about an 80% better SG statistic (Table A.2). Figure 4.1 shows heatmaps of the actual, predicted, and residuals of the intensities by region for the best statewide model. The density plot of the residuals is shown in Figure 4.1d. The heatmaps show that the predictions are a close match to the actual intensities, with the exception of 2 white region cells. The test data set did not have points in these regions, therefore there are no predictions for them. There were few fires to begin with in these regions, so it is not expected to significantly affect the modeling performance. Additionally, the density curve of the residuals is mainly centered around 0, with some overestimates.

For the natural-caused fires, the best performing model incorporates the background rate, average wind speed, and maximum temperature. This model has a 43% better mean absolute residual and about 48% better SG statistic value (Table A.3). Figure 4.2 shows the true values of the conditional intensities of natural-caused fires. Figure 4.3 demonstrates a progression of improvement in residuals as variables are added to or subtracted from the model.

For unnatural-caused fires, the model with background rate, a constant term, seasonal precipitation, and maximum temperature has the lowest residual (Table A.4). Among this subset of data, this model improves the constant model's residual by 83% and the SG statistic by 70% (Table A.4). The average residual, without counting the constant model's residual, for natural fires is 835 and 656 for unnatural fires. The average SG statistic was .02496 for natural versus .02795 for unnatural. There were 1799 more data points, however, for unnatural caused fires.



(a) All fires true values

(b) All fires best model predictions

(c) All fires best model residuals

(d) Residual density

Figure 4.1: Best performing model predictions, residuals, and residual density curve for statewide fire rates.

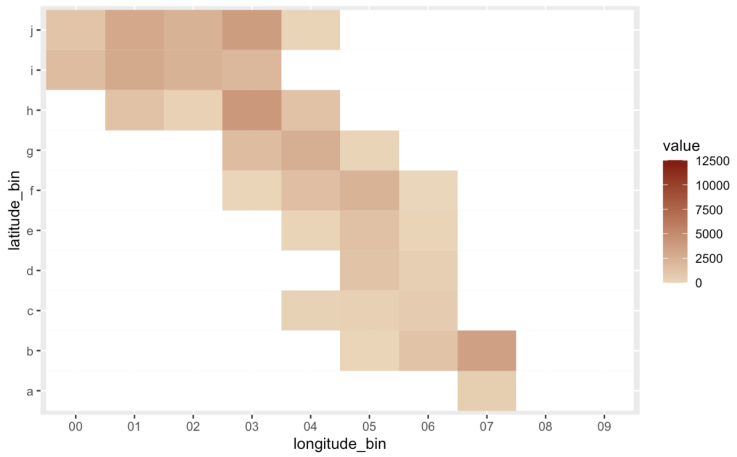
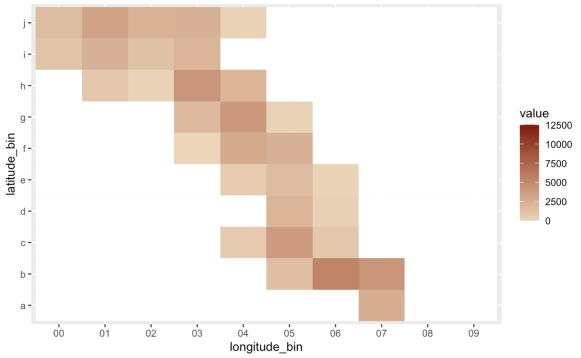
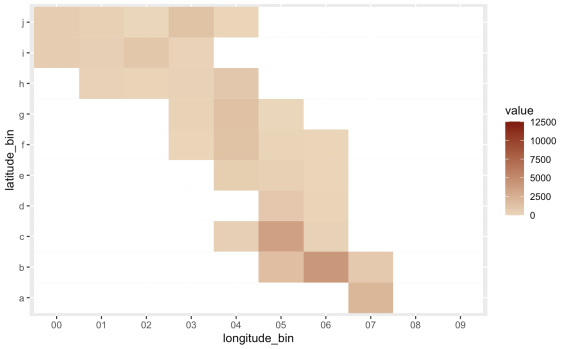


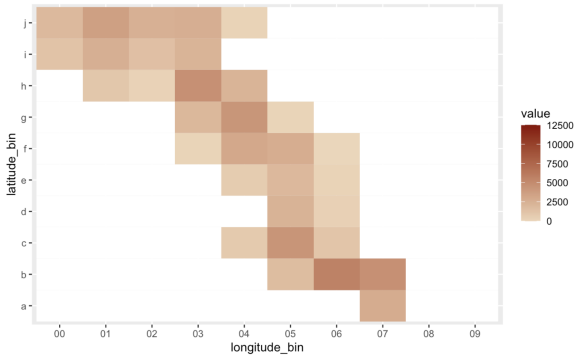
Figure 4.2: True values for all natural fires.



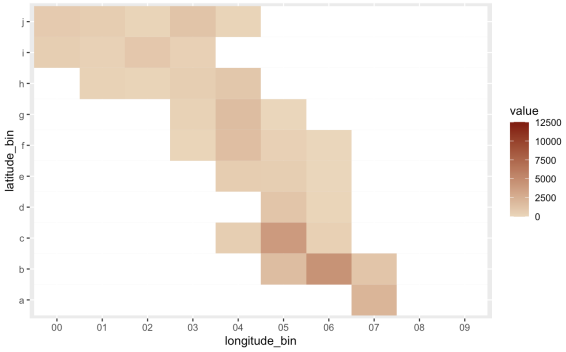
(a) all natural best model



(b) Mean Abs Residual: 773

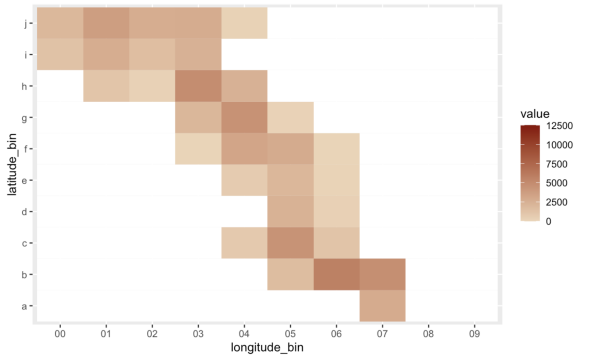


(c) all natural just BKGD model

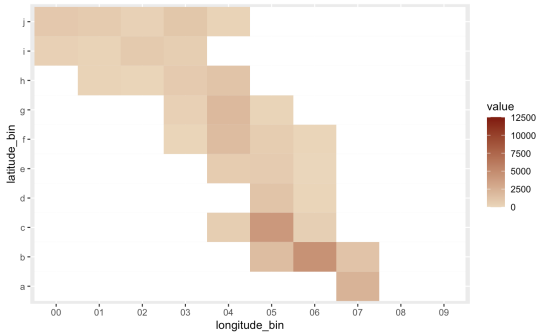


(d) Mean Abs Residual: 852

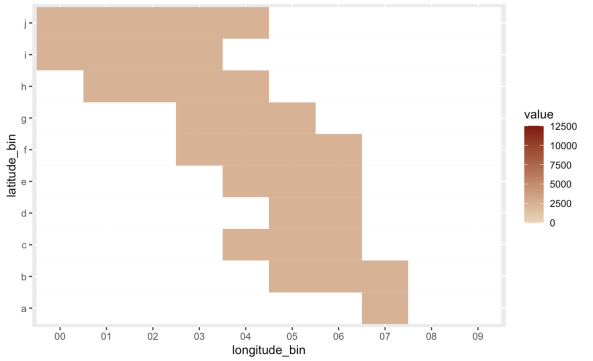
Figure 4.3: Sample of models for statewide natural subset (continued on next page)



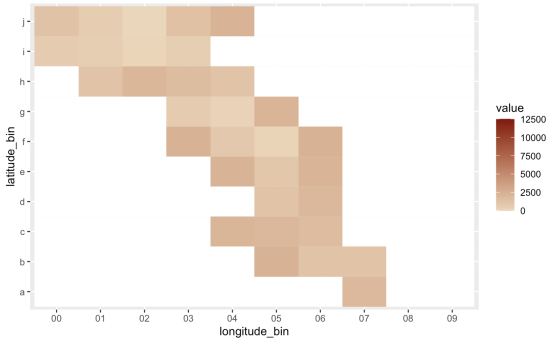
(e) all natural poor model



(f) Mean Abs Residual: 917



(g) all natural constant model



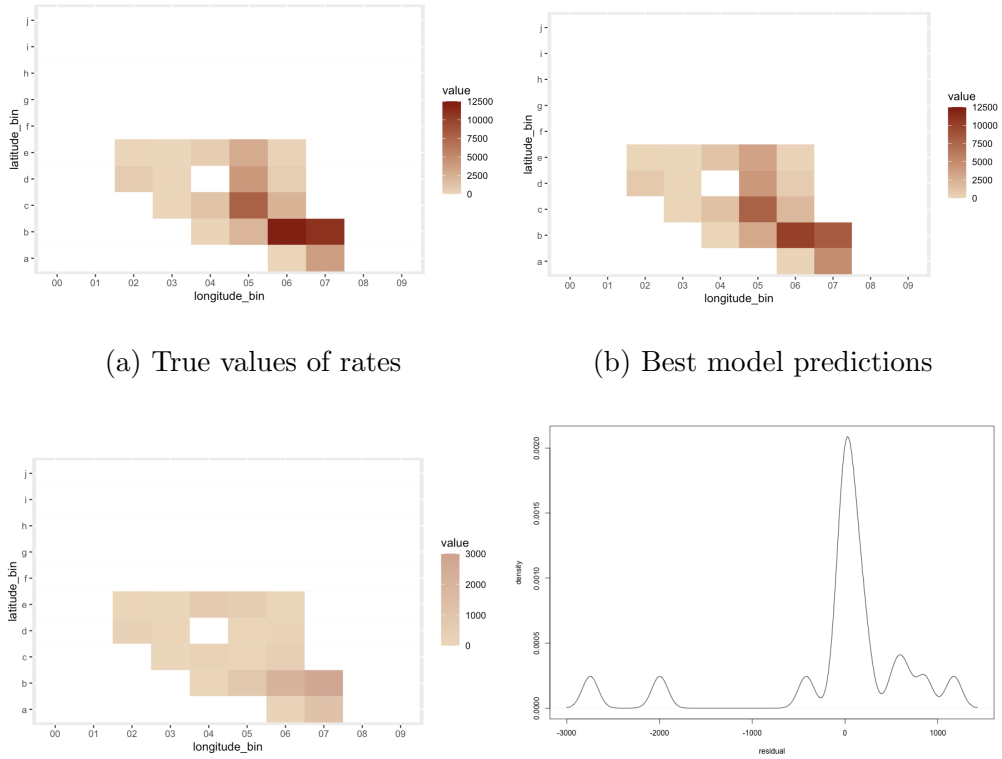
(h) Mean Abs Residual: 1368

Figure 4.3: (continued)

4.2 Rate of Southern CA Fires

Of the models that predict fire ignitions for southern California, the model using the background rate, average wind speed, maximum temperature, evaporation, minimum temperature, seasonal precipitation, and a constant term performs the best (Table A.5). The residual for this model is 488 fires per megameter per year, which improves upon the constant model's residual by 92%. This model improves on the constant model's SG statistic by 92%. Figure 4.4 shows heatmaps of the actual and predicted conditional intensities, as well as their

residuals. Residuals are mostly centered around 0, but there seem to be a few outliers that are experiencing underestimation (Figure 4.4d).



(a) True values of rates

(b) Best model predictions

(c) Residuals

(d) Residuals density

Figure 4.4: Best performing model predictions, residuals, and residual density curve for Southern CA fire rates.

For the subset of southern CA natural-caused fires, the model with the lowest residual (602) incorporates the background rate, evaporation and a constant term (Table A.6). This model has a 55% better residual and 52% better SG statistic than the constant model. Among the southern CA unnatural-caused fires, background rate, a constant term, average wind speed, maximum temperature, and seasonal precipitation proved to be the most effective covariates (Table A.7). The resulting residual for this model was 493 and the SG stat was .00916, which is 90% and 80% better, respectively, than the constant model. The mean residual for all of the southern CA unnatural fire models (516) tend to be lower than the

mean residual for the natural fire models (720). The SG statistics also tend to be slightly lower for unnatural fires (.01163 vs .01172).

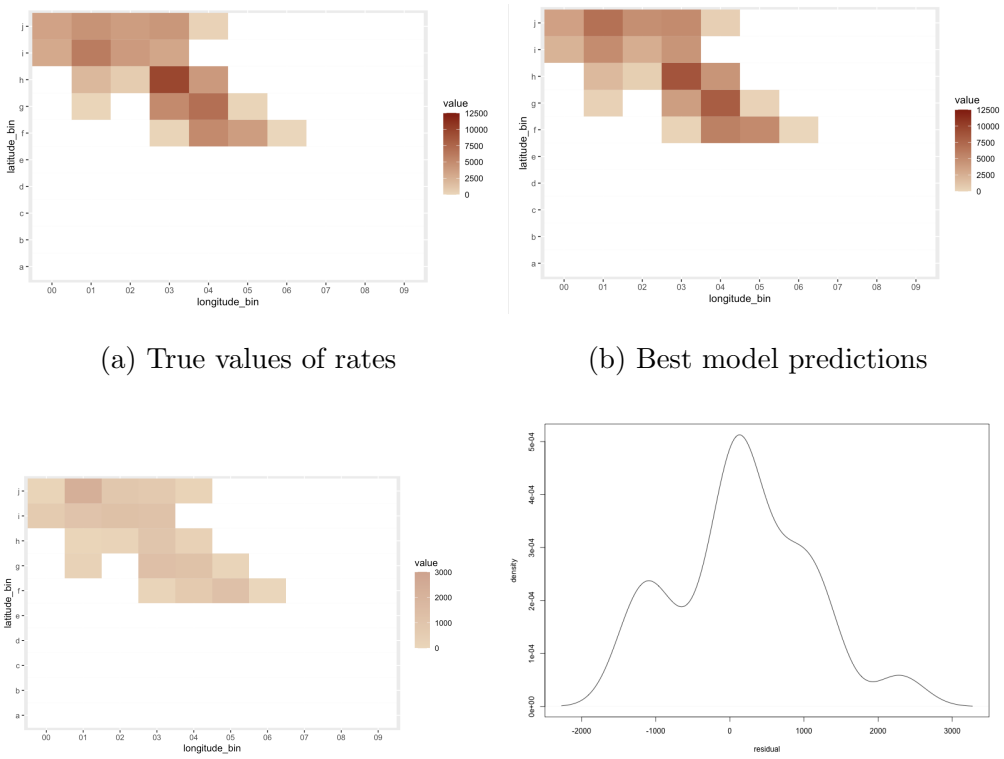
4.3 Rate of Northern CA fires

For the subset of northern California fire data, the model with just the background rate performs the best with a residual of 704 and SG statistic of 0.00812 (Table A.8). The heatmaps in Figure 4.5 show the difference in performance between the realized conditional intensities versus the predicted intensities according to this model. The density of residuals (Figure 4.5d) are centered around 0, with some weight around -1000, suggesting some underestimation. Compared to the constant model, the background rate model has a 73% better residual value and 70% better SG statistic.

In the case of natural-caused fires in northern CA, the model with the lowest residual (547) incorporates the background rate, average wind speed, and maximum temperature (Table A.9). The SG statistic for this model is 0.00801. Conversely, the most effective covariates to model unnatural-caused fires seem to be the background rate, seasonal precipitation, maximum temperature, and a constant (Table A.10). This model had a residual of 636 and an SG stat of 0.01073. The unnatural fire models have a slightly higher mean residual and SG stat of 666 and 0.01039, respectively, as compared to the natural fire models that have a mean residual of 610 and SG stat of 0.00777.

4.4 Rate of Large Fires

When sub setting the data for fires that grew larger than 1 acre, the model that shows the lowest residual (217) incorporates the background rate, a constant term, average wind speed, evaporation rate, and maximum temperature (Table A.11). The SG stat for this model is 0.02863. The heatmaps show a close resemblence between the acutal and predicted



(a) True values of rates

(b) Best model predictions

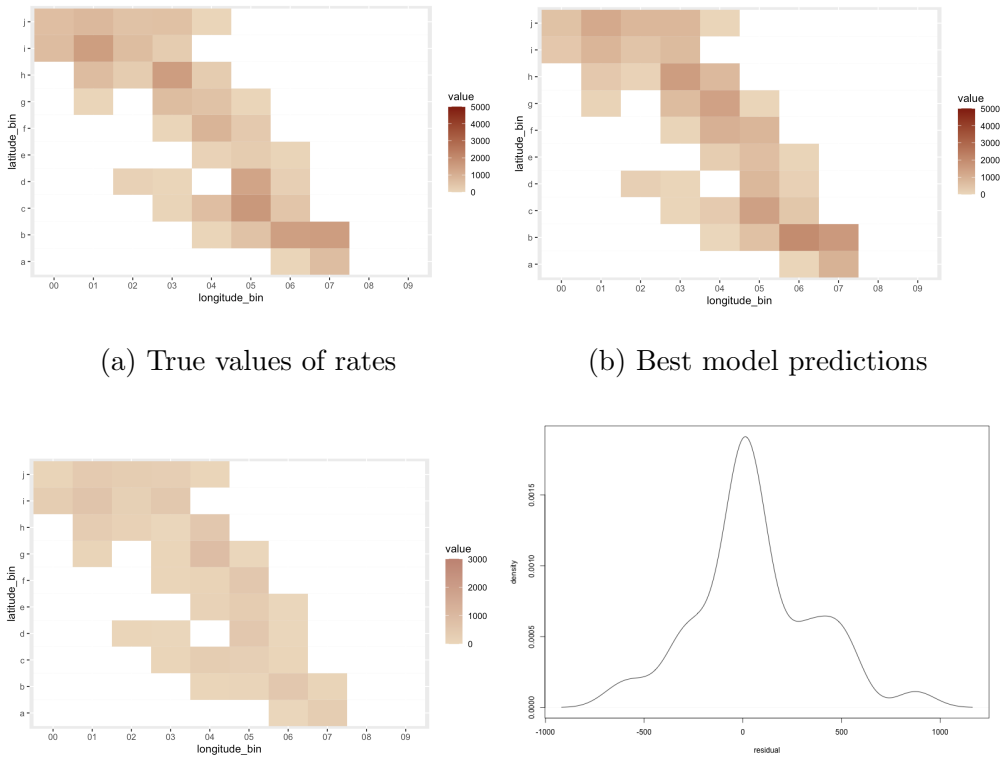
(c) Residuals

(d) Residual density

Figure 4.5: Best performing model predictions, residuals, and residual density curve for Northern CA fire rates.

intensities (Figure 4.6). The residuals appear normally distributed (Figure 4.6d).

Among fires that were larger than 10 acres, the best model uses the background rate, a constant, seasonal precipitation, and maximum temperature covariates (Table A.12). This model resulted in a mean absolute residual of 119 fire ignitions per megameter per year and an SG stat of 0.03609. The residuals appear to be reasonable for this subset as well (Figure 4.7).



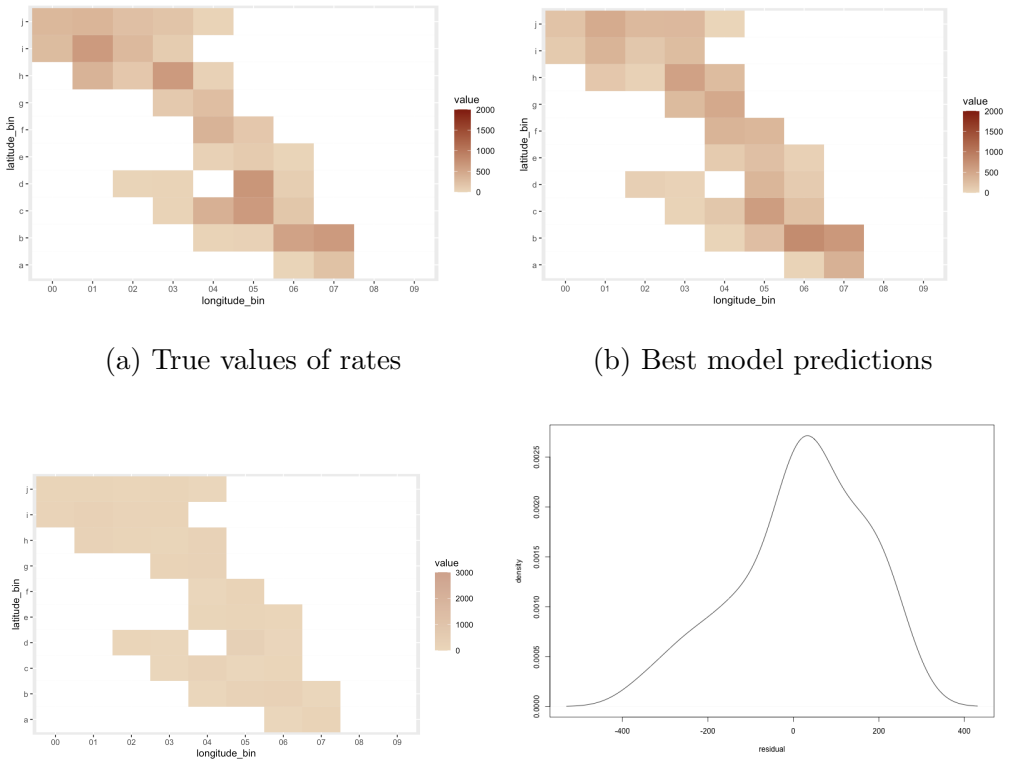
(a) True values of rates

(b) Best model predictions

(c) Residuals

(d) Residual density

Figure 4.6: Best performing model predictions, residuals, and residual density curve for > 1 acre fire rates.



(a) True values of rates

(b) Best model predictions

(c) Residuals

(d) Residual density

Figure 4.7: Best performing model predictions, residuals, and residual density curve for > 10 acre fire rates.

CHAPTER 5

Discussion

It is interesting how influential the background rate is for all of the models, carrying most of the weight in out of all of the covariates. This is unsurprising, however, considering that the background rate was calculated using a min-max normalization of counts of fire occurrences data from 2005-2007, and therefore encapsulates seasonal and regional patterns of the rates of fire ignitions.

Another notable result is that the best models for all subsets of unnatural fire models incorporate seasonal precipitation and maximum temperature covariates. This could be attributed to factors like camping accidents, which may be more likely to happen in better (warmer, drier) weather for camping, or accidental ignitions (from discarded cigarette butts, fireworks, etc.).

In the case of natural fires, both in northern CA and statewide, the best models incorporate average wind speed and maximum temperature covariates. This may indicate that either these conditions are more conducive to lightning or these conditions are more conducive to fire ignitions resulting from lightning. Southern CA natural fires, on the other hand, seem to be better predicted by evaporation. There are a greater percentage of natural fires in northern CA (50%) than in southern California (20%) in this data set, suggesting that lightning caused fires may be more common in northern CA.

Other than the greater than 1 and 10 acre subsets, the model seems to predict southern CA fires most accurately, potentially because unnatural fires make up a large majority of fires (80%) in that region. Perhaps the model captures conditions conducive to igniting or

creating fire-prone conditions in forested areas in southern California.

On a different note, there might be better ways to generate a test data set to evaluate the models. Generating 100 random fire locations and points for each region might lead to over or under-estimating fire counts for certain regions. Moreover, the years 2007-2009 were some of California's driest years in the state's hydrologic record [Jon10], which might influence the weather data and, subsequently, impact model accuracy. Another potential source of inaccuracy stems from the scarcity of fires larger than 1 acre in this data set, which may result in some bias when predicting smaller fires. Additionally, evaporation readings may be inaccurate for most fires since only 17 weather stations recorded evaporation.

Finally, it is worth noting the constant model is not a perfect benchmark, since the constant model's residual and SG statistic might already be performing significantly better (larger than 10 acre subset) or worse (Southern CA fire subset) to begin with.

CHAPTER 6

Conclusion

By incorporating randomness and spatial patterns, point process theory offers an intuitive way to model the complex spatial dynamics of forest fire behavior. The SG statistic seems to be an easy and accurate tool for modeling forest fire point processes based on the relatively low residuals of all of the above models. In terms of models, models for the SoCal, SoCal unnatural, and larger fires seem to perform notably well. Modeling with the computationally effective SG statistic could help inform decision-making among insurance agencies, fire departments, and land management organizations.

Potential avenues for future research could include implementing more recent or diverse data that are not restricted to forest system lands. Furthermore, integrating additional covariates that are suggested in literature to be important to forest fire ignition could be implemented in modeling, such as fuel composition or topography [SRK07]. An analysis of the performance of the SG statistic against the MLE for a variety of point process models could also be insightful. It also would be interesting to further investigate why the SG statistic performs well in modeling and if there are advantages in model selection over the MLE method. Additionally, there is research to be done to examine other ways to divide, including non-uniform divisions, regions in question [Coe23].

APPENDIX A

Subset	Variables	Parameters	Mean Abs Residual	SG Statistic
All Fires (n=7179)	Bkgd, constant, tmax	9831, 1, 1	556	0.01545084
All Natural (n=2690)	Bkgd, awnd, tmax	6555, 1, 1	773	0.02499571
All Unnatural (n=4489)	Bkgd, constant, prcp, tmax	6822, 48, -.1, 1	620	0.02548996
SoCal (n = 2982)	Bkgd, awnd, tmax, evap, tmin, prcp, constant	9888, 21, .1, -524, -2, -.3, 210	488	0.004326582
SoCal Natural (n = 600)	Bkgd, constant, evap	3064, -312, 946	602	0.01156501
SoCal Unnatural (n = 2382)	Bkgd, constant, awnd, tmax, prcp	7724, -332, -4, 5, 1	493	0.009156331
NorCal (4197)	Bkgd	9831	704	0.008119111
NorCal Natural (n = 2090)	Bkgd, awnd, tmax	6052, .3, 1	547	0.008012237
NorCal Unnatural (n = 2107)	Bkgd, constant, prcp, tmax	6115, -69, .04, 1	636	0.01072884
All > 1 acre (n = 1195)	Bkgd, constant, awnd, evap, tmax	2273, -99, -1, 498, 1	217	0.02863146
All > 10 acres (n = 496)	Bkgd, constant, prcp, tmax	1102, 6, .1, 1	119	0.03608812

Table A.1: Summary of best performing models of all the subsets.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	9831	610	0.01905628
Bkgd, awnd	10435, .9	594	0.01885406
Bkgd, awnd, tmax	10435, .9, 1	591	.01873597
Bkgd, awnd, tmax, evap	10296, -1, .4, 1	585	0.01772847
Bkgd, awnd, tmax, evap, tmin	10286, -3, .5, -57, 1	595	0.01562562
Bkgd, awnd, tmax, evap, tmin, prcp	10568, -9, .2, 8, .6, 1	594	0.0158948
Bkgd, awnd, tmax, evap, tmin, prcp, constant	10568, -9, .2, 8, .6, 1, 1	581	0.01356592
Bkgd, constant	9831, 1	595	0.01887556
Bkgd, constant, awnd,	10443, 5, 1	595	0.01884705
Bkgd, constant, awnd, evap	10443, 5, 1, 1	601	0.01739696
Bkgd, constant, awnd, evap, tmax	10374, -82, 25, -142, 1	561	0.01291623
Bkgd, constant, awnd, tmax	10374, -82, 25, 1	564	0.01510645
Bkgd, constant, prcp	10443, 5, 1	597	0.01744089
Bkgd, constant, tmax	9831, 1, 1	556	0.01545084
Bkgd, constant, tmax, prcp	9831, 1, 1, 1	580	0.01470446
constant	6555	4304	0.07451696

Table A.2: Performance of statewide models. Green highlight indicates the best performing model.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	6555	852	0.02632275
Bkgd, awnd	6555, 1	828	0.02607548
Bkgd, tmax	6555, 1	824	0.02604629
Bkgd, awnd, tmax	6555, 1, 1	773	0.02499571
Bkgd, awnd, tmax, evap	6555, 1, 1, 1	806	0.02482011
Bkgd, awnd, tmax, evap, tmin	6555, 1, 1, 1, 1	815	0.02484679
Bkgd, awnd, tmax, evap, tmin, prcp	6555, 1, 1, 1, 1, 1	871‘	0.02394823
Bkgd, awnd, tmax, evap, tmin, prcp, constant	5902, 9, 2, -5, -10, .3, 230	917	0.02406679
Bkgd, constant	5654, 1	825	0.02606407
Bkgd, constant, awnd,	5654, 1, 1	815	0.02603648
Bkgd, constant, evap	5654, 1, 1	844	0.02601103
Bkgd, constant, tmax	5654, 1, 1	831	0.02421392
Bkgd, constant, awnd, evap	5654, 1, 1, 1	806	0.02396486
Bkgd, constant, awnd, evap, tmax	5654, 1, 1, 1, 1	789	0.02392683
Bkgd, constant, awnd, tmax	5654, 1, 1, 1	806	0.02396486
Bkgd, constant, prcp	5654, 1, 1	886	0.02477208
Bkgd, constant, prcp, tmax	5769, -147, .1, 1	915	0.02424074
constant	2459	1368	0.04800243

Table A.3: Performance of statewide natural fire models. Green highlight indicates the best performing model.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	9831	800	0.03760277
Bkgd, awnd	8100, 1	643	0.03007842
Bkgd, tmax	8100, 1	643	0.02995469
Bkgd, tmin	8100, 1	642	0.03073728
Bkgd, awnd, tmax	6653, 4, 1	643	0.02995108
Bkgd, awnd, tmax, evap	6660, -.8, .4, 1	647	0.02675935
Bkgd, awnd, tmax, evap, tmin	6751, -2, 2, -414, 1	644	0.02662729
Bkgd, awnd, tmax, evap, tmin, prcp	6701, 1, 2, -333, -.5, 1	700	0.02548808
Bkgd, awnd, tmax, evap, tmin, prcp, constant	6701, 1, 2, -333, -.5, 1, 1	686	0.02132167
Bkgd, awnd, tmax, evap, prcp	6660, -.8, .4, 1, 1	659	0.02504448
Bkgd, constant	8100, 1	644	0.0302903
Bkgd, constant, awnd,	6671, 28, 1	644	0.0294622
Bkgd, constant, evap	6671, 28, 1	661	0.02865668
Bkgd, constant, tmax	6671, 28, 1	622	0.02583569
Bkgd, constant, awnd, evap	6590, -114, 27, 1	644	0.02944663
Bkgd, constant, awnd, evap, tmax	6591, -130, 29, 24, 1	623	0.02578562
Bkgd, constant, awnd, tmax	6590, -114, 27, 1	623	0.02576577
Bkgd, constant, prcp	6671, 28, 1	640	0.02914911
Bkgd, constant, prcp, tmax	6822, 48, -.1, 1	620	0.02548996
constant	4916	3639	0.0852278

Table A.4: Performance of statewide unnatural fire models. Green highlight indicates the best performing model.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	10658	508	0.01093702
Bkgd, awnd	10658, 1	549	0.009708471
Bkgd, tmax	10658, 1	550	0.009599583
Bkgd, tmin	10658, 1	541	0.009966676
Bkgd, evap	10658, 1	550	0.009667786
Bkgd, awnd, tmax	9493, 3, 1	549	0.009551603
Bkgd, awnd, evap	9493, 3, 1	550	0.009623112
Bkgd, awnd, evap, tmin	9468, -4, 170, 1	595	0.007365078
Bkgd, awnd, tmax, evap, tmin	9924, .4, 2, -467, 1	501	0.005661551
Bkgd, awnd, tmax, evap, tmin, prcp	9795, 15, 1, -355, 1, -.2	509	0.006362882
Bkgd, awnd, tmax, evap, tmin, prcp, constant	9888, 21, .1, -524, -2, -.3, 210	488	0.004326582
Bkgd, awnd, evap, tmax	9468, -4, 170, 1	513	0.007898588
Bkgd, awnd, evap, tmax, prcp	9718, -.02, -408, 1, 1	572	0.00996795
Bkgd, constant	10658, 1	547	0.009764503
Bkgd, constant, awnd,	9520, 17, 1	549	0.009696463
Bkgd, constant, evap	9520, 17, 1	558	0.006815458
Bkgd, constant, tmax	9520, 17, 1	540	0.005975145
Bkgd, constant, awnd, evap	9487, -14, 5, 1	546	0.005827478
Bkgd, constant, awnd, evap, tmax	9487, -14, 5, 1, 1	561	0.004854959
Bkgd, constant, awnd, tmax	9487, -14, 5, 1	531	0.005821689
Bkgd, constant, prcp	9520, 17, 1	528	0.00838603
Bkgd, constant, prcp, tmax	9653, 46, -.1, 1	533	0.005452304
constant	7717	5930	0.04476622

Table A.5: Performance of Southern CA fire models. Green highlight indicates the best performing model.

Variables	Parameter Input	Residual	SG
Bkgd	4916	823	0.01457826
Bkgd, tmax	4072, 1	677	0.01326278
Bkgd, tmin	4072, 1	681	0.01334229
Bkgd, evap	4072, 1	638	0.0125644
Bkgd, awnd	4072, 1	679	0.01307194
Bkgd, awnd, tmax	3004, 17, 1	742	0.01215888
Bkgd, awnd, evap	3004, 17, 1	616	0.0122512
Bkgd, awnd, evap, tmin	2899, -21, 705, 1	784	0.01111701
Bkgd, awnd, evap, tmin, tmax	3111, 125, 357, -14, 1	767	0.01100117
Bkgd, awnd, evap, tmin, tmax, prcp	3274, 74, 100, -23, 12, -.3	708	0.009768093
Bkgd, awnd, evap, tmin, tmax, prcp, constant	3274, 74, 100, -23, 12, -.3, 1	711	0.009841118
Bkgd, awnd, evap, tmax	2899, -21, 705, 1	729	0.01107663
Bkgd, awnd, evap, tmax, prcp	2844, 126, 892, -11, 1	807	0.009460894
Bkgd, constant	4072, 1	681	0.01321258
Bkgd, constant, awnd,	3039, 108, 1	720	0.01235648
Bkgd, constant, evap	2961, -123, 612	602	0.01156501
Bkgd, constant, tmax	3039, 108, 1	725	0.01307052
Bkgd, constant, evap, awnd	3064, -312, 946, 1	614	0.01073742
Bkgd, constant, evap, awnd, tmax	3064, -312, 946, 1, 1	607	0.01085555
Bkgd, constant, awnd, tmax	2709, -562, 126, 1	921	0.01168934
Bkgd, constant, prcp	3039, 108, 1	744	0.01080448
Bkgd, constant, prcp, tmax	3039, 108, 1, 1	874	0.01008116
constant	2049	1323	0.02378105

Table A.6: Performance of Southern CA natural caused fire models. Green highlight indicates the best performing model.

Variables	Parameter Input	Residual	SG
Bkgd	9831	500	0.01698546
Bkgd, tmax	9662, 1	524	0.01324631
Bkgd, tmin	9662, 1	521	0.0138267
Bkgd, evap	9662, 1	522	0.01342258
Bkgd, awnd	9662, 1	524	0.01341743
Bkgd, awnd, tmax	7558, 4, 1	523	0.0132095
Bkgd, awnd, evap	7558, 4, 1	530	0.01249217
Bkgd, awnd, evap, tmin	7547, 15, -209, 1	530	0.01238891
Bkgd, awnd, evap, tmin, tmax	7547, 15, -209, 1, 1	495	0.0078457
Bkgd, awnd, evap, tmin, tmax, prcp	7470, 22, -527, -4, 1, 1	518	0.01220914
Bkgd, awnd, evap, tmin, tmax, prcp, constant	7547, 15, -209, 1, 1, 1, 1	546	0.008587935
Bkgd, awnd, evap, tmax	7547, 15, -209, 1	516	0.01018514
Bkgd, awnd, evap, tmax, prcp	7820, -18, -265, 1, .5	468	0.01035117
Bkgd, constant	9662, 1	525	0.01347154
Bkgd, constant, awnd,	7580, 27, 1	524	0.01341715
Bkgd, constant, evap	7580, 27, 1	527	0.01181268
Bkgd, constant, tmax	7580, 27, 1	500	0.01007124
Bkgd, constant, awnd, evap	7558, -2, 5, 1	533	0.01114382
Bkgd, constant, awnd, evap, tmax	7762, -299, -6, -229, 5	508	0.009204567
Bkgd, constant, awnd, tmax	7558, -2, 5, 1	499	0.01001216
Bkgd, constant, awnd, tmax, prcp	7724, -332, -4, 5, 1	493	0.009156331
Bkgd, constant, prcp	7580, 27, 1	530	0.01233982
Bkgd, constant, prcp, tmax	7748, -376, -.1, 5	509	0.008579723
constant	6555	5147	0.04626734

Table A.7: Performance of Southern CA unnatural caused fire models. Green highlight indicates the best performing model.

Variables	Parameter Input	Residual	SG
Bkgd	9831	704	0.008119111
Bkgd, tmax	10681, 1	723	0.007605743
Bkgd, tmin	10681, 1	723	0.007617798
Bkgd, evap	10681, 1	720	0.007659412
Bkgd, awnd	10681, 1	721	0.007619417
Bkgd, awnd, tmax	10755, -14, 1	710	0.007676337
Bkgd, awnd, evap	10862, -1, 1	745	0.006885035
Bkgd, awnd, evap, tmin	11159, -18, 171, 1	869	0.006534318
Bkgd, awnd, evap, tmin, tmax	11159, -18, 171, 1, 1	823	0.00696118
Bkgd, awnd, evap, tmin, tmax, prcp	11159, -18, 171, 1, 1, 1	752	0.004578604
Bkgd, awnd, evap, tmin, tmax, prcp, constant	11159, -18, 171, 1, 1, 1, 1	785	0.003429229
Bkgd, awnd, evap, tmax	11045, -23, 111, 1	809	0.006411216
Bkgd, awnd, evap, tmax, prcp	11159, -18, 171, 1, 1	750	0.005227454
Bkgd, constant	10681, 1	723	0.007585825
Bkgd, constant, awnd,	10884, -9, 1	751	0.007148185
Bkgd, constant, evap	10884, -9, 1	778	0.00628336
Bkgd, constant, tmax	10884, -9, 1	716	0.00643593
Bkgd, constant, awnd, evap	10884, -9, 1, 1	786	0.006261589
Bkgd, constant, awnd, evap, tmax	11763, -198, 3, 262, 1	737	0.006187247
Bkgd, constant, awnd, tmax	10884, -9, 1, 1	735	0.006145212
Bkgd, constant, awnd, tmax, prcp	10884, -9, 1, 1, 1	727	0.005265315
Bkgd, constant, prcp	10884, -9, 1	717	0.005396933
Bkgd, constant, prcp, tmax	11052, -152, .3, 1	752	0.005324268
constant	4916	2560	0.02707852

Table A.8: Performance of Northern CA fire models. Green highlight indicates the best performing model.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	6555	621	0.009087815
Bkgd, tmax	6130, 1	606	0.008989532
Bkgd, tmin	6130, 1	606	0.008988054
Bkgd, evap	6130, 1	607	0.008987192
Bkgd, awnd	6130, 1	608	0.008995568
Bkgd, awnd, evap	6052, .3, 1	646	0.007670369
Bkgd, awnd, tmax	6052, .3, 1	547	0.008012237
Bkgd, awnd, evap, tmin	6330, -25, 259, 1	587	0.007438812
Bkgd, awnd, evap, tmin, tmax	5984, -31, 125, 2, 1	608	0.007485533
Bkgd, awnd, evap, tmin, tmax, prcp	6242, -28, 143, -1, 1, .2	596	0.006489052
Bkgd, awnd, evap, tmin, tmax, prcp, constant	5927, 4, -153, -2, 1, .3, -8	585	0.006474877
Bkgd, awnd, evap, tmax	6330, -25, 259, 1	616	0.007610472
Bkgd, awnd, evap, tmax, prcp	6156, -27, 204, 1, 1	592	0.006488753
Bkgd, constant	6130, 1	607	0.008999138
Bkgd, constant, awnd,	6047, 2, 1	613	0.008987556
Bkgd, constant, evap	6047, 2, 1	641	0.008824857
Bkgd, constant, tmax	6047, 2, 1	607	0.006965281
Bkgd, constant, awnd, evap	6086, -11, 2, 1	667	0.007586513
Bkgd, constant, awnd, evap, tmax	6463, -46, -18, 265, 1	554	0.006800176
Bkgd, constant, awnd, tmax	6086, -11, 2, 1	594	0.00688513
Bkgd, constant, awnd, tmax, prcp	6077, -236, -3, 3, 1	602	0.006798284
Bkgd, constant, prcp	6047, 2, 1	648	0.007101054
Bkgd, constant, prcp, tmax	6047, 2, 1, 1	706	0.008212981
constant	2459	1200	0.02305346

Table A.9: Performance of Northern CA natural fire models. Green highlight indicates the best performing model.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	6555	643	0.01406388
Bkgd, tmax	5791, 1	646	0.01214027
Bkgd, tmin	5791, 1	647	0.01247978
Bkgd, evap	5791, 1	645	0.01254081
Bkgd, awnd	5791, 1	647	0.01214036
Bkgd, awnd, evap	5791, 1, 1	668	0.009863116
Bkgd, awnd, tmax	6319, 17, -2	659	0.009160935
Bkgd, awnd, evap, tmin	6511, 16, -683, 1	659	0.009413558
Bkgd, awnd, evap, tmin, tmax	6511, 16, -683, 1, 1	653	0.009209737
Bkgd, awnd, evap, tmin, tmax, prcp	6511, 16, -683, 1, 1, 1	656	0.00828473
Bkgd, awnd, evap, tmin, tmax, prcp, constant	6754, 2, -799, -2, 1, .2, 24	653	0.008252987
Bkgd, awnd, evap, tmax	6636, 16, -611, 1	660	0.009148306
Bkgd, awnd, evap, tmax, prcp	6424, 16, -11, -2, 5225	959	0.008854719
Bkgd, constant	5791, 1	643	0.01236474
Bkgd, constant, awnd	6224, -60, 1	647	0.009718151
Bkgd, constant, evap	6218, -52, -25	647	0.01038015
Bkgd, constant, tmax	6224, -60, 1	653	0.01195231
Bkgd, constant, awnd, evap	6670, -191, 18, 1	653	0.009446194
Bkgd, constant, awnd, evap, tmax	6518, -178, 25, -338, 1	663	0.009131983
Bkgd, constant, awnd, tmax	6670, -191, 18, 1	662	0.009143235
Bkgd, constant, awnd, tmax, prcp	6670, -191, 18, 1, 1	677	0.008205568
Bkgd, constant, prcp	6224, -60, 1	639	0.01223331
Bkgd, constant, prcp, tmax	6115, -69, .04, 1	636	0.01072884
constant	3278	1793	0.03105843

Table A.10: Performance of Northern CA unnatural fire models. Green highlight indicates the best performing model.

Variables	Parameter Input	Residual	SG
Bkgd	3278	619	0.04949834
Bkgd, tmax	3602, 1	312	0.03434171
Bkgd, tmin	3602, 1	309	0.03437456
Bkgd, evap	3602, 1	309	0.03433921
Bkgd, awnd	3602, 1	331	0.03519787
Bkgd, awnd, evap	2428, 4, 1	301	0.03423737
Bkgd, awnd, tmax	2428, 4, 1	302	0.0342134
Bkgd, awnd, evap, tmin	2288, -2, 146, 1	300	0.03423358
Bkgd, awnd, evap, tmin, tmax	2303, -2, 65, -.2, .4	303	0.03418714
Bkgd, awnd, evap, tmin, tmax, prcp	2385, -10, 30, -.1, .3, .2	305	0.03056667
Bkgd, awnd, evap, tmin, tmax, prcp, constant	2385, -10, 30, -.1, .3, .2, 1	305	0.02917121
Bkgd, awnd, evap, tmax	2300, -2, 22, .4	301	0.03420666
Bkgd, awnd, evap, tmax, prcp	2305, -3, 239, -.2, -440	295	0.03293188
Bkgd, constant	3602, 1	324	0.03496247
Bkgd, constant, awnd	2388, 25, 1	325	0.03495536
Bkgd, constant, evap	2388, 25, 1	295	0.03196619
Bkgd, constant, tmax	2388, 25, 1	265	0.02961159
Bkgd, constant, awnd, evap	2396, 21, 1, 1	296	0.03190561
Bkgd, constant, awnd, evap, tmax	2273, -99, -1, 498, 1	217	0.02863146
Bkgd, constant, awnd, tmax	2396, 21, 1, 1	250	0.0294631
Bkgd, constant, awnd, tmax, prcp	2466, -23, -8, .5, .2	303	0.03036383
Bkgd, constant, prcp	2388, 25, 1	295	0.03192359
Bkgd, constant, prcp, tmax	2252, -26, .2, 1	245	0.03100999
constant	1230	626	0.06083515

Table A.11: Performance of models for fires $>$ 1 acre. Green highlight indicates the best performing model.

Variables	**Parameter Input**	**Residual**	**SG**
Bkgd	3278	945	0.07453827
Bkgd, tmax	3460, 1	193	0.04092866
Bkgd, tmin	3460, 1	185	0.04042579
Bkgd, evap	3460, 1	206	0.0415448
Bkgd, awnd	3460, 1	225	0.04402666
Bkgd, awnd, evap	1150, 6, 1	202	0.04126217
Bkgd, awnd, tmax	1150, 6, 1	158	0.03889707
Bkgd, awnd, evap, tmin	1085, -3, 172, 1	160	0.03745831
Bkgd, awnd, evap, tmin, tmax	837, -11, 26, 2, .4	153	0.03740829
Bkgd, awnd, evap, tmin, tmax, prcp	978, -6, -4, 2, -.2, .03	162	0.03755906
Bkgd, awnd, evap, tmin, tmax, prcp, constant	978, -6, -4, 2, -.2, .03, 1	149	0.03514591
Bkgd, awnd, evap, tmax	1085, -3, 172, 1	158	0.03889697
Bkgd, awnd, evap, tmax, prcp	877, -11, 1, 2, 1	143	0.03717759
Bkgd, constant	3460, 1	207	0.04177178
Bkgd, constant, awnd	1074, 39, 1	183	0.04002182
Bkgd, constant, evap	1074, 39, 1	206	0.04150975
Bkgd, constant, tmax	1074, 39, 1	148	0.03721344
Bkgd, constant, awnd, evap	918, 131, -12, 1	183	0.04001638
Bkgd, constant, awnd, evap, tmax	918, 131, -12, 1, 1	141	0.03679683
Bkgd, constant, awnd, tmax	918, 131, -12, 1	134	0.0368862
Bkgd, constant, awnd, tmax, prcp	840, -211, -6, 4, 1	132	0.03632122
Bkgd, constant, prcp	1074, 39, 1	208	0.04016326
Bkgd, constant, prcp, tmax	1102, 6, .1, 1	119	0.03608812
constant	436	256	0.05223884

Table A.12: Performance of models for $>$ 10 acre fires. Green highlight indicates the best performing model.

REFERENCES

[al17] Donald J Wuebbles et al. "Climate Science Special Report." Technical report, U.S. Global Change Research Program, Washington DC, 2017.

[Autnd] No Author. "Fire Hazard Severity Zone Methods." Technical report, Cal Fire, Sacramento CA, n.d.

[BFL14] Michael Billmire, Nancy H F French, Tatiana Loboda, R Chris Owen, and Marlene Tyner. "Santa Ana winds and predictors of wildfire progression in southern California." *International Journal of Wildland Fire*, **23**(8):1119–1129, 2014.

[BTM05] A Baddley, R Turner, J Moller, and M Hazelton. "Residual analysis for spatial point processes (with discussion)." *Royal Statistical Society*, **67**(5):617–666, 2005.

[CL18] O Cronie and M N. M. Van Lieshout. "A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions." *Biometrika*, **105**(2):455–462, 2018.

[Coe23] Abigail Coelho. *"Estimation of Wildfire Ignition Conditional Intensity Parameters Via the Stoyan-Grabarnik Statistic."*. Master's thesis, University of California Los Angeles, 2023.

[GSG96] M Goulard, A Sarkaa, and Pavel Grabarnik. "Parameter Estimation for Marked Gibbs Point Processes through the Maximum Pseudo-Likelihood Method." *Scandinavian Journal of Statistics*, **23**(3):365–379, 1996.

[JMS12] P Juan, J Mateu, and M Saez. "Pinpointing spatio-temporal interactions in wildfire patterns." *Stochastic Environmental Research and Risk Assessment*, **26**:1131–1150, 2012.

[Jon10] Jeanine Jones. "California's Drought of 2007-2009." Technical report, California Department of Water Resources, Sacramento CA, 2010.

[KS23] Conor Kresin and Frederic Schoenberg. "Parametric estimation of spatial–temporal point processes using the Stoyan–Grabarnik statistic." *The Institute of Statistical Mathematics*, **75**(6):887–909, 2023.

[KSF09] Jon E Keeley, Hugh Safford, C J Fotheringham, Janet Franklin, and Max Moritz. "The 2007 Southern California Wildfires: Lessons in Complexity." *Journal of Forestry*, **107**(6):287–296, 2009.

[LDC18] Yi-Chin Liu, Pingkaun Di, Shu-Hua Chen, and John DaMassa. "Relationships of Rainy Season Precipitation and Temperature to Climate Indices in California: Long-Term Variability and Extreme Events." *American Meteorological Survey*, **31**(5):1921–1942, 2018.

[Nat24] National Oceanic and Atmospheric Association. "Climate Data Online.", 2024.

[NSK13] Kevin Nichols, Frederic Paik Schoenberg, Jon E. Keeley, Andrew Bray, and David Diez. "The application of prototype point processes for the summary and description of California wildfires." *Journal of Time Series Analysis*, **32**(4):420–429, 2013.

[Oga98] Yosihiko Ogata. "Space-Time Point-Process Models for Earthquake Occurrences." *Annals of the Institute of Statistical Mathematics*, **50**:379–402, 1998.

[PSW05] Roger D Peng, Frederik Paik Schoenberg, and James A Woods. "A Space–Time Conditional Intensity Model for Evaluating a Wildfire Hazard Index." *American Statistical Association*, **100**(469):26–35, 2005.

[QDM21] Jose J Quinlan, Carlos Diaz-Avalos, and Ramses H Mena. "Modeling wildfires via marked spatio-temporal Poisson processes." *Environmental and Ecological Statistics*, **28**:549–565, 2021.

[Rei05] A Reinhart. "A review of self-exciting spacio-temporal point processes and their applications." *Statistical Science*, **33**(3):299–318, 2005.

[SG91] Dietrich Stoyan and Pavel Grabarnik. "Second-order Characteristics for Stochastic Structures Connected with Gibbs Point Processes." *Mathematische Nachrichten*, **151**(1):95–100, 1991.

[SJA20] Adam J P Smith, Matthew W Jones, John T Abatzoglou, Josep G Canadell, and Richard A Betts. "Climate Change Increases the Risk of Wildfires." *Science Brief*, 2020.

[SJV13] Laura Serra, Pablo Juan, Diego Varga, Jorge Mateu, and Marc Saez. "Spatial pattern modelling of wildfires in Catalonia, Spain 2004–2008." *Environmental Modelling and Software*, **40**:235–244, 2013.

[SRK07] Alexandra D Syphard, Volker C Radeloff, Nicholas S Keuler, Robert S Taylor, Todd J Hawbaker, Susan I Stewart, and Marray K Clayton. "Predicting spatial patterns of fire on a southern California landscape." *International Journal of Wildland Fire*, **17**(5):602–613, 2007.

[SSM13] Laura Serra, Marc Saez, Jorge Mateu, Diego Varga, Pablo Juan, Carlos Diaz-Avalos, and Harvard Rue. "Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994–2008." *Environmental and Ecological Statistics*, **21**:531–563, 2013.

[USD24] USDA Forest Service. "FIRESTAT Fire Occurrence - Yearly Update.", 2024.

[WHC06] A L Westerling, H G Hidalgo, D R Cayan, and T W Swetnam. "Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity." *Science*, **313**(5789):940–943, 2006.

[XS11] Haiyong Xu and Frederic Schoenberg. "Point process modeling of wildfire hazard in Los Angeles County, California." *Annals of Applied Statistics*, **5**(2A):684–704, 2011.

[YWD15] Jian Yang, Peter J Weisberg, Thomas E Dilts, E Louis Loudermilk, Robert M Scheller, Alison Stanton, and Carl Skinner. "Predicting wildfire occurrence distribution with spatial point process models and its uncertainty assessment: a case study in the Lake Tahoe Basin, USA." *International Journal of Wildland Fire*, **24**:380–390, 2015.