
# NAVAL POSTGRADUATE SCHOOL

**MONTEREY, CALIFORNIA**

# THESIS

**A SPATIAL-TEMPORAL POINT PROCESS MODEL FOR ESTIMATING PROBABILITY OF WILDFIRES IN LOS ANGELES COUNTY**

by

Wook Yi

March 2022

Thesis Advisor: Robert A. Koyak
Second Reader: Javier Salmeron-Medrano

**Approved for public release. Distribution is unlimited.**

THIS PAGE INTENTIONALLY LEFT BLANK

| REPORT DOCUMENTATION PAGE | | | *Form Approved OMB No. 0704-0188* |
|---|---|---|---|

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

| **1. AGENCY USE ONLY** *(Leave blank)* | **2. REPORT DATE** March 2022 | **3. REPORT TYPE AND DATES COVERED** Master's thesis |
|---|---|---|

**4. TITLE AND SUBTITLE**
A SPATIAL-TEMPORAL POINT PROCESS MODEL FOR ESTIMATING PROBABILITY OF WILDFIRES IN LOS ANGELES COUNTY

**5. FUNDING NUMBERS**

**6. AUTHOR(S)** Wook Yi

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Naval Postgraduate School
Monterey, CA 93943-5000

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
N/A

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES** The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release. Distribution is unlimited.

**12b. DISTRIBUTION CODE**
A

**13. ABSTRACT (maximum 200 words)**

In Los Angeles County, wildfires are among the most catastrophic environmental events caused by regional characteristics and climate change. In this study, we develop a point process model to estimate the probability of wildfires based on historical weather data and past wildfires data from Los Angeles County from 2004 to 2018. First, we partition Los Angeles County into small rectangular regions, called voxels, with daily temporal resolution. Then, we use random forests and generalized additive models to obtain estimated probabilities on a training data set. In addition to daily weather and fuel-condition measurements, our models incorporate seasonal and geographical effects. Because measurements on weather and fuel conditions are available only from a fixed set of remote automated weather stations, their data must be averaged to relate them to the voxel level, and the way this is done is a factor in modeling. Through the developed model, it is possible to obtain localized, estimated probabilities of wildfires. Ultimately, this tool can aid Los Angeles County Fire Department in improving its capability and effectiveness.

| **14. SUBJECT TERMS** | **15. NUMBER OF PAGES** |
|---|---|
| point process models, wildfire probabilities, Los Angeles County Fire Department, random forests, generalized additive models. | 59 |
| | **16. PRICE CODE** |

| **17. SECURITY CLASSIFICATION OF REPORT** | **18. SECURITY CLASSIFICATION OF THIS PAGE** | **19. SECURITY CLASSIFICATION OF ABSTRACT** | **20. LIMITATION OF ABSTRACT** |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UU |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release. Distribution is unlimited.**

# A SPATIAL-TEMPORAL POINT PROCESS MODEL FOR ESTIMATING PROBABILITY OF WILDFIRES IN LOS ANGELES COUNTY

Wook Yi
So-ryeong, Republic of Korea Air Force
BSB, Korea Airforce Academy, 2006

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN OPERATIONS RESEARCH**

from the

**NAVAL POSTGRADUATE SCHOOL**
**March 2022**

Approved by: Robert A. Koyak
Advisor

Javier Salmeron-Medrano
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

In Los Angeles County, wildfires are among the most catastrophic environmental events caused by regional characteristics and climate change. In this study, we develop a point process model to estimate the probability of wildfires based on historical weather data and past wildfires data from Los Angeles County from 2004 to 2018. First, we partition Los Angeles County into small rectangular regions, called voxels, with daily temporal resolution. Then, we use random forests and generalized additive models to obtain estimated probabilities on a training data set. In addition to daily weather and fuel-condition measurements, our models incorporate seasonal and geographical effects. Because measurements on weather and fuel conditions are available only from a fixed set of remote automated weather stations, their data must be averaged to relate them to the voxel level, and the way this is done is a factor in modeling. Through the developed model, it is possible to obtain localized, estimated probabilities of wildfires. Ultimately, this tool can aid Los Angeles County Fire Department in improving its capability and effectiveness.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| AOM | Augmentation Optimization Model |
| AOMS | Augmentation Optimization Model with Simulation |
| AUC | Area Under the Curve |
| BI | Burning Index |
| CALFIRE | California Department of Forestry and Fire Protection |
| DFM | Dead Fuel Moisture |
| ERC | Energy Release Component |
| GAM | Generalized Additive Model |
| GLM | Generalized Linear Model |
| KBDI | Keetch-Byram Drought Index |
| LACoFD | Los Angeles County Fire Department |
| LFM | Live Fuel Moisture |
| LGCP | Log-Gaussian Cox Process |
| NPS | Naval Postgraduate School |
| P-value | Probability Value |
| RAWS | Remote Automated Weather Station |
| RH | Relative Humidity |
| ROC | Receiver Operating Characteristic |
| SC | Spread Component |

THIS PAGE INTENTIONALLY LEFT BLANK

# EXECUTIVE SUMMARY

Wildfires in Los Angeles County are among the most catastrophic environmental hazards caused by regional characteristics and climate change. In a real-life situation, the commander of the Los Angeles County Fire Department (LACoFD) usually decides when and where to augment equipment or where to pre-position firefighters. In many cases, such decisions depend on the commander's judgement. Ideally, a commander would operate rationally, based on an objective analysis of the entire risk situation.

Our objectives are to describe spatial and temporal dependence that may lead to the development of improved models for predicting wildfires. In this research project, we develop a point process model to estimate wildfire probabilities in accordance with historical weather and past wildfires data observed in Los Angeles County from 2004 to 2018. This spatio-temporal point process modeling approach not only contributes to helping LACoFD unit commanders make more informed decisions but also reveals new information to LACoFD. For instance, based on historical weather data, it is feasible with this point-process model to estimate the occurrence of future wildfires stochastically.

The wildfire probabilities estimating model proposed in this thesis is based on point process modeling using Remote Automated Weather Stations (RAWS) data and historical wildfires occurrence data. RAWS data contains the Burning Index (BI), temperature, relative humidity, etc. The data was provided to the Naval Postgraduate School by the LACoFD. Our research was confined to the Los Angeles County area and the specific time from May 4, 2004, to December 31, 2018.

By partitioning Los Angeles County into small rectangular regions (25km$^2$ each), called "voxels," with daily temporal resolution, we use random forests and generalized additive model to obtain estimated probabilities on a training data set. To estimate the probabilities for wildfires from the RAWS data, we utilize the latitude and longitude coordinates of center of voxels in Los Angeles County to find the nearest RAWS.

In addition to daily weather and fuel condition measurements, our models incorporate seasonal and geographical effects. Because measurements on weather and fuel-

condition are available only at a fixed set of stations, their data must be averaged to relate them to the voxel level, and the way this is done is a factor in modeling. For each date, we assign predictor variables to each voxel. Weather and environmental variables, however, are not measured for each voxel. Instead, we calculate a set of weights for each voxel to apply to the 21 RAWS stations, based on the distances of a voxel center to each of those stations. We start by creating a matrix of weights: each row is a voxel, and each column is a RAWS. Only RAWS that are present can be used in averaging. RAWS stations that are closer to a voxel center are assigned more weight than those farther away, but there is no unique way to do this. We explore a series of formulas that range from applying the same weight to each RAWS (straight averaging) to applying a large weight to the closest RAWS and small weights to all others.

We measure the performance of the point-process model by dividing the past wildfires occurrence data of 256 voxels into the training data and test data. We estimate model parameters using training data consisting of the period from May 4, 2004, to December 31, 2014, and evaluate the results on test data from the period of January 1, 2015, to December 31, 2018. For model selection we use the area under the receiver operating characteristic curve as a measure of predictive capability. Our research demonstrates the existence of these effects in two ways: first, by noting that four predictor variables that capture past and nearby incidences of wildfires into our models contribute measurably to the predictive quality of the models; and second, by finding notable spatio-temporal correlations in the residuals that we derive from those models. Although our models are an improvement over those that do not include spatio-temporal predictors, other models that more fully capture spatio-temporal effects may be more accurate. Through this, we obtain localized, estimated probabilities of wildfires occurring, and this information can serve as one component of the process through which the LACoFD can improve its capability and effectiveness.

# ACKNOWLEDGMENTS

I think I could never graduate this glorious Naval Postgraduate School Operations Research master's degree course without the help of the people around me. First, I would like to express my gratitude to Professor Robert A. Koyak. It was a pleasure working under his leadership on the research. He always kindly informed me what I did not know well and devotedly guided me even though I had language barriers as an international student. I have learned a lot from him as a professor, mentor, and senior. Also, I would like to thank not only my wife but also the best friend in my life, Bora, and my lovely children, Leo and Lia, for being healthy and happy in the United States. My family was the driving force behind me to endure hard times and always gave me hope and courage. I want to thank second reader, Professor Salmerón, for carefully advising me on my thesis. I also want to take this opportunity to express my gratitude to Jung hwan Kim who helped me academically and emotionally. Finally, I would like to thank Dong-geon Lim who came here from Korea to study together with me. Above all, I would like to thank the Republic of Korea Air Force and Korea National Defense University for providing me with this challenging educational opportunity. Again, as a fighter pilot who returns to the field, I hope that I can apply and develop what I have learned here to help my organization.

THIS PAGE INTENTIONALLY LEFT BLANK

# I. INTRODUCTION

## A. BACKGROUND

In Los Angeles County area, wildfires are among the most catastrophic environmental events that adversely affect resident life. Wildfires have frequently occurred in Los Angeles County due to regional characteristics and climate change. As this danger increases, wildfire prevention and management has emerged as a critical issue. A summary of all 2021 wildfires in California, reported 8,835 wildfires and a total of 2,568,948 acres burned (California Department of Forestry and Fire Protection [CALFIRE] 2021). This data reveals a dramatic and troubling increase from 2020, when wildfires burned 61,850 acres within Los Angeles County (Los Angeles County Fire Department 2021). Although wildfires can occur due to natural causes and as well as human carelessness, large-scale wildfires are clearly affected by meteorological factors such as wind, humidity, and drought. However, rapid response through correct prediction can protect the people's property and lives from the risk of wildfires. For effective fire management, it is essential to identify the spatio-temporal variability in wildfire intensity. In this thesis research, historic weather data is used for model fitting, and it comes from the Remote Automated Weather Station (RAWS) system for the Los Angeles County area. Figure 1 is an image of RAWS. The weather data used consist of 21 RAWS data sets for Los Angeles County, each representing one day from May 4, 2004, to December 31, 2018.

In a real-life situation, the commander of the Los Angeles County Fire Department (LACoFD) usually decides when and where to augment equipment or where to pre-position firefighters. In many cases, such decisions depend on the commander's judgement. Ideally, a commander would operate rationally, based on an objective analysis of the entire risk situation. Such analysis would rely on quantitative factors such as the Burning Index, temperature, and relative humidity.

Figure 1. A Remote Automated Weather Station (RAWS) in Idaho. Source: raws.nifc.gov (2021).

## B. RESEARCH OBJECTIVES

The goal of our research is to identify the spatial and temporal properties of wildfire occurrences in Los Angeles County. Our approach to achieving this goal is to develop a point-process model for fires that allows us to investigate these properties over an extended period. We use geocoded data on daily fire occurrences together with RAWS data collected from May 4, 2004, to December 31, 2018, for this purpose. By partitioning Los Angeles County into small rectangular regions (25km$^2$ each), called "voxels," with daily temporal resolution, we employ random forests and generalized additive models to obtain estimated probabilities on a training data set, and we evaluate performance on a test data set. By examining the structure of the estimated models and residuals obtained from them, we obtain insights that allow us to achieve our research objectives.

## C. THESIS STRUCTURE

This thesis is organized as follows. In Chapter II, we explore the literature on wildfire probabilities estimation models. We review literature not only on wildfire research, but also include a spatio-temporal analysis of another phenomenon. Chapter III

explains the weather data processing to build our model from an array of 21 RAWS data from 2004 to 2018 and the detailed process of fitting a model using the point process algorithm. Chapter IV presents the results obtained by applying the model and analyzes the results from various aspects. Chapter V discusses the conclusions drawn from our analysis and offers suggestions for potential extensions requiring further research.

THIS PAGE INTENTIONALLY LEFT BLANK

# II. LITERATURE REVIEW

There are two previous studies at the Naval Postgraduate School (NPS) that have used the LACoFD data, and there is extensive literature on the topic of predicting events from historical data. The literature review is carried out in chronological order. In the examination of the progress of this research, parts that need to be studied in more depth are suggested.

Preisler et al. (2004) present a model for estimating the probability of wildfire occurrence in their study that focuses on Oregon. They use a spatio-temporal nonparametric logistic regression model for this purpose. The authors group their data by voxels that provide daily measurements on a grid of squares that are $1km^2$ in size. In our thesis, we adopt a similar approach to obtain estimated probabilities of wildland fires in Los Angeles County. Using the results of their model the authors produce a monthly maps of wildland fire risk in Oregon based on the predicted probability. The total number of fires per month can be visualized for each voxel and compared with the actual number of fires in a specific period. Nonetheless, the authors note their model's limitations due to the high variability of estimates. Accordingly, it was suggested that more weather variables and topographic data be collected to improve the estimates. Consequently, we used more variables than the ones in the mentioned study to predict the probability of wildfires occurring in Los Angeles County.

Genton et al. (2006) use clustering of wildfire events in the St. Johns River Water Management District in Florida between the years 1981 and 2001 to explain the irregular distributions of wildfires. Those researchers use factors such as human resources and fuel to optimize resources for fire suppression. By conducting an analysis of the structure of clustering using three stages of modeling (pure-spatial, pure-temporal and spatio-temporal), the authors identify lightning and arson as the main causes of wildfires in that region. The authors also use spatial visualization to describe the risk of wildfires, a technique that we also employ.

Xu and Schoenberg (2010) present a point process modeling of wildfire hazard in Los Angeles County using data from 1975 to 2000. Although the Burning Index (BI) is often used as a predictor of wildland fires, the authors find that it is less effective in predicting wildland fires in Los Angeles County than directly employing the same variables used to construct the BI. The authors investigate the predictive ability of the model using a set of covariates that include seasonal forest fire trend, past spatial image pattern, and weather-related variables to fit the point process models. They find that a multiplicative model, which directly uses weather variables, provided significantly improved predictions relative to models that account for the covariates using BI.

Diggle et al. (2013) present the Log-Gaussian Cox processes (LGCP) model for studying spatio-temporal phenomena such as lung cancer mortality or wildland fire risk in a geographical region over a particular period. Although the authors do not consider wildland fires specifically, the applications that they consider (describing spatial point patterns for the occurrence of hickory trees in a forested region, for the occurrence of bovine tuberculosis in cattle herds, and for lung cancer mortality in a region of Spain) point to the potential usefulness of LGCPs to model a variety of geostatistical paradigms including wildland fires.

McEvoy et al. (2019) analyze the relationship between severe drought and wildland fire damage in California and Nevada between 2012 and 2015 and confirm the effectiveness of using drought indices to predict large wildland fires. It can be seen by season that there is a strong correlation between the four drought indices and fire risk. The more severe the drought, the greater the chances of large wildland fires. Their findings were tested and checked with a research team in Northern California in 2018. Initial feedback was received that fire management could be effectively performed by utilizing drought indices such as the Evaporative Demand Drought Index.

Scholz (2019) presents two predictors necessary for developing an Augmented Optimization Model (AOM) through a statistical method for the LACoFD. First, the probability of fire for the Los Angeles County area was estimated using logistic regression. Second, Scholz estimated the expected wildfire area using a multiple linear regression model. Based on these two variables, an optimized model that includes available staff and

equipment, the cost of enacting off-duty firefighters and moving equipment was proposed under a limited budget.

Opitz, Bonneu and Gabriel (2020) analyze the occurrence of forest fires through Bayesian stochastic modeling to prevent forest fires and estimate the probabilities of forest fires occurring in the Mediterranean Sea region from 1995 to 2018. They identify the mechanisms affecting the intensity of forest fires and quantify them. Then, they develop a point-process framework for the observed forest fire ignition point and fit the spatio-temporal log-Gaussian Cox process models. Finally, they implement the solution by specifying covariates and count values using a package in R that does approximate Bayesian Inference for the Integrated Nested Laplace Approximation method. They also use frequency-based inference techniques to estimate fixed-effect and random-effect hyperparameters, thereby realizing statistical inference and enabling the prediction of wildfires.

Seeberger (2020) presents Augmented Optimization Models with Simulation (AOMS) based on a mathematical decision-making tool for the efficient placement of resources during the initial outbreak of a wildfire, using simulations to identify problems with the estimates of the AOM presented by Scholtz (2019). Using feedback from the LACoFD, the author proposes a new solution evaluation that incorporates accessibility, terrain slope, and hand-crew resources. The AOMS are upgraded using these enhancements, and a more efficient objective function for optimization of resources is presented.

THIS PAGE INTENTIONALLY LEFT BLANK

# III. METHODOLOGY

In this chapter we describe the process of producing an analyzable data set from the wildfire data provided to us by the LACoFD. We also describe several models that we use to analyze the data, and the methods we employ for addressing the study questions using our proposed models.

## A. DATA PREPARATION

### 1. Description of Data

Our approach is based on point-process modeling using RAWS data and historical wildfires occurrence data for Los Angeles County from May 4, 2004, to December 31, 2018. The RAWS variables include weather and information on vulnerability to wildfires due to the condition of vegetation in a particular area. A Burning Index (BI) that combines weather and environmental data into a measure of vulnerability to fire, is also included. A detailed description of the variables is given in Scholz (2019) and is reproduced below in Table 1.

Table 1. RAWS variables collected in Los Angeles County.
Source: LACoFD (2021).

| Predictor | Description |
|---|---|
| Date | Date of RAWS data |
| BI | Burning Index |
| Temperature | Temperature in Fahrenheit |
| RH | Relative Humidity |
| Wind | Wind speed (mph) |
| LFM | Live Fuel Moisture. It is missed about 27% |
| KBDI | Keetch-Byram Drought Index |
| DFM | Dead Fuel Moisture |
| ERC | Energy Release Component |
| SC | Spread Component |

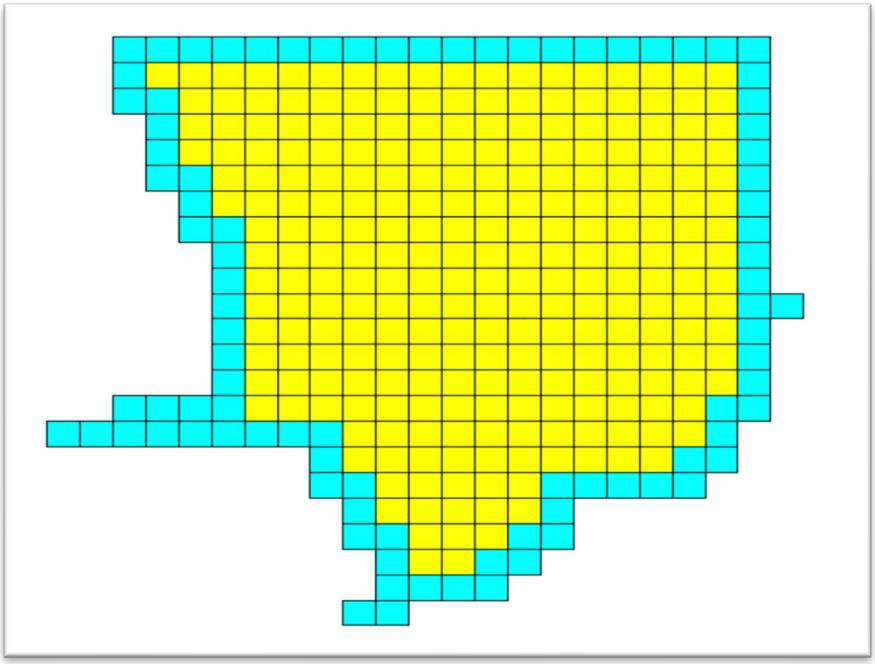
For data analysis we use R (R Core Team 2021), which is a statistical language and environment for computing and graphics. R is capable of handling large amounts of RAWS data for data processing. Moreover, R is also effective as a tool for visualizing wildfires data on maps. The weather data used in this thesis consists of 21 RAWS data sets in Los Angeles County, each representing one day from May 4, 2004, to December 31, 2018. Although we do not use them in our analyses, LACoFD recognizes five climatic zones in Los Angeles County identified as Santa Monica Mountains, Santa Clarita Valley, High County, Los Angeles Basin, and Antelope Valley. Figure 2 shows these zones and the locations of the 21 RAWS (Brown et al. 2021).


Figure 2. Los Angeles County climatic zones and RAWS locations. Source: Brown et al. (2021).

## 2. Formation of Voxels

In order to conduct a spatio-temporal analysis of wildfire occurrences, we divide Los Angeles County into a grid of small, square-shaped regions called voxels, each with an area of 25 $km^2$. Each voxel is observed daily for the occurrence of wildfires, which is

the basis of our analysis. Figure 3 shows the subdivision of Los Angeles County into 347 voxels that are contained entirely within its boundaries. We count the number of wildfire events in each voxel over the time frame of our study. To account for the spatial association between voxels, we further limit our analysis to only those voxels that have a full set of eight neighbors (north, south, east, west, and diagonally incident), which reduces the region of interest to 256 voxels. These voxels are shown in yellow in Figure 3. In each voxel and on each day, we record the variable Fire that is equal to zero if no wildfire occurred and equal to one if at least one wildfire occurred. Due to the size of the voxels, it is unusual to observe more than one fire in the same voxel on the same day.



The 256 voxels shown in yellow have a full set of eight neighboring voxels

Figure 3. Partitioning of Los Angeles County into square (5 km × 5 km) voxels

## 3. Cleaning of RAWS Data

It is not unusual in data analysis to address data quality issues, including analyses that use RAWS data. For example, Live Fuel Moisture (LFM) is missing on nearly 25 percent of days, which explains our decision not to use this variable in our analyses. The

weather variables contain a small number of obvious errors or outliers, such as temperatures that are below zero or above 120 degrees Fahrenheit, and relative humidity measurements that are negative or above 100 percent. We use imputation to replace these values with measurements from other stations on the same day.

## 4. Assignment of RAWS Data to Voxels

Our models incorporate seasonal and geographical effects including daily weather and fuel condition measurements. Because measurements on weather and fuel-condition are available only at a fixed set of stations, their data must be averaged to relate them to the voxel level, and the way this is done is a factor in modeling. For each date, we assign predictor variables to each voxel. However, weather, and environmental variables are not measured for each voxel. Instead, we calculate a set of averaging weights for each voxel to apply to the 21 RAWS stations, based on the distances of a voxel center to each of those stations, which we depict in Figure 4. There are many ways to determine a set of weights that have the desired properties., and we explore a series of formulas that range from applying the same weight to each RAWS (straight averaging) to applying a large weight to the closest RAWS and small weights to all others. Figure 5 describes the process that we use for defining the weights, which depends on a nonnegative parameter $a$ that regulates the degree to which the weights are diffused depending on distances between a voxel center and the set of RAWS locations. The derived weights, which we denote as $w_{ij}$, are used for averaging RAWS variables across stations (subscript $j$) to produce an interpolated set of variables at a given voxel (subscript $i$). If $a = 0$ all weights are equal and simple averaging is used; but as $a$ becomes larger, the weight of the RAWS that is closest to the voxel approaches 1, and all other weights approach 0.

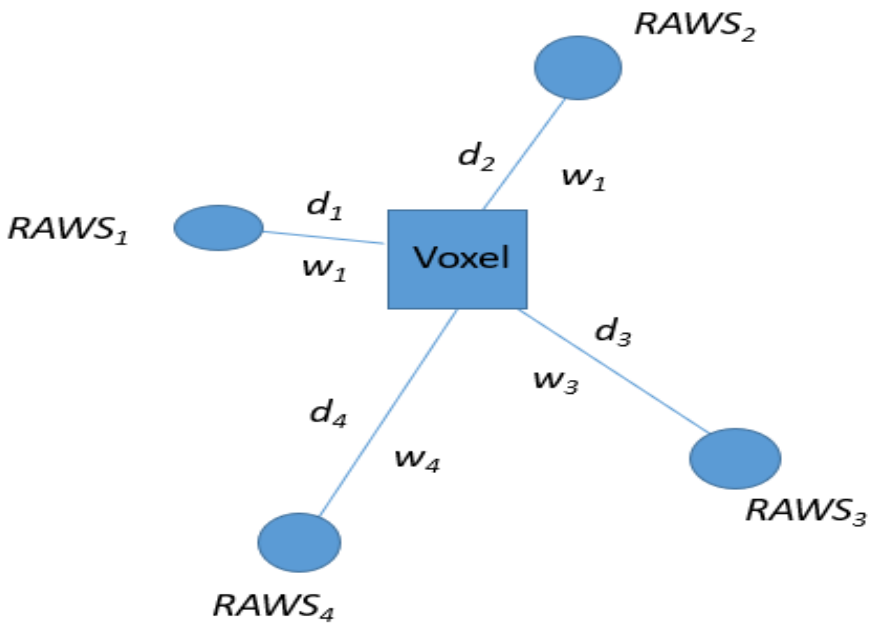



Figure 4. Distance and weights between the voxel center and RAWS

$n =$ total number of RAWS

$n = 21$ in Los Angeles County

$d_{ij} =$ distance from center of voxel $i$ to RAWS $j$

$d_i^* =$ minimum distance of RAWS to voxel $i$

$a =$ nonnegative parameter

$$S_{ij} = \exp\left(-ad_{ij} / d_i^*\right)$$

$$w_{ij} = S_{ij} / \sum_{k=1}^{n} S_{ik}$$

Figure 5. Procedure for calculating voxel dependent weights

There are days during the study period for which some or all RAWS measurements are not available. Therefore, we adopt the following rule: weighted averaging of measurements from available RAWS stations is used on a given day if at least 15 stations report measurements; otherwise, averaging is not used and that day is removed from our analyses. The procedure outlined in Figure 5 is modified accordingly. The time frame of

our analysis (May 4, 2004, to December 31, 2018) comprises 5,355 days, of which 3,653 satisfy our rule for inclusion.

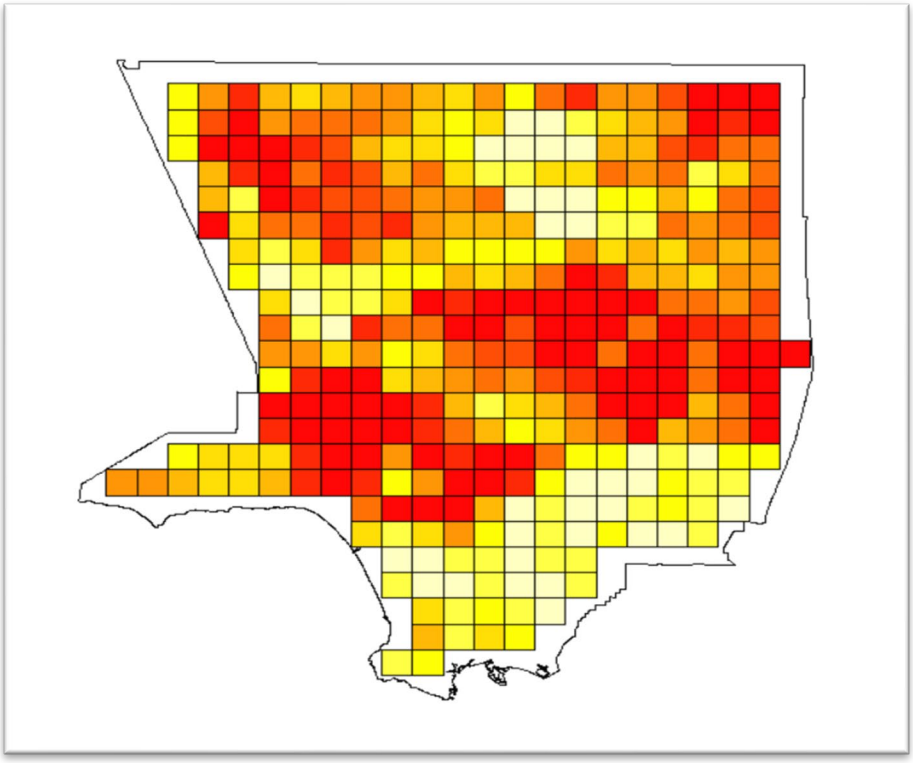
## 5. Derivation of Time-Lagged Variables

We construct the following additional predictors for point process models. Variable Spatial is the total number of fires that occurred in the voxel during the time spanned by the learning data set (May 4, 2004, to December 31, 2015). Figure 6, which shows a heat map for the Spatial variable, indicates which areas of Los Angeles County are more likely to experience wildfires due to their locations. The variable Fire.prior indicates whether a fire occurred in the same voxel during the prior day, and Fire.last5 indicates whether a fire occurred during the last five days. The variables Nghb.prior and Nghb.last5 are indicators of fire occurrences in at least one of the eight neighboring voxels on the prior day or during the prior five days, respectively. Variable Week (which takes on the values 1 to 52) is used to detect seasonal effects.

Table 2. Additional predictors for point process modeling

| Predictor | Description |
|---|---|
| Spatial | Total number of fires that occurred in a voxel during the training period (5/4/2004 to 12/31/2015) |
| Fire.prior | Binary: if fire occurred on the prior day in the voxel, the value is 1; otherwise, 0 |
| Fire.last5 | Binary: if fire occurred during the prior five days in the voxel, the value is 1; otherwise, 0 |
| Nghb.prior | Binary: if fire occurred on the prior day in at least one of the neighboring eight voxels, the value is 1; otherwise, 0 |
| Nghb.last5 | Binary: if fire occurred during the prior five days in neighboring voxels the value is 1; otherwise, 0 |
| Week | Week of the year (taking on the values 1 to 52) |

Figure 6 shows a heatmap of locations of wildfires occurrences in Los Angeles County from 2004 to 2014 by voxel level.



Deeper red coloring implies higher occurrence of fires

Figure 6. Heatmap of locations of wildfires occurrences in Los Angeles County from 2004 to 2014 by voxel level

## 6. Designation of Training and Test Data Sets

The data that we use for our research consists of daily measurements for each of 256 voxels during the time frame of the study, excluding days on which sufficient RAWS measurements are not available. We partition the data into a training data set (which we use to fit the model) and a test dataset (which we use to evaluate model performance). The training dataset comprises the period from May 4, 2004, to December 31, 2015, and the test data set comprises the period from January 1, 2016, to December 31, 2018. Additionally, we examine model performance on "high fire risk" voxel-day combinations in the test data set. These instances consist of the 128 voxels with the highest incidence of

fires in the training data (variable Spatial) during weeks 23 to 48 which correspond approximately to the six months with the highest fire risk (June through November). Table 3 gives a brief summary of the training and test data sets. The proportions of observations with fires are shown in parentheses.

Table 3. A summary of the training and test data sets

| **Data Set** | **Period** | **Observations** | **Voxels with Fire** |
|---|---|---|---|
| Training | 5/4/2004 to 12/31/2014 | 622,848 | 1,820 (0.29%) |
| Test | 1/1/2015 to 12/31/2018 | 312,320 | 782 (0.25%) |
| *High-risk test subset* | | *82,176* | *400* (*0.64%*) |
| TOTAL | 5/4/2004 to 12/31/2018 | 935,168 | 2,602 (0.27%) |

## B. MODEL FORMULATION

It is natural to treat the occurrence of wildfires in Los Angeles County as a time- and spatially dependent stochastic process. The inclusion of predictor variables allows this process to account for inherent spatial variability, seasonal effects, time-lagged effects, and environmental conditions (including weather). By discretizing time into days and the spatial aspect into voxels it becomes possible to employ well-known estimation methods such as Poisson or logistic regression to the counts of fires in small spatio-temporal regions. Because the number of fires on a given day in any fixed 25 km$^2$ region of Los Angeles County most likely is equal to zero, and when not zero most likely is equal to one, Poisson or logistic regression could be used almost interchangeably to produce probability estimates or to assess the influence of predictor variables. To illustrate, let $\lambda$ denote the mean of a Poisson random variable $X$ for which the probability that $X$ is greater than 1 is negligibly small, and let $Y = min(X, 1)$. Then, $E(Y) = P(Y = 1) = p$, where

$$p = 1 - exp(-\lambda) \approx \lambda \quad (1)$$

if $\lambda$ is a small, positive number. A linear predictor based on a set of explanatory variables is estimated with both types of generalized linear regression models. With Poisson regression, the linear predictor estimates $log(\lambda)$, and with logistic regression it estimates $\text{logit}(p) \overset{\text{def}}{\equiv} log(p/1-p)$, which Equation (1) implies is approximately the same as $log(\lambda)$. Therefore, either type of model can be used to produce essentially the same analysis. We use logistic regression or extensions of it for consistency, as did other authors, including Scholz (2019).

An alternative to a model-based approach is to use a machine-learning method to estimate the probability of fire in a voxel from a set of explanatory variables. There are several advantages to this approach: the form of a model does not have to be specified in advance, and complex interactions between variables can be incorporated into the predictor. The main disadvantages are that the black-box nature of a machine learner does not naturally lend insight into how the explanatory variables influence the estimates, and the final product is not easily ported into other applications such as a resource-optimization model used by Scholz (2019) or Seeberger (2021). Despite these shortcomings, we use random forests (Breiman 2001) to identify important predictors for our model-based approach.

### 1. Random Forests

Random forest is a widely used machine-learning algorithm that uses bootstrap aggregation of decision trees to develop predictors of an outcome variable. In addition, random forests can be used for various analysis purposes such as selecting important predictor variables, expressing interactions between those variables, and obtaining accurate metrics for model performance. For our research, we implement the software of random forests in the R package **ranger** (Wright et al. 2017), and we use functions from the package **randomForestExplainer** (Paluszynska et al. 2020) to interpret the results of fitting a random forest to data.

## 2. Generalized Additive Models

A generalized additive model (GAM) is an extension of a generalized linear model (GLM) which is used to develop regression-like prediction models for exponential families. A GLM expresses the relationship between the mean $\mu$ of the outcome variable and the predictor variables using a function of the following form

$$g(\mu) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \tag{2}$$

where $g(\mu)$ is the link function, and $X_1, \ldots X_k$ are predictor variables. The expression on the right side of Equation (2) is the response function, which in this case is a linear function of the predictor variables. The link function depends on the type of GLM that is being fit. For Poisson regression $g(\mu)$ is the natural logarithm and for logistic regression it is the logit function introduced earlier. These models are fit to data using maximum likelihood estimation. GAM extends the GLM framework by allowing the response function to be additive in the predictors

$$g(\mu) = s_1(X_1) + \cdots + s_k(X_k), \tag{3}$$

where $s_1, \ldots, s_k$ are possibly unspecified functions of the predictors. These functions are estimated using smoothers such as splines or kernel smoothers. The reader is referred to Faraway (2016) for details on fitting a GLM or GAM to data.

We use GAMs to estimate wildfire probabilities in Los Angeles County based on predictor variables that need transformations. For example, because the variable Week is cyclic its transformation should be nonlinear and agree at the endpoints. Similarly, meteorological and environmental variables should not be assumed to have linear effects. We use the R software function **gam** in the **mgcv** package (Wood 2006) to fit GAMs.

## 3. Predictor Variable Selection

The predictor variables that we consider can be grouped into three categories. The first category represents "static" effects that are well known in advance to influence the probability of a fire in a particular voxel or on a particular day. The variables Spatial and Week belong to this category. The second category, which represents spatio-temporal

effects, includes the variables Fire.prior, Nghb.prior, Fire.last5, and Nghb.last5. The third category consists of the daily RAWS variables: BI, Temperature, RH, Wind, LFM, KBDI, DFM, ERC, and SC. For the models that we consider we always include variables from the first and second categories as predictors, which we refer to as the "default" variables. For variables in the third category, we adopt the approach used by Xu and Schoenberg (2010) to compare the effect of BI relative to all other RAWS variables excluding BI. The reasoning is that BI is already a function of the other RAWS variables. It is of interest to examine whether BI is effective in capturing the information from those variables. For Model 1, we use default variables and BI. For Model 2, we use default variables and all RAWS variables except BI. Figure 7 shows the variables selection for building a point process model. The same variables are used to make comparisons between a GAM and random forests.

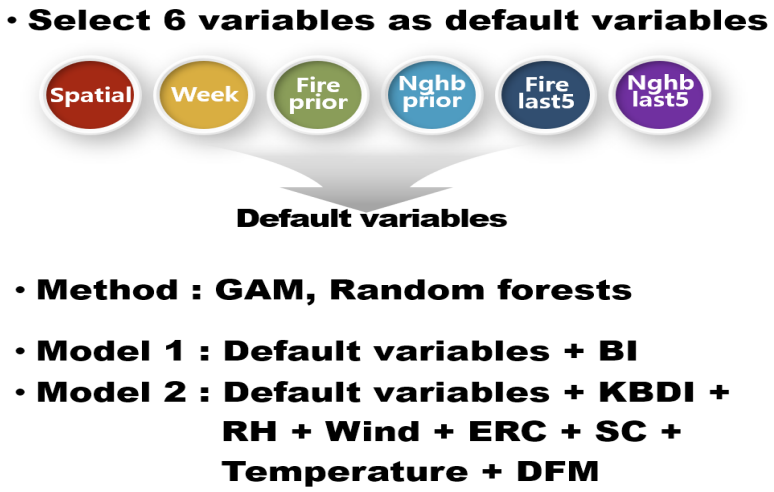



Figure 7. Variables selection for building point process models

## C. MODEL ASSESSMENT

In order to address our study objectives, it is necessary to evaluate and compare the performance of different models that we fit to the data. An important consideration is that the probability of a fire in any voxel on a particular day is certain to be small. Typical performance metrics for classifiers such as false positive or false negative rates do not lend

much insight to this assessment. We do not, in any case, make claims about the ability of our models to accurately predict the occurrence of a fire in a small location on a given day. Instead, we focus on aspects of fire generation that are detectable with our models, such as spatio-temporal dependence and the influence of predictor variables.

In classification problems, the area under the receiver operating characteristic (ROC) curve (AUC) is often used as a metric for model performance. The ROC curve is a plot of the sensitivity (detection of true effects) of a classifier versus its specificity (non-detection of spurious effects). If AUC is equal to 1 the classifier is essentially perfect, while a value of .5 suggests a classifier that does not perform better than guessing. We compare models fit to the training data by calculating their AUC values on the test data.

The influence of predictor variables in models that we consider can be measured in various ways. Random forests measure importance of a predictor variable by calculating the average depth of splitting of the bootstrapped decision trees that involve that variable. Smaller values indicate greater importance. The estimated smoothing function of a variable obtained from a GAM can be examined for evidence that it is not consistent with a constant. Statistical measures such as P-values for variable effects are also used, although they do not measure the impact of a predictor variable in determining outcomes.

Assessment of spatial or temporal effects poses challenges due to the fact that the probabilities of fire are not homogeneous in either respect. In regression problems this is overcome by examining residuals from the fitted model. In a logistic regression model, deviance residuals are used which take the form (Faraway 2016)

$$r_i = sign(y_i - \hat{p}_i) \cdot \sqrt{-2(y_i \text{logit}(\hat{p}_i) + log(1 - \hat{p}_i))}, \quad (4)$$

where $y_i = 1$ if a fire occurred in observation $i$ and $y_i = 0$ otherwise; and $\hat{p}_i$ is the estimated probability of a fire. We emphasize that these residuals are formed by estimating the model on the training data and using it to predict the test data. Because the probabilities of fire in a voxel on a single day are typically very small, the vast majority of residuals are negative numbers with small magnitudes; but on the relatively few occasions where fires occur, the residuals tend to be positive numbers with large magnitudes. Although useful

inferences with these residuals is possible by averaging large numbers of them, the effect of single fires as "outliers" remains to some extent.

We assess spatial and temporal effects by "stacking" the residuals in Equation (4) by voxel, which creates 256 time series over the time span. To reduce outlier influences we replace each residual in a series by its rank, where the smallest value is assigned a value of 1, the second largest a value of 2, etc. To detect spatial effects, we produce the 256 × 256 rank-correlation matrix across voxels and average the correlations for voxels that share a common distance between voxel centers. To detect temporal effects, we calculate the rank-autocorrelation function for each voxel and average these values across voxels at common time-lag values. We calculate standard errors under a null hypothesis of no correlation by permuting the 256 time series separately 1000 times and recalculating the respective statistics. The use of rank-based measures of correlation in a time series context is discussed in Hallin and Puri (1991).

THIS PAGE INTENTIONALLY LEFT BLANK

# IV. RESULTS

Our research is to further develop estimating the probability of wildfires in Los Angeles County using the daily weather variables recorded at RAWS and past wildfire occurrences. We start with a basic model that utilizes the smoothing effect, which is implemented through the R function **gam** in the **mgcv** package for generalized additive models, and the R function **ranger** in the **ranger** package for random forests. This model allows us to find out to what degree the seasonal and geographical effects influence the probabilities of wildfires in Los Angeles County. In addition, we are expanding the model so that it can estimate the probability of wildfires using other variables recorded at RAWS. We also conduct in-depth research on how much RAWS data we use for each voxel based on distance to find appropriate weights. Through this, we develop the best model that can also check the intensity of the probabilities of wildfire at a specific location. We also study what value to estimate by interpolating missing data for meteorological and environmental variables. Our analysis consists largely of two parts. First, we discuss the model fitting and model comparisons, using the area under the receiver operating curve (AUC). We look specifically at how well the model estimates predict wildfires in voxel-day combinations that are most prone to wildfires based on historical data. Second, we analyze dependence with respect to time and space that remains after fitting our models.

## A. MODEL FITTING AND ASSESSMENT

Based on the probability values estimated for each voxel, we can find the voxels with the highest probabilities of fire occurrence for resource allocation. Scholz (2019) selected areas to be allocated resources based on a partitioning of Los Angeles County into twenty-one areas; unlike his study, we consider 256 areas (voxels) that comprise smaller areas (25 $km^2$). We begin by presenting a summary of the point process models that we fit to the training data and analyze their performance on the test data.

### 1. Random Forests and Variable Importance

To use a spatio-temporal point process modeling approach, we divide Los Angeles County into small rectangular region called “voxels,” as explained in Chapter III. We

consider the occurrence of fires within each voxel as a function of time which we take as the outcome variable. We would like to find the best combination of variables for estimating wildfire probabilities. To do this, we fit a random forest to the training data with all of the predictor variables described in Chapter III using the R function **ranger**. We also include the variable Weekend, which is equal to 1 if the day falls on a Saturday or Sunday and is equal to 0 otherwise. We then use the R function **randomForestExplainer** on the estimated model to visualize the variable importance. As explained in Chapter III, high importance of a predictor is revealed by decision trees that contain splits on the variable at the earliest stages, which is measured by depth taking on smaller values. Figure 8 shows the result of this analysis. The variable Weekend is the least important of the predictor variables, and it is substantially weaker than the second least important predictor (Wind), which justifies our decision to exclude this variable from further analyses.

Figure 8 shows that Spatial is the most important variable: it indicates where fires are likely to occur solely from the geographical positioning of a voxel. This is not surprising: some voxels are in areas with large amounts of fuel for wildfires, others are not. The next three most important variables (Fire.prior, Nghb.prior, and Fire.last5) belong to our spatio-temporal set of predictors. Knowledge that a fire occurred in the same voxel the prior day, in a neighboring voxel the prior day, and in the same voxel during the last five days, stand out as important; and we note that these variables would be available to decision makers. The remaining spatio-temporal variable (Nghb.last5) is in the middle of the range of variable importance.

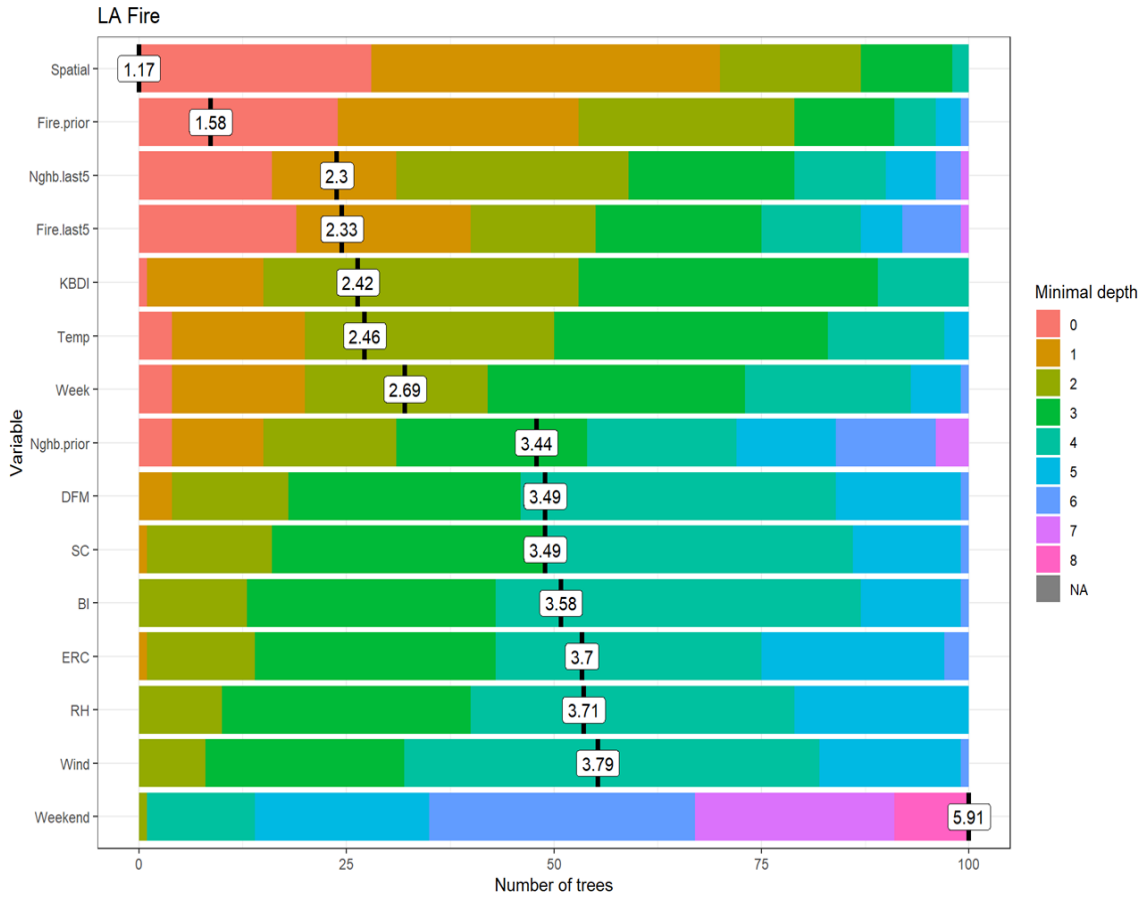



Figure 8. Variable importance using the R function **randomForestExplainer**

## 2. Comparisons of Models

Results of model-fitting are summarized in order to assess the quality of the estimated models, and to achieve our research objectives that are focused on the spatial and temporal properties of wildfires in Los Angeles County. Using the variable selection method described in Chapter III Section A, we fit various models using training data, then evaluate their performance using test data by calculating and plotting their AUC values. Table 4 shows AUC values with various setting of the parameter $a$ which determines how weighted averages of the RAWS variables are calculated. GAM Model 1 has the highest AUC values but GAM Model 2 and Random Forests Model 1 perform similarly. It is noteworthy that AUC changes little as a function of $a$, which may be due to low variability of RAWS variables over Los Angeles County on a given day. As expected, all models tend

to perform less well on the high-risk test set where fires are more frequent. Random Forests Model 2 shows the largest degradation in performance in this respect.

The effects of spatio-temporal variables (Fire.prior, Fire.last5, Nghb.prior, and Nghb.last5) on model performance can be seen by comparing Model 1 with GAM and random forests on the high-risk test set including and excluding these variables. In both instances their exclusion degrades model performance to a small extent, but the effect is greater with random forests.

Table 4. AUC values of GAM and random forest models under various scenarios

| | $a = 0.5$ | $a = 1.0$ | $a = 1.5$ | $a = 2.0$ | $a = 2.5$ | $a = 3.0$ |
|---|---|---|---|---|---|---|
| **GAM Model 1** | | | | | | |
| Full test set | 0.917 | 0.917 | 0.917 | 0.917 | 0.917 | 0.917 |
| High-risk test set | 0.837 | 0.837 | 0.837 | 0.837 | 0.837 | 0.837 |
| High-risk no S-T predictors | 0.832 | 0.832 | 0.832 | 0.832 | 0.832 | 0.832 |
| **GAM Model 2** | | | | | | |
| Full test set | 0.915 | 0.915 | 0.915 | 0.915 | 0.915 | 0.916 |
| High-risk test set | 0.838 | 0.837 | 0.838 | 0.837 | 0.837 | 0.837 |
| **Random Forests Model 1** | | | | | | |
| Full test set | 0.908 | 0.909 | 0.908 | 0.909 | 0.908 | 0.908 |
| High-risk test set | 0.834 | 0.832 | 0.834 | 0.834 | 0.833 | 0.834 |
| High-risk no S-T predictors | 0.815 | 0.816 | 0.817 | 0.815 | 0.816 | 0.818 |
| **Random Forests Model 2** | | | | | | |
| Full test set | 0.881 | 0.883 | 0.885 | 0.887 | 0.887 | 0.887 |
| High-risk test set | 0.814 | 0.816 | 0.815 | 0.818 | 0.815 | 0.813 |

S-T = Spatio-temporal

In order to give visual context to the AUC values presented in Table 4, Figure 9 shows box plots of estimated probabilities from Random Forest Model 1 on the high-risk test data ($a = 1.5$) separated by whether a fire occurred on a voxel-day combination. The

AUC value is 0.834 which suggests that the probabilities should show good separation, which is apparent in the box plots.

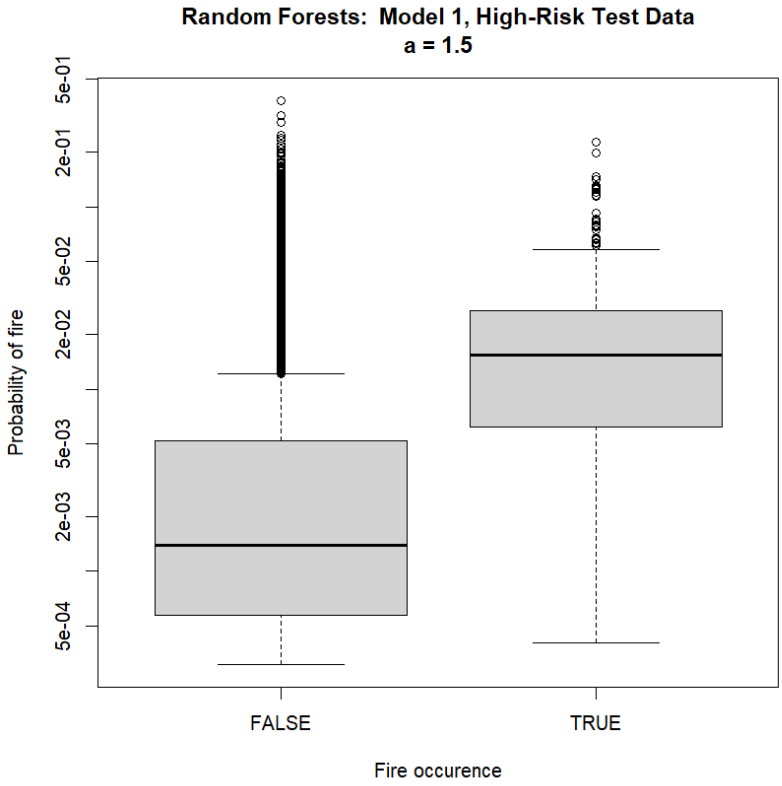



Figure 9. Box plots of fire probabilities estimated from random forests Model 1 high-risk test data ($a$ =1.5)

Figure 10 summarizes the results of fitting GAM Model 1 to the training data. Due to the large sample size and the variable importance shown in Figure 8, it is not surprising that all of the predictors included in the model have statistically significant effects. All coefficients on the spatio-temporal predictors are positive, which suggests that a recent occurrence of fire in the same or neighboring voxel increases the probability of fire.

Figure 11 shows the transformations that are estimated for the numerical predictors. For Week the smoother is constrained to be periodic so that Week = 1 follows Week = 52. The grey area indicates 95% confidence bounds. A useful interpretation of these plots is to see if a completely horizontal line, or a line with nonzero slope, can remain inside the

confidence bounds. It would indicate that the predictor variable is not significant in the first case, and that a nonlinear transformation is not needed in the second case. In none of the plots shown do either of these conditions hold. In terms of Week, wildfires are most frequent during summer in Los Angeles County. The smooth of Spatial suggests an increasing effect on the probability of fire that tapers off for larger values, and a similar interpretation applies to BI.

```
Parametric coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    -8.15999    0.09974 -81.813  < 2e-16 ***
Fire.priorTRUE  0.87021    0.10612   8.200 2.40e-16 ***
Nghb.priorTRUE  0.32376    0.07665   4.224 2.40e-05 ***
Fire.last5TRUE  0.30232    0.07672   3.941 8.13e-05 ***
Nghb.last5TRUE  0.25942    0.05473   4.740 2.14e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
             edf Ref.df Chi.sq p-value
s(Spatial) 8.950  8.999 1998.8  <2e-16 ***
s(Week)    4.966  9.000  215.5  <2e-16 ***
s(BI)      6.675  7.841   50.3  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 10. A summary of GAM Model 1 ($a$ = 1.5)

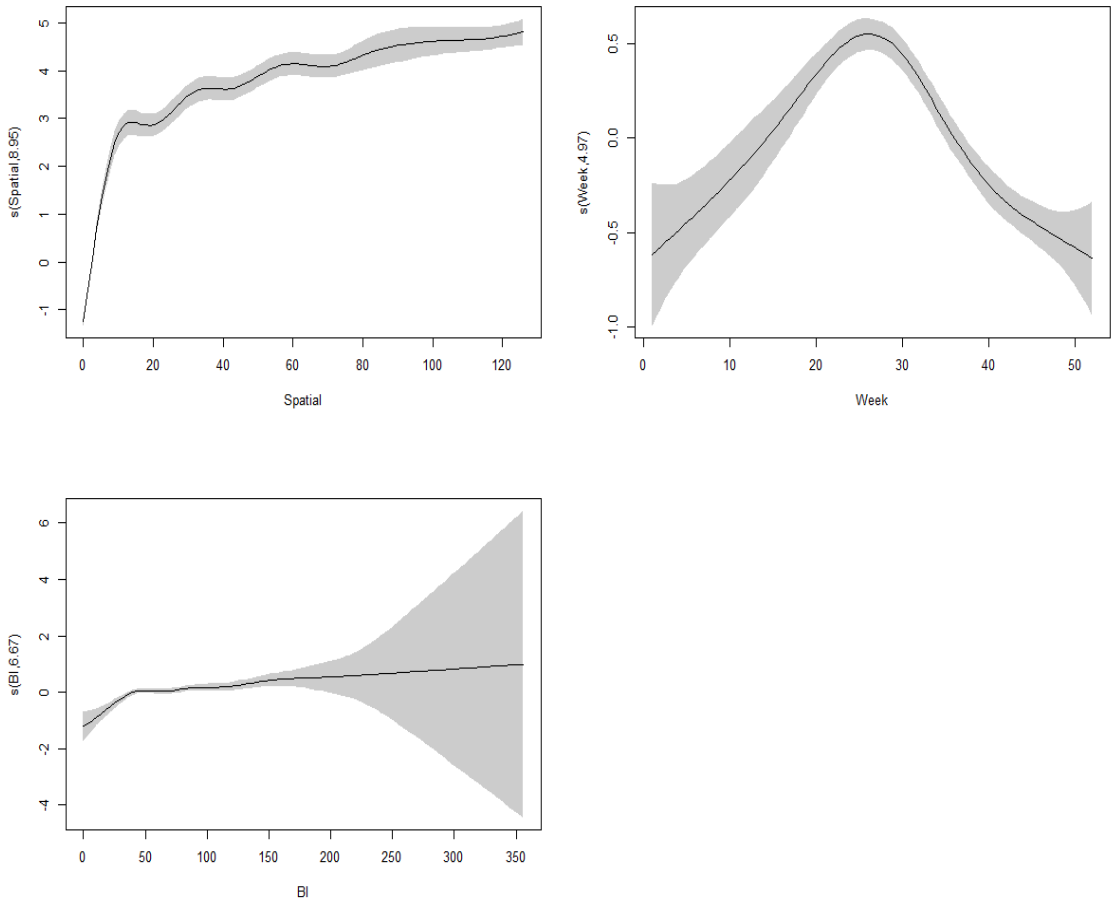



Figure 11. Plots of smoothed transformations for GAM Model 1 ($a$ = 1.5)

We produce a heatmap of voxels for July 24, 2018, that show voxels with higher probabilities in darker shades of red. It demonstrates that GAM-based, spatial temporal point process modeling could be used to create plausible wildfires probability maps from historic weather and fire occurrence data. They can be used to inform resource-allocation decisions similar to the approach used by Scholz (2019) by using a finer geographic partition of Los Angeles County that takes spatio-temporal factors into consideration.

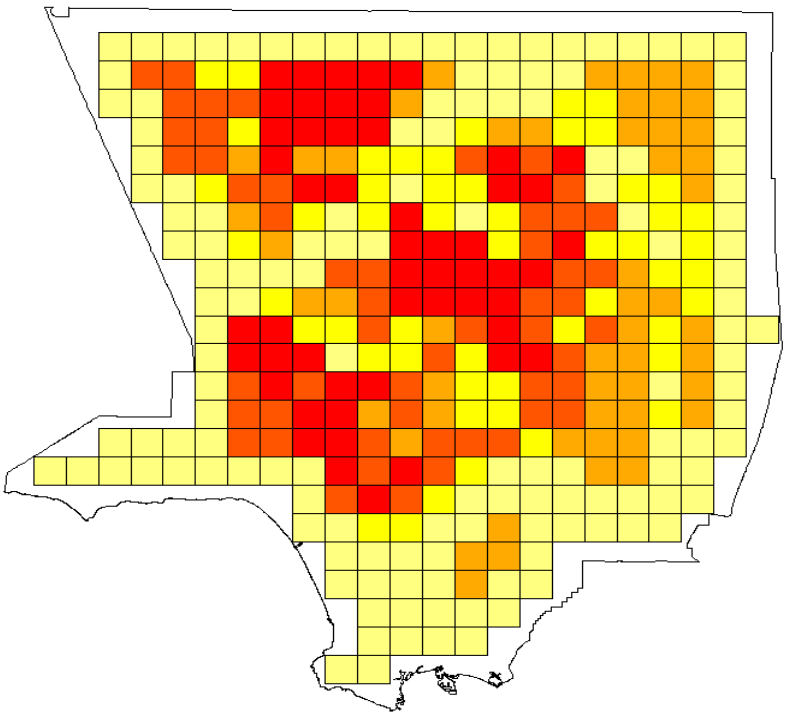


Figure 12. Heatmap of voxels with high probability of wildfires on a specific date (July 24, 2018)

Figure 13 compares the number of actual fires with the number of predicted number of fires with daily temporal resolution from June to November 2018. Wildfires occurred in 14 voxels on July 4, 2018, and thus the predicted number of voxels with fires increased on the following day due to the use of Fire.prior as a predictor variable. It should be noted that July 4 is a national holiday that experiences many wildfires due to its association with fireworks. From 2005 to 2018, July 4 had more fires than any other day of the year, and more than twice as many as July 5, which had the second most fires.

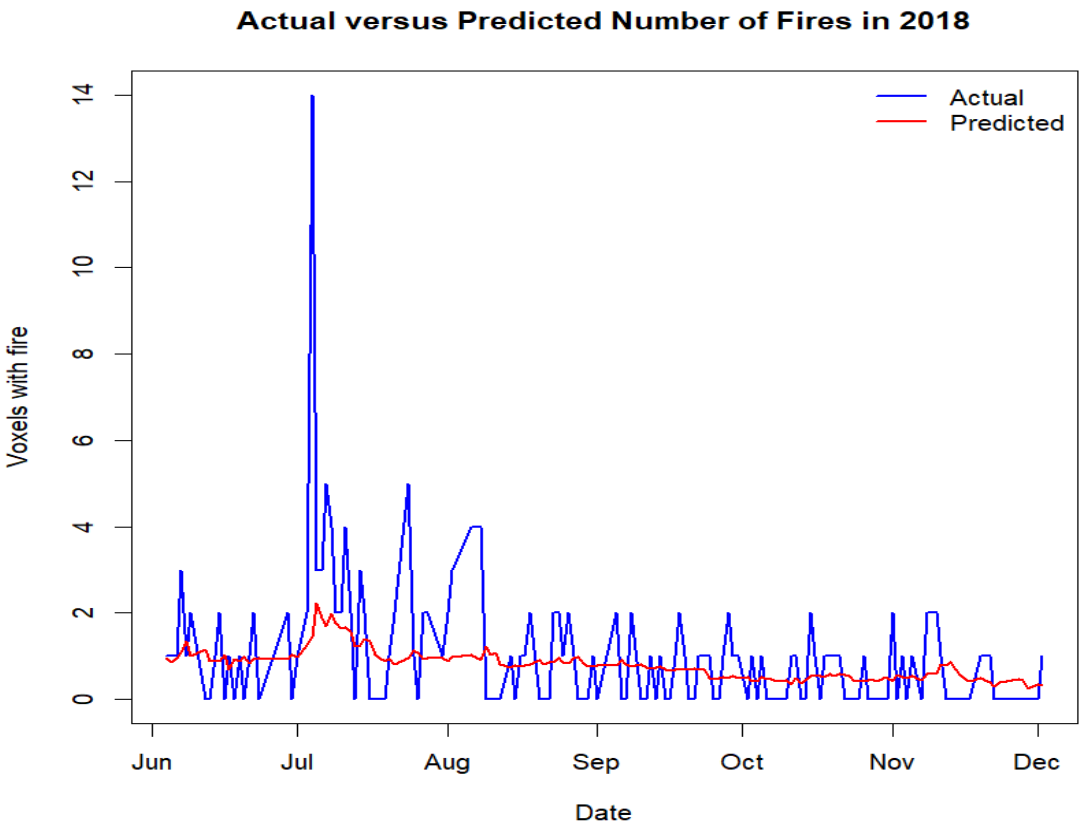



Figure 13. Actual versus predicted number of fires in 2018

## B. ANALYSIS OF SPATIAL AND TEMPORAL PROPERTIES

In the previous section we find that the spatio-temporal variables Fire.prior, Fire.last5, Nghb.prior, and Nghb.last5 contribute measurably to the quality of our models, which directly addresses our research objectives. In this section, we consider whether spatio-temporal effects remain after accounting for these variables. To do this, we analyze deviance residuals from the training data in GAM Model 1 to assess spatial and time dependence. We do this for the full training data and for a high-risk set that corresponds to the 128 voxels having the largest number of fires restricted to the months of June through November.

### 1. Spatial Analysis of Deviance Residuals

To examine residual spatial dependence, we calculate a 256 × 256 rank correlation matrix of deviance residuals from GAM Model 1 corresponding to voxels, and average the correlations for voxel pairs sharing a common distance between centroids. Directly neighboring voxels, for instance, have a distance of 5.0 km; diagonally neighboring voxels share a distance of 7.07 km, etc. We repeat this analysis on the high-risk training set. Figure 14 shows the rank correlations by distance with the full training set and the high-risk training set.

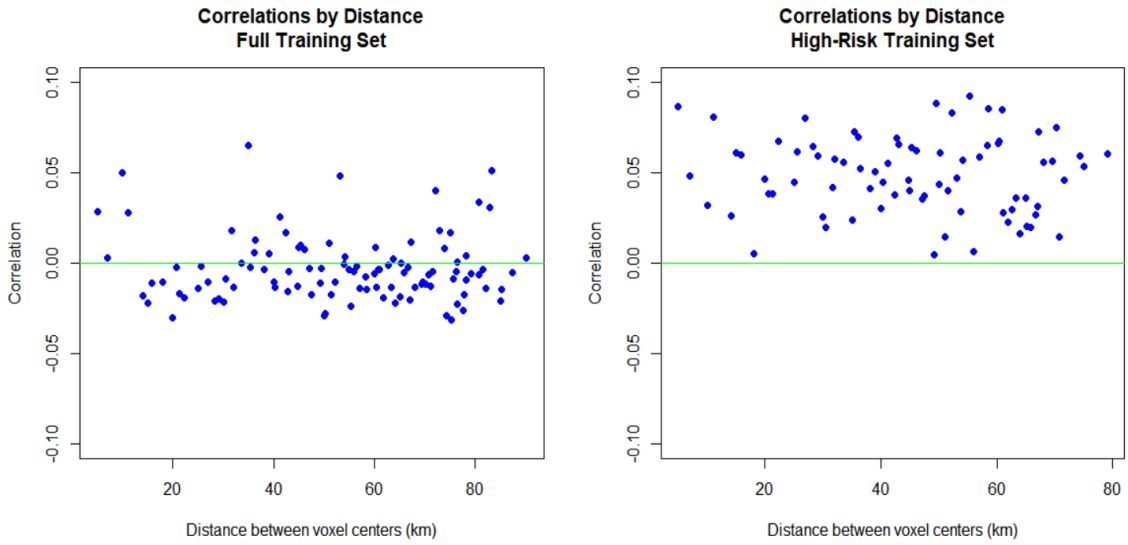



Figure 14. Rank correlations by distance with full training set and high-risk training set

Correlations in the full training set are small in magnitude and take on both positive and negative values with an average of about –.002, although the four smallest distances produce positive spatial correlations. Restricting the analysis to the 128 voxels with the highest occurrence of fires, and the months June through November when fires are most frequent, presents a somewhat different picture. In this case the spatial autocorrelations are all positive and average near .05. This analysis suggests that our models do not fully capture

the spatial relationship in fire events, and that it may not be the same across all voxel-day combinations.

### 2. Temporal Analysis of Residuals

To examine time-based dependence, we use the **acf** function in R to calculate rank autocorrelations with the deviance residuals from GAM Model 1 with lags up to 20 days for each voxel, and then average the results across voxels for each lag value. As in the previous sub-section we do this for the full training data and for the high-risk training data separately. Figure 15 shows rank autocorrelations for the two cases, both of which point to the presence of autocorrelation in the residuals despite the use of time-lagged predictors in the model.

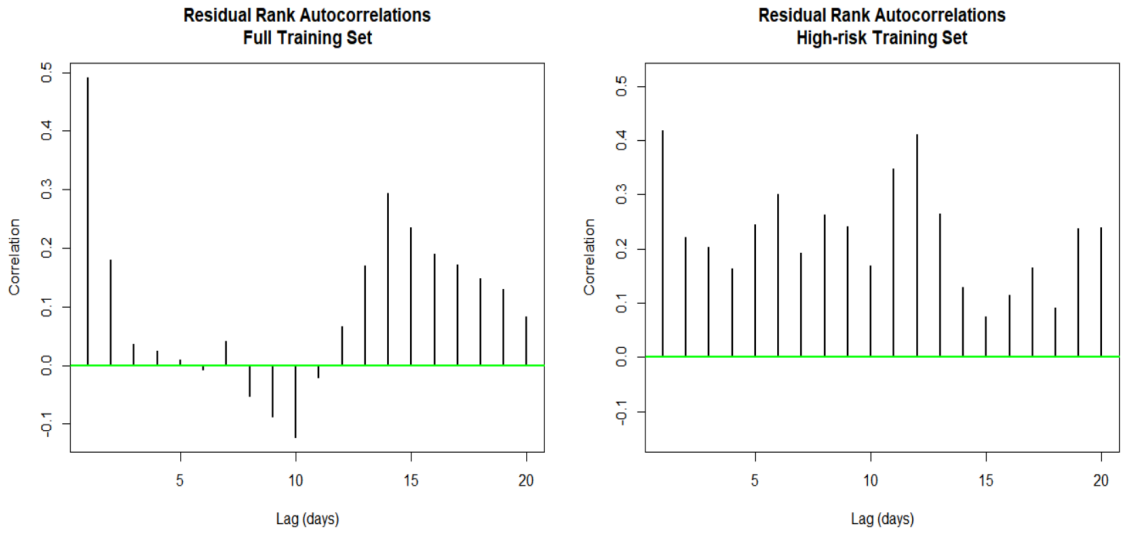



Figure 15. Residual rank autocorrelations with full training set and high-risk training set

THIS PAGE INTENTIONALLY LEFT BLANK

# V. CONCLUSION AND RECOMMENDATIONS

## A. CONCLUSION

We use historical RAWS data and past fire occurrence data to estimate the probability of wildfires in Los Angeles County. Our objectives are to describe spatial and temporal dependence that may lead to the development of improved models for predicting wildfires. Our research demonstrates the existence of these effects in two ways: first, by noting that four predictor variables that capture past and nearby incidences of wildfires into our models contribute measurably to the predictive quality of the models; and second, by finding notable spatio-temporal correlations in the residuals that we derive from those models. Although our models are an improvement over those that do not include spatio-temporal predictors, other models that more fully capture spatio-temporal effects may do even better.

Although generalized additive models (GAMs) and random forests perform similarly, GAMs hold a small advantage. Using probability estimates from either model could reduce the cost and time it takes to augment and position personnel and equipment.

## B. RECOMMENDATIONS

We believe it is worthwhile to continue research into improving models for wildfire prediction. It becomes an increasingly challenging problem as finer details of geography including urbanization, vegetation, fuel type, etc., are made available for use in models.

In addition, the model would be enhanced by more data that could be gained from the installation of additional RAWS to measure the weather and environmental data in Los Angeles County. This enhancement would be affordable and practicable as a RAWS station is inexpensive to equip and easy to install. Ideally, if there were more RAWS stations in Los Angeles County, we could set weights to come up with a better way to reflect the weather information from RAWS for each voxel.

Similarly, in terms of model expansion, we have created models and evaluated their application for only the Los Angeles County area. The opportunity also exists to apply this our approach to other fire-prone areas, assuming similar data exists.

# LIST OF REFERENCES


Breiman L (2001) Random forests. *Machine Learning* 45(1):5–32, https://doi.org/10.1023/A:1010933404324.

Brown GG, Koyak RA, Salmerón J, Scholz Z (2021) Optimizing prepositioning of equipment and personnel for Los Angeles County Fire Department to fight wildland fires. *INFORMS Journal on Applied Analytics,* 51(6), https://doi.org/10.1287/inte.2021.1084.

CALFIRE (2020) 2021 Incident archive. Accessed February 3, 2022, https://www.fire.ca.gov/incidents/2021/.

Diggle PJ, Moraga P, Rowlingson B, Taylor BM (2013) Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statistical Science* 28(4), https://doi.org/10.1214/13-STS441.

Faraway JJ (2016) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models* (Chapman and Hall, New York, NY).

Genton MG, Butry DT, Gumpertz ML, Prestmon JP (2006) Spatio-temporal analysis of wildfire ignitions in the St. Johns River Water Management District, Florida. *International Journal of Wildland Fire* 15(1), https://doi.org/10.1071/WF04034.

Hallin M, Puri ML (1991) Rank tests for time series analysis: A survey. *New Directions in Time Series Analysis* (Springer-Verlag, New York, NY), 111–154.

LACoFD (2021) 2020 statistical summary, Accessed November 5, 2021, https://fire.lacounty.gov/wp-content/uploads/2021/06/2020-Statistical-Summary-FINAL-DRAFT.pdf.

McEvoy DJ, Hobbins M, Brown TJ, VanderMolen K, Wall T, Huntington JL, Svovoda M (2019) Establishing relationships between drought indices and wildfire danger outputs: A test case for the California-Nevada drought early warning system. *Climate* 7(4), https://doi.org/10.3390/cli7040052.

Opitz T, Bonneu F, Gabriel E (2020) Point-process based Bayesian modeling of space-time structures of forest fire occurrences in Mediterranean France. *Spatial Statistics* 40, https://doi.org/10.1016/j.spasta.2020.100429.

Preisler HK, Brillinger DR, Burgan RE, Benoit JW (2004) Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire* 13(2):133-142, https://doi.org/10.1071/WF02061.

RAWS (2021) Remote automated weather stations. Accessed March 11, 2022, https://raws.nifc.gov/.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria), https://www.R-project.org/.

Scholz ZT (2019) Optimizing resource augmentation for wildland fires. Master's thesis, Department of Operations Research, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/62835.

Seeberger RA (2020) A new simulation-optimization model for wildland fire resource pre-positioning. Master's thesis, Department of Operations Research, Naval Postgraduate School, Monterey, CA. http://hdl.handle.net/10945/66715.

Paluszynska A, Biecek P, Jiang Y (2020) randomForestExplainer: Explaining and visualizing random forests in terms of variable importance. R package version 0.10.1. Accessed January 10, 2022, https://CRAN.R-project.org/package=randomForestExplainer.

Wood SN (2006) mgcv: *Generalized Additive Models: An Introduction with R* (Chapman and Hall, New York, NY).

Wright M, Ziegler A (2017) ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77(1):1–17, https://doi.org/10.18637/jss.v077.i01.

Xu H, Schoenberg FP (2011) Point process modeling of wildfire hazard in Los Angeles County, California. *The Annals of Applied Statistics* 5(2a):684–704.

# INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California