UNIVERSITY OF CALIFORNIA

Los Angeles

Spatial Modeling of the Los Angeles County Fires

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Ashwin Ramaseshan

2025

The dissertation of Ashwin Ramaseshan is approved.

Nicolas Christou

Guang Cheng

Rick Schoenberg, Committee Chair

University of California, Los Angeles

2025

Dedication

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

(Acknowledgments omitted for brevity.)

CHAPTER 1

Introduction

Wildfires have emerged as one of the most pressing environmental and social challenges of the twenty-first century, with devastating consequences for ecosystems, communities, and the economy. In recent years, Los Angeles County has experienced increasingly frequent and severe fires, driven by a combination of climate variability, urban expansion, and fuel accumulation. Understanding the spatio-temporal dynamics of wildfire occurrence is critical not only for scientific insight but also for informing insurance models, risk mitigation policies, and resource allocation for fire management agencies.

This thesis aims to develop a statistical framework for modeling the occurrence of wildfires in Los Angeles County by integrating both temporal and spatial covariates. Leveraging fire detection data from NASA's VIIRS and MODIS satellites, combined with meteorological variables and topographic features, I investigate the extent to which predictors such as maximum temperature, vapor pressure deficit, wind speed, and precipitation explain the likelihood of fire occurrence. Particular attention is given to the incorporation of space-time interactions, specifically measures of the time since a location or nearby location last burned.

The methodology builds on the Stoyan–Grabarnik (SG) space-time point process model, a flexible approach for capturing self-exciting processes such as wildfires. By comparing kernel density estimates with SG-based conditional intensity models, I compare both the explanatory power of covariates and the predictive ability of the framework.

1.1 Abstract

Wildfire risk has become an urgent concern across the western United States, with California in particular experiencing unprecedented fire seasons in the past two decades. According to the California Department of Forestry and Fire Protection (Cal Fire), seven of the ten largest wildfires in California history have occurred since 2017, underscoring a rapidly escalating crisis. The human and economic toll is immense: in addition to lives lost and homes destroyed, insurers face billions of dollars in claims each year, while local governments struggle to allocate limited resources for fire suppression and prevention. These events highlight the need for rigorous, data-driven models that can improve our understanding of where and when fires are most likely to occur.

Los Angeles County represents a particularly compelling study area. Its unique geography encompasses coastal regions, dense urban development, and mountainous terrain, creating highly heterogeneous fire regimes. The combination of Santa Ana winds, prolonged droughts, and increasing development at the wildland–urban interface (WUI) amplifies both the frequency and destructiveness of fires. High-profile events such as the 2019 Getty Fire, the 2021 Palisades Fire, and more recent incidents in 2025 have demonstrated how rapidly wildfires can ignite and spread within this region, often overwhelming response capabilities. Understanding spatial variation in fire risk within Los Angeles County is therefore crucial for both scientific inquiry and practical risk assessment.

Wildfires present a rich but challenging problem for statistical modeling. Fires are inherently spatio-temporal phenomena: they occur at specific locations and times, but their risk is influenced by both past events and environmental conditions. Unlike purely temporal models, which treat events as independent over space, or purely spatial models, which ignore the dynamic evolution of risk through time, space-time statistical models allow for joint treatment of location and timing. This makes them particularly well suited to wildfire applications, where recurrence at the same site are fundamental in this process.

In this thesis, I first develop exploratory analyses that provide descriptive insight into the patterns of fire occurrence, using kernel density estimates and summary statistics to

highlight spatial hotspots and temporal trends. These baseline methods establish benchmarks against which more advanced models can be evaluated. Second, I apply and extend the Stoyan–Grabarnik (SG) space-time model, which explicitly incorporates covariates to estimate the conditional intensity of wildfire occurrence. A particular focus is placed on time-since-burn metrics: the time since a given location last burned, and the time since nearby locations burned. These measures capture fuel dynamics and spatial spillover, both of which are critical in understanding fire recurrence.

1.2 Background and Literature Review

Wildfires are complex natural phenomena driven by the interaction of climate, vegetation, topography, and human activity. The scientific literature on wildfire dynamics spans multiple disciplines, including ecology, atmospheric science, engineering, and statistics. In this section, I review three main areas of prior work that inform this thesis: (i) ecological and climatic drivers of wildfire occurrence, (ii) statistical and machine learning models for fire prediction, and (iii) spatio-temporal point process methods, particularly the Stoyan–Grabarnik framework, that provide a foundation for the modeling approach adopted here.

1.2.1 Ecological and Climatic Drivers of Wildfire Occurrence

A large body of ecological research emphasizes the role of fuel availability and climate in shaping fire regimes. Vegetation type and density determine the amount of combustible material, while climate conditions influence both fuel moisture and ignition likelihood. In California, the combination of Mediterranean climate patterns (characterized by cool, wet winters and hot, dry summers) and prolonged droughts has created conditions favorable to frequent large-scale fires. Studies have consistently identified temperature, precipitation, relative humidity, and wind speed as critical predictors of fire ignition and spread. In particular, vapor pressure deficit (VPD) has emerged as a robust indicator of fuel aridity, with recent work linking increasing VPD levels under climate change to more extreme fire seasons.

Topography also plays an important role. Elevation, slope, and aspect shape microcli-

mates and influence fire spread rates, while canyons and ridgelines can funnel winds, amplifying fire intensity. In Los Angeles County, the Santa Monica and San Gabriel Mountains create sharp gradients in fire risk across short distances, with urban neighborhoods often in high-risk zones. Human activity further complicates these dynamics: most ignitions in California are human generated, resulting from power line failures, equipment sparks, or intentional acts. The wildland–urban interface (WUI) is thus a critical zone of study, where natural and human factors interact to heighten risk.

1.2.2 Statistical and Machine Learning Approaches

Efforts to model and predict wildfire occurrence span a wide methodological spectrum. Traditional regression-based approaches have long been employed to link fire activity with climate variables. For example, logistic regression models have been used to estimate the probability of ignition given environmental covariates, while Poisson and negative binomial regressions model fire counts or burned area. These models are interpretable and can highlight key drivers, but they often struggle with complex, nonlinear interactions.

Machine learning methods, including random forests, gradient boosting, and neural networks, have gained traction in recent years. These approaches excel at capturing nonlinearities and interactions, and they can integrate high-dimensional datasets such as satellite imagery. For instance, researchers have successfully applied boosted regression trees to predict fire occurrence in Mediterranean ecosystems and neural networks to detect fires directly from remote sensing data. However, the main limitation of these methods is their "black box" nature and limited interpretability. While they often deliver strong predictive accuracy, they do not always provide clear statistical inference about underlying processes.

A separate but related branch of the literature focuses on hazard models and survival analysis, which treat time-to-fire as the outcome of interest. These models emphasize recurrence intervals and are useful for capturing the influence of covariates on the probability of burning at different times. Nonetheless, survival approaches are typically temporal in scope and do not fully integrate spatial dependence.

1.2.3 Spatio-Temporal Statistical Models

Wildfires are inherently spatio-temporal processes, and statistical models that jointly account for space and time have become increasingly important. Spatial point process models, originally developed in ecology and epidemiology, are particularly relevant. These models treat fire ignitions as "events" that occur in continuous space and time, allowing for the estimation of conditional intensities that vary with location, time, and covariates.

The Stoyan–Grabarnik (SG) framework extends these ideas by integrating spatial kernels and temporal decay functions into a unified conditional intensity model. In SG models, the probability of fire occurrence at a given location and time depends on both external covariates (e.g., temperature, wind speed, precipitation) and internal components (e.g., time since last burn at that site or in neighboring sites). This makes the SG model particularly appealing for wildfire analysis. Although SG models have been applied in forestry and pest infestation studies, their use in wildfire contexts remains relatively limited, especially at the county level. This thesis therefore contributes to an emerging but underdeveloped area of wildfire modeling.

1.3 Sources

The data for this thesis was collected from a combination of satellite remote sensing products, publicly available climate reanalysis datasets, and official geographic boundary files. Because wildfire research relies on consistent spatial and temporal coverage, all data were chosen for their open availability, high resolution, and reliability. This section describes the sources used and how they were accessed.

1.3.1 Fire Detection Data

The primary source of wildfire occurrence data was NASA's **Fire Information for Resource Management System (FIRMS)**. FIRMS provides near real-time active fire detections from the MODIS and VIIRS satellites. For this study, I focused on the VIIRS 375

m product due to its finer spatial resolution and higher sensitivity, which allows for the detection of smaller fires that are common in Los Angeles County.

Data were obtained through the FIRMS website and API (https://firms.modaps.eosdis.nasa.gov/), where I specified the geographic bounding box for Los Angeles County and requested all available records from January 2015 through January 2025. The downloaded files contained variables such as acquisition date and time, latitude and longitude, Fire Radiative Power (FRP), brightness, and confidence score.

1.3.2 Climate and Meteorological Data

Daily climate data were sourced from the **GRIDMET** gridded surface meteorology dataset, developed by the University of Idaho and hosted by the University of California Merced's Climate Toolbox and Google Earth Engine (https://climate.northwestknowledge.net/). GRIDMET provides high-resolution (4 km) meteorological variables including precipitation, temperature, humidity, and wind speed across the contiguous United States.

I used Google Earth Engine (GEE) to access this data. GEE allows users to query specific datasets and export subsets for a chosen region and time period. Using GEE's interface, I selected the GRIDMET collections, filtered them by date range (2015–2025) and by the Los Angeles County boundary, and exported daily values in CSV format. This approach provided a consistent and continuous record of weather conditions relevant to fire activity.

1.3.3 Geographic Boundaries

To define the spatial extent of the study, I used the official **Los Angeles County boundary shapefile** available from the U.S. Census Bureau's TIGER/Line database (https://www.census.gov/geo files/time-series/geo/tiger-line-file.html). This shapefile provided an accurate and standardized polygon representing the county, which was used to clip and filter all other datasets.

In addition, I accessed the **USGS National Land Cover Database (NLCD)** (https://www.usgs.g land-cover-database) to obtain categorical land cover information for the region. Although primarily used for descriptive purposes, these data provide context on vegetation types and

the wildland–urban interface, both of which are important to understanding the fire environment.

1.3.4 Data Access and Tools

All datasets used in this thesis are publicly available and were accessed through their respective websites or APIs. NASA FIRMS offers both a web-based map viewer for interactive exploration and direct download options for CSV or shapefile formats. The GRIDMET data were accessed via Google Earth Engine, which supports reproducible, script-based queries and large-scale data exports. Boundary shapefiles and land cover data were downloaded manually from official government data portals.

Together, these sources form a comprehensive and transparent foundation for wildfire research. The use of publicly accessible, well-documented datasets ensures that the work presented here can be replicated and extended in future studies.

graphicx float [section]placeins

1.4 Exploratory Data Analysis

This section provides an exploratory overview of the wildfire and climate data used in the study. The goal of this step is to visualize the spatial and temporal patterns of fire detections across Los Angeles County, identify areas with higher concentrations of fire activity, and understand the general distribution of fire radiative power (FRP). Each figure in this section is accompanied by a detailed caption explaining the key patterns observed. Together, these figures provide context for the modeling work that follows and motivate the inclusion of specific predictors in the statistical models.

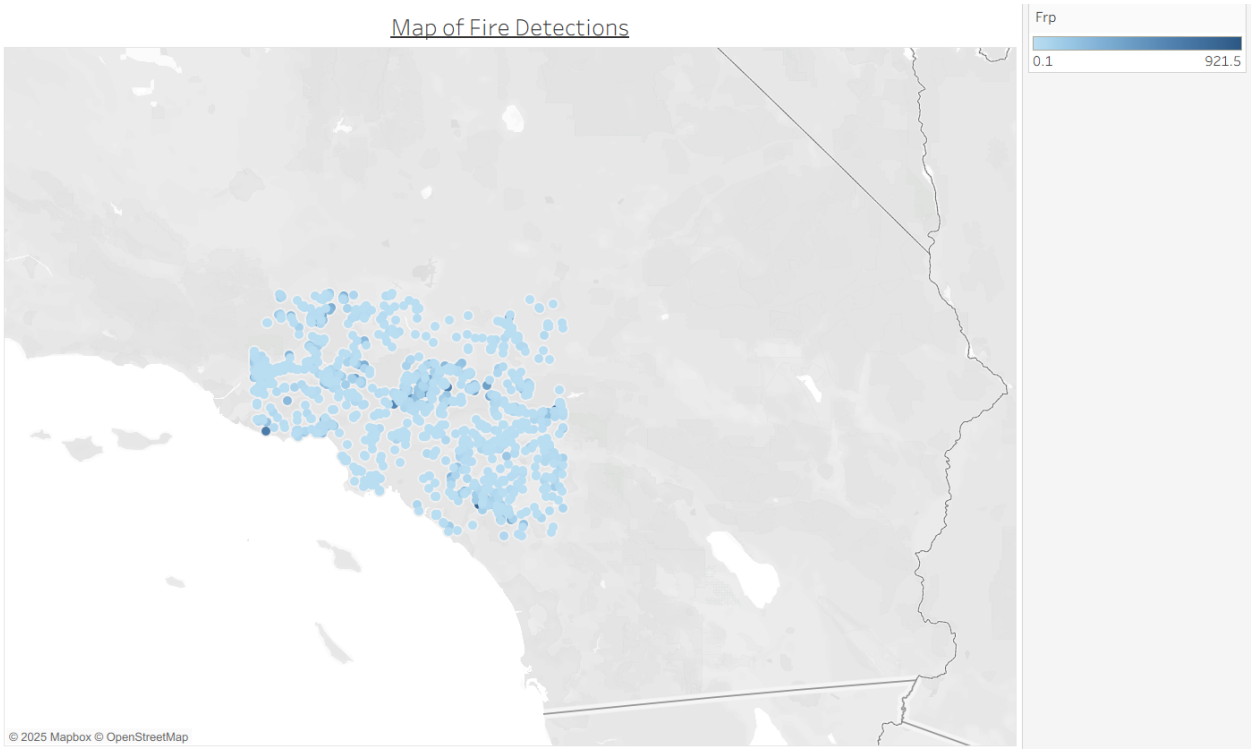

Figure 1.1: **Map of Fire Detections (View 1).** This map shows all fire detections recorded in Los Angeles County between 2015 and 2024, based on VIIRS satellite data obtained from NASA FIRMS. Each point represents a fire detection, with color intensity scaled by Fire Radiative Power (FRP). The plot highlights the clustering of fire activity in the northern and northwestern portions of the county, particularly near the Santa Monica Mountains and the Angeles National Forest. Coastal and urban areas show relatively lower detection density, consistent with reduced fuel loads and active fire suppression efforts in these regions.

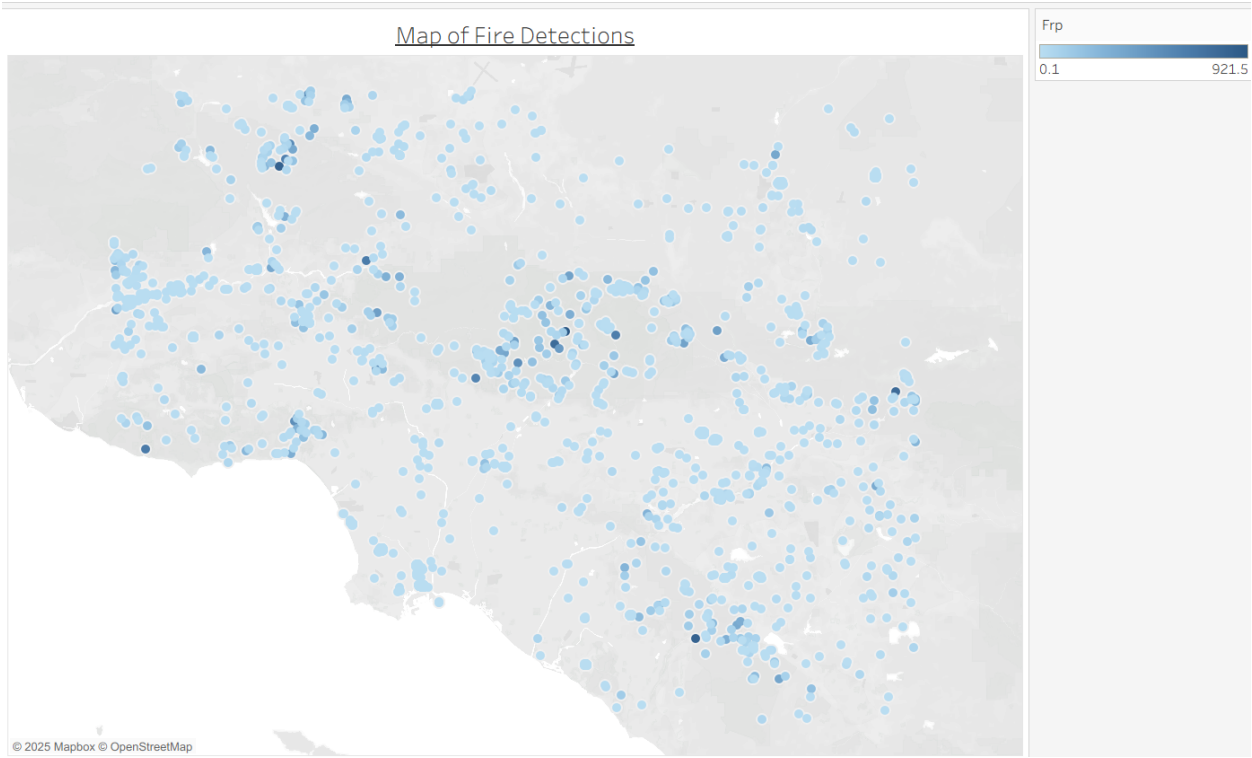

Figure 1.2: **Map of Fire Detections (View 2).** A second map of fire detections with a slightly different visualization style to confirm spatial patterns. The use of a continuous blue color scale helps emphasize FRP intensity, showing that while most fire detections have relatively low FRP values, a few high-intensity events stand out and are spatially concentrated. These hotspots likely correspond to major wildfire events over the study period.

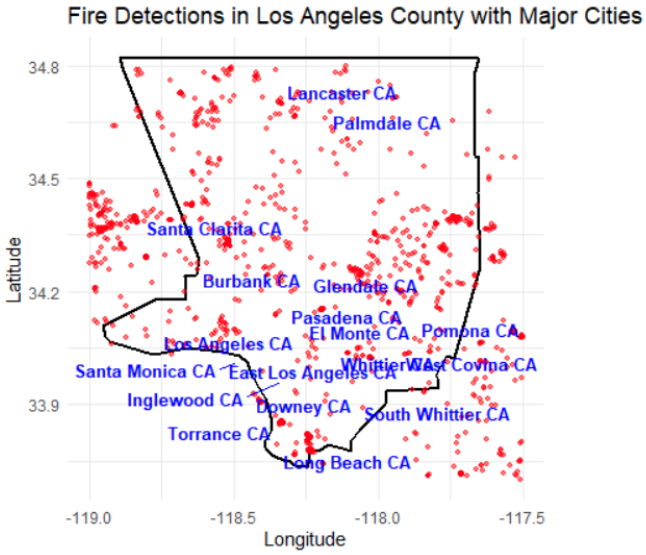

Figure 1.3: **Fire Detections with Major Cities.** This plot overlays fire detections (red points) with the major cities of Los Angeles County (blue labels) to contextualize the proximity of fire activity to population centers. The figure shows that many detections occur near the wildland–urban interface, especially in northern cities like Santa Clarita, Pasadena, and Burbank. This visualization underscores the importance of fire risk management in areas where residential communities and wildland vegetation meet.

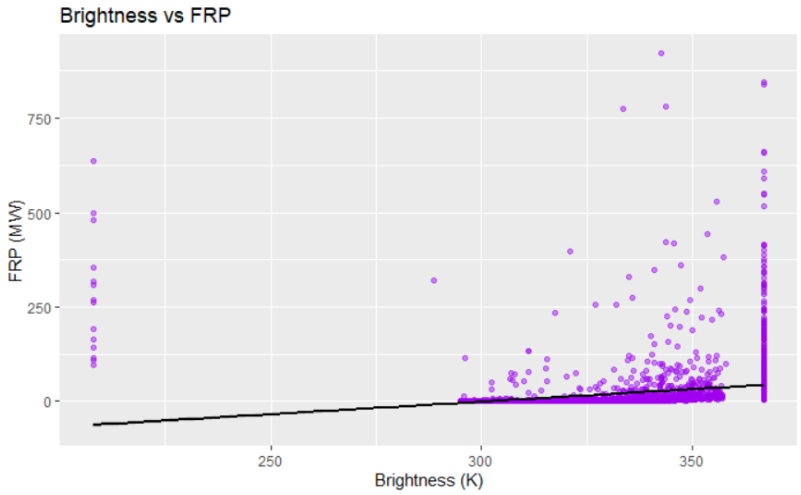

Figure 1.4: **Brightness vs. Fire Radiative Power (FRP).** This scatterplot compares the brightness temperature of detected fire pixels (in Kelvin) with their corresponding Fire Radiative Power (in MW). A generally positive relationship is observed, where higher brightness values tend to correspond to higher FRP, though there is substantial spread. The black line represents a fitted linear trend, which captures the overall upward association despite the presence of outliers and heteroscedasticity at high FRP values.

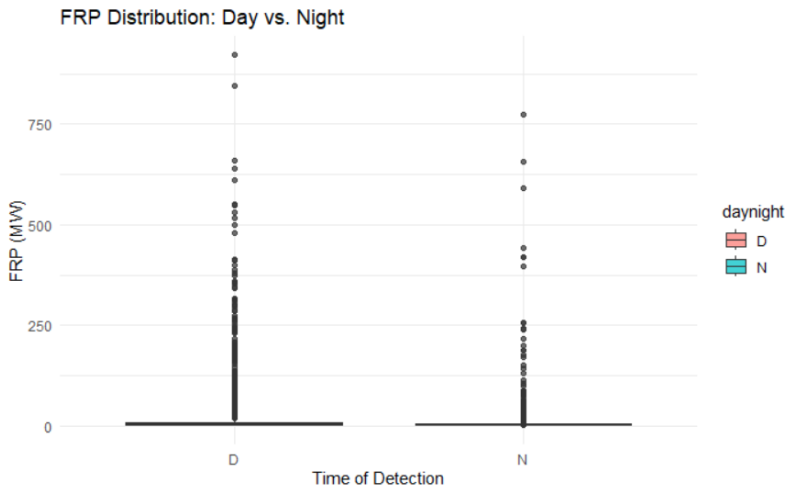

Figure 1.5: **FRP Distribution by Day vs. Night Detections.** This boxplot shows the distribution of FRP values for fire detections during the day (D) and night (N) over the study period. Although the median FRP appears similar across both groups, the spread is slightly wider for night detections, suggesting that some high-intensity events are more likely to be observed at night when background thermal noise is lower.

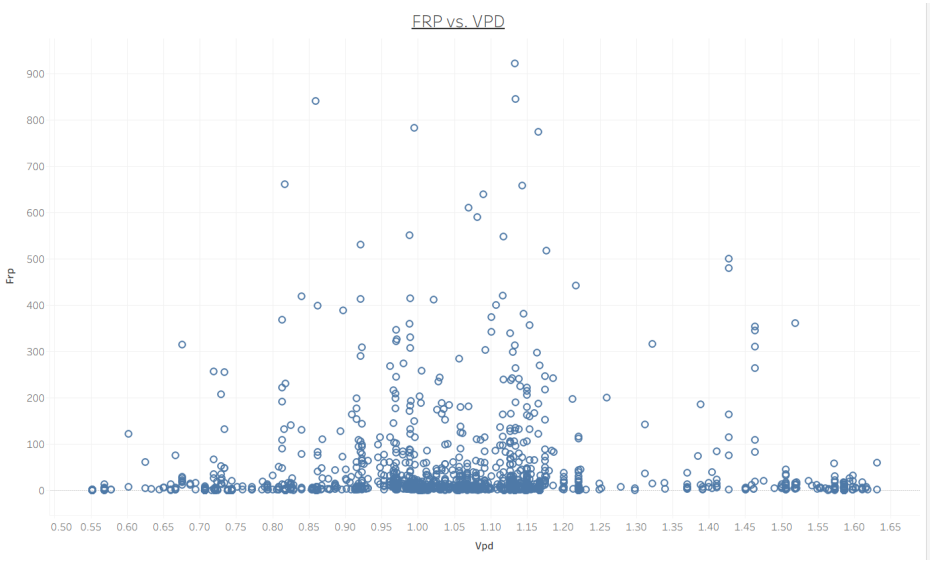

Figure 1.6: **FRP vs. Vapor Pressure Deficit (VPD).** This plot compares fire radiative power to daily vapor pressure deficit (VPD), a key indicator of atmospheric dryness. While the relationship is noisy, there is a visible clustering of higher FRP values at elevated VPD levels, indicating that more intense fires tend to occur under drier atmospheric conditions, which is consistent with existing wildfire literature.

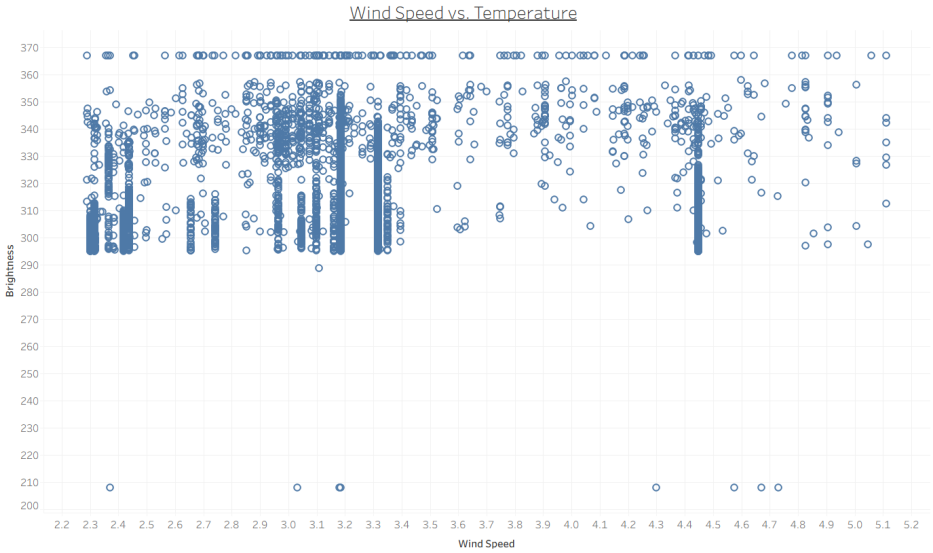

Figure 1.7: **Wind Speed vs. Brightness Temperature.** This scatterplot explores the relationship between surface wind speed and brightness temperature for fire detections. Although no strong linear relationship is apparent, the majority of points cluster around wind speeds between 2.5 and 3.5 m/s. These plots help confirm that wind speed variation is moderate across detections and may still play a role as a covariate when modeling fire behavior.

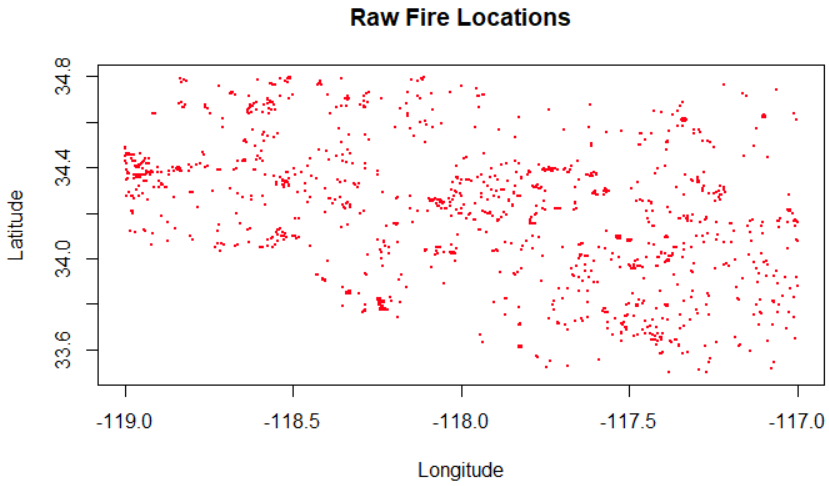

Figure 1.8: **Raw Fire Locations.** This figure plots the raw latitude–longitude coordinates of all fire detections in Los Angeles County from 2015–2024, without any spatial aggregation. The plot shows a wide spatial distribution with clear clusters along the northern and northeastern parts of the county, reflecting areas of higher fire frequency such as chaparral–dominated mountain regions.

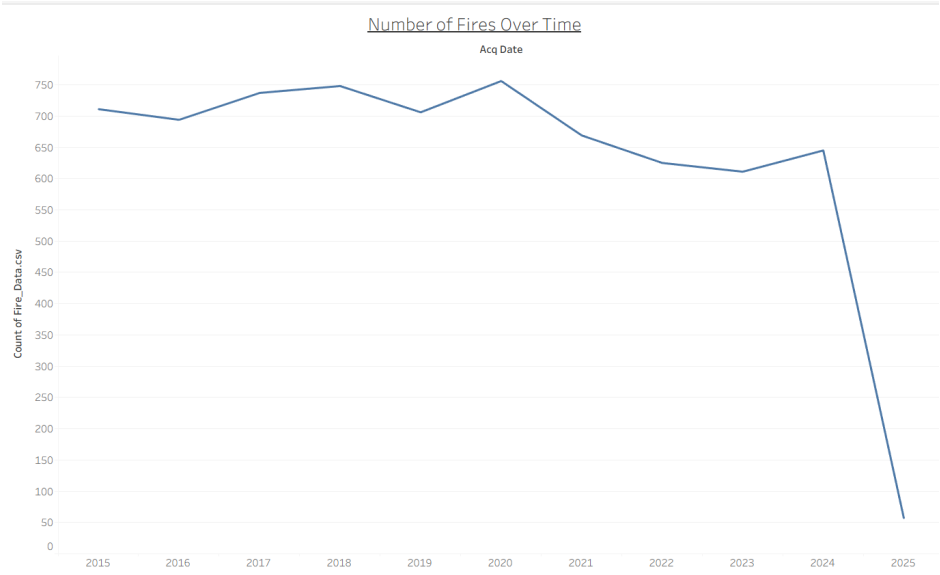

Figure 1.9: **Number of Fires Over Time.** This line chart shows the annual count of fire detections across the ten-year study window. Fire activity peaks around 2020, followed by a gradual decline in recent years, though the sharp drop in 2025 likely reflects incomplete data coverage rather than an actual reduction in fire events.

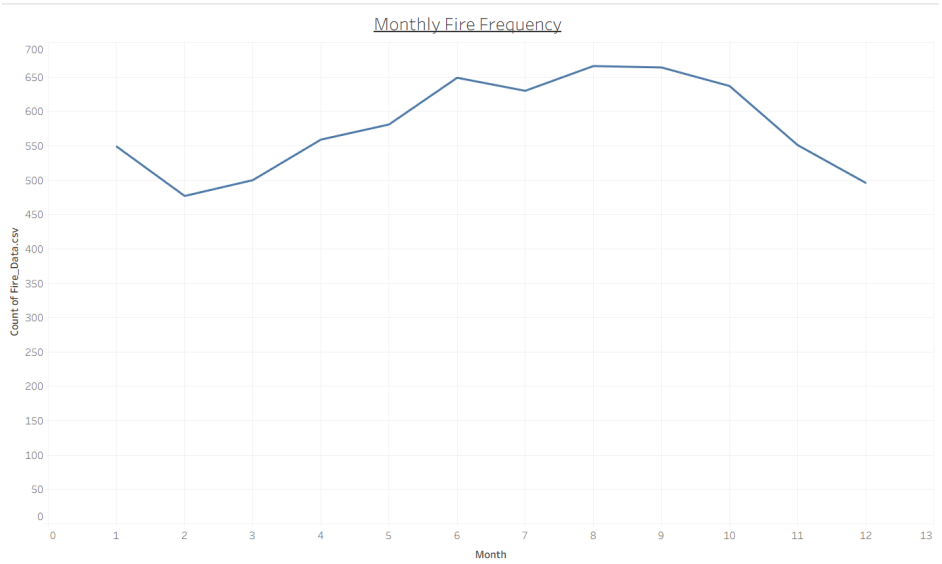

Figure 1.10: **Monthly Fire Frequency.** This figure displays the distribution of fire counts by month across all years. The highest fire frequencies occur between July and September, aligning with the peak of the dry season in Southern California, while the lowest frequencies are observed in the winter months. This confirms the strong seasonal nature of fire occurrence in the region.

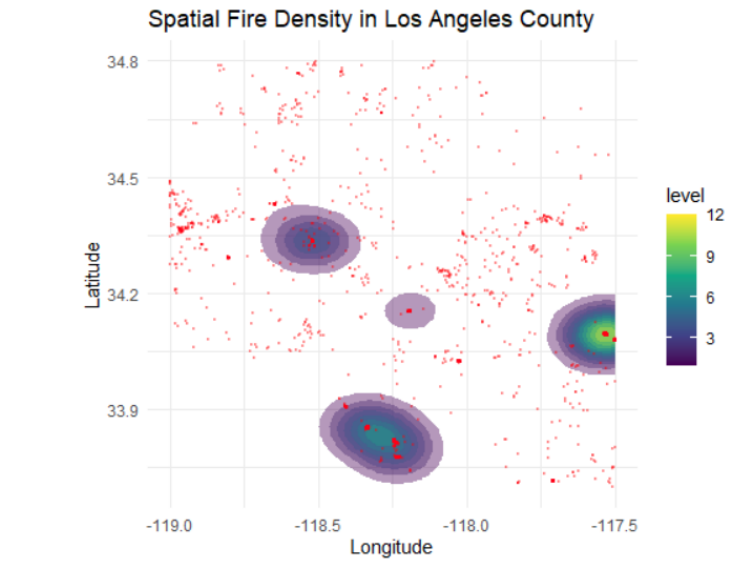

Figure 1.11: **Spatial Fire Density in Los Angeles County.** Kernel-smoothed heat map of fire detections highlighting localized hotspots across the county. Warmer colors (higher levels) indicate higher relative point density. Four prominent clusters emerge: (i) the northwestern Santa Monica Mountains, (ii) foothill areas along the San Gabriel range, (iii) interior chaparral south of the San Fernando Valley, and (iv) the northeastern edge near the Angeles National Forest. The concentration of detections along these wildland–urban interface zones aligns with fuel availability and topographic funneling of winds.

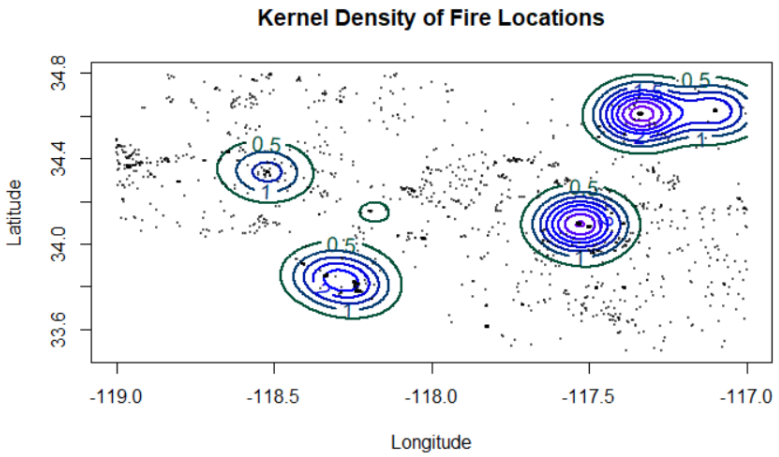

Figure 1.12: **Kernel Density of Fire Locations (Contour View).** Contour lines of the spatial kernel density estimate provide an alternative view of the same hotspots, with nested rings indicating increasing density toward local maxima. Labeled contour levels (e.g., 0.5, 1.0) help compare relative intensity across clusters and confirm the persistence of the northwestern, central, and northeastern foci observed in the heat map. The contour representation is useful for referencing specific ridgelines and boundaries in later map figures.

1.5 Data Preparation and Preprocessing

Before fitting statistical models to investigate fire radiative power (FRP) and its relationship with climatic and spatiotemporal drivers, it was necessary to carefully prepare and preprocess the dataset. Wildfire data is inherently complex because it involves spatiotemporal observations, heterogeneous variables such as climate, fire history, and brightness, and the presence of both missing and extreme values. Without proper preprocessing, such issues could bias model estimates or prevent convergence altogether. This section details the steps undertaken to clean, transform, and structure the dataset for reliable modeling.

The starting dataset (`panel_work_ready`) consisted of fire detections and associated attributes collected at the cell–day level across Los Angeles County. Each row corresponded to a fire observation within a spatial cell (10×10 km grid) on a given day, and included outcome variables such as the mean fire radiative power for that cell–day (`cell_frp_mean`) and a binary indicator for whether a fire occurred (`y`). Predictors included the number of days since the last fire in neighboring cells (`days_since_near_burn_queen`), average brightness temperature (`cell_bright_mean`), vapor pressure deficit (`cell_vpd_mean`), precipitation level (`cell_pr_mean`), and average wind speed (`cell_wind_mean`). Metadata such as cell identifiers, latitude, longitude, and date were also included. While comprehensive, this raw dataset contained several challenges, including missing values, extreme outliers caused by very intense fires, and structural imbalances across cells, as many contained few or no fire events.

Because the severity models focused on predicting FRP conditional on fire occurrence, the dataset was restricted to rows where `y = 1`. This reduced the sample size but ensured that the response variable was meaningful and strictly positive. For occurrence models, which modeled the probability of fire ignition, the full dataset including days without fires was retained. Several predictors, particularly `days_since_near_burn_queen`, contained missing values. These arose because some cells had never burned within the observation window, leaving the variable undefined. To address this, missing values were replaced with the global median across observed cells. At the same time, a binary indicator variable (`ds_missing`) was

created to flag imputed cases, which allowed the models to account for systematic differences between observed and imputed values. Missing values for climate variables such as vapor pressure deficit, precipitation, and wind were relatively rare, but where they occurred, they were replaced with cell-level averages.

Certain preprocessing steps such as log transformations required strictly positive data. To ensure stability, infinite values were replaced with missing values and subsequently dropped, FRP values of zero were excluded in severity models, and predictors with zero variance were set to zero during scaling to avoid undefined values. Extreme outliers were also a concern, as FRP and related predictors sometimes exhibited values several orders of magnitude larger than the majority of observations. These extreme values reflect the presence of rare but massive fire events. Instead of removing them, which would discard important ecological information, light winsorization was applied to reduce their undue influence. Specifically, continuous predictors including days since burn, brightness, vapor pressure deficit, precipitation, and wind were capped at the 1st and 99th percentiles. This procedure preserved the rank order of predictors while reducing the leverage of extreme observations.

To ensure comparability across predictors, all continuous variables were standardized into z-scores. Each predictor was transformed by subtracting its mean and dividing by its standard deviation. This transformation allowed regression coefficients to be interpreted in terms of the effect of a one-standard-deviation change in the predictor on the outcome. Standardization was applied to all continuous variables used in the severity and occurrence models, including days since burn, brightness, vapor pressure deficit, precipitation, and wind speed. This step improved interpretability while also stabilizing estimation.

For the occurrence models, which were based on Poisson regression, fire counts were aggregated at the cell level. A fire count variable representing the number of fire events per cell was created, along with an exposure term representing the number of days observed in each cell. This allowed the Poisson model to estimate differences in fire frequency across cells while accounting for unequal exposure times. In contrast, the severity models retained the fire-event level data, since the outcome of interest was the magnitude of FRP on days when fires occurred.

After preprocessing, three modeling-ready datasets were created. The first was the climate-only dataset at the cell level, which modeled fire counts as the outcome with vapor pressure deficit, precipitation, and wind speed as predictors, along with an exposure offset. The second was the days-only dataset at the fire-event level, which modeled FRP using only days since last burn as the predictor. The third was the severity dataset, also at the fire-event level, which modeled FRP with the full set of predictors, including days since burn, brightness, vapor pressure deficit, precipitation, and wind speed. Each dataset was checked for non-finite values, multicollinearity, and balance across predictors prior to model fitting.

The rationale for this preprocessing strategy was vital. First, it preserved interpretability by using z-scores and light winsorization rather than heavy transformations that would obscure the meaning of coefficients. Second, it balanced robustness and realism by reducing the leverage of extreme FRP values without removing them entirely. Third, it enabled stable estimation by ensuring that models converged and produced meaningful coefficients. In summary, preprocessing transformed a noisy spatiotemporal dataset into structured, model-ready inputs. The pipeline involved filtering to fire events where appropriate, imputing missing values and flagging them with indicators, standardizing continuous variables, and constructing cell-level aggregates. This careful preparation allowed subsequent models, including Poisson generalized linear models for occurrence and Gamma or log-normal regressions for severity, to focus on the underlying relationships between FRP and environmental drivers rather than artifacts of the raw data.

1.6 Modeling Framework

1.6.1 Introduction to Modeling

The modeling framework in this thesis is designed to capture the drivers of wildfire activity in Los Angeles County, focusing specifically on both the occurrence of fires and their severity once ignited. Fire behavior is shaped by a combination of antecedent conditions, climatic

drivers, and spatial heterogeneity across landscapes. Understanding these relationships requires statistical models that are both flexible and interpretable. The central outcome of interest in this study is fire radiative power (FRP), which serves as a proxy for fire severity and energy release. Because FRP is only defined when a fire occurs, the analysis must distinguish between the probability of fire occurrence and the intensity of events conditional on occurrence. For this reason, the modeling strategy is divided into two complementary components: occurrence models that estimate the likelihood of fires across space and time, and severity models that examine the distribution of FRP values during fire events.

The occurrence component addresses the question of where and when fires are most likely to occur. This is particularly important for risk assessment and prevention, as it provides insights into how climatic conditions, fuel buildup, and other factors contribute to the initiation of fires. Occurrence modeling was conducted at the cell level, where the study region was divided into a grid of 10 km by 10 km cells for the Los Angeles County. For each cell, the number of fire events detected over the study period was aggregated, along with information about exposure (the number of days during which fire activity could have been observed). This cell-based formulation allowed fire activity to be modeled as a spatial point process while controlling for variation in observation time. The statistical models for occurrence were fit using the Poisson family with a log link, an approach that directly relates the expected count of fire events in a cell to climatic predictors. In this setting, the log of the expected number of fires per cell is modeled as a linear function of predictors such as vapor pressure deficit, precipitation, and wind, with the log of exposure included as an offset.

The severity component, in contrast, focuses on variation in FRP values. Unlike occurrence, which is naturally discrete, severity requires a continuous modeling framework that can account for positive and highly skewed outcomes. In this thesis, two related approaches were used to estimate severity: Gamma regression with a log link and log-normal regression based on linear modeling of the logarithm of FRP. Each of these approaches provides a different perspective. The Gamma regression is theoretically appropriate for skewed positive data but can suffer from convergence issues. The log-normal approach is the most robust in practice, as the logarithm of FRP tends to approximate a Gaussian distribution, and it

also offers the advantage of familiar diagnostics such as residual plots and R^2 values. By comparing across these three frameworks, it is possible to assess the robustness of results and ensure that conclusions are not driven by a single modeling assumption.

1.6.2 The SG Estimation Method

A key component of the modeling framework is the use of the Stoyan–Grabarnik (SG) method for estimating fire risk across space. The SG method originates in the field of spatial point process theory, where the intensity of events is modeled as a function of location and time. In this setting, the intensity function $\lambda(s, t)$ describes the expected number of fire events per unit area and time, conditional on environmental covariates. The SG method estimates this intensity function by combining kernel smoothing with covariate-adjusted regression, thereby capturing both the local density of fire occurrences and the influence of explanatory variables.

The SG estimator is particularly well suited for wildfire data because fire events are sparse and clustered, and their likelihood depends on a complex mixture of local and regional factors. In practice, the SG method works by splitting the study region into grid cells and aggregating information within each cell. Each cell is treated as an observational unit, and covariates such as days since last burn, brightness, and climate measures are averaged or summarized within the cell. The estimator then fits a log-linear model of the form

$$\lambda(s, t; \theta) = \exp\{f(s, t; \theta)\},$$

where $f(s, t; \theta)$ is a linear combination of covariates indexed by parameters θ. The exponential transformation ensures that the intensity is always positive, while the linear predictor retains interpretability in terms of additive effects of covariates.

An important strength of the SG framework is that it accounts for spatial heterogeneity in event risk. By splitting the region into discrete cells, the model estimates an intensity surface across Los Angeles County, identifying cells with systematically higher or lower risk. Moreover, the method allows incorporation of both spatial covariates (such as climate averages within a cell) and temporal covariates (such as days since last burn). This dual capac-

ity is crucial for wildfire modeling, where both long-term fuel accumulation and short-term weather patterns play significant roles.

1.6.3 Grid Construction and Cell-level Splits

To implement the SG method, the study area was partitioned into a uniform grid of 10 km by 10 km cells, covering the entire extent of Los Angeles County. This resolution strikes a balance between capturing meaningful spatial variation and maintaining sufficient fire events per cell to support statistical estimation. Each fire detection was assigned to its corresponding grid cell based on its latitude and longitude. By aggregating fire detections in this way, it became possible to calculate cell-level averages for covariates, as well as counts of fire events. This cell-based design reduced the complexity of the spatial point process while allowing direct linkage between fire activity and environmental conditions.

Before introducing the grid, we first fit point-level models using every detected fire event as its own observation (i.e., $y = 1$ for each detection) with covariates joined at the event's latitude–longitude and date. No averaging or spatial aggregation was performed at this stage. This direct event-level specification isolates relationships at the native resolution of the detections and ensures that subsequent grid-based analyses are clearly distinguished from the initial point-process modeling.

To implement the SG method and enable spatial summaries and visualization, the study area was then partitioned into a uniform grid of 10 km by 10 km cells covering Los Angeles County. This resolution balances meaningful spatial variation with sufficient events per cell for stable estimation. Each fire detection was linked to its corresponding grid cell based on geographic coordinates. With this structure, we could compute cell-level covariate averages and fire counts, which serve different modeling goals than the event-level specification above.

This grid framework also supports derived variables. For example, "days since last burn" is computed per cell using the most recent fire event in that cell (or its neighbors), and cell-level means of VPD, precipitation, and wind are formed by aggregating daily climate records. These cell-level summaries provide stable covariates for grid-based models and

enable predicted intensities ($\hat{\lambda}$) to be mapped back onto geographic space, while remaining complementary to the initial point-level analyses.

The dataset was then split by cell identifiers (`cell_id`), which served as the spatial index for modeling. For occurrence models, each `cell_id` had an associated fire count, exposure measure, and climate averages. For severity models, each `cell_id` contained multiple fire events, each with its own FRP measurement and corresponding predictor values. By structuring the data in this way, it was possible to model both cross-sectional variation across cells and within-cell variation across fire events. Furthermore, the use of a grid allowed for direct visualization of model outputs, as predicted intensities ($\hat{\lambda}$ values) could be mapped back onto the geographic space of Los Angeles County.

1.6.4 Integration of SG Estimation with Regression Models

The integration of SG estimation with regression modeling created a powerful framework for understanding wildfire dynamics. The SG method provided a means to estimate the baseline intensity of fires across the grid, capturing spatial variation independent of covariates. Regression modeling then explained deviations in FRP severity using predictors such as days since burn, brightness, and climate. Together, these methods allowed both a descriptive and explanatory understanding of wildfire behavior. The descriptive aspect highlighted regions and times of systematically higher risk, while the explanatory aspect quantified the contribution of specific environmental factors.

In summary, the modeling introduction combines the SG method for spatial intensity estimation with regression models for fire severity. By partitioning Los Angeles County into grid cells and structuring the data by `cell_id`, the models were able to capture both spatial patterns and predictor-driven effects. This framework provides the foundation for the results presented in subsequent sections, where the performance of different model specifications is evaluated and their implications for understanding fire risk are discussed.