Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Regional utility and non-utility wildfire ignition models across space and time

Gabriela Calana Somoza

gcalana@stanford.edu

Stanford University

Neetesh Sharma

Florida State University

Jack W. Baker

Stanford University

Research Article

Keywords: power utility, wildfire ignitions, risk, regional analysis, network, inhomogeneous Poisson point process, linear network

Posted Date: October 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10978797/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Regional utility and non-utility wildfire ignition models across space and time

Gabriela Calana Somoza[1*], Neetesh Sharma[2] and Jack W. Baker[1]

[1*]Civil and Environmental Engineering, Stanford University, Stanford, 94305, California, USA.

[2]Civil and Environmental Engineering, Florida State University, Tallahassee, 32306, Florida, USA.

*Corresponding author(s). E-mail(s): gcalana@stanford.edu;
Contributing authors: nsharma@eng.famu.fsu.edu;
bakerjw@stanford.edu;

Abstract

In California, electrical infrastructure has been implicated in about 40% of the state's most destructive wildfires. Existing models treat ignition sources in aggregate or focus on system-level failures, limiting their applicability to regional risk assessment, where utility ignitions in particular are a major driver of destructive events. This manuscript presents an integrated wildfire ignition modeling framework that treats utility ignitions and non-utility ignitions (other human-related and naturally caused ignitions) as parallel components with distinct spatial structures, coupled with temporal components that account for weather conditions. Utility ignitions are modeled as an inhomogeneous Poisson point process on a linear network representing the electrical infrastructure, with intensity depending on infrastructure and vegetation covariates; non-utility ignitions are an inhomogeneous Poisson point process across the region, with intensity depending on land cover and proximity to roads. Separability testing helps justify the assumption of space and time independence. The results show that voltage and canopy stand height are significant covariates for utility ignitions, whereas fuel type and road proximity are significant for non-utility ignitions, with a road proximity of 60 meters or less associated with a 3.28-fold increase in ignition rate. The framework supports calibration to empirical ignition records and stochastic simulation of ignitions. By disaggregating ignition sources, stakeholders can better understand wildfire risk and prepare targeted mitigation actions.

Keywords: power utility, wildfire ignitions, risk, regional analysis, network, inhomogeneous Poisson point process, linear network

1 Introduction

Wildfires represent a growing threat to communities in the wildland-urban interface (WUI), driven by climate change, expanding development, and inadequate vegetation and infrastructure management. These compounding factors have significantly increased the vulnerability of both natural ecosystems and human communities.

Wildfire ignitions can arise from various sources, generally divided into two broad categories: human-caused and naturally caused. Human-caused ignitions include wildfires caused by direct or indirect anthropogenic sources, such as electrical power infrastructure, vehicles, agricultural burning, fireworks, cigarettes, and arson, among others (Ganteaume and Syphard 2018). Naturally caused ignitions are those with no human involvement; their main sources include lightning, volcanism, gas emissions, and meteor strikes (Ganteaume and Syphard 2018). Although both ignition categories contribute to wildfire occurrence, the spatial patterns, environmental drivers, and resulting impacts can differ between them.

In California, northern counties such as Sonoma, Lake, and Napa have experienced large wildfires after 2000 (Li and Banerjee 2021). Ignition patterns are changing throughout the state, with an increase in small human-caused fires occurring earlier in the year and extending the duration of the fire season (Li and Banerjee 2021). Furthermore, utility-related ignitions have become an important driver of destructive wildfires. A study by Mitchell (2013) found that power-line-caused fires are, on average, approximately 10 times larger than other fires. In California, 40% of the 20 most destructive wildfires in the state's history were caused by electrical infrastructure failures (CAL FIRE 2023). This includes the Camp Fire, which remains the deadliest wildfire in California's history. Similar concerns have emerged outside California, where power utility-related ignitions have historically contributed to catastrophic wildfire events in Chile, Australia, Mediterranean Europe, and Canada. During the 2023 Maui fires, bare electrical wires and failed utility poles were identified as likely ignition sources (PBS NewsHour 2023), while investigations into the Eaton Fire in Altadena have also identified potential sparking from distribution lines (NPR 2025). Wildfires caused by power utility infrastructure are often particularly destructive because they occur during extreme wind events, which contribute to rapid fire spread (Vahedi et al. 2025). In addition, studies have reported a notable increase in powerline failures associated with wildfire ignitions over the past 70 years, driven in part by inadequate maintenance, aging infrastructure, and continued grid expansion (Vahedi et al. 2025). For example, Pacific Gas and Electric Company reported 468 ignitions associated with its transmission and distribution systems in 2022 alone (California Public Utilities Commission). Existing power-utility ignition models such as Muhs et al. (2020) and Bayani et al. (2022) quantify utility ignition probability but only at the system level, considering wind speed and vegetation.

Furthermore, terrain, fuel conditions, and weather influence not only whether an ignition occurs but also whether it develops into a large and destructive wildfire. Previous studies have shown that fuel-dominated events are more strongly associated with naturally caused ignitions, whereas wind-dominated events are more strongly associated with human-caused ignitions (Keeley and Syphard 2019). Wind-dominated fires, in particular, have been linked to power lines and can result in more destructive

wildfires than those ignited by lightning (Keeley and Syphard 2019). These findings suggest that utility-related ignitions require specific attention because they are shaped by interactions among infrastructure exposure, local environmental conditions, and extreme weather.

Existing wildfire ignition models provide useful approaches for estimating ignition likelihood, but they often do not distinguish between ignition sources in ways that support regional comparisons or infrastructure-specific risk assessments. For instance, Costafreda-Aumedes et al. (2017) reviews human-caused ignition models designed to predict wildfire ignitions over both short-term (daily) and long-term (annually or longer). These models generally produce two types of outputs: binary ignition occurrence or the number of fires. Common methods for binary occurrence include logistic regression and classification and regression trees (CART), whereas common methods for predicting the number of fires include Poisson regression and multiple linear regression. Similarly, Arndt et al. (2013) uses logistic regression to identify significant predictors of human-caused fire occurrence and to model ignition probability. These approaches are useful for identifying broad drivers of human-caused ignitions, but they generally treat human-caused ignitions as a single category rather than distinguishing among specific anthropogenic sources, such as power utility infrastructure, and how these sources affect regional ignition patterns.

Models of naturally caused ignitions, by contrast, often focus on lightning as the dominant natural ignition source. For example, Read et al. (2018) developed a lightning-caused wildfire ignition model that produces daily, spatially explicit forecasts using logistic regression and covariates such as maximum temperature, relative humidity, precipitation, elevation, and vegetation type. The study identifies maximum temperature, 3 p.m. relative humidity, precipitation, elevation, grassland, and other variables as statistically significant and influential predictors of lightning-caused ignition. Other variables, including 850 hPa dewpoint depression, drought factor, and wind speed, are statistically significant but less influential for this ignition type. The model output is a regional probability map of lightning ignition for a specific date at a spatial resolution of 20 km (Read et al. 2018). Other studies have modeled naturally caused wildfires using generalized linear models (GLMs) and generalized additive models (GAMs), including the Montana case study presented in Jiménez-Ruano et al. (2022). Although these studies provide important insights into lightning and naturally caused ignitions and improve the predictability of such events, they generally do not disaggregate ignition sources beyond the broad categories of human-caused and naturally caused ignitions. This limits their ability to identify which specific ignition sources contribute most to wildfire risk.

In general, spatio-temporal point process models have also been used to estimate wildfire ignition intensity across space and time. For example, Poisson point process models can predict ignition intensity using spatio-temporal formulations together with weather and terrain characteristics (D'Angelo et al. 2025). However, this study paid less attention to specific ignition mechanisms, including those attributed to power utility infrastructure along power line networks. In general, this article assumes that ignition events arise from an intensity function defined over a two-dimensional spatial domain. This assumption is appropriate for ignition processes that can occur broadly

across the landscape, but it is less suitable for infrastructure-related ignitions, whose spatial support is restricted to linear networks. For power utility ignitions in particular, the relevant point process should be defined along the power line network, with ignition intensity varying as a function of line-segment attributes, surrounding terrain, vegetation, and weather conditions. Furthermore, studies have used Gaussian Cox Processes to model wildfire occurrences considering terrain and climate covariates (Gabriel et al. 2017), (Koh et al. 2023). However, Gabriel et al. (2017) provides a space-yearly resolution of wildfire occurrences, and Koh et al. (2023) considers only the percentage of forested area as a spatial covariate, missing the type of land cover that can affect the likelihood of wildfire occurrences. Other studies have applied spatial point process approaches to wildfire ignitions in general and to lightning ignitions specifically, but these studies primarily provide kernel-based estimates of ignition intensity and do not incorporate weather or terrain covariates that may influence ignition likelihood (Podur et al. 2003) (Turner 2009). Thus, although existing spatial ignition models can characterize where ignitions are likely to occur, they provide limited insight into how ignition risk varies by source type, especially for infrastructure-related ignitions distributed along linear network systems.

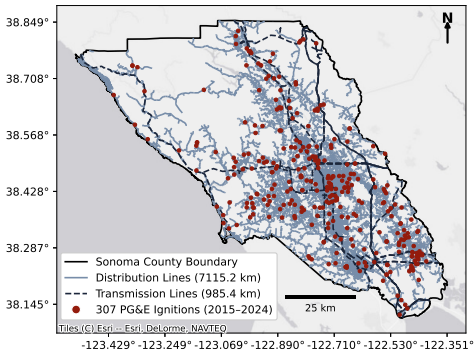
Therefore, to address the gaps identified above, this article proposes a decoupled spatial and temporal wildfire ignition model that can be applied across different ignition mechanisms and used to identify which ignition sources pose the highest risk. Wildfire ignitions are divided into two categories: utility and non-utility ignitions. Non-utility ignitions encompass all types of ignitions not attributed to power-utilities, which allows the risk from utility ignitions to be disaggregated. The spatial modeling component uses a Poisson point process model with covariates describing vegetation, infrastructure attributes, and proximity to major roads to estimate ignition likelihood. For power utility-related ignitions, the model represents ignition occurrence along linear infrastructure networks, whereas for non-utility ignition sources, it represents ignition occurrence across the broader landscape. The temporal modeling component uses a Negative Binomial model that estimates the daily ignition rate as a function of maximum temperature and wind speed. To justify the separability assumption for space and time, the article includes a likelihood ratio test. Given weather forecasts or climate projections, the proposed framework can evaluate the impacts of changing temperature and wind patterns on wildfire ignition rates. Together, these components enable the simulation of wildfire ignitions in space and time.

The rest of this paper continues as follows: Section 2 provides the methodology; Section 3 describes the case study used to calibrate the ignition models; Section 4 presents the results of the spatial and temporal analysis, including example simulations; Section 5 provides the discussion and conclusions.

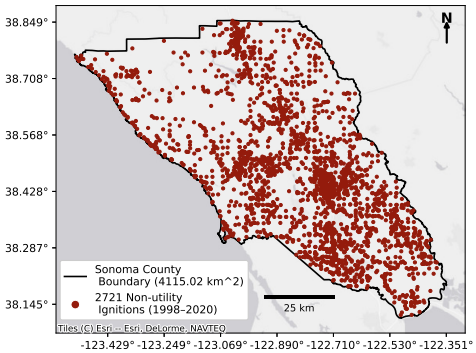
2 Methods

The methodology for predicting spatial ignitions across a region is based on inhomogeneous Poisson point process models. Depending on the wildfire ignition type, these can occur across an area or along a network system like roads or power-utility infrastructure lines. The following parametric models estimate the expected rate of ignitions

[λ] across a network or area as a function of spatial parameters. To model wildfire ignitions over time, a negative binomial regression estimates the daily ignition rate using historical weather and ignition data.



(a) A map of historical power-utility ignitions is shown along the power-utility network.



(b) A map of historical non-utility wildfire ignitions is shown across the region: ignitions attributed to sources other than power utilities.

Fig. 1: Maps of historical wildfire ignitions are shown in Sonoma County.

2.1 Ignition prediction in a linear network

Wildfire ignitions that occur along linear networks, such as power utility systems, can be modeled as an inhomogeneous Poisson point process on such networks. The model form is

$$\lambda_{l,\theta}(u) = \exp\left(\theta_{l,0} + \boldsymbol{\theta}_l^{\mathsf{T}} \mathbf{Z}_l(u)\right) \tag{1}$$

where $\lambda_{l,\theta}(u)$ estimates the expected rate of ignitions per unit length as a function of the spatial characteristics at location u, $\theta_{l,0}$ is the intercept, $\boldsymbol{\theta}_l$ is the parameter vector of the spatial covariates, and $\mathbf{Z}_l(u)$ is the known spatial covariate vector at location u. The subscript l indicates that the quantity of interest corresponds to the linear network, and bold symbols indicate vector quantities. The subscript θ indicates that the quantity is a parametric estimate.

To develop this model, first, a linear network must be generated to represent the infrastructure of interest. This linear network G is the union of N line segments l_i such that $G = \bigcup_{i=1}^{N} l_i$ and $N < \infty$. The line segments l_i consist of $[u_i, v_i]$ where u_i and v_i are the endpoints of each line segment l_i (Baddeley et al. 2021).

After the linear network is established, the point pattern $\mathbf{x}$ along the network G can be generated. The point pattern is defined as $\mathbf{x} = \{x_1, \ldots, x_n\}$, where each x_i represents the location of a wildfire ignition on the network G. Modeling $\mathbf{x}$ as

an inhomogeneous Poisson point process imposes two properties. First, for any sub-network $B \subseteq G$, the number of ignitions $N(B)$ follows a Poisson distribution with mean $\int_B \lambda_{l,\theta}(u)\, du$, the integral of the intensity in equation 1 over B. Second, the numbers of ignitions in disjoint sub-networks are independent (Baddeley et al. 2021).

The goal is to have a parametric model that predicts wildfire ignitions as a function of spatial covariates, therefore, the log-linear function shown in equation 1 shows how the intensity $\lambda_{l,\theta}(u)$ depends on the spatial function $\mathbf{Z}_l(u)$ and its corresponding parameter vector $\boldsymbol{\theta}_l$, while $\theta_{l,0}$ is the intercept. The covariate values $\mathbf{Z}_l(u)$ must be available along the network G as a callable function or a discretized image. Depending on the spatial covariate of interest, the most appropriate type of object should be selected to represent its value as closely as possible along the network. For example, covariates such as voltage along the network are better represented by a function object that maps each location along the network. On the other hand, covariates such as canopy stand height, which are often recorded in raster file format, are better represented as pixel image objects. The resolution of the covariates used is essential to characterize the variables of interest along G.

This inhomogeneous Poisson point process with parametrically specified intensity can be fitted to a point pattern by maximizing the log-likelihood, $L_l(\boldsymbol{\theta})$ shown in equation 2 using the method of Berman and Turner (1992). The fitted inhomogeneous Poisson point process must include dummy points that define locations of no ignitions. The spacing of the dummy points must be specified to estimate the likelihood. Berman and Turner's method approximates the integral in equation 2 by a finite sum using any quadrature rule, where $u_j, j = 1, ...m$ are data points and dummy points in G and $w_j > 0$ are quadrature length weights adding to the total line length, $|G|$, as shown in equation 3 (Baddeley et al. 2015).

$$\log L_l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \lambda_{l,\theta}(x_i) - \int_G \lambda_{l,\theta}(u)\, du \tag{2}$$

$$\int_G \lambda_{l,\theta}(u)\, du \approx \sum_{j=1}^{m} \lambda_{l,\theta}(u_j) w_j \tag{3}$$

Since the list of points $u_j, j = 1, .., m$ includes all data points $x_i, i = 1, ...n$, then the log-likelihood equation yield to the sum over quadrature points where $\lambda_{l,j} = \lambda_{l,\theta}(u_j)$ and I_j is 1 if u_j is a data point and 0 if u_j is a dummy point, resulting in Berman and Turner likelihood estimation shown in equation 4, where $y_j = I_j / w_j$, allowing it to be maximized using standard software for fitting a generalized linear model (Baddeley et al. 2015).

$$\log L_l(\boldsymbol{\theta}) \approx \sum_{j=1}^{m} (y_j \log \lambda_{l,j} - \lambda_{l,j} w_j) \tag{4}$$

This fitted model can be used to quantify the expected number of ignitions at any location along the network G when the relevant covariates are present, and to simulate realizations of potential wildfire ignitions.

2.2 Ignition prediction across space

Wildfire ignitions across a region of interest can be modeled using an inhomogeneous Poisson point process in two-dimensional space. The form is

$$\lambda_{\omega,\theta}(s) = \exp\left(\theta_{\omega,0} + \boldsymbol{\theta}_{\omega}^{\mathsf{T}}\mathbf{Z}_{\omega}(s)\right) \tag{5}$$

where $\lambda_{\omega,\theta}(s)$ estimates the expected number of ignitions per unit area as a function of the spatial covariate vector $\mathbf{Z}_{\omega}$ measured at s. In this model, $\theta_{\omega,0}$ represents the intercept and $\boldsymbol{\theta}_{\omega}$ is the parameter vector of the spatial covariates. For this model, it is important to define the region of interest as $\Omega \subset \mathbb{R}^2$, the point pattern $\mathbf{x}$, where each x_i represents the location of a wildfire ignition in space, and the spacing of the dummy points representing no ignition locations. In this model, covariates $\mathbf{Z}_{\omega}(s)$ must be callable at every location, s in region Ω, and represented as a dataframe, images, a function, windows, tessellations, or single numbers (Baddeley et al. 2015). This model can also be fitted by maximizing the log-likelihood following Berman and Turner's method, but integrating over the region Ω rather than the network G. The subscript ω indicates that the quantity of interest corresponds to the region of interest Ω.

2.3 Temporal ignition prediction model

The temporal ignition prediction model estimates the number of ignitions, y_i, occurring in the time interval t_i, given the weather covariate vector $\mathbf{W}_{\tau}(t)$. To obtain y_i, we partition the time period covering the set of ignitions $\{x_i\}$ into n_t time intervals (e.g., days), where t_i indicates one such interval, so that $|\mathbf{x}| = \sum_{i=1}^{n_t} y_i$. We model y_i using a Negative Binomial parameterized by the mean rate of occurrences per time interval, $\lambda_{\tau}(t)$, and the dispersion parameter, α, such that the variance of the number of ignitions per time interval is $\lambda_{\tau,\theta}(t) + \alpha(\lambda_{\tau,\theta}(t))^2$. The proposed methodology selects a Negative Binomial model because the variability in daily ignitions is expected to exceed the average number of ignitions per day. The weather covariates considered are drawn from historical weather datasets for the time period corresponding to the historical ignition data. Matching the weather covariates of interest to the day on which each historical ignition occurred, while also recording these covariates for days with no ignition, yields a complete dataset.

The expected rate of ignitions per day, $\lambda_{\tau,\theta}(t)$, uses the following form similar to equations 1 and 5:

$$\lambda_{\tau,\theta}(t) = \exp\left(\theta_{\tau,0} + \boldsymbol{\theta}_{\tau}^{\mathsf{T}}\mathbf{W}_{\tau}(t)\right) \tag{6}$$

where $\lambda_{\tau,\theta}(t)$ estimates the expected rate of ignitions per day, as a function of the weather covariate vector $\mathbf{W}_{\tau}(t)$ at time t, the corresponding parameter vector $\boldsymbol{\theta}_{\tau}$, and the intercept $\theta_{\tau,0}$.

The parameters α, $\theta_{\tau,0}$ and $\boldsymbol{\theta}_{\tau}$ are estimated by means of Maximum Likelihood estimation as per equation 7, where $f_Y(y_i)$ is the Negative Binomial probability mass function.

$$
\begin{aligned}
L(\alpha, \theta_{\tau,0}, \boldsymbol{\theta}_\tau) &= \prod_{i=1}^{n_t} f_Y(y_i) \\
&= \prod_{i=1}^{n_t} \frac{\Gamma\left(y_i + \alpha^{-1}\right)}{\Gamma\left(\alpha^{-1}\right)\Gamma\left(y_i + 1\right)} \left(\frac{1}{1 + \alpha\lambda_{\tau,\theta}(t_i)}\right)^{1/\alpha} \left(\frac{\alpha\lambda_{\tau,\theta}(t_i)}{1 + \alpha\lambda_{\tau,\theta}(t_i)}\right)^{y_i}
\end{aligned} \tag{7}
$$

where, $\Gamma(\cdot)$ is the Gamma function.

Since the temporal model is decoupled from the spatial model, this can be applied to any ignition source considered, where $\lambda_{l,\tau,\theta}$ is used for wildfire ignitions modeled along a linear network and $\lambda_{\omega,\tau,\theta}$ for wildfire ignitions modeled across space.

Furthermore, this fitted model can be used to simulate ignition days, including the expected number of ignitions and weather covariates. This can be done using the latest available daily weather dataset or a future daily weather forecast, with the necessary weather covariates to predict the expected number of ignitions for each day in the new dataset. Lastly, once the wildfire ignition days are simulated with the corresponding number of ignitions, these can be paired with the spatially simulated ignitions to complete the wildfire ignition model in space and time.

2.4 Coupling space and time models

After the space and time models are fitted using the methodology described, using the independence assumption, the resultant rate of ignition in space per day takes the form of equation 8,

$$
\lambda(s,t) = \frac{\lambda_{\omega,\theta}(s)}{\sum_{s\in\Omega} \lambda_{\omega,\theta}(s)} \times \lambda_{\tau,\theta}(t) \tag{8}
$$

where the spatial density is normalized over the region or network and, following the independence assumption, is multiplied by the expected number of daily ignitions. In this equation, $\lambda_{\omega,\theta}(s)$ represents the spatial intensity of wildfire ignition as a function of the corresponding spatial covariates, and $\lambda_{\tau,\theta}(t)$ is the temporal intensity as a function of weather covariates at time t. Using this equation, the expected number of ignitions per day at any location in the region can be computed using the appropriate spatial and weather covariates. Following this, we randomly sample ignitions in space and assign these to the sampled ignitions per day.

2.5 Spatial/temporal separability assessment

To evaluate the suitability of a decoupled model, we compare nested Poisson point-process models using likelihood-ratio tests. The ignition data for each ignition source is partitioned temporally in two ways: first, into two half-year periods, and second, into four calendar quarters. Then, the point patterns for each temporal period are jointly fitted as replicated point patterns.

For each ignition source, we create a separable model with a period-specific intercept and the corresponding spatial covariate effects, which are constrained to be

constant across periods. Hence, the model allows the overall ignition intensity to vary among temporal periods while assuming that the relative effects of the spatial covariates remain unchanged. In this case, $p = 1, \ldots, P$ represents the temporal period, where $P = 2$ for the half-year analysis and $P = 4$ for the quarterly analysis. The period-specific intercept is represented by ϵ_p, with the first period serving as the reference category. Under this parameterization, coefficients for subsequent periods represent differences in baseline log-intensity relative to the first period, holding the spatial covariates constant. Equation 9 represents the general separable model for wildfire ignitions.

$$\lambda_{sep,\theta}(s,p) = \exp\left(\epsilon_p + \sum_{i=1}^{m} \theta_i Z_{\omega,i}(s)\right) \tag{9}$$

On the other hand, the alternative model allows the effects of the spatial covariates to vary among temporal periods by including period $\times$ covariate interactions. Equation 10 represents the alternative model for wildfire ignitions. The interaction parameter δ is zero for the reference period; for subsequent periods, it represents changes in the corresponding spatial covariate effect relative to the reference period.

$$\lambda_{alt,\theta}(s,p) = \exp\left(\epsilon_p + \sum_{i=1}^{m} (\theta_{\omega,i} + \delta_{\omega,p,i}) Z_{\omega,i}(s)\right) \tag{10}$$

In this case, the null hypothesis (H_0) is that all period $\times$ covariate interaction coefficients are zero, implying that spatial covariate effects are invariant across temporal periods. The alternative hypothesis (H_A) is that at least one spatial covariate effect differs among periods. Therefore, the likelihood-ratio test is performed by comparing the separable and alternative models. For the Poisson point-process models, the likelihood-ratio statistic is calculated as

$$D = 2\left[\ell(\mathcal{M}_{\text{alt}}) - \ell(\mathcal{M}_{\text{sep}})\right],$$

and compared with a chi-squared distribution with degrees of freedom equal to the difference in the number of estimated parameters between the two models. The separable and alternative models equations are the same for wildfire ignitions along a linear network and across space.

3 Case study

This section describes the steps for calibrating the proposed model using Sonoma County as an example. Calibrating this wildfire ignition model requires five types of data. First, regional boundaries must be defined through either polygon shapefiles or coordinates that establish the spatial window for the spatial Poisson point process model. Second, linear network data from utility companies or public datasets (e.g., roads, power lines) are needed as shapefiles with line geometry to generate a linear network. Third, historical ignition records from nationally available datasets or utility-reported incidents with geospatial coordinates and occurrence dates are important for enabling both spatial and temporal modeling. Fourth, spatial covariates describing

the road proximity, fuel type, canopy height, and canopy coverage at the best possible resolution from remote sensing sources are essential for spatially characterizing the prediction model. Lastly, meteorological data from a local weather station spanning the same temporal period as the wildfire ignition dataset are required to develop the decoupled temporal model.

3.1 Region data

Sonoma County was used as the study region to calibrate this model, as it is vulnerable to wildfire hazards and has experienced multiple catastrophic wildfires. The region boundary was obtained from the Sonoma County GIS website (County of Sonoma GIS Central 2025).

3.2 Infrastructure data

Sonoma County has experienced multiple wildfire events attributed to power-utility infrastructure; therefore, the linear network selected for the study is PG&E's electrical infrastructure system, the operating utility company in Sonoma County. The infrastructure data was obtained from the Integrated Capacity Analysis and Distributed Resource Planning data for PG&E (Pacific Gas and Electric Company 2026a).

3.3 Ignition data

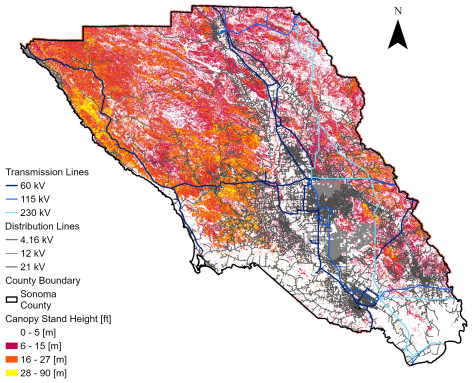
A total of 309 power-utility wildfire ignitions were reported by PG&E in Sonoma County from 2015 to 2024; however, only 307 were used as the historical wildfire points across the network because two of the ignitions occurred at the same location. These wildfire ignitions were reported by PG&E to the California Public Utilities Commission and obtained from their website (Pacific Gas and Electric Company 2026b).

A total of 2721 non-utility wildfire ignitions in Sonoma County were obtained from the sixth version of the Fire Program Analysis Fire-Occurrence Database from the US Department of Agriculture and processed through the Interactive Visualization Portal for FPA-FOD Attributes Dataset (Short 2022), (Boise State University). The positional precision of these records varies by reporting agency, and the database standard requires only that a fire be located to a nominal area of one square mile. This dataset excludes any wildfire ignitions attributed to power-utility infrastructure. These historical ignition datasets are shown in Figure 1.

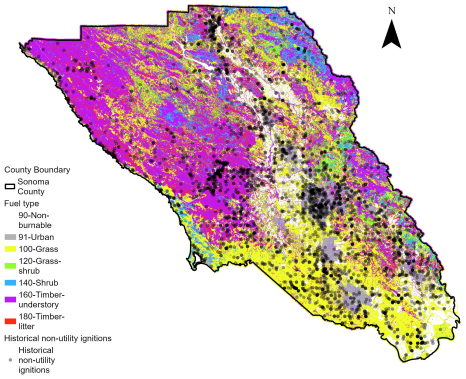
3.4 Spatial covariates

Wildfire ignitions result from triggers influenced by the spatial and environmental components. For example, when power lines are close to trees and wind speeds are high, the probability of a wildfire igniting increases as the interaction between the power lines and the flammable vegetation around them becomes inevitable. The following spatial covariates were tested to quantify the predictive power of spatial parameters in wildfire ignitions: power-line voltage, fuel type, canopy stand height, canopy base height, canopy bulk density, canopy cover, and distance to the closest road. The power-line voltage was obtained from PG&E's infrastructure data, the terrain covariates

were included in the Sonoma County landscape file available on the Sonoma County GIS website (Sonoma County Vegetation Mapping and LiDAR Program 2019), and the road distance was calculated using the road Tiger Files by Census (U.S. Census Bureau, Geography Division 2020) and the Sonoma GIS website to retain federally eligible network and significant rural network (County of Sonoma n.d.) and filter major roads. The power-line voltage values correspond to each power line, the spatial covariates provided by the Sonoma County GIS website are in raster files with a pixel resolution of 5 meters, and the road distance consists of a raster file with a pixel resolution of 10 meters that depicts the distance to the closest major road.



(a) The spatial distribution of power lines and canopy stand height is shown in Sonoma County.



(b) The spatial distribution of fuel type and non-utility ignitions is shown in Sonoma County.

Fig. 2: The spatial distributions of canopy stand height, power lines, fuel types, and non-utility ignitions are shown in Sonoma County.

The spatial covariates describing terrain characteristics were compared to assess correlations among these variables. Since the fuel type was a categorical covariate, the correlation ratio η^2 was used to capture how the fuel types explain the other covariates. On the other hand, the correlation coefficient was used to understand the relationship among the other numerical covariates.

Figure 3 shows how the spatial covariates relate; values from 5,000,000 random pixels are plotted, rather than the larger comprehensive raster file. Covariates such as canopy stand height and canopy base height are highly correlated as they provide similar metrics at the same location. Canopy stand height and canopy cover are highly correlated as well, indicating that these variables should not be used together. Figure 4a and 4b show the spatial distribution of wildfire ignitions and the major roads considered to compute the road proximity. Here, we observe that the road network system follows the power-line network system closely; therefore, it will not be considered in the utility ignition model, whereas for non-utility ignitions, proximity to the roads can affect the chances of an ignition. Lastly, Figure 5 shows how the fuel type total area

relates to the number of non-utility ignitions per fuel type to better understand how the fitted intensity should behave with respect to this categorical covariate.

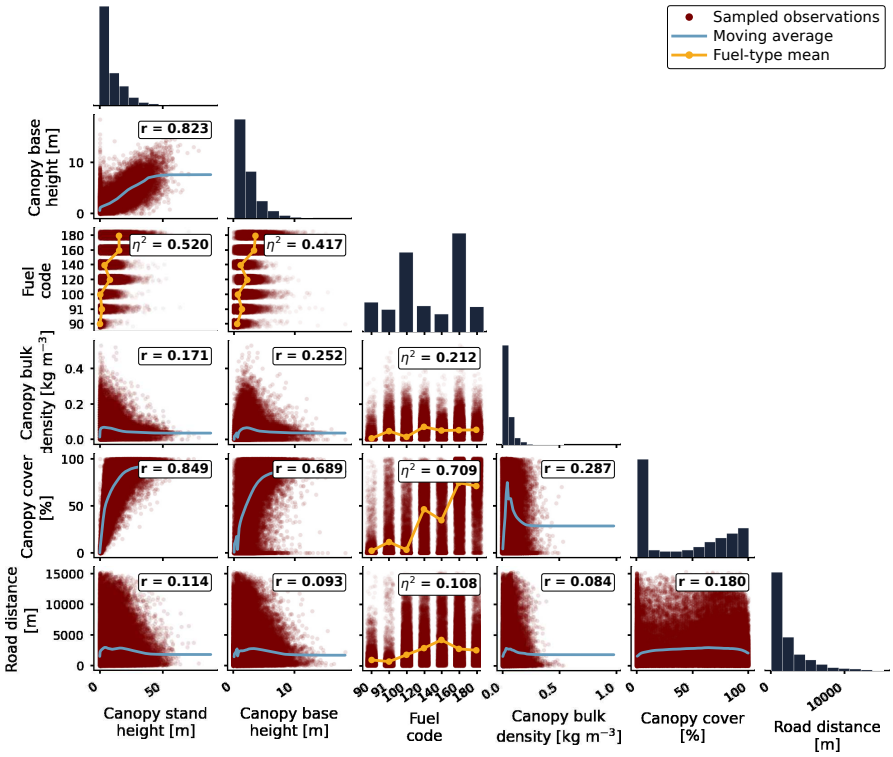
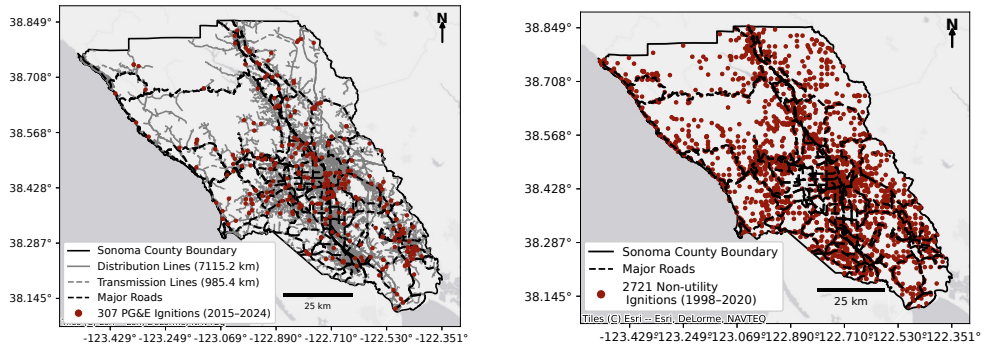


Fig. 3: The scatter plot matrix shows the spatial covariates that were tested.

3.5 Weather covariates

The NOAA (National Oceanic and Atmospheric Administration) station at the Santa Rosa Airport served as the weather condition data source to calibrate the prediction of ignition days. The weather covariates obtained from the NOAA Climate Data website (Menne et al. 2012) were the daily maximum temperature and the fastest 5-second wind speed, similar to other studies (D'Angelo et al. 2025). This single station was used to characterize the weather covariates on the day of ignition.



(a) A map of power-utility ignitions and power-line infrastructures with major roads is shown in Sonoma County.

(b) A map of non-utility ignitions with major roads is shown in Sonoma County.

Fig. 4: Major roads and wildfire ignitions are shown in Sonoma County.

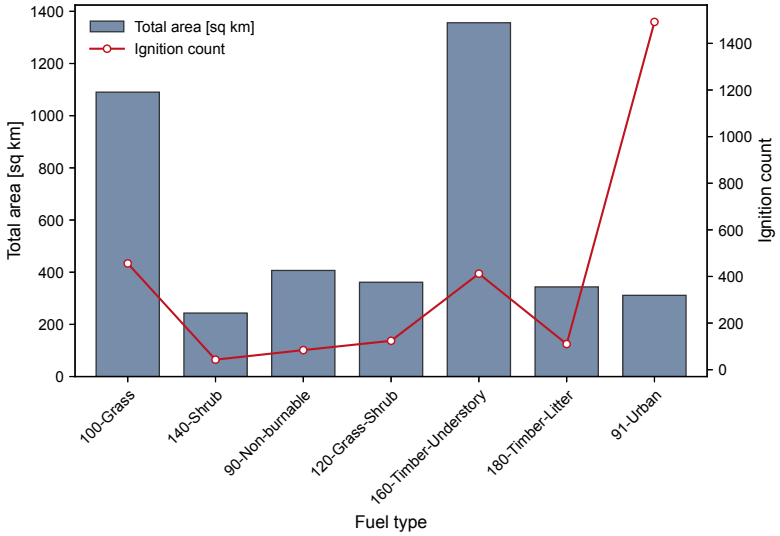


Fig. 5: The estimated area and distribution of non-utility ignitions is shown for each fuel type.

4 Results

4.1 Power-utility ignitions in a linear network

An inhomogeneous Poisson point process on a linear network was computed for power-utility ignitions as a function of voltage and canopy stand height. The inhomogeneous linear K-function and pair correlation function based on Euclidean distance

provide no evidence against the inhomogeneous Poisson point process assumption, as the functions remained within their simulation envelopes across the full distance range. Although the K-function was generally above the simulated mean and the pair-correlation function shows some clustering at very short distances, both patterns remained within the variation expected under the fitted model. The results of this assessment are shown in Appendix C.

For wildfire ignitions occurring along power-line infrastructure voltage, canopy stand height, canopy base height, fuel type, canopy bulk density, and canopy cover were the spatial covariates tested. In Figure 3, the correlations among the terrain characteristics were checked to avoid collinearity problems. Since voltage is a property of the infrastructure independent of vegetation, and not a terrain characteristic, it was not included in this plot. This plot shows how the terrain characteristics relate to each other. All of these covariates were tested individually, and only voltage, canopy stand height, and canopy bulk density were statistically significant. Therefore, canopy base height, fuel type, and canopy cover were omitted. The null model and model combinations of voltage plus canopy stand height, voltage plus canopy bulk density, canopy bulk density plus canopy stand height, and voltage plus canopy bulk density plus canopy stand height were tested, and the AIC score was used to find the best model. The AIC scores and log-likelihood values shown in Table A1 demonstrate that the voltage and canopy stand height model was the best model. Lastly, the distance to the closest road was not considered for this type of ignition since the road network overlaps with the power-line infrastructure in many regions. Hence, canopy stand height and voltage were selected to characterize ignition rates across the power-line infrastructure. The canopy stand height represents the full height of trees, which can be obtained by LiDAR or other remote sensing methods, and voltage is characteristic of the power lines. By choosing these two standard covariates, the model could be more adaptable to different regions in the future.

Using the identified spatial covariates, an inhomogeneous point process model (equation 1) was fitted to the power-utility network of Sonoma County. The dummy point spacing was set to approximately 610 meters because the fitted coefficients converged at this value. All covariates used are statistically significant, as shown in Table 1. Equation 11 shows the fitted functional form of the expected number of power-utility ignitions per meter representative of 10 years of utility ignition data.

$$\lambda_{l,\theta}(u) = \exp\left(-10.1287 - 0.0068V\left(u\right) + 0.0271H\left(u\right)\right) \tag{11}$$

Table 1: The calibrated power-utility wildfire ignition model is presented for Sonoma County.

Covariate	Estimate	Standard Error	Z-value	P-value
(Intercept)	-10.1287	0.0774	-130.82	< 0.001
V: Voltage [kV]	-0.0068	0.0021	-3.26	0.001
H: Canopy stand height [m]	0.0271	0.0078	3.47	< 0.001

The results from the Poisson point process analysis shown in Table 1 demonstrate that the rate of power-utility ignitions in the network is a function of voltage and canopy stand height. These covariates are statistically significant, as shown by Z-tests with p-values ≤ 0.001. Table 1 shows the relationship that each covariate has with the rate of ignitions along the power-utility network. First, as canopy stand height increases, the ignition rate also increases. This relationship can be attributed to interactions between overhead power lines and surrounding trees, which can lead to ignitions during extreme winds. Furthermore, as voltage increases, the ignition rate decreases, suggesting an association between lower-voltage distribution lines and higher observed ignition rates. These lower-voltage lines are also more commonly located in areas closer to where people live.

The result of this fitted model is an intensity function $\lambda_{l,\theta}(u)$ that estimates the rate of ignition per meter at any location u in the power network as a function of voltage and canopy stand height, where $V(u)$ is the voltage at location u in kilovolts and $H(u)$ is the canopy stand height at location u in units of meters. Figure 6 shows the intensity map of power-utility ignitions along the infrastructure as the expected number of ignitions per kilometer. This map utilizes equation 11, but the resultant intensity is multiplied by 1000 to obtain the expected number of ignitions per kilometer. Here we can see that most distribution lines have higher ignition rates, between 0.045 and 0.074 ignitions per kilometer, whereas transmission lines have lower ignition rates, between 0.01 and 0.017 ignitions per kilometer. The highest rate of ignitions seems to be concentrated in distribution lines in the central-west portion of the county, where lower voltage and high canopy stand height coincide, as shown in Figure 2a.

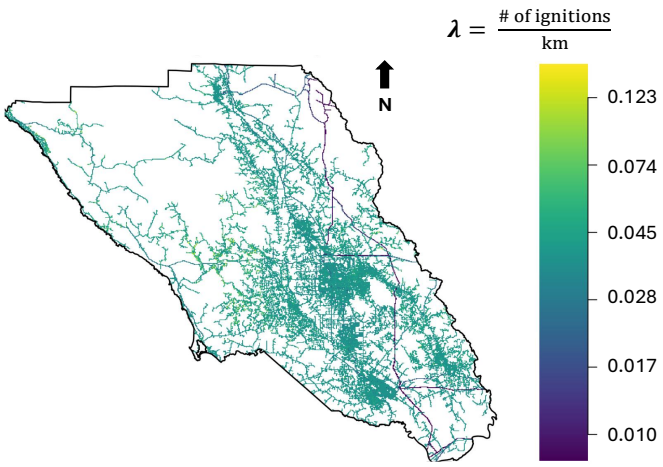


Fig. 6: The map shows the intensity of power-utility wildfire ignitions in Sonoma County.

Figure 7 shows some examples of how the fitted model from equation 11 compares with binned data from particular voltages and canopy stand heights. As canopy stand height increases, the ignition rate also increases. As the voltage increases, the ignition rate decreases.

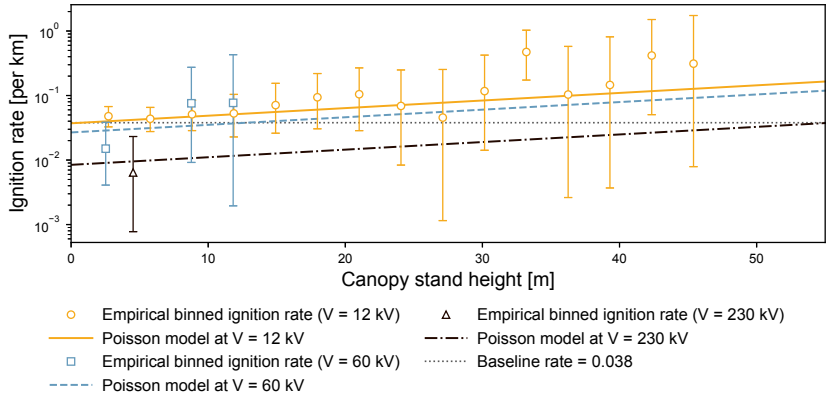


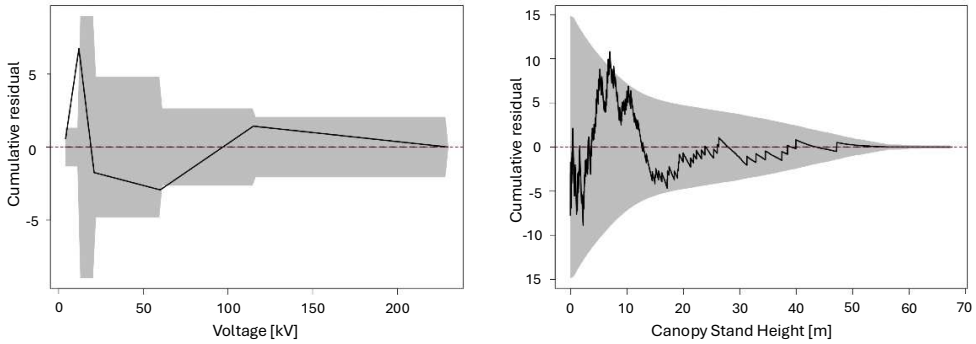
Fig. 7: The fitted Poisson point process function is compared with raw data intensities by voltage group across canopy stand heights.

Lastly, Figure 8 shows the lurking plots for the covariates analyzed. This plot shows, for every covariate value Z on the x-axis, the cumulative residual discrepancy between the observed and model-predicted number of points across all locations where the covariate value Z is less than or equal to z (Baddeley et al. 2015). Positive residuals indicate an under-prediction of ignitions, whereas negative residuals indicate an over-prediction. The solid line is the cumulative residual of the fitted model, while the dashed line is the theoretical cumulative residual. These figures show that for most covariate values, the residuals lie within the 5% significance bands.

4.2 Non-utility ignitions across space

An inhomogeneous Poisson point process across space was computed for non-utility ignitions as a function of fuel type and road proximity. The inhomogeneous K-function and pair correlation function provide no evidence against the fitted inhomogeneous Poisson point process model for non-utility ignitions, although the pair-correlation function indicates some short-range clustering. The excess is concentrated at the smallest separations and largely disappears at larger separation distances. The results of this assessment are shown in Appendix C.

For non-utility wildfire ignitions, canopy stand height, canopy base height, canopy cover, canopy bulk density, and fuel type were tested individually. Voltage was not tested since power-utility ignitions are excluded. In this case, fuel type, canopy stand height, canopy bulk density, and canopy cover were statistically significant on their own, whereas canopy base height was not. However, fuel type and canopy cover have a



(a) The lurking plot shows the effect of voltage.

(b) The lurking plot shows the effect of canopy stand height.

Fig. 8: The residual plots show trends across the covariates with the grey area representing the 5% significance bands.

high correlation ratio η^2, as shown in Figure 3. Therefore, the model combinations of fuel type plus canopy stand height and fuel type plus canopy bulk density were tested. The AIC scores for both model combinations were very similar, as shown in Table A2; however, when considering the fuel type and the canopy stand height, the canopy stand height was no longer statistically significant. This may be due to the moderate correlation ratio η^2 shown in Figure 3. The AIC scores of the model using fuel type alone and fuel type plus canopy bulk density were very similar as well; however, canopy bulk density is not always available across regions. Therefore, for practical reasons and due to similar AIC scores among models, fuel type was the only terrain covariate selected.

Furthermore, to improve the model performance, a range of road distance thresholds were tested as spatial covariates, given that Figure 4b shows the concentration of ignitions near major roads. Exploratory analysis indicated that road adjacency had a binary effect on ignitions, rather than a continuous effect as a function of distance to a road. Hence, road distance thresholds ranging from 2000 meters to 15 meters were tested. The AIC and log-likelihood scores were used to select the most predictive model using the fuel type and the road distance threshold. Table A2 in Appendix A reports results from the full sweep, and shows that adding road proximity improves on the fuel-only model at every threshold tested, with AIC declining monotonically from 2000 m to a minimum at 60 m before rising again at shorter distances. The interior optimum indicates that the effect is not an artifact of the positional precision of the ignition coordinates: if those coordinates were effectively snapped onto the road geometry, the narrowest bands would fit best, whereas here the fit degrades below 60 m. A band of a few tens of meters is instead consistent with roadside ignition sources operating at vehicle scale, such as sparks, hot exhaust components, and discarded

cigarettes, rather than with a diffuse accessibility effect, which would have favored the broader thresholds.

Therefore, the selected model uses a distance threshold of 60 meters. This covariate was statistically significant on its own, and its correlation ratio with fuel type was only 0.108. This covariate is a binary indicator with a 60-meter threshold to classify ignitions close enough to roads. These models were also compared with a homogeneous Poisson point process model, which yielded a higher AIC score.

The inhomogeneous Poisson point process fitted for these wildfire ignitions also used a dummy point spacing of 610 meters. The results of the fitted model are shown in Table 2, where most of the fuel categories and distance to road are statistically significant. The functional form of this model is shown in equation 12. It shows that the fuel type of grass and distance to a major road greater than 60 meters is the baseline of this wildfire ignition intensity which equals to $\exp(-14.69)$ while the intensity of the other fuel types is relative to the intensity of the grass fuel as $\exp(-14.69 + \theta_{\omega,F})$ meaning the coefficients $\theta_{\omega,F}$ are the difference in effect between level i and the first level in log scale. In the case that the location s is within 60 meters of the road $\theta_{\omega,R}$ is also considered. Equation 12 has intensity units of expected number of ignitions per square meter representative of 23 years of the ignition data.

$$\lambda_{\omega,\theta}(s) = \exp\left(-14.69 + \theta_{\omega,F}F(s) + \theta_{\omega,R}R(s)\right) \tag{12}$$

Table 2: The calibrated non-utility wildfire ignition model is presented for Sonoma County.

Covariate	Estimate	Standard Error	Z-value	P-value
(Intercept, baseline fuel: 100-Grass)	-14.69	0.046	-318.35	< 0.001
Fuel: 140-Shrub	-0.77	0.154	-4.97	< 0.001
Fuel: 90-Non-burnable	-0.70	0.117	-5.96	< 0.001
Fuel: 120-Grass-Shrub	-0.25	0.103	-2.40	0.017
Fuel: 160-Timber-understory	-0.30	0.066	-4.53	< 0.001
Fuel: 180-Timber-litter	-0.18	0.101	-1.75	0.08
Fuel: 91-Urban	1.58	0.056	28.08	< 0.001
Road distance less than 60 meters	1.19	0.046	25.79	< 0.001

The result of this fitted model is a rate of ignition per square meter at any location s as a function of the categorical fuel type $\theta_{\omega,F}$ and road distance threshold of 60 meters $\theta_{\omega,R}$. All of the considered covariates are statistically significant relative to the baseline, except for the timber-litter fuel type, as shown by the Z-test and its corresponding p-values. Since the baseline of this model is a grass fuel type and a distance from a road greater than 60 meters, the influence of other covariate types can be interpreted with respect to the baseline. For example, with regard to shrub fuel types ($\exp(-0.77) = 0.46$), the rate of wildfire ignitions in this type of fuel and occurring at a distance from a road greater than 60 meters is 0.46 times the rate in grass fuel types and with distances greater than 60 meters. On the other hand,

locations that are 60 meters or less from the nearest road have a 3.28-fold increase in the ignition rate ($\exp(1.19) = 3.28$).

Figure 9 shows the intensity map of non-utility ignitions in units of ignitions per square kilometer as per equation 12 multiplied by 1000^2, demonstrating the regions in the county that are more prone to experiencing a wildfire ignition. This map represents the ignition intensity representative of 23 years of data. This Figure shows how ignition rates increase in areas near roads and urban fuel sources, as these drive higher intensities. Lastly, the fitted function can be used to predict future ignitions as a function of fuel type and distance to the nearest road.

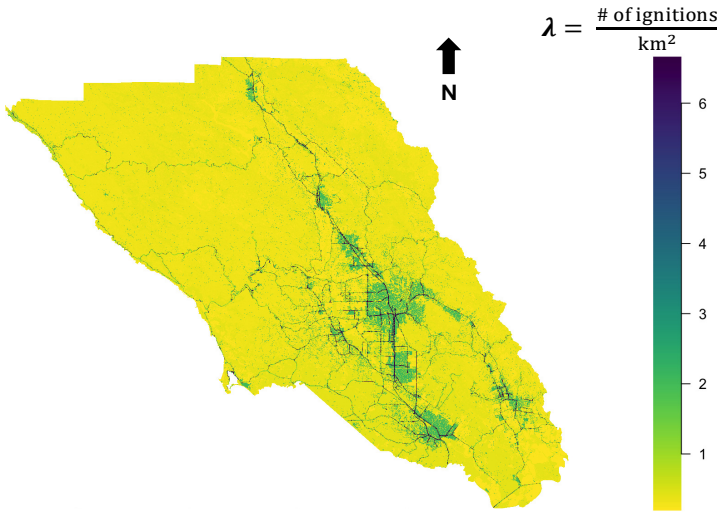


Fig. 9: The map shows the intensity of non-utility wildfire ignitions in Sonoma County.

To understand how the rate of ignitions in space relates to the different fuel types, Figure 5 shows the estimated area per fuel type along with the number of ignitions per category. This Figure shows that wildfire ignitions in urban fuel types are higher than in other fuel types, given the high number of ignitions and the smaller total area in Sonoma County. Also, Figure 2b shows the spatial distribution of the fuel categories and the number of wildfire ignitions across each category, showing that urban regions have higher occurrences, although they represent a smaller area of the entire county, and other fuel types like grass, which have larger areas, also contain a higher number of ignitions. Figure 10 shows the locations where the distance to the nearest road is less than 60 meters in red, and the locations of the non-utility ignitions in black, indicating the correspondence of ignitions with road locations. The resolution of the red area is 10 meters.

To assess the performance of the fitted inhomogeneous Poisson point process, the residual measure of the fitted model was computed and plotted to visualize the discrepancy between the fitted intensity and the raw data. Figure 11 shows the smoothed residuals across the region. The raw residuals are calculated using equation 13 where

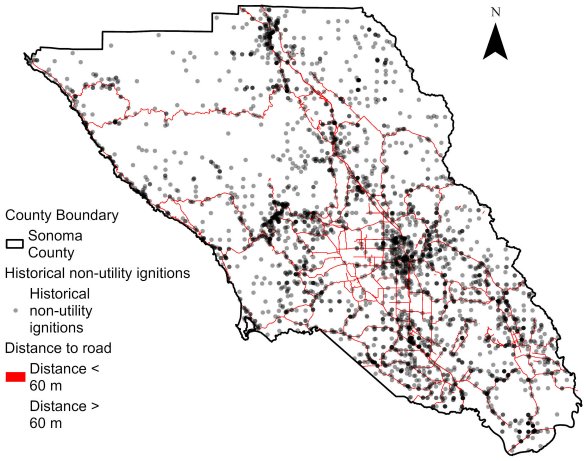


Fig. 10: The road distance threshold and non-utility ignitions is shown in Sonoma County.

z_j is equal to 1 for observed data points and 0 for dummy points, w_j is the quadrature weight, which approximates the pixel area around the point (if a pixel gets multiple points, the pixel area is divided among them), and λ_j is the fitted conditional intensity. The smoothed version of this residual applies a kernel to the raw residuals obtaining equation 14, where $\tilde{\lambda}(u)$ is the nonparametric estimate of the intensity and $\lambda^{\dagger}(u)$ is the smoothed version of the parametric estimate of the intensity as per equation 15 where $e(u)$ is the edge correction defined in equation16 (Baddeley et al. 2015). In order to smooth the kernel κ, an optimized bandwidth for this ignition pattern was generated using likelihood cross-validation, which selects the bandwidth that best predicts point occurrence using leave-one-out intensity estimates, as shown in equation 17 (RDocumentation 2020).

$$r_j = z_j - w_j \lambda_j \tag{13}$$

$$s(u) = \tilde{\lambda}(u) - \lambda^{\dagger}(u) \tag{14}$$

$$\lambda^{\dagger}(u) = \frac{1}{e(u)} \int_{\Omega} \kappa(u - v)\, \lambda_{\hat{\theta}}(v)\, dv \tag{15}$$

$$e(u) = \int_{\Omega} \kappa(u - v)\, dv \tag{16}$$

$$LCV(\sigma) = \sum_{i} \log \hat{\lambda}_{-i}(x_i) - \int_{\Omega} \hat{\lambda}(u)\, du \tag{17}$$

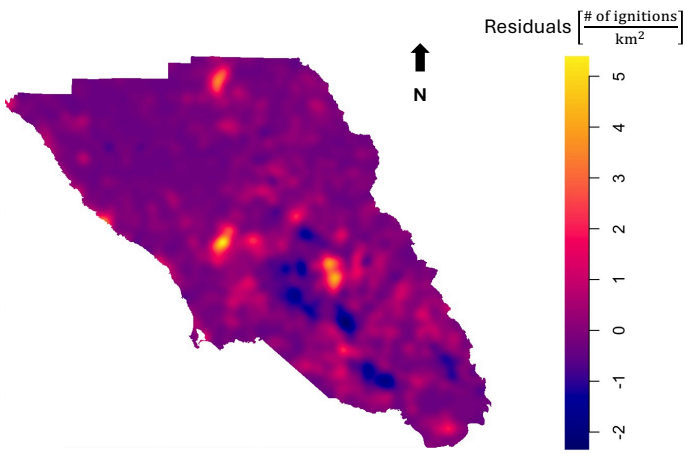


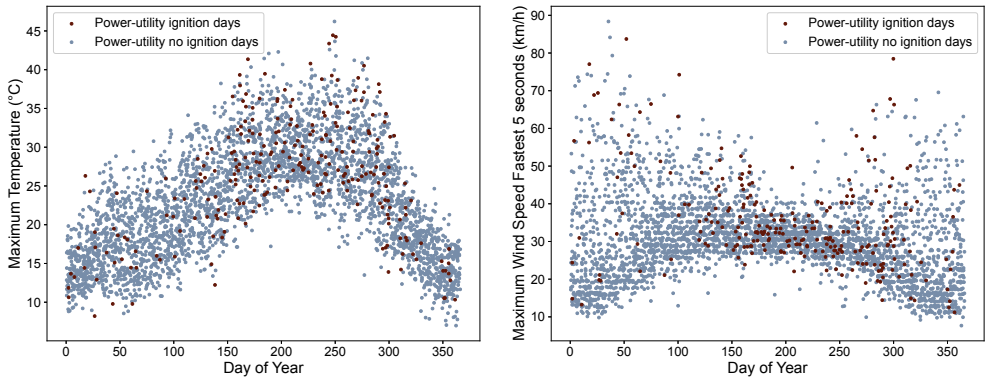
Fig. 11: The residual map shows non-utility wildfire ignitions in units of the expected number of ignitions per square kilometer.

Figure 11 demonstrates how the model predicts wildfire ignitions across Sonoma County using the optimized smoothing sigma of 974 meters. Positive residuals indicate areas where the fitted model underpredicts wildfire ignitions, while negative residuals indicate areas where the model overpredicts ignitions. Overall, the model performs well across the region. Given the average spatial intensity is 0.66 ignitions per square kilometer, the bright yellow areas show an underprediction difference of 0.34 to 4.34 ignitions per square kilometer, and the dark blue areas show an overprediction difference of 0.34 to 1.34 ignitions per square kilometer. The spots with overprediction indicate regions where there is little observed ignition. However, the fuel in this region is an urban fuel type, and there is also a concentration of roads; therefore, ignition intensity is high, and an overprediction effect is observed. On the other hand, the area of underprediction appears to have fuel types such as timber understory and lower road proximity, which confirms the lower intensity in these areas. Lastly, the underprediction hotspot near the center of the county corresponds to regions with mixed grass and non-burnable fuels and is close to urban fuel types, which lead to a lower ignition rate than observed in this region.

4.3 Ignitions across time

A Negative Binomial model was computed as a function of maximum temperature and wind speed to represent ignitions across time. Following the methodology described in Section 2.3, the first step in calibrating this ignition model over time is to collect historical weather data for the period covered by the ignition datasets. For example, for power-utility ignitions, weather data was collected from 2015 to 2024, whereas for non-utility ignitions, weather data was collected from 1998 to 2020, corresponding to each ignition dataset. Using daily weather data from the NOAA station at Santa Rosa Airport, the selected weather conditions were the recorded maximum temperature and

the maximum 5-second wind speed, given that they are statistically significant and that other studies use similar weather covariates in temporal models. These weather covariates were tracked for ignition and non-ignition days. For example, Figure 12 shows how the maximum temperature and maximum wind speed have varied over 10 years and how these weather covariates differ on days with power-utility-related ignitions compared to days without such ignitions. Similarly, Figure 13 shows the same covariates but for non-utility ignitions, which range from 1998 to 2020.



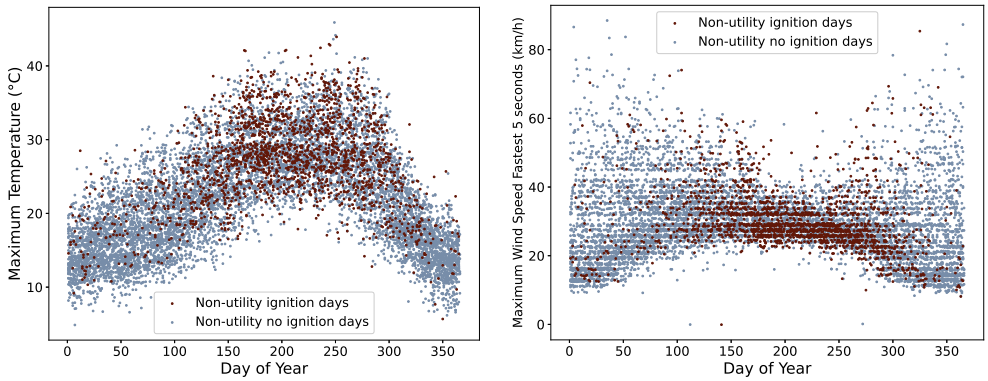
(a) The maximum temperature history is shown for power-utility ignitions in Sonoma County from 2015 to 2024.

(b) The maximum 5-second wind speed history is shown for power-utility ignitions in Sonoma County from 2015 to 2024.

Fig. 12: The distribution of weather covariates is shown for power-utility ignitions in Sonoma County.

The scatter plots of maximum temperature in figures 12 and 13 show that the density of ignition days increases with temperature. Regarding the maximum wind speed, we observe clustering in the middle of the year, towards the summer months, where ignition density is higher than on non-ignition days. These patterns and the methodology described earlier allowed us to describe the rate of ignition days as a function of maximum temperature and maximum wind speed.

After combining the ignition and daily weather datasets, count regression models were tested to predict the rate of daily ignitions as a function of temperature and wind. The variance of the number of ignitions per day in both ignition source datasets slightly exceeded the mean ($\sigma^2_{l,\tau} = 0.1088$, $\mu_{l,\tau} = 0.0841$, $\sigma^2_{\omega,\tau} = 0.5229$, $\mu_{\omega,\tau} = 0.3302$), indicating mild overdispersion. This overdispersion indicated that a Negative Binomial model would perform better than a traditional Poisson regression, since it accounts for it by including an extra term, α, in the variance. The Negative Binomial and the Poisson model's AIC score were compared and the Negative Binomial was selected for ignition sources. The results from the Negative Binomial model are shown in Tables 3 and 4 with their respective fitted equations, where $\lambda_{l,\tau,\theta}(t)$ is the estimated rate



(a) The maximum temperature history is shown for non-utility ignitions in Sonoma County from 1998 to 2020.

(b) The maximum 5-second wind speed history is shown for non-utility ignitions in Sonoma County from 1998 to 2020.

Fig. 13: The distribution of weather covariates is shown for non-utility ignitions in Sonoma County.

of power-utility ignitions per day and $\lambda_{\omega,\tau,\theta}(t)$ is the estimated rate of non-utility ignitions per day.

$$\lambda_{l,\tau,\theta}(t) = \exp\left(-6.0879 + 0.0807T(t) + 0.0483WS(t)\right) \tag{18}$$

Table 3: The calibrated power-utility wildfire ignition days model is presented for Sonoma County.

Covariate	Estimate	Standard Error	Z-value	P-value
(Intercept)	-6.0879	0.342	-17.798	< 0.001
T: Maximum Temperature [C]	0.0807	0.009	8.880	< 0.001
WS: Maximum Wind Speed [km/hr]	0.0483	0.005	8.982	< 0.001
α	1.6136	0.431	3.742	< 0.001

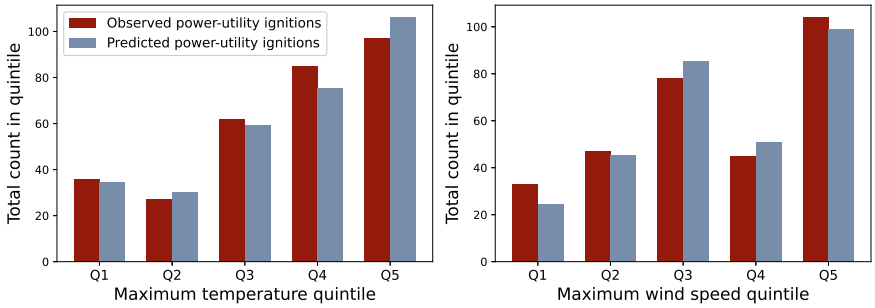
$$\lambda_{\omega,\tau,\theta}(t) = \exp\left(-4.5090 + 0.1092T(t) + 0.0231WS(t)\right) \tag{19}$$

To assess the performance of the fitted models, these were used to predict the number of ignitions per temperature and wind speed quintiles. The predicted number of ignitions and the observed number of ignitions per quintile were compared as shown in Figures 14a and 14b. For the more extreme conditions, represented by quintile 5 with a temperature range of 28 to 46 degrees Celsius, the number of predicted power-utility ignitions exceeds the observed number, whereas the model slightly underpredicts the

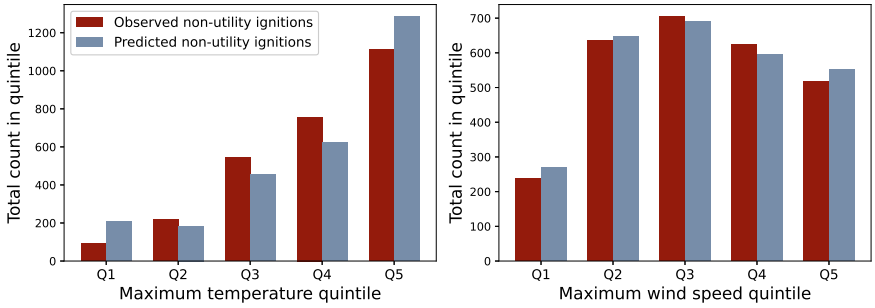
Table 4: The calibrated non-utility wildfire ignition days model is presented for Sonoma County.

Covariate	Estimate	Standard Error	Z-value	P-value
(Intercept)	-4.5090	0.122	-37.021	< 0.001
T: Maximum Temperature [C]	0.1092	0.003	31.790	< 0.001
WS: Maximum Wind Speed [km/hr]	0.0231	0.002	9.742	< 0.001
α	0.7094	0.066	10.779	< 0.001

number of ignitions with extreme wind speeds of 35 to 88 kilometers per hour. On the other hand, the model slightly overpredicts for both extreme weather covariates for non-utility ignitions. Appendix B presents a complementary check, comparing predicted and observed ignition rates by decile and as continuous functions of maximum temperature and wind speed. Finally, using the latest available data or daily weather forecasts, the expected number of ignitions per day can be computed using maximum temperature and wind speed as predictors using this fitted model.



(a) Predicted and observed power-utility ignitions are shown across quintiles of maximum temperature and wind speed.



(b) Predicted and observed non-utility ignitions are shown across quintiles of maximum temperature and wind speed.

Fig. 14: The predicted number of ignitions from the Negative Binomial model is compared with the observed number of ignitions across quintiles of maximum temperature and wind speed.

4.4 Spatial/temporal separability assessment

To evaluate the suitability of a decoupled model, we compare nested Poisson point-process models using the likelihood-ratio tests of Section 2.5. Tables 5 through 8 summarize the likelihood-ratio test results.

Table 5: The analysis of deviance summarizes the power-utility ignition model divided by year halves.

Model	Npar	Df	Deviance	$\Pr(>\chi^2)$
$X \sim \text{period} + \text{V} + \text{H}$	4	–	–	–
$X \sim \text{period} * (\text{V} + \text{H})$	6	2	0.205	0.9026

Table 6: The analysis of deviance summarizes the power-utility ignition model divided by quarters.

Model	Npar	Df	Deviance	$\Pr(>\chi^2)$
$X \sim \text{period} + \text{V} + \text{H}$	6	–	–	–
$X \sim \text{period} * (\text{V} + \text{H})$	12	6	9.3131	0.1567

Table 7: The analysis of deviance summarizes the non-utility ignition model divided by year halves.

Model	Npar	Df	Deviance	$\Pr(>\chi^2)$
$Q \sim \text{period} + \text{F} + \text{R}$	9	–	–	–
$Q \sim \text{period} * (\text{F} + \text{R})$	16	7	10.288	0.1728

The likelihood-ratio tests did not provide evidence of period-dependent spatial covariate effects for three of the four model comparisons. For power-utility ignitions, the interaction model did not improve model fit when the data were divided into half-year periods (D=0.205, df=2, p=0.9026) or quarters (D=9.313, df=6, p=0.157) as shown in Tables 5 and 6. Similarly, for non-utility ignitions divided into half-year periods, the separable model was not rejected (D=10.288, df=7, p=0.173) as shown in Table 7. However, when non-utility ignitions were divided into quarters, the period $\times$ covariate interactions improved the model (D=41.063, df=21, p=0.0055) as shown in Table 8. This result provides evidence that at least one spatial covariate effect varies seasonally for non-utility ignitions at the quarterly temporal resolution. The observed covariates that varied substantially by period were the fuel types of urban, shrub, and timber-litter. In this case, the urban fuel type coefficient had higher ignition rates in

Table 8: The analysis of deviance summarizes the non-utility ignition model divided by quarters.

Model	Npar	Df	Deviance	$\Pr(>\chi^2)$
$Q \sim \text{period} + \text{F} + \text{R}$	11	–	–	–
$Q \sim \text{period} * (\text{F} + \text{R})$	32	21	41.063	0.005508

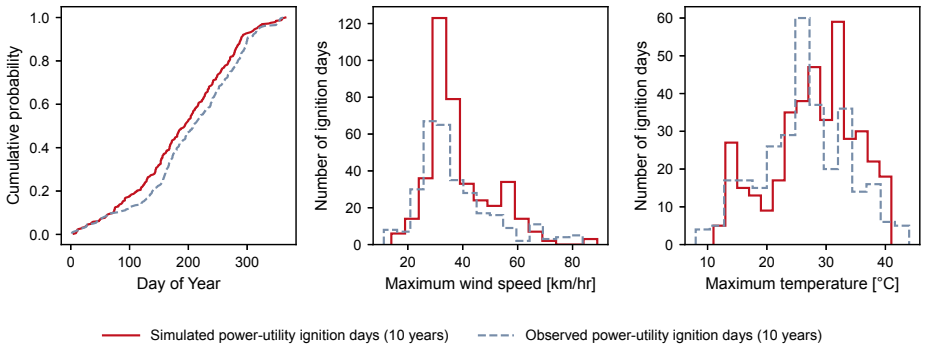
quarters 1 and 4 than during other quarters. Overall, the results were consistent with period-invariant spatial covariate effects in three of the four analyses, but indicated a departure from separability for non-utility ignitions at the finer quarterly temporal resolution.

Lastly, the separability from kernstadapt R package was not completed because it addresses a different statistical question from the covariate-based models considered here (González and Moraga 2022). This test asks whether the spatial pattern is constant in time, which is dominated by the absolute rate of events, whereas our temporal model accounts for that. This test also does not incorporate the spatial covariates used in the ignition models or directly test whether their estimated effects vary among temporal periods. Therefore, we conducted a likelihood-ratio test for both ignition sources to support this assumption.

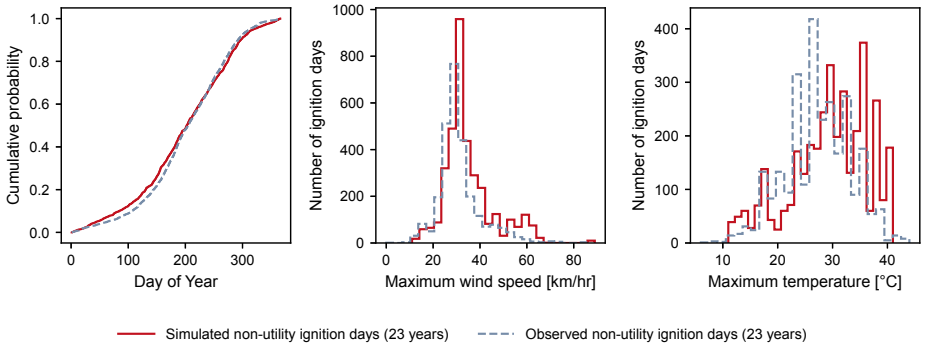
4.5 Simulated ignitions

Having a calibrated wildfire ignition model for Sonoma County across its power utility network, county area, and time, allows for modeling wildfire ignition locations and ignition days. The spatial models described in Section 4.1 and 4.2 can be used to simulate the spatial location of future ignitions, whereas the Negative Binomial model described in Section 4.3 can be used to sample the ignition days with the corresponding number of ignitions, maximum temperature, and maximum wind speed for power-utility and non-utility wildfire ignitions. Using the latest year available daily weather dataset collected at the Santa Rosa Airport station (2024), we can simulate ignition days. Figure 15 compares the observed ignition days and simulated ignition days for power-utility and non-utility ignitions. Using the 2024 weather dataset results in ignition days with maximum temperatures and wind speed close to the observed dataset, producing realistic ignitions under the calibrated conditions.

To illustrate how the framework can be used to evaluate ignition rates under altered weather conditions, we conducted a sensitivity analysis in which the 2024 weather dataset was perturbed to represent hotter and windier conditions than those observed. Each day's maximum temperature was increased by a uniform random number between 0 and 4, and each day's maximum wind speed was increased by a uniform random number between 0 and 9. These perturbation ranges are illustrative rather than calibrated to any specific climate projection; the analysis is intended to demonstrate how the fitted temporal model responds to shifted weather distributions, not to forecast future ignition rates. If downscaled climate projections were available for the study region, those inputs could be substituted directly. Using the perturbed weather



(a) Distribution of ignition day simulations is shown for power-utility ignitions using the 2024 weather data.

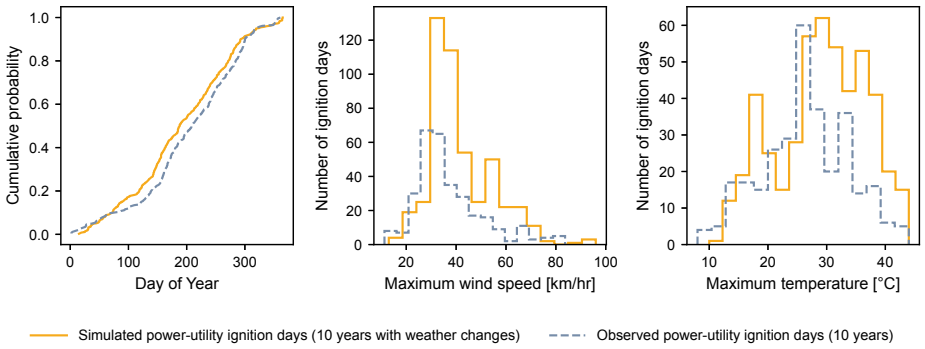


(b) Distribution of ignition day simulations is shown for non-utility ignitions using the 2024 weather data.

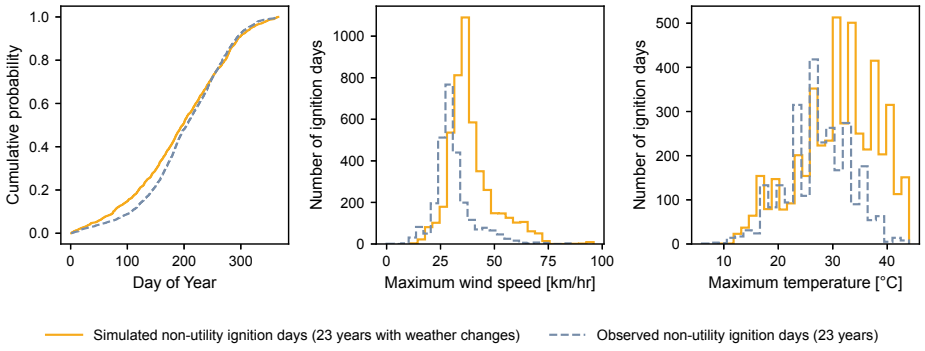
Fig. 15: Ignition-day simulations are shown using the 2024 weather dataset.

dataset and the fitted Negative Binomial model, ignition days and their associated number of ignitions were simulated. Figure 16 shows the results.

Figure 16 shows how the sampled ignition days, with their corresponding maximum temperatures and wind speeds, compare with the historical dataset of power-utility and non-utility ignitions. In general, ignition days throughout the year are well represented in the simulations, with most ignitions occurring in the summer and fall. With the random increase in temperature and wind speed, these figures show how the number of ignition days shifts towards more extreme weather covariates that tend to lead to more catastrophic events. For both ignition sources, the maximum wind speed on ignition days shifts toward windier days, which also corresponds to a higher number



(a) Distribution of ignition day simulations is shown for power-utility ignitions using the modified weather dataset with hotter and windier conditions.



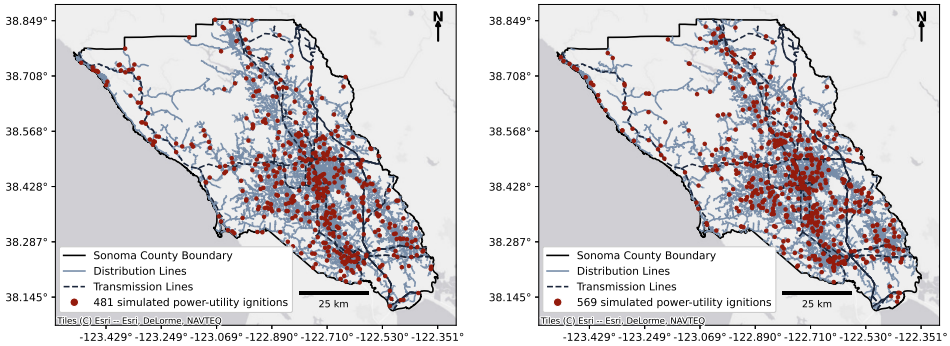
(b) Distribution of ignition day simulations is shown for non-utility ignitions using the modified weather dataset with hotter and windier conditions.

Fig. 16: Ignition day simulations are shown with a modified weather dataset.

of ignitions. On the other hand, a similar trend is observed for maximum temperature, which, in combination with wind speed, will result in an overall increase in the number of ignitions as these weather covariates change over time.

After modeling the spatial location and ignition days of both ignition sources considered, they were coupled by randomly assigning spatial location to the number of ignitions per simulated ignition days while keeping the corresponding weather conditions of maximum wind speed and temperature. This way, the wildfire ignition model provides complete spatial coordinates and weather conditions as initial inputs to wildfire spread models. Given that the catalog duration of the utility spatial model differs from the non-utility spatial model, the spatial ignition simulations obtained from these models correspond to their respective catalog durations. For example, Figure 17 shows

examples of two different sets of simulated power utility ignitions along the power utility network of Sonoma County corresponding to ten years of wildfire ignitions. This Figure shows that ignitions along power lines concentrate in areas where the ignition rate λ shown in Figure 6 is higher. Comparing the simulated ignitions with Figure 1, we can observe that the regions where fewer ignitions occurred tend to have fewer simulated ignitions, and regions with dense distribution-line clusters, where historical occurrences were more numerous, also show more simulated ignitions. On the other hand, Figure 18 also shows two simulated examples of non-utility ignition sets across Sonoma County, corresponding to 23 years of wildfire ignitions, with higher ignition rates in regions where Figure 9 shows higher intensity.

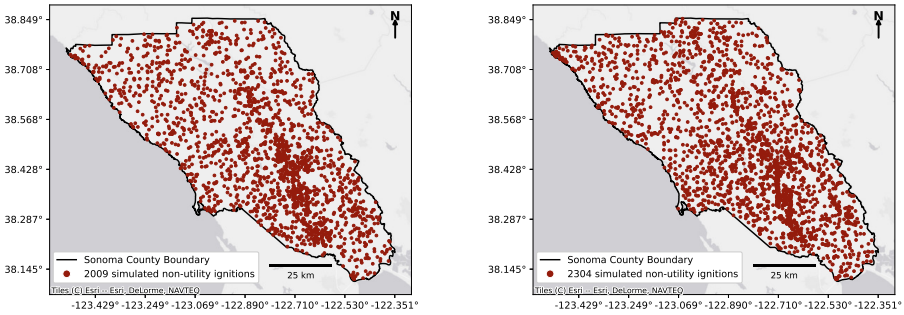


(a) Example 1 shows ten years' worth of simulated power-utility ignitions.

(b) Example 2 shows ten years' worth of simulated power-utility ignitions.

Fig. 17: The spatial distribution of simulated power-utility ignitions is shown for Sonoma County.

Lastly, we analyze the weather covariates observed for both ignition sources to better understand what drives each. Figure 19 shows how simulated power-utility ignitions are driven by high-wind conditions, whereas non-utility ignitions are driven more by high-temperature days. These indicators and maps of potential future ignition locations can help community leaders better assess their wildfire risk and implement informed mitigation actions.



(a) Example 1 shows 23 years' worth of simulated non-utility ignitions.

(b) Example 2 shows 23 years' worth of simulated non-utility ignitions.

Fig. 18: The spatial distribution of simulated non-utility ignitions is shown for Sonoma County.

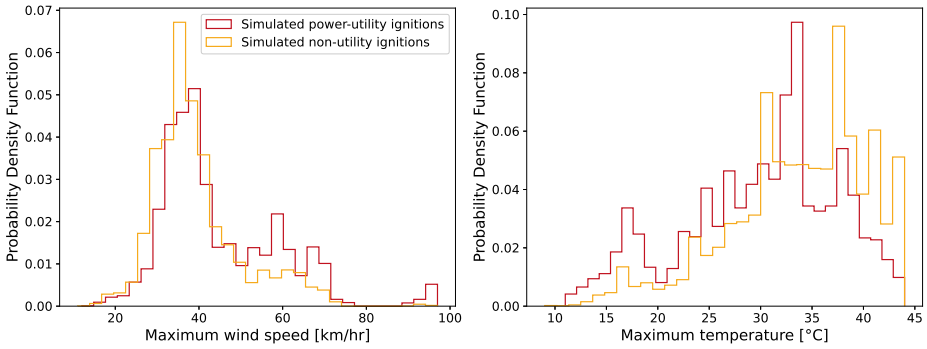


Fig. 19: The maximum wind speed and temperature distributions is shown for non-utility and power-utility ignitions.

5 Discussion and conclusions

This manuscript proposed an integrated wildfire ignition model that accounts for differences in the spatial patterns of utility-caused and non-utility ignitions. The spatial and temporal models are developed separately, assuming that the rates of ignition in space and time are independent.

The spatial patterns of ignitions were examined across different temporal intervals to assess whether the spatial rate of ignitions is affected by the time of year. Models of spatial patterns of ignition occurrences are coupled with temporal rates of occurrence. The mathematical forms of the models are described, and empirical data from Sonoma County, California, are used to calibrate the models and to produce stochastic simulations of ignitions in time and space. In Sonoma County, the utility ignitions occur only along power lines, and depend on power-line voltage and canopy stand height; non-utility ignitions throughout the county depend on the fuel type of the landscape and proximity to roads. For temporal ignition models, maximum temperature and wind determine the ignition intensity for the two sources considered.

The separability assumption underlying this decoupling was tested using likelihood-ratio tests comparing models with and without period-specific spatial covariate effects. For power-utility ignitions, the assumption was not rejected at either half-year or quarterly resolution; for non-utility ignitions, it was not rejected at half-year resolution but was rejected at quarterly resolution ($p = 0.0055$). The relative influence of fuel type and road proximity on where non-utility ignitions occur therefore varies seasonally. The decoupled model reproduces the seasonal variation in the number of non-utility ignitions, but holds their spatial distribution fixed across the year, so simulations at quarterly or finer resolutions will understate seasonal shifts in non-utility ignition geography. Utility ignitions showed no evidence of comparable seasonal variation, which is consistent with a process governed by a fixed physical network rather than by seasonally varying human activity, although the smaller number of utility ignitions means that detecting variations is more difficult.

The proposed models show that power-utility ignitions are more strongly influenced by wind speed than non-utility ignitions, consistent with Keeley and Syphard (2019). The selected spatial and weather covariates, and their estimated significance, are broadly consistent with those of D'Angelo et al. (2025). This article differs in modeling utility ignitions on the power-line network itself, so that their spatial support is the linear system on which they occur. It also differs in decoupling the spatial and temporal components, which keeps the framework tractable and straightforward to recalibrate in a new region. Other studies, such as Muhs et al. (2020) and Bayani et al. (2022), calculate utility ignition probability at the system level using covariates relevant at the regional scale, including wind speed and vegetation.

The proposed models have some limitations. The model relies heavily on region-specific data, and without it, it would not be possible to predict ignition likelihood. In the future, this model will be calibrated in other regions of California to demonstrate its applicability across different areas and to account for other ignition sources. By applying this model across different regions in California, it will become more generalizable and applicable to multiple regions with less reliance on local data. Furthermore, the weather data used to calibrate the temporal models were collected only at one

station. In the future, if more stations are available in the same region, they should be used to better represent weather conditions on ignition and non-ignition days near the ignition location.

The proposed methodology is general and can be applied to different sources of ignition, provided data are available to simulate ignitions along a linear corridor or in space. These spatial models can produce intensity maps that help communities better understand wildfire ignition risk at the regional scale. The fitted weather model indicates that weather covariates, such as temperature and wind speed, can increase the intensity of wildfire ignition on a given day, enabling communities to better prepare during wildfire season. We also demonstrated how these decoupled models can be connected to enable stochastic simulations of wildfires that communities can use to assess their risk from different wildfire ignition sources and to prepare for potential scenarios associated with this hazard. Furthermore, the outputs of these models can be used to simulate wildfire ignitions across linear networks and regions, as shown in Section 4.5, serving as input parameters for wildfire spread simulations in regional risk assessments.

Data and code availability. The spatial Poisson point process models were developed using the R package spatstat (Baddeley et al. 2015) while the temporal models were developed using Statsmodel in Python. Code implementing the analysis described above is available at https://github.com/gcalana/Wildfire-ignition-modeling upon publication.

Acknowledgments. We thank Adrian Baddeley for his feedback and assistance developing the Poisson point Process model on a linear network. We also thank Nikola Blagojevic, Simona Meiler, Emily Mongold, Mia Lochhead, Gregory Deierlein, and Tinger Zhu for helpful discussions and feedback. This work was funded in part by the UPS Endowment Fund and by the Stanford Urban Resilience Initiative. Some of the computing for this project was performed on the Sherlock cluster. We thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

Declarations

The authors have no relevant financial or non-financial interests to disclose.

References

Arndt, N., Vacik, H., Koch, V., Arpaci, A., Gossow, H.: Modeling human-caused forest fire ignition for assessing forest fire danger in austria. iForest—Biogeosciences and Forestry **6**(6), 315 (2013)

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., Davies, T.M.: Analysing point patterns on networks—a review. Spatial Statistics **42**, 100435 (2021)

Boise State University: FPA-FOD-Plus Visualizer. https://fpafod.boisestate.edu/About. Accessed: April 20, 2026

Baddeley, A., Rubak, E., Turner, R.: Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London (2015)

Berman, M., Turner, T.R.: Approximating point process likelihoods with glim. Journal of the Royal Statistical Society: Series C (Applied Statistics) **41**(1), 31–38 (1992)

Bayani, R., Waseem, M., Manshadi, S.D., Davani, H.: Quantifying the risk of wildfire ignition by power lines under extreme weather conditions. IEEE Systems Journal **17**(1), 1024–1034 (2022)

Costafreda-Aumedes, S., Comas, C., Vega-Garcia, C.: Human-caused fire occurrence modelling in perspective: a review. International Journal of Wildland Fire **26**(12), 983–998 (2017)

California Public Utilities Commission: Fire Ignition Data. Accessed August 2, 2026. https://www.cpuc.ca.gov/industries-and-topics/wildfires#fire-ignition-data

CAL FIRE: Top 20 Most Destructive California Wildfires. https://34c031f8-c9fd-4018-8c5a-4159cdff6b0d-cdn-endpoint.azureedge.net/-/media/calfire-website/our-impact/fire-statistics/top20_destruction.pdf. Accessed: 2025-05-05 (2023)

County of Sonoma GIS Central: County Boundary. County of Sonoma. Date updated: March 21, 2025. Accessed May 27, 2026 (2025). https://gis.sonomacounty.ca.gov/datasets/county-boundary/about

County of Sonoma: Sonoma County Interactive GIS Map. Esri. ArcGIS item ID: 1a27ccfca0cb469f8dbd14bfacbec63f (n.d.). https://sonomacounty.maps.arcgis.com/apps/instant/sidebar/index.html?appid=1a27ccfca0cb469f8dbd14bfacbec63f Accessed 2026-07-27

D'Angelo, N., Albano, A., Gilardi, A., Adelfio, G.: Non-separable spatio-temporal poisson point process models for fire occurrences. Environmental and Ecological Statistics **32**(1), 347–381 (2025)

González, J.A., Moraga, P.: Kernstadapt: Spatio-Temporal Adaptive Kernel Estimators for Intensities. (2022). R package version 0.0.2

Gabriel, E., Opitz, T., Bonneu, F.: Detecting and modeling multi-scale space-time structures: The case of wildfire occurrences. Journal de la Société Française de Statistique **158**(3), 86–105 (2017)

Ganteaume, A., Syphard, A.D.: Ignition sources. In: Manzello, S. (ed.) Encyclopedia of Wildfires and Wildland-Urban Interface (WUI) Fires. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-51727-8-43-1

Jiménez-Ruano, A., Jolly, W.M., Freeborn, P.H., Vega-Nieva, D.J., Monjarás-Vega,

N.A., Briones-Herrera, C.I., Rodrigues, M.: Spatial predictions of human and natural-caused wildfire likelihood across montana (usa). Forests **13**(8), 1200 (2022)

Koh, J., Pimont, F., Dupuy, J.-L., Opitz, T.: Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. The Annals of Applied Statistics **17**(1), 560–582 (2023) https://doi.org/10.1214/22-AOAS1642

Keeley, J.E., Syphard, A.D.: Twenty-first century california, usa, wildfires: fuel-dominated vs. wind-dominated fires. Fire Ecology **15**, 24 (2019) https://doi.org/10.1186/s42408-019-0041-0

Li, S., Banerjee, T.: Spatial and temporal pattern of wildfires in california from 2000 to 2019. Scientific Reports **11**, 8779 (2021)

Menne, M.J., Durre, I., Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R.S., Gleason, B.E., Houston, T.G.: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3. NOAA National Climatic Data Center. [Indicate subset used]. Accessed May 27, 2026 (2012). https://doi.org/10.7289/V5D21VHZ . https://www.ncei.noaa.gov/cdo-web/search?datasetid=GHCND

Mitchell, J.W.: Power line failures and catastrophic wildfires under extreme weather conditions. Engineering Failure Analysis **35**, 726–735 (2013)

Muhs, J.W., Parvania, M., Nguyen, H.T., Palmer, J.A.: Characterizing probability of wildfire ignition caused by power distribution lines. IEEE Transactions on Power Delivery **36**(6), 3681–3688 (2020)

NPR: Eaton Fire investigation. https://www.npr.org/2025/10/02/g-s1-90701/california-wildfire-altadena-socal-edison-investigation. Accessed: 2026-05-11 (2025)

Pacific Gas and Electric Company: Grid Resource Integration Portal (GRIP): Integration Capacity Analysis (ICA) Data. https://grip.pge.com/. Online data portal. Accessed 2026-05-30 (2026)

Pacific Gas and Electric Company: PG&E Fire Incident Data. California Public Utilities Commission. Fire Ignition Data. Accessed May 27, 2026 (2026). https://www.cpuc.ca.gov/industries-and-topics/wildfires

PBS NewsHour: Bare electrical wire, leaning utility poles seen as possible cause of deadly Maui fires. https://www.pbs.org/newshour/nation/bare-electrical-wire-leaning-utility-poles-seen-as-possible-cause-of-deadly-maui-fires. Accessed: 2025-05-05 (2023)

Podur, J., Martell, D.L., Csillag, F.: Spatial patterns of lightning-caused forest fires in ontario, 1976–1998. Ecological Modelling **164**(1), 1–20 (2003)

RDocumentation: bw.ppl: Likelihood Cross Validation Bandwidth Selection for Kernel

Density. RDocumentation, (2020). RDocumentation. Documentation for the R package spatstat, version 1.64-1. https://www.rdocumentation.org/packages/spatstat/versions/1.64-1/topics/bw.ppl Accessed 2026-07-28

Read, N., Duff, T.J., Taylor, P.G.: A lightning-caused wildfire ignition forecasting model for operational use. Agricultural and Forest Meteorology **253**, 233–246 (2018)

Short, K.C.: Spatial wildfire occurrence data for the United States, 1992–2020 [FPA_FOD_20221014]. Forest Service Research Data Archive, Fort Collins, CO (2022). https://doi.org/10.2737/RDS-2013-0009.6

Sonoma County Vegetation Mapping and LiDAR Program: Sonoma County 5-meter Fuel Model (Map Service, v. 2/2021). ArcGIS Online. Available: https://www.arcgis.com/home/item.html?id=2d194202dfee4ab5b157e978cd0e4901. Accessed: Apr. 20, 2026 (2019)

Turner, R.: Point patterns of forest fire locations. Environmental and Ecological Statistics **16**(2), 197–223 (2009)

U.S. Census Bureau, Geography Division: 2020 TIGER/Line Shapefiles: Roads. All Roads shapefile for Sonoma County, California; file: tl_2020_06097_roads.shp (2020). https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2020&layergroup=Roads Accessed 2026-07-27

Vahedi, S., Zhao, J., Pierre, B., Lei, F., Anagnostou, E., He, K., Jones, C., Wang, B.: Wildfire and power grid nexus in a changing climate. Nature Reviews Electrical Engineering **2**(4), 225–243 (2025)

Appendix A Model comparison summary

Tables A1, A2, A3 show the resultant AIC scores for all the spatial models considered and mentioned in section 4.1 and 4.2 as well as the temporal models mentioned in 4.3.

Table A1: The AIC and log-likelihood values for the power-utility ignition spatial models.

Model and Covariates	AIC	Loglikelihood
Homogeneous	6866.809	-3432.404
Voltage	6851.13	-3423.56
Canopy stand height	6857.21	-3426.61
Canopy base height	6867.16	-3431.58
Fuel type	6870.28	-3428.14
Canopy bulk density	6864.33	-3430.17
Canopy cover	6865.67	-3430.84
Voltage + canopy stand height	6842.97	-3418.484
Voltage + canopy bulk density	6849.90	-3421.95
Voltage + canopy stand height + bulk density	6843.36	-3417.678
Canopy stand height + canopy bulk density	6856.83	-3425.412

Table A2: The AIC and log-likelihood values for the non-utility ignition spatial models. 'RDxx' indicates road distance threshold as a predictor variable, with "xx' indicating the threshold distance in meters.

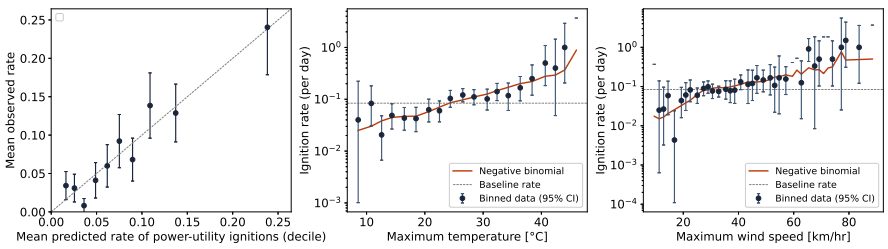
Model and Covariates	AIC	Loglikelihood
Homogeneous	82877.1	-41437.55
Canopy stand height	82588.2	-41292.1
Canopy base height	82842.2	-41419.1
Canopy cover	82411.7	-41203.8
Canopy bulk density	82808.4	-41402.2
Fuel	79785.9	-39886.0
Fuel + canopy stand height	79785.8	-39884.88
Fuel + canopy bulk density	79774.4	-39879.19
Fuel + RD2000	79473.3	-39728.63
Fuel + RD1000	79347.8	-39665.88
Fuel + RD500	79312.2	-39648.09
Fuel + RD250	79296.9	-39640.44
Fuel + RD125	79216.3	-39600.12
Fuel + RD60	79168.3	-39576.15
Fuel + RD30	79223.1	-39603.53
Fuel + RD15	79252.1	-39618.03

Table A3: The AIC and log-likelihood values for the temporal models tested for power-utility and non-utility ignitions.

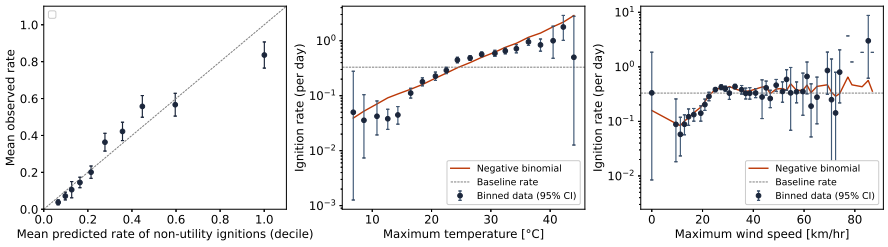
Model	AIC	Loglikelihood
Power-utility, Poisson log-linear	2037.1	-1015.5
Power-utility, Negative binomial	2007.5	-999.7
Non-utility, Poisson log-linear	11333.2	-5663.6
Non-utility, Negative binomial	11097.8	-5544.9

Appendix B Negative binomial fitting check

The fitted Negative binomial model was further assessed by comparing it with the observed data. Here, the rate of ignition per day was plotted as a function of maximum temperature and wind speed. As shown in figures B1a and B1b, the estimated rate of ignitions closely aligns with the observed rate of ignitions for both ignition sources. The ignition rate trend from maximum temperature and wind speed is also shown in this figure, where the negative binomial can closely capture the increase in the rate as these weather covariates increase. The middle and right plots indicate the model's average prediction for all days in each temperature/wind speed bin, showing close predictions for both covariates.



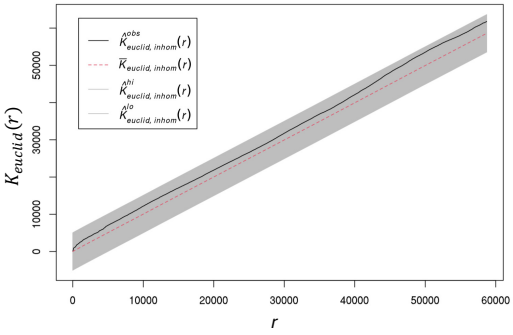
(a) Power-utility ignitions.



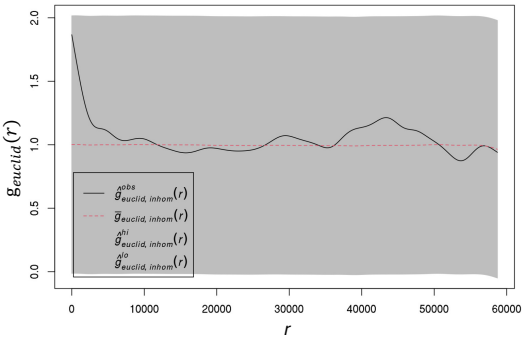
(b) Non-utility ignitions.

Fig. B1: The fitted Negative Binomial rate is compared with the observed rate as a function of maximum temperature and wind speed.

Appendix C Inhomogeneous Poisson point process assumption test

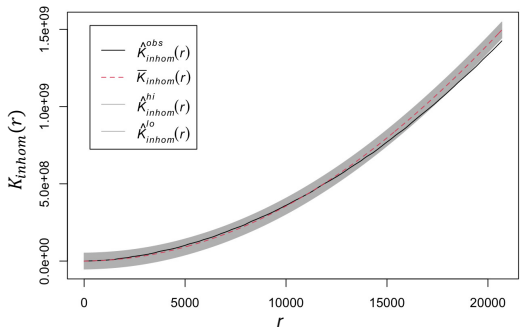


(a) The inhomogeneous linear K-function based on Euclidean distances is shown for the fitted spatial utility ignition model, where r is in meters .

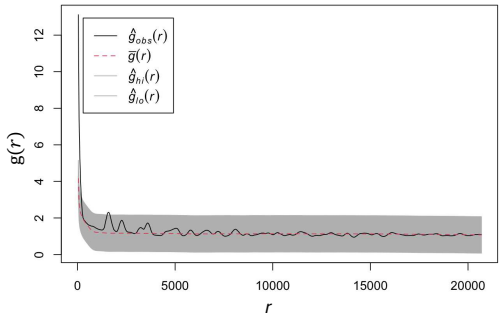


(b) The inhomogeneous pair correlation function based on Euclidean distances is shown for the fitted spatial utility ignition model, where r is in meters.

Fig. C2: The inhomogeneous Poisson assumption of the fitted utility ignition point-process model is assessed using the inhomogeneous linear K-function and pair correlation function. Simulation envelopes are constructed from 199 realizations of the fitted model.



(a) The inhomogeneous K-function is shown for the fitted spatial non-utility ignition model, where r is in meters.



(b) The inhomogeneous pair correlation function is shown for the fitted spatial non-utility ignition model where r is in meters.

Fig. C3: The inhomogeneous Poisson assumption of the fitted non-utility ignition point-process model is assessed using the inhomogeneous K-function and pair correlation function. Simulation envelopes are constructed from 199 realizations of the fitted model.