The Annals of Applied Statistics
2020, Vol. 14, No. 1, 339–356
https://doi.org/10.1214/19-AOAS1309

MODELING WILDFIRE IGNITION ORIGINS IN SOUTHERN CALIFORNIA USING LINEAR NETWORK POINT PROCESSES

BY MEDHA UPPALA* AND MARK S. HANDCOCK†

*Department of Statistics, University of California Los Angeles, *umedha@ucla.edu; †handcock@ucla.edu*

This paper focuses on spatial and temporal modeling of point processes on linear networks. Point processes on linear networks can simply be defined as point events occurring on or near line segment network structures embedded in a certain space. A separable modeling framework is introduced that posits separate *formation* and *dissolution* models of point processes on linear networks over time. While the model was inspired by spider web building activity in brick mortar lines, the focus is on modeling wildfire ignition origins near road networks over a span of 14 years. As most wildfires in California have human-related origins, modeling the origin locations with respect to the road network provides insight into how human, vehicular and structural densities affect ignition occurrence. Model results show that roads that traverse different types of regions such as residential, interface and wildland regions have higher ignition intensities compared to roads that only exist in each of the mentioned region types.

1. Introduction. One of the most pressing, large-scale disasters affecting the State of California and other parts of the country are wildfires. California is specifically prone to high-intensity fires due to its dry summers and vegetation. Southern California also faces high-speed winds, called the Santa Ana winds, that originate inland and bring hot, dry air with them which further create critical fire conditions. The California Wildland Fire Coordinating Group reports that "approximately 95% of all wildfires in California have human-related origins." Balch et al. (2017) analyze 1.5 million government records of wildfires between 1992 to 2012 and report that about 84% of all U.S. wildfires have human-related origins; humans are also extending the length of fire seasons. Human-related origins of wildfires include equipment use, vehicular fire, campfire, debris burning, smoking, children and arson.

Wherever a human-related wildfire originates, an individual needed to have access to the area, most likely through a network of roads, paths and trails. Moreover, there has been a rapid expansion of the wildland-urban interface (WUI) regions in the country, increasing the chances of harmful human activity near fuel-rich areas (fuel rich indicates low-moisture vegetation). Given the high frequency of human-related ignitions and growing WUI regions in the country, it is useful to understand how human density and movement affect the spatial and temporal distributions of wildfire ignitions. The separable temporal linear point process model (STLPPM) introduced in this paper is ideal for modeling wildfire ignition locations with respect to road and trail networks. It would be immensely useful if we could identify wildfire ignition hot spots that are highly correlated with human and vehicular traffic.

Following the precedence of Syphard and Keeley (2015), the Santa Monica Mountains region (Los Angeles County, California) is used as the study area in this paper, and the wildfire ignition locations between the years 2000 and 2013 are shown on the road network plot in Figure 1.

Received April 2019; revised October 2019.
Key words and phrases. Point processes, linear network, spatiotemporal modeling, pseudolikelihood, Berman–Turner method, spider webs, wildfires, ignition origins, road networks.

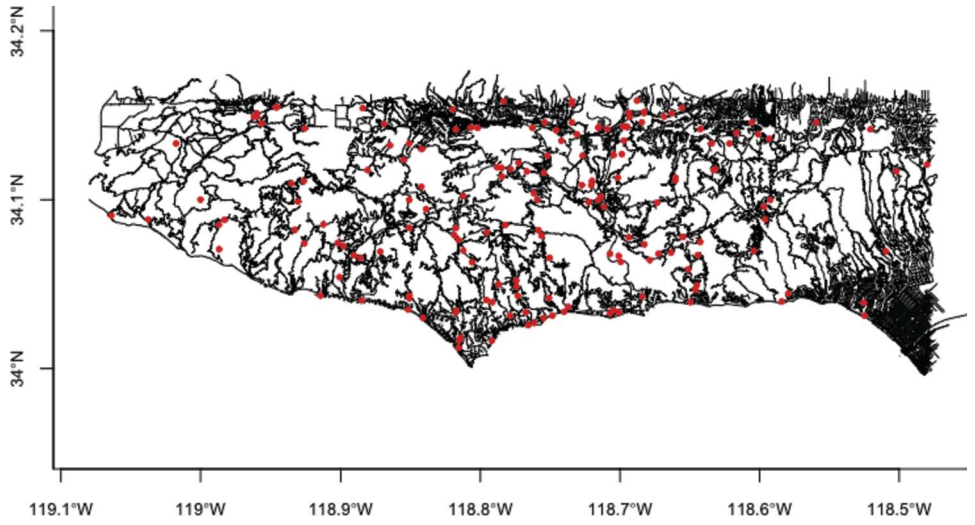

FIG. 1. *Locations of* 2000–2013 *wildfire origins in the Santa Monica Mountains region.*

2. Background. While spatial point process theory is well established, spatial analysis of events along networks is not common. A large portion of this analysis has been spearheaded by Okabe and collaborators, with applications to boutique openings along Tokyo streets and acacia tree populations along road networks (Okabe and Okunuki (2001), Spooner et al. (2004)). The only textbook that addresses this area, by Okabe and Sugihara (2012), introduces multiple examples and expands on computational methods appropriate to network processes.

Current work extends the point process theory to network processes with applications to street crimes in Chicago and spine formation on dendrite networks of neurons (Ang, Baddeley and Nair (2012), Baddeley, Jammalamadaka and Nair (2014)). Ang, Baddeley and Nair (2012) expand on the (empirical) "network K function," introduced by Okabe and Yamada (2001), to incorporate network generality resulting in a more interpretable network K function. The *geometrically corrected* empirical K function weights the shortest path distance between a pair of points (x_i, x_j), with the reciprocal of the number of points on the network situated at the same distance from x_i as x_j is. The corrected network K function is

$$\hat{K}_L(r) = \frac{|L|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1\{d_L(x_i, x_j) \leq r\}}{m(x_i, d_L(x_i, x_j))}, \tag{2.1}$$

where n is the total number of points in the process, $|L|$ is the total length of the network, $d_L(x_i, x_j)$ denotes the shortest path distance and $m(x_i, d_L(x_i, x_j))$ is the number of points on L lying at the exact distance $d_L(x_i, x_j)$ away from point x_i. This reciprocal weighting is similar to Ripley's isotropic edge correction, as it considers the number of neighbors on the ball of radius $d_L(x_i, x_j)$ centered on x_i. These results are further extended to an "inhomogeneous" network K function and the pair correlation function. Ang, Baddeley and Nair (2012) fit the network K function and the inhomogeneous version to a single quadrant of spider data (see Appendix A for details); both reveal only insignificant clustering at larger scales, attributable to spatial inhomogeneity.

Ang, Baddeley and Nair (2012) also apply the methods to street crime data from an area close to the University of Chicago. The inhomogeneous network K function showed no evidence of clustering; however, the authors believe there is confounding of inhomogeneity and clustering at short distances and suggest further study.

Baddeley, Jammalamadaka and Nair (2014) apply the corrected network K function to spines on dendrite networks of neurons resulting in an extension to multitype points. The data used in the paper is of an observed spine pattern on dendrite network of a rat neuron. At a specific resolution, the spines can be categorized into three types based on their size. It is necessary to study the spatial distribution of spines to understand normal function and disease processes. Since the spines occur only on a dendrite network, event analysis along linear networks is a natural choice. Fitting the inhomogeneous network K function did not reveal significant clustering, but the authors present numerous caveats to interpreting these results in a biological context.

3. Motivation. Thus far, the existing literature presents computational and descriptive methods complementary to the point process literature adapted to linear networks. There has been little work that consciously models the spatial and temporal progression of point processes on network structures. Given the advent of this new data structure, the methodology presented in this paper is an effort to adapt and to integrate spatial analysis and network structure and to produce a compounded modeling framework. The motivation is to harness insights from spatial geometry imposed by network structures to better understand the distribution of event occurrence.

The original motivation for this model was derived from the spider-web data (outlined in Appendix A) that provides the locations of webs made by the small *Oecobius navus* in the crevices of mortar lines on brick walls. The indented mortar lines provide the perfect structural support for web building, while the overall network of lines delineate and guide web formation over the whole wall. Initially pursued as a curious application, the spider-web data was used as a guide to build a spatial and temporal model of point processes on linear networks. The idea was to produce a modeling framework that incorporated the brick mortar line's spatial influence on the web building activity. This dataset's influence is more apparent in the presentation of the separable temporal framework in Section 5. Besides this spider data, many processes that are analogous to this data structure include events along stream networks, events along Interstate high-tension wires, water and gas pipeline networks, etc.

The two applications that have caught attention that can largely benefit from such modeling, with extensions to prediction, are street crimes and human related wildfires originating near road and trail networks. Mapping and predicting "hot spots" for street crimes and human related wildfire origins could greatly aid police and firefighter efforts on ground. Crimes in cities follow a categorical, spatial pattern depending on the type of target and crime; the distribution of residential and commercial areas naturally influence crime occurrence, and spatial temporal modeling of crimes over city road networks is the logical next step (Hering and Bair (2014)).

In the case of wildfires, the vast majority of California wildfires have human related origins; as most wildfire origins are in nonurban settings, we want to map and understand their interaction with the extensive road and trail network that exists in most park and forest areas of California. With the right and sufficient covariate information, mapping wildfire origins with respect to the road network could reveal high-intensity network regions that require greater fire management. The various types of roads also act as the perfect proxy for human, vehicular and structural density in the surrounding region; this provides information on road accessibility and fuel availability that further affect fire spread. The methodology presented below is a step toward achieving greater insight into such areas that require a rigorous amalgamation of both point process and spatial network theories.

The next section defines point processes on linear networks, and the following section outlines the separable temporal model and inference. Section 6 presents the model's application to the wildfire ignition locations, and Section 7 covers model assessment through comparison to a simpler Euclidean model and residual analysis.

4. Point processes on linear networks. A *point process* is a set of unordered points $\mathbf{x} = \{x_1, \ldots, x_i, \ldots, x_n\}$, such that $n \geq 0$ and $x_i \in W$ where W is a spatial window in d-dimensional space $\mathcal{R}^d$, $d \geq 1$. The window W is assumed to have finite positive volume $|W|$.

A *line segment* in a plane with endpoints u and v is written as $[u,v] = \{tu + (1-t)v : 0 \leq t \leq 1\}$. A *linear network* L is defined as a union $L = \bigcup_{i=1}^{n} \ell_i$ of a finite collection of line segments $\ell_1, \ldots, \ell_n$ in the plane. The total length of all the line segments in L is denoted by $|L|$. Another way to represent a *linear network* L is by a set of vertices $v_1, \ldots, v_n$ and edges $e_{ij} = [v_i, v_j]$, such that the intersection of two different edges occurs at only one point.

A *point process on a linear network* is a point process that occurs *on* or *near* a linear network, generally in a two-dimensional space. The linear network L occurs within a specific window W. The occurrence of the point process on/near the linear network could be due to the structural limitations of the point generating process, such as road accidents on a street network, or due to a contextual relationship between the point process and the environment in which it occurs. For instance, the occurrence of human settlements near and around river networks demonstrates the synergic interaction between a point process and a preexisting linear network structure. While the majority of the examples illustrated in the present paper are physical and social events occurring on a plane, the occurrence of such data structures is not limited to these fields. The occurrence of spines on a neuron dendrite network is one such case.

4.1. *Papangelou conditional intensity.* The first established step in understanding a point process (on a linear network or otherwise) is estimating an intensity function (first moment measure) for the points. The intensity function $\lambda(x)$ is proportional to the probability of a point occurring at infinitesimal space around x; otherwise, it is the expected infinitesimal rate of a point in the observed window W. For a marked or unmarked point process, the ideal case is to estimate an inhomogeneous intensity function $\lambda_\theta(u)$ to capture the spatial trend. Ideally, this can be done through formulating a likelihood function and estimating the parameters through maximum likelihood. However, formulating and evaluating a likelihood function comprising individual intensities of the points $P(\mathbf{X}|\Theta)$, where $\mathbf{X} = \{x_1, \ldots, x_n\}$, is highly intractable. A convenient approximation to this is the pseudolikelihood function that is a product of conditional intensity at each point.

The first step in formulating the pseudolikehood function is to estimate the (Papangelou) conditional intensity at every location $u \in \mathbf{x} \cap W$ as

$$\lambda(u; \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x} \setminus \{u\})}. \tag{4.1}$$

The Papangelou conditional intensity (Papangelou (1973/74)), loosely speaking, is the conditional intensity of observing a point at u, given the rest of the process $\mathbf{x}$ in the window W.

In this paper, we assume a general Gibbs point process model with the Papangelou conditional intensity taking a log linear form (Baddeley and Turner (2000)). Using the covariate information on point locations, one way to model the Papangelou conditional intensity is

$$\lambda_\theta(u; \mathbf{x}) = \exp\left(\theta^T S(u; \mathbf{x})\right), \tag{4.2}$$

where $S(u; \mathbf{x})$ is a matrix of spatial and network covariates defined at each point $u \in \mathbf{x} \cap W$, while $\theta \in \Theta$ is the vector of parameters to be estimated.

5. Separable temporal model on linear networks. While the Papangelou conditional intensity function is informative, it does not capture the temporal progression and death of the point process. To understand and capture the various factors that affect the evolution of a point process on a linear network, we introduce the separable temporal model of point processes on linear networks. The goal of this model is to capture the dynamic action of new

and existing points entering and exiting the process. As in the case of the spider web building data, such modeling is apt, as over time, new webs are built while existing webs persist and disintegrate over time. The primary concept is to estimate separable (conditionally independent) *formation* and *dissolution* intensity functions that capture the change in the point process over a single time step. Conditioned on time step t, the product of the *formation* and *dissolution* models estimates the point process at time step $t+1$. The idea is that, conditioned on the previous time step, the point *formation* and *dissolution* models are independent and might have different factors influencing them. The *formation* intensity captures the rate of new point events forming after each time step, and the *dissolution* model captures the rate of persistence of an existing point into the next time step. The model can be extended to be conditioned on all the previous time steps, allowing for generalized *formation* and *dissolution* models that could be used for prediction.

The impetus for this separable temporal model was gained from the separable temporal exponential random graph model (STERGM) framework for dynamic networks developed by Krivitsky and Handcock (2014).

5.1. *Model specification.* Given an observed point process on a linear network $\mathbf{x}^t$ at time step t, the *formation point process* $\mathbf{X}^+$ includes the new points along with the existing points at t, and the *dissolution point process* $\mathbf{X}^-$ includes only the points that persisted on from time t. The realized counterparts of these processes are $\mathbf{x}^+$ and $\mathbf{x}^-$. Given $\mathbf{x}^+$, $\mathbf{x}^-$ and $\mathbf{x}^t$, the point process $\mathbf{x}^{t+1}$ can be evaluated as follows:

$$\mathbf{x}^{t+1} = \mathbf{x}^- \cup (\mathbf{x}^+ \backslash \mathbf{x}^t),$$

where $\mathbf{x}^+ = \mathbf{x}^t \cup \mathbf{x}^{t+1}$ and $\mathbf{x}^- = \mathbf{x}^t \cap \mathbf{x}^{t+1}$. As $\mathbf{X}^+$ is conditionally independent of $\mathbf{X}^-$ given $\mathbf{X}^t$, then

$$P(\mathbf{X}^{t+1} = \mathbf{x}^{t+1} | \mathbf{X}^t = \mathbf{x}^t; \boldsymbol{\theta}) = P(\mathbf{X}^+ = \mathbf{x}^+ | \mathbf{X}^t = \mathbf{x}^t; \boldsymbol{\theta}) P(\mathbf{X}^- = \mathbf{x}^- | \mathbf{X}^t = \mathbf{x}^t; \boldsymbol{\theta}).$$

Here, *separability* is defined as $\mathbf{X}^+$ being conditionally independent of $\mathbf{X}^-$ given $\mathbf{X}^t$, such that the parameter space $\boldsymbol{\theta}$ is a product of the individual parameter spaces $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$. This indicates that the point formation and the dissolution processes do not interact, once the point process at the beginning of the time step is observed.

Continuing from (4.2), we use a log-linear model to estimate the *Markovian* Papangelou conditional intensity for the *formation* and *dissolution* models. That is, we are estimating a Papangelou conditional intensity of a process ($\mathbf{x}^+$ and $\mathbf{x}^-$) that is also conditioned on the previous time step ($\mathbf{x}^t$). As a result, the point *formation* model is

$$P(\mathbf{x}^+ | \mathbf{x}^t) = \prod_{u \in W} \lambda(u; \mathbf{x}^+) = \prod_{u \in W} \exp\{(\boldsymbol{\theta}^+)^T S^+(u; \mathbf{x}^+, \mathbf{x}^t)\},$$

and the point *dissolution* model is

$$P(\mathbf{x}^- | \mathbf{x}^t) = \prod_{u \in W} \lambda(u; \mathbf{x}^-) = \prod_{u \in W} \exp\{(\boldsymbol{\theta}^-)^T S^-(u; \mathbf{x}^-, \mathbf{x}^t)\},$$

where u indicates every point on the whole window W on which the linear point process was observed. $S^+(u; \mathbf{x}^+, \mathbf{x}^t)$, and $S^-(u; \mathbf{x}^-, \mathbf{x}^t)$ represent the sufficient statistics, based on $\mathbf{X}^t = \mathbf{x}^t$, which attempt to capture the formation and tenacity of point events over time.

5.2. *Pseudolikelihood inference.* The conditional intensity of the separable model for a single time step can be written as follows:

$$\begin{aligned}\lambda_\theta(\mathbf{x}^{t+1}|\mathbf{x}^t) &= \prod_{u\in W}\lambda(u;\mathbf{x}^+)\prod_{u\in W}\lambda(u;\mathbf{x}^-)\\ &= \prod_{u\in W}\exp\{(\boldsymbol{\theta}^+)^T S^+(u;\mathbf{x}^+,\mathbf{x}^t)\}\prod_{u\in W}\exp\{(\boldsymbol{\theta}^-)^T S^-(u;\mathbf{x}^-,\mathbf{x}^t)\}\\ &= \exp\Big\{\sum_{u\in W}((\boldsymbol{\theta}^+)^T S^+(u;\mathbf{x}^+,\mathbf{x}^t))+\sum_{u\in W}((\boldsymbol{\theta}^-)^T S^-(u;\mathbf{x}^-,\mathbf{x}^t))\Big\}.\end{aligned}$$

Given the conditional intensity at a single time step, the pseudolikelihood function of the separable temporal linear point process model at a single time step (Besag and Diggle (1977)) is

$$\mathrm{PL}(\boldsymbol{\theta}|\mathbf{x}^{t+1}) = \Big(\prod_{x_i\in W}\lambda_\theta(\mathbf{x}^{t+1}|\mathbf{x}^t)\Big)\exp\Big(-\int_W\lambda_\theta(\mathbf{x}^{t+1}|\mathbf{x}^t)\,du\Big).$$

Assuming a first order Markov property, the pseudolikelihood function of the complete separable temporal linear point process model, over all the k time steps is

$$\mathrm{PL}(\Theta|\mathbf{x}_1,\ldots,\mathbf{x}_k) = \prod_{t=1}^{k-1}\mathrm{PL}(\boldsymbol{\theta}|\mathbf{x}^{t+1}), \tag{5.1}$$

and the corresponding log pseudolikehood function is

$$\begin{aligned}\log\mathrm{PL}(\Theta|\mathbf{x}_1,\ldots,\mathbf{x}_k) &= \sum_{t=1}^{k-1}\log\mathrm{PL}(\boldsymbol{\theta}|\mathbf{x}^{t+1})\\ &= \sum_{t=1}^{k-1}\Big(\sum_{i=1}^{n}\log\lambda_\theta(\mathbf{x}_i^{t+1}|\mathbf{x}^t)-\int_W\lambda_\theta(\mathbf{x}^{t+1}|\mathbf{x}^t)\,du\Big).\end{aligned} \tag{5.2}$$

The numerical solution to the above expression needs iterative algorithms. Instead, we use the estimation method used by Baddeley and Turner (2000) in R's *spatstat* package; it is an adaptation of a technique introduced by Berman and Turner (1992). The Berman–Turner method, with the help of a quadrature rule, discretizes the integral calculation to a weighted poisson likelihood. We approximate the integral in (5.2) using the following quadrature rule:

$$\int_W\lambda_\theta(u|\mathbf{x}^t,\boldsymbol{\theta})\,du \approx \sum_{j=1}^{m}\lambda_\theta(u_j|\mathbf{x}^t,\boldsymbol{\theta})w_j,$$

where u_j, $j=1,\ldots,m$, are quadrature points in W and $w_j>0$ are quadrature weights summing to $|W|$. This produces an approximation to the log-pseudolikelihood in (5.2) as

$$\log\mathrm{PL}(\Theta|\mathbf{x}_1,\ldots,\mathbf{x}_k) \approx \sum_{t=1}^{k-1}\Big(\sum_{i=1}^{n}\log\lambda_\theta(\mathbf{x}_i^{t+1}|\mathbf{x}^t)-\sum_{j=1}^{m}\lambda_\theta(u_j|\mathbf{x}^t,\boldsymbol{\theta})w_j\Big). \tag{5.3}$$

Here, consider that the list of quadrature points $\{u_j, j=1,\ldots,m\}$ subsume all the observed data points $\{x_i, i=1,\ldots,n\}$. Then, (5.3) can be rewritten as

$$\log\mathrm{PL}(\Theta|\mathbf{x}_1,\ldots,\mathbf{x}_k) \approx \sum_{t=1}^{k-1}\Big(\sum_{j=1}^{m}(y_j\log\lambda_j-\lambda_j)w_j\Big), \tag{5.4}$$

where $\lambda_j = \lambda_\theta(u_j)$ and $y_j = z_j/w_j$, and

$$z_j = \begin{cases} 1 & \text{if } u_j \text{ is a data point, } u_j \in \mathbf{x}, \\ 0 & \text{if } u_j \text{ is a dummy point, } u_j \notin \mathbf{x}. \end{cases}$$

The expression in (5.4) is similar to the log likelihood of independent poisson variables Y_k with means λ_k taken with weights w_k. As a result, a GLM framework software that is capable of handling weighted likelihoods and noninteger response values was used to fit the seperable temporal linear point process model (STLPPM) framework. However, as the GLM setup assumes independent poisson variables, which is not the case here, the standard errors produced by the software are invalid. As an alternative, moving block bootstrap (MBB) standard errors are presented. For more details on MBB standard errors, see Appendix B.

6. Application to wildfire ignitions.

6.1. *Current literature on wildfire modeling*. Most work in wildfire literature focusses on modeling burn area *centroidal* locations and not wildfire *ignition* origin locations. However, since the majority of California wildfires have human-related origins, it is important to understand the spatial and temporal distribution of human-related wildfire ignitions. Besides Balch et al. (2017), Syphard and Keeley (2015) is one of the few papers that focusses on modeling wildfire ignition locations. The paper aims to identify whether different ignition sources cause distinct spatial or intraannual temporal patterns of wildfires. The authors use ignition data from 2006 to 2010 and focus on two subregions of Southern California (Santa Monica Mountains and southwestern region of San Diego County). They model each ignition type separately with a maximum entropy model and conclude that, while equipment fires were the most frequent, arson and power line ignitions caused the most area burned in the Santa Monica Mountains region. They also identify that the distance to roads and structures predominantly affects ignition probability.

While our goal is similar to Syphard and Keeley (2015), we do not model ignition types separately nor do we use the ignition type as a point covariate in our model. This is because we want to map and model original ignition locations with respect to the road network and capture if and to what degree different road types affect ignition intensities. The idea is that different road types act as proxy for the amount of human and structural density in that region. Our goal is to fit a *formation* model to 13 years of wildfire ignition location data, so as to understand the spatial and temporal distributions of high ignition intensity roads in the Santa Monica Mountains region.

6.2. *Data*: *Road network*. To aid comparison with Syphard and Keeley (2015), the Santa Monica Mountains region is chosen to model the wildfire origins near the road network. The linear network was extracted using the "osmar" R package (Schlesinger and Eugster (2010)) from OpenStreetMap[1] (https://www.openstreetmap.org). The different road types have been consolidated into five broad categories depending on the vehicular and foot activity, and whether they are commercial or residential.[1] These five road types can be seen in Figure 2 and their relative proportions in Table 1. Road1 represents the major motorways, trunk, primary and their link roads. Road2 represents the secondary, tertiary, service and link roads. Paths represents cycleways, footways, paths, bridleways, tracks and pedestrian paths. Residential and path road types make up most of the linear network at 39% and 30%, respectively.

[1]It is assumed that the road network over the Santa Monica Mountains region has remained the same over the 13 years of the model. It is also assumed that the road classification in this region has not drastically changed for the broad categories used in the present paper to be affected.

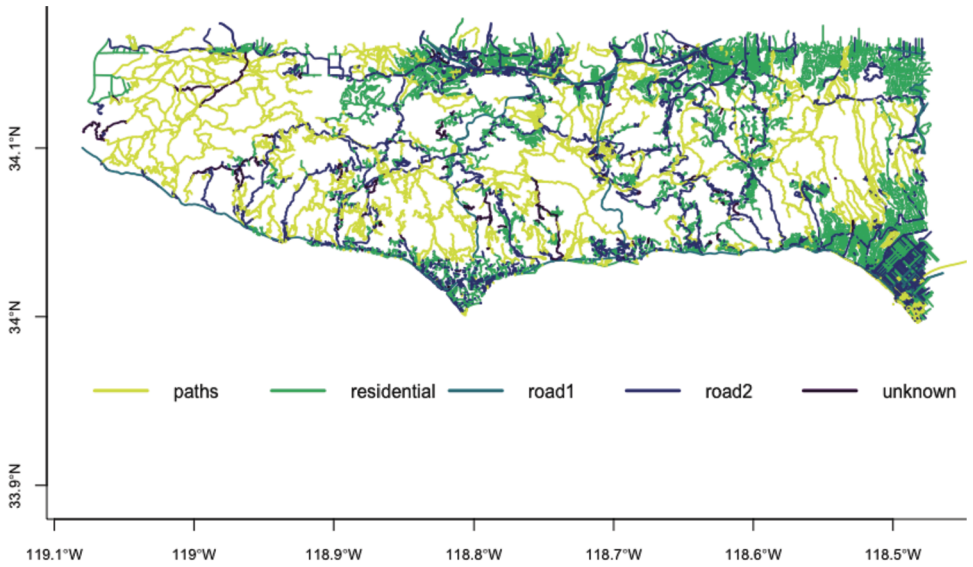

FIG. 2. *Road type in the Santa Monica Mountains region.*

6.3. *Data*: *Wildfire ignition data*. The wildfire ignition data is sourced from Short (2015), and subsetted to the Santa Monica Mountains region. This resulted in 175 unique wildfires in the region between 2000 and 2013. The database provides discovery date, final fire size, ignition cause and an origin point location. The ignition cause breakdown for the 175 fires is shown in Table 2. Despite the availability, the ignition cause variable is *not* used in model fitting as the main goal of the linear networks methodology is to understand and measure the effect of human activity (including vehicular and structural density) on where wildfire ignitions are occurring.

As the methodology requires the point events to occur *on* the linear network, the 175 wildfire origin locations are projected to the nearest network road and are shown in Figure 4. The mean projected distance for the 175 wildfires is 66 meters and 150 (85%) of the 175 wildfires are within 100 meters to the nearest road on the network. All projection lengths for the 175 wildfires are under a mile in length, and a histogram of the distances can be seen in Figure 3.

6.4. *Data*: *RAWS data*, *elevation and wildland-urban interface*. Following the lead of Xu and Schoenberg (2011), we extract daily meteorological variables from the nearest four remote automated weather stations (RAWS shown in Figure 4 in blue) around the region to use as covariates in the model. Daily values of total precipitation, average humidity, average temperature and average wind speed from 2000 to 2012 were extracted and spatially and

TABLE 1
Breakdown of road types in the Santa Monica Mountains region

Road type	% of Road type by length in the network
Paths	29.73
Residential	38.69
Road1	5.42
Road2	23.89
Unknown	2.24

TABLE 2
Santa Monica Mountains 2000–2013 *wildfire ignition cause breakdown*

Ignition Cause	# of Fires between 2000–2013	Acres burned
Arson	10	26.51
Campfire	4	4916.2
Children	4	42.21
Debris Burning	4	13.3
Equipment Use	36	109.66
Fireworks	3	3.1
Lightning	1	0
Miscellaneous	23	134.91
Missing/Undefined	69	4732.33
Powerline	12	904.9
Smoking	8	1.81
Structure	1	2

temporally smoothed using generalized additive models (Hastie and Tibshirani (1990)) over the linear network window. Elevation (from USGS' https://nationalmap.gov) and change in elevation are additional covariates included in the model as they affect the presence of roads, infrastructure and human activity.

Finally, the road network was tagged by wildland-urban interface (WUI) categories sourced from Radeloff et al. (2017).[2] A WUI region is an area where houses meet or intermingle with wildland vegetation, and these regions provide large amounts of fire fuel. The WUI categories used to tag the road network are "Interface, Intermix" and "Non-WUI." It is important to note that the non-WUI category refers to both structurally dense and vegetation dense regions as they are neither interface or intermix regions. So an urban region and a forest land would both be categorized as non-WUI. The WUI regions over Santa Monica Mountains region can be seen at silvis.forest.wisc.edu/data/wui_change/.

6.5. *Results*: *Formation model of wildfire ignition origins*. The 14 years of wildfire ignition data result in 13 *formation* point processes that form the basis of the *formation* model.

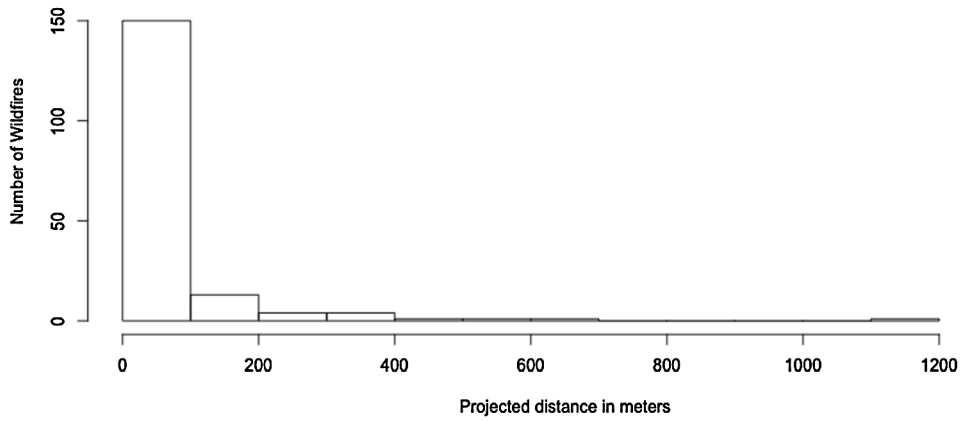

FIG. 3. *Projection distance in meters of the* 2000–2013 *wildfire origin locations to the nearest road.*

[2]We have looked closely at the change of WUI from 2000 to 2010 and given the minute differences in the region, decided to use to 2010 WUI data for all the years of the formation model in the paper.

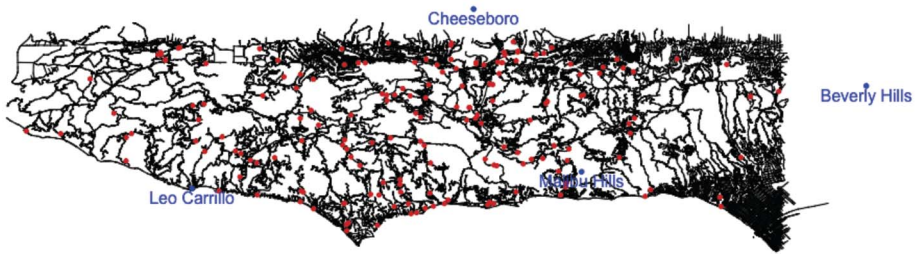

FIG. 4. *Remote automated weather stations in blue in the Santa Monica Mountains region and the projected locations of* 2000–2013 *wildfires in red.*

For example, the *formation point process* between the years 2000 and 2001 is:

$$\mathbf{x}^{+}_{2000-01} = \mathbf{x}_{2000} \cup \mathbf{x}_{2001}. \tag{6.1}$$

Each *formation point process* is assigned with dummy points that are placed equidistantly on the linear network and weighted accordingly. The data points (wildfire origin locations) are also assigned weights based on the length of the network segment on which they lie. Through the Berman–Turner device (Berman and Turner (1992)) introduced in Section 5.2, the weighted data and dummy points along with their corresponding spatial covariate values (meteorological and network) produce the following estimate to the wildfire ignition *formation* model (conditional intensity of ignition occurrence), seen in Table 3.

TABLE 3
Formation intensity of wildfire ignitions on the Santa Monica Mountains region. Roadtype-Residential, August and Intermix are the reference levels

Parameter	Coefficient Estimate	MBB SE	MBB 95% CI
(Intercept)	0.127	2.407	(−4.535, 3.565)
Roadtype-Paths	0.410	0.193	(0.290, 0.813)
Roadtype-Road1	2.013	0.080	(2.001, 2.233)
Roadtype-Road2	0.808	0.228	(0.612, 1.291)
Roadtype-Unknown	1.244	0.284	(0.694, 1.564)
January	−1.005	0.636	(−2.331, 0.047)
February	−1.736	0.640	(−3.518, −0.924)
March	−1.072	0.637	(−2.802, −0.805)
April	−1.158	0.518	(−2.213, −0.344)
May	−0.470	0.443	(−1.387, 0.205)
June	−0.074	0.422	(−0.952, 0.394)
July	0.298	0.100	(−0.069, 0.296)
September	−0.178	0.268	(−0.969, 0.075)
October	−0.441	0.326	(−1.255 − 0.162)
November	−0.671	0.539	(−1.579, 0.368)
December	−1.845	0.479	(−2.473, −0.460)
Total Precipitation	−0.143	0.085	(−0.305, −0.038)
Ave. Temperature	−0.011	0.068	(−0.128, 0.113)
Ave. Humidity	−0.011	0.004	(−0.012, 0.003)
Ave. Wind Speed	−0.016	0.230	(−0.470, 0.220)
Elevation	−0.0002	0.001	(−0.002, 0.002)
Slope	0.009	0.002	(−0.001, 0.007)
Interface	0.861	0.182	(0.726, 1.285)
Non-WUI	0.282	0.200	(0.164, 0.749)

6.6. *Discussion*. The estimates for all roadtypes, compared to the residential road type, indicate higher ignition intensities than the residential roads in the network. This is apparent as most human-related wildfires would not spread to become wildfires if originating in high density residential areas. "Paths, Road1" and "Road2" categories are all likely to increase wildfire ignition intensity by at least a factor of 1.5 when compared to residential roads. "Road1," which represents that high traffic highways and motorways, has the largest effect on ignition intensity at a factor of 7.5. This is possibly due to drier fuels along the high-traffic highways. The "Unknown" road type also has a relatively high effect on ignition intensity. As they indicate the associated human activity and structural density in that area, the road-type categories are crucial to understanding wildfire origin distribution.

If we reconsider the roadtype map in Figure 2, "Paths" are generally in areas of higher elevation and mostly experience foot traffic in this region. Residential roads frame a large part of the region and generally lead into highly urbanized areas, reducing the likelihood of wildfire ignitions. "Road1" and "Road2" types generally traverse the whole window, covering urban and interface/intermix regions in the window. This indicates that roads that see a varying levels of human activity (vehicular and otherwise) and structures, like "Road1" and "Road2," have higher ignition intensities than the mostly foot trafficked paths and densely residential areas.

The monthly effects of ignition intensity vary over the year, with August as the reference month. August is a relatively dry month in Southern California that also receives the dry, high-speed Santa Ana winds from the east, making it a high fire hazard month. Compared to August, almost every other month has a lower ignition intensity at varying degrees. December has the lowest rate of ignition intensity while July has the highest rate of ignition intensity over the year. Overall, the late summer and early fall months, June to September, indicate higher ignition intensities. Despite the monthly effect, there has been a sustained expansion in the length of fire seasons due to increase in human-related causes (Balch et al. (2017)). While this expansion in the fire season is not captured due to the time inhomogeneity of the model, the inclusion of the road network as a human proxy does account for the consistent human involvement in wildfire ignition intensity over the 14 year time period.

The meteorological variables, elevation and slope have modest effects on wildfire ignition intensity. While the main goal of the separable temporal linear point process (STLPP) model is to understand human involvement and contribution to ignition occurrence, the meteorological variables account for the highly fire prone climate of Southern California. The dry summer and fall seasons, dense highly-flammable chaparral vegetation and seasonal Santa Ana winds all combine to produce high fire hazard conditions (Xu and Schoenberg (2011)) in this region.

Finally, the wildland-urban interface regions are 2.3 times more likely to have higher ignition intensities than the intermix regions. The non-WUI regions are also more likely than intermix regions, to experience higher ignition intensities. However, as previously mentioned, non-WUI regions include densely urbanized *and* purely wildland regions. The next step would be to categorize these regions separately, to isolate intensities on wildland regions. Once again, WUI regions have been systematically growing in the United States, increasing fire risk. This has mostly been due to increase in housing in these regions in the past three decades (Mann et al. (2014), Radeloff et al. (2018)).

While the results in Table 3 only provide indications on how network and weather covariates affect ignition intensity, the great benefit of the STLPP model is that we can study the spatial distribution of the estimated ignition intensities on the road network. Figure 5 shows the conditional ignition intensities for the last three *formation* years over the Santa Monica Mountains region.

The plotted intensities fall into five different quantiles with the lowest 20% quantile represented by the darkest color and the top 20% quantile represented by the lightest color; the

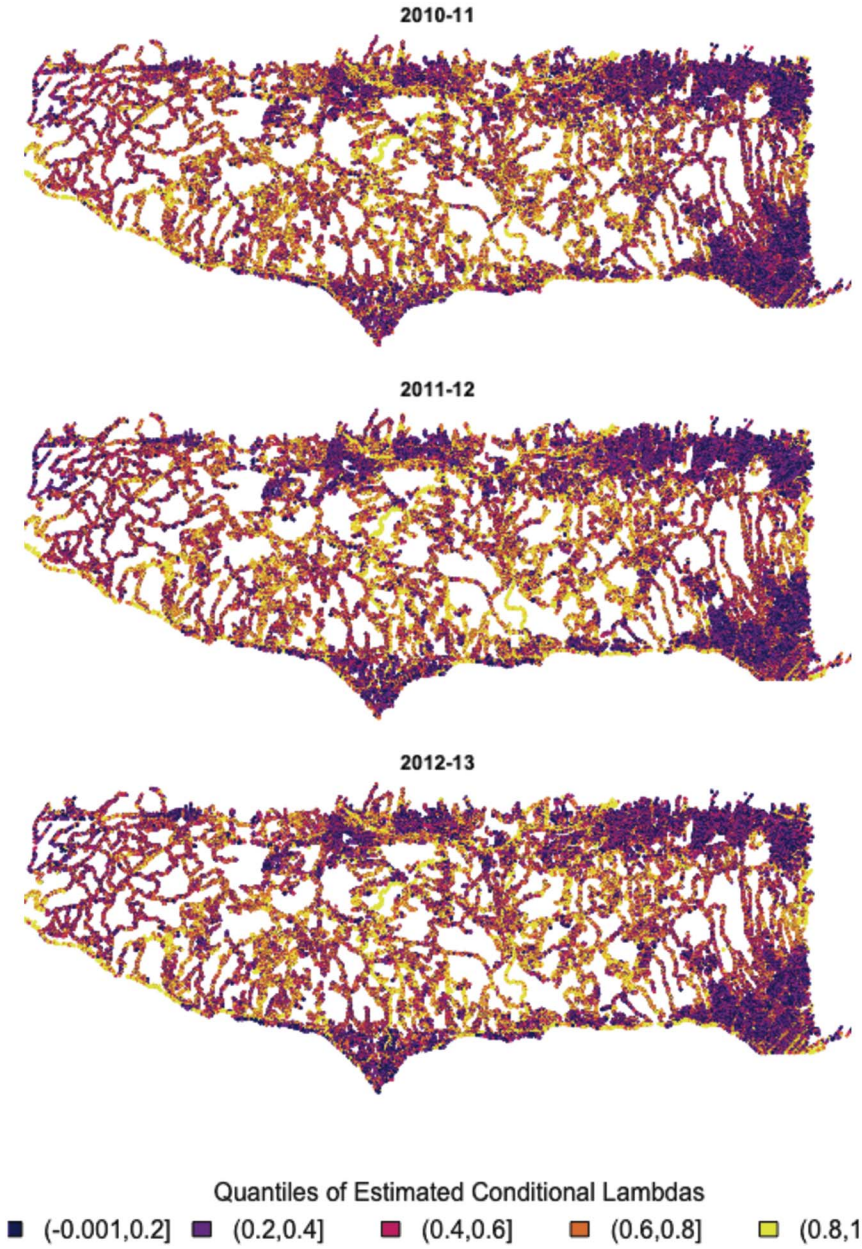

FIG. 5. *Estimated conditional ignition intensities over the Santa Monica Mountains region.*

lighter the color, the larger the conditional ignition intensity estimate. An initial glance indicates that the intensities (from 2010 to 2013) have stayed relatively similar over the region. The highly urbanized regions in the south (Malibu), southeast (Pacific Palisades), and the north/northeast (Woodland Hills, Tarzana and Encino) have consistently lower ignition intensities. These regions are predominantly residential and are also categorized as interface and non-WUI regions (Radeloff et al. (2017)). Even though the model indicates high-ignition intensities for interface regions, the stated regions are highly residential or central urban localities possibly leading to lower consistent ignitions over the 14 years. The model would probably disentangle the urban vs. wildland effects better if the non-WUI category was broken down to identify them separately.

Intensities begin to grow as we move inward into the mountainous areas of the network. This region is a mix of non-WUI vegetated and WUI-intermix regions, the roads here are also mostly "Road2" and "Paths;" this indicates that houses and building structures are intermixed with large swaths of wildland areas. This makes this region highly prone to fast spreading fires due to the availability of vegetation and increasing elevation. Interestingly enough, the model estimates varying and increasing ignition intensities in this region. The smaller, nonresidential roads and paths act as a perfect proxy for the sparser amount of human and structural presence in the nonurbanized region; such regions are more wildfire prone than highly urbanized regions due to the availability of burn fuel.

6.7. 2017–2018 *Southern California wildfires*. Southern California has experienced some of the largest, most damaging fires in its history in the years 2017 and 2018. The Woolsey fire of November 2018 burned through most of the network window but has its origin location above it. The Skirball fire of December 2017 has its origin location within the current window. We focus in on this region to observe how the ignition intensities have changed over the last three years of data. The Skirball fire origin is shown in red, and the neighboring region can been seen Figure 6.

The Skirball fire in Los Angeles occurred in December 2017 and originated due to illegal cooking at a homeless encampment beside the high traffic 405 freeway. The origin location (in the northeast corner of our study window) is away from the residential roads but relatively close to the city's main traffic artery and some other narrower dirt roads and paths. The area of the origin location is also categorized as wildland-urban intermix, where the ratio of wildlands to structural density is higher. The STLPP model's estimates are generally consistent and only slightly variable over the last three years. A closer look reveals that regions with higher density of residential roads have consistently lower ignition intensities, while WU intermix regions with a variety of road types have higher, more variable ignition intensities. This reiterates the point that nonurbanized regions that flank urbanized regions are prone to higher ignition intensities due to human activity and vegetation availability.

The more recent Woolsey fire of November 2018 originated on a laboratory complex that is in a vegetated region with very low to no housing density. While the origin cause is being investigated, a circuit outage was reported in the area. The region is relatively inaccessible and, due to the nearby Hill fire, fire fighter response was delayed. The confluence of all these factors along with the Santa Ana winds and the surrounding vegetated canyons led the

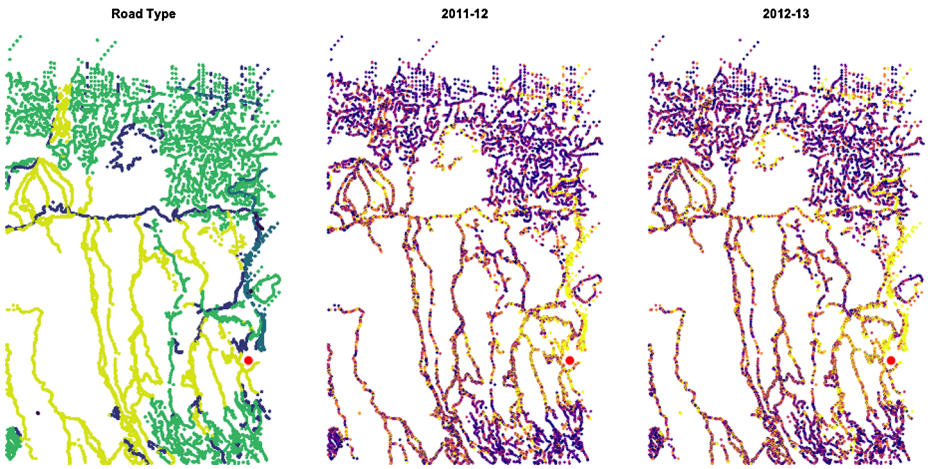

FIG. 6. *Ignition location of Skirball fire of December* 2017 *and estimated ignition intensities around the Skirball fire location from* 2011 *to* 2013. (*Colors correspond to the legends seen in Figures* 2 *and* 5.)

Woolsey fire to spread very quickly. A close look at the final burn area reveals that most of the areas burned through are vegetation rich WUI regions that are also relatively inaccessible. This is still consistent with our model outlook that low structural density, vegetated, interface/intermix regions have high ignition intensities and increasing construction in these regions only adds to the fire risk.

7. Model assessment.

7.1. *Comparison to the Euclidean distance model.* To better understand the benefits of the STLPPM framework of modeling point processes on linear networks, a comparison to a simple Euclidean distance model is presented. The spatial model takes into account the nearest distance to different road types from each point (including the wildfire locations) in the window as spatial covariates. Elevation over the window is also included into the model as a spatial covariate. This model presents the original (nontemporal) Papangelou conditional intensity, and the main goal is to discern how distance to different roadtypes affects the ignition intensity. The spatial model presented in Table 4 is

$$\log \lambda_\theta(u) = \theta_0 + \sum_{i=1}^{5} \theta_i d_i(u, R) + \theta_6 S(u), \tag{7.1}$$

where $d_i(u, R)$ is the nearest distance from point u to the ith type road and $S(u)$ is the elevation at point u.

The results from the Euclidean distance model indicate that "Unknown, Residential" and "Road1" types have the highest effect, that is, the lower the distance to these roads, the higher the ignition intensity. Alternatively, "Road2" and "Path" types have extremely low effects on ignition intensities. Elevation has about the same effect as it did in the STLPPM. While the model is too simplistic and does not include many important weather and land classification variables, it is difficult to gain satisfactory insight from this model. However, despite the obvious improvements to the Euclidean model, a direct contrast with the STLPPM framework does reveal a few drawbacks that are absent in the STLPPM framework.

A major drawback with the Euclidean model is that it adopts a purely isotropic view of the land's geography while the separable temporal model is geographically responsive as it models along the linear network. The Santa Monica Mountains region is geographically variable, with many canyons and hills. While the Euclidean model adopts a straight line distance to road types, the STLPPM framework is geographically aware of how many of the road types, especially "Road1, Road2" and "Paths," traverse different terrains in the region. This possibly explains the contradictory effects of "Road2" and "Unknown" roadtypes in the Euclidean model and STLPPM.

TABLE 4
Euclidean distance model

Parameter	Coefficient Estimate	MBB SE	MBB 95% CI
(Intercept)	9.098	13.804	(−20.302, 9.055)
Distance to Paths	−92.409	548.844	(−74.331, 55.264)
Distance to Road1	−9.784	313.218	(−62.257, 7.795)
Distance to Road2	−169.766	174.700	(−386.5, 4.15e−12)
Distance to Residential	−5.937	78.495	(−71.06, 1.40e−12)
Distance to Unknown	−2.157	382.801	(−32.104, 17.055)
Elevation	0.00115	0.014	(−6.76e−05, 0.021)

Another drawback of the Euclidean model is that the spatial trend plot does not provide the same level of granularity that the STLPPM does. The STLMMP identifies specific roads that exhibit higher ignition intensities and knowing specific lengths of roads is more beneficial to land management solutions. Overall, while we do agree that the presented Euclidean distance model can be improved upon, the comparison was mainly to reiterate STLPPM's ability to present a more nuanced view of ignition distribution, which is implicitly dependent on human and vehicular activity.

7.2. *Residual analysis*. Aside from the moving block bootstrap standard errors provided with the model estimates, residual analysis, outlined in Baddeley et al. (2005), was performed. While the residuals were relatively small, they were not very informative in terms of model improvement. Another technique by Clements, Schoenberg and Veen (2012), called super-thinning, was also employed to assess the fit of the seperable temporal model. Super-thinning is a combination method of thinning and superposition that produces a homogeneous poisson process with a certain rate k if and only if the estimated intensity ($\hat{\lambda}$) is equal to the actual intensity (λ) almost everywhere. The super-thinned process is then examined for homogeneity to assess the goodness of fit of $\hat{\lambda}$.

The mean of the estimated conditional intensity from the formation model is around 0.66. For the first attempt of super-thinning, the tuning parameter $k = 1$ was chosen. Figure 7 shows the original wildfires and the super-thinned residuals over the Santa Monica Mountains region.

In general, there are more super-thinned residuals than there are original wildfires. To assess the goodness-of-fit of the formation model, the super-thinned residual wildfires should be similar to a homogeneous poisson process with rate $k = 1$. It is important to note that super-thinning was limited to the linear network over the Santa Monica Mountains region,

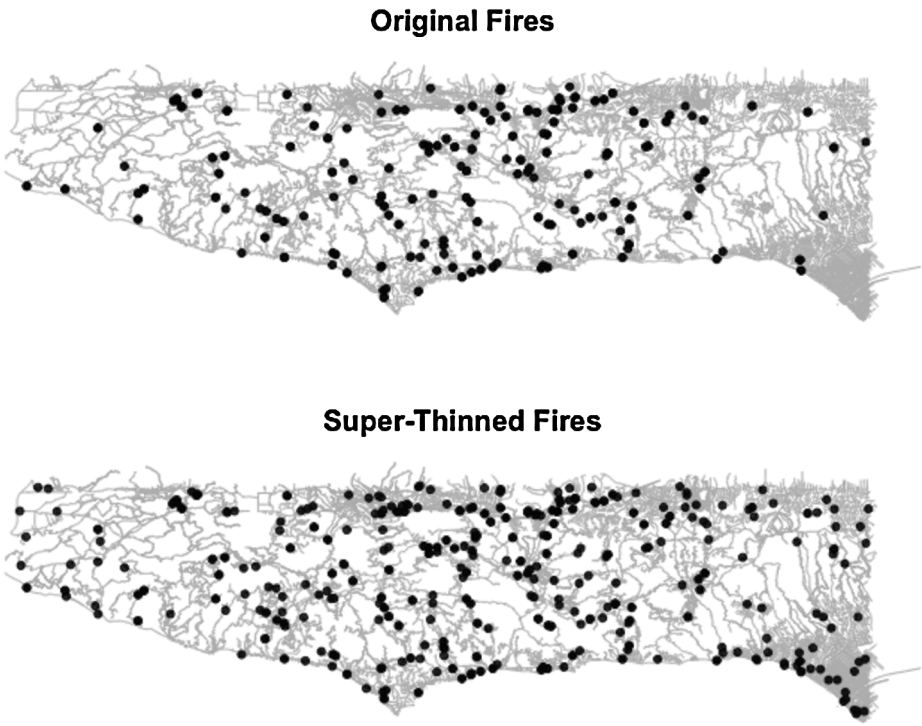

FIG. 7. *The original and super-thinned residual wildfires over the Santa Monica Mountains region with tuning parameter* $k = 1$.

and the uniformity of the super-thinned process should be assessed with respect to the network density. As a result, the super-thinned residual fires are much more common in high-density network regions of Santa Monica and along the 101 Freeway in the north. Visually, the super-thinned residuals seem to be fairly uniform given the linear network density around them; that is, high network density regions have more fires than low network density regions; this is in keeping with a homogeneous poisson process on a linear network. This indicates that, in general, the estimated *formation* model has captured the wildfire origin intensity decently well.

8. Conclusion. This paper presents a new approach to modeling point process data combined with linear networks. We present a novel and critical application to the modeling of wildfire origin locations on road networks. Aside from the novelty of using linear network data, most of the current wildfire research models wildfire burn area *centroids* rather than *origin* locations. The presented seperable temporal model not only models *origin* locations but also models their relationship to road networks. This provides a more nuanced look at how human movement and activity affect wildfire ignitions in Southern California. The temporal aspect of the model also provides a medium to long term spatial view of where most of the ignitions have occurred, allowing greater insight into handling and running fire management programs. The model, combined with burning and fire hazard indices, can aid and direct park rangers and fire fighters to better implement programs such as park closures, fuel management etc.

The model's strongest suit is in providing a spatial distribution of conditional intensities dependent on the density and type of roads. This feature has also brought to light the importance of wildland-urban interface and intermix (WUI) areas and their high proclivity to fire ignitions (due to high fuel abundance and volatile human activity).

APPENDIX A: SPIDER WEB LOCATIONS ON BRICK MORTAR LINES

This data was identified by entomologist Sasha Voss of University of Western Australia and manually digitized by Mark S. Handcock. The dataset records the *Oecobius navus* spider web positions in the mortar lines of brick walls. The dataset consists of weekly recordings from multiple sites, observed over a six week period from May to July 1999. In Figure 8 is a view of a single quadrant's data with covariate information.

APPENDIX B: MOVING BLOCK BOOTSTRAP FOR SPATIAL DATA

As the Papangelou conditional intensity of web or wildfire occurrence (on a linear network) is spatially dependent, one cannot draw random samples from the dataset to produce bootstrap samples. As an alternative, the moving block bootstrap (MBB), introduced independently by Künsch (1989) and Liu and Singh (1992), is employed here to produce bootstrapped standard errors. The idea of applying moving block bootstrap to spatial data was also introduced by Hall (1985).

Assume that $X_1, X_2, \ldots$ is a sequence of stationary random variables, and let $\mathbf{X_n} = (X_1, \ldots, X_n)$ denote the observations. In the wildfires case, $\mathbf{X_n}$ represents the complete set of data and dummy points from the window. Given the data, blocks of size l are defined as $B_j = (X_j, \ldots, X_{j+l-1})$, $j = 1, \ldots, N$ where $N = n - l + 1$ denotes all the possible overlapping blocks in $\mathbf{X_n}$. The bootstrap blocks are obtained by selecting a random sample of b blocks from the full set of overlapping blocks $\{B_1, \ldots, B_N\}$.

For the standard errors in Table 3, the MBB method was implemented with a block size of 1.3 million data points taken from the full set of 2.3 million points.

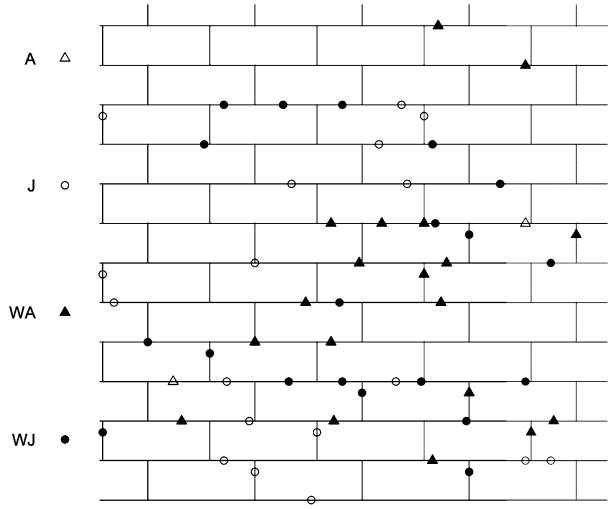

FIG. 8. *Site D, Quadrant Q2 on* 10 *June* 1999. *A—Adult spider, J—Juvenile spider, WA—Web with Adult spider, WJ—Web with Juvenile spider.*

Acknowledgments. We thank Sasha Voss for allowing us to use her spider web data for this paper. We also thank Rick Schoenberg for information on data sources, useful insights and discussion of the wildfire origins application and spatial residual analysis methods. We also thank the reviewers for their comments and suggestions which led us to include the comparison to the Euclidean model.

REFERENCES

ANG, Q. W., BADDELEY, A. and NAIR, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scand. J. Stat.* **39** 591–617. MR3000837 https://doi.org/10.1111/j.1467-9469.2011.00752.x

BADDELEY, A., JAMMALAMADAKA, A. and NAIR, G. (2014). Multitype point process analysis of spines on the dendrite network of a neuron. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63** 673–694. MR3269407 https://doi.org/10.1111/rssc.12054

BADDELEY, A. and TURNER, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Stat.* **42** 283–322. MR1794056 https://doi.org/10.1111/1467-842X.00128

BADDELEY, A., TURNER, R., MØLLER, J. and HAZELTON, M. (2005). Residual analysis for spatial point processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 617–666. MR2210685 https://doi.org/10.1111/j.1467-9868.2005.00519.x

BALCH, J. K., BRADLEY, B. A., ABATZOGLOU, J. T., NAGY, R. C., FUSCO, E. J. and MAHOOD, A. L. (2017). Human-started wildfires expand the fire niche across the United States. *Proc. Natl. Acad. Sci. USA* **114** 2946–2951.

BERMAN, M. and TURNER, R. (1992). Approximating point process likelihoods with GLIM. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **41** 31–38.

BESAG, J. and DIGGLE, P. J. (1977). Simple Monte Carlo tests for spatial pattern. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **26** 327–333.

CLEMENTS, R. A., SCHOENBERG, F. P. and VEEN, A. (2012). Evaluation of space-time point process models using super-thinning. *Environmetrics* **23** 606–616. MR3020078 https://doi.org/10.1002/env.2168

HALL, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20** 231–246. MR0808159 https://doi.org/10.1016/0304-4149(85)90212-1

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HERING, A. S. and BAIR, S. (2014). Characterizing spatial and chronological target selection of serial offenders. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **63** 123–140. MR3148272 https://doi.org/10.1111/rssc.12029

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. MR3153932 https://doi.org/10.1111/rssb.12014

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) (R. Lepage and L. Billard eds). 225–248. Wiley, New York. MR1197787

MANN, M. L., BERCK, P., MORITZ, M. A., BATLLORI, E., BALDWIN, J. G., GATELY, C. K. and CAMERON, D. R. (2014). Modeling residential development in California from 2000 to 2050: Integrating wildfire risk, wildland and agricultural encroachment. *Land Use Policy* **41** 438–452.

OKABE, A. and OKUNUKI, K. I. (2001). A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. *Trans. GIS* **5** 209–220.

OKABE, A. and SUGIHARA, K. (2012). *Spatial Analysis Along Networks*: *Statistical and Computational Methods*. Wiley, Chichester.

OKABE, A. and YAMADA, I. (2001). The K-function method on a network and its computational implementation. *Geogr. Anal.* **33**.

PAPANGELOU, F. (1973/74). The conditional intensity of general point processes and an application to line processes. *Z. Wahrsch. Verw. Gebiete* **28** 207–226. MR0373000 https://doi.org/10.1007/BF00533242

RADELOFF, V. C., HELMERS, D. P., KRAMER, H. A., MOCKRIN, M. H., ALEXANDRE, P. M., BAR MASSADA, A., BUTSIC, V., HAWBAKER, T. J., MARTINUZZI, S. et al. (2017). The 1990–2010 wildland-urban interface of the conterminous United States—geospatial data [2nd ed.].

RADELOFF, V. C., HELMERS, D. P., KRAMER, H. A., MOCKRIN, M. H., ALEXANDRE, P. M., BAR-MASSADA, A., BUTSIC, V., HAWBAKER, T. J., MARTINUZZI, S. et al. (2018). Rapid growth of the US wildland-urban interface raises wildfire risk. *Proc. Natl. Acad. Sci. USA* **115** 3314–3319. https://doi.org/10.1073/pnas.1718850115

SCHLESINGER, T. and EUGSTER, M. J. A. (2010). osmar: OpenStreetMap and R. *R J.*

SHORT, K. C. (2015). Spatial wildfire occurrence data for the United States, 1992–2013 [FPA_FOD_20150323] (3rd ed.).

SPOONER, P. G., LUNT, I. D., OKABE, A. and SHIODE, S. (2004). Spatial analysis of roadside Acacia populations on a road network using the network K-function. *Landsc. Ecol.* **19** 491–499.

SYPHARD, A. D. and KEELEY, J. E. (2015). Location, timing and extent of wildfire vary by cause of ignition. *Int. J. Wildland Fire* **24** 37–47.

XU, H. and SCHOENBERG, F. P. (2011). Point process modeling of wildfire hazard in Los Angeles County, California. *Ann. Appl. Stat.* **5** 684–704. MR2840171 https://doi.org/10.1214/10-AOAS401