

Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20


# Semiparametric Multinomial Logistic Regression for Multivariate Point Pattern Data


**Kristian Bjørn Hessellund, Ganggang Xu, Yongtao Guan & Rasmus Waagepetersen**

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–16: Theory and Methods
https://doi.org/10.1080/01621459.2020.1863812



# Semiparametric Multinomial Logistic Regression for Multivariate Point Pattern Data

Kristian Bjørn Hessellund*[a], Ganggang Xu*[b], Yongtao Guan[b], and Rasmus Waagepetersen[a]

[a]Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark; [b]Department of Management Science, University of Miami, Coral Gables, FL

**ABSTRACT**
We propose a new method for analysis of multivariate point pattern data observed in a heterogeneous environment and with complex intensity functions. We suggest semiparametric models for the intensity functions that depend on an unspecified factor common to all types of points. This is for example well suited for analyzing spatial covariate effects on events such as street crime activities that occur in a complex urban environment. A multinomial conditional composite likelihood function is introduced for estimation of intensity function regression parameters and the asymptotic joint distribution of the resulting estimators is derived under mild conditions. Crucially, the asymptotic covariance matrix depends on ratios of cross pair correlation functions of the multivariate point process. To make valid statistical inference without restrictive assumptions, we construct consistent nonparametric estimators for these ratios. Finally, we construct standardized residual plots, predictive probability plots, and semiparametric intensity plots to validate and to visualize the findings of the model. The effectiveness of the proposed methodology is demonstrated through extensive simulation studies and an application to analyzing the effects of socio-economic and demographical variables on occurrences of street crimes in Washington DC. Supplementary materials for this article are available online.

**ARTICLE HISTORY**
Received November 2019
Accepted December 2020

**KEYWORDS**
Conditional likelihood; Cross pair correlation functions; Multivariate point process

## 1. Introduction

Multivariate point pattern data with many types of points are becoming increasingly common. Ecologists collect large datasets on locations and species of plants and animals, while police authorities gather ever-increasing datasets on times, locations, and types of crimes. In epidemiology, multivariate point pattern datasets concern geo-referenced occurrences of different types of disease or bacteria. While the literature of bivariate point patterns is fairly well-developed (see, e.g., the review in Waagepetersen et al. (2016)), much less work has been done on the statistical analysis of point patterns with more than two types of points. Diggle, Zheng, and Durr (2005) and Baddeley, Jammalamadaka, and Nair (2014) considered four- and six-variate multivariate Poisson processes and more recently Jalilian et al. (2015) and Waagepetersen et al. (2016) considered five- and nine-variate multivariate Cox processes. Rajala, Murrell, and Olhede (2018) and Choiruddin et al. (2020) considered penalized estimation for, respectively, multivariate Gibbs and log Gaussian Cox point processes for datasets containing locations of more than 80 species of rain forest trees.

This article is concerned with statistical modeling of the first-order intensity functions of a multivariate spatial point process with an arbitrary number of types of points. For clarity of exposition, we discuss our proposal in relation to the specific problem of street crime analysis where we focus on the spatial aspects of street crimes aggregated over a time span of interest, see also the data example in Section 6. To model street crime activities as a multivariate point process poses three major challenges: (1) to handle the high complexity of the first-order intensity function for each type of points; (2) to relate the street crime locations to available spatial covariates; (3) to take into account spatial correlations within and between different types of crimes. The first challenge arises because street crime activities depend in a complicated way on the layout of the city (streets, squares, malls, etc.) as well as the typically unknown population density at any location. Moreover, the intensity of crime activities may also change abruptly from one area to neighboring areas. The second challenge arises because it is of great interest to police and criminologists to gain information on how street crime occurrences are related to demography, socio-economic variables, and other covariates. Such information is, for example, helpful to assess the validity of competing theories concerning the causes of the occurrence of crime in space (Weisburd et al. 1993; Cohen, Gorr, and Olligschlaeger 2007; Haberman 2017), see also Section 3.1. To properly assess the effects of covariates it is necessary to take into account the spatial correlation between street crimes, which leads to the third challenge.

To address the aforementioned first two challenges, we propose a semiparametric regression model for the first-order intensity functions. Specifically, we propose a multiplicative model where the intensity function for each type of points is a product of a nonparametric component common to all types of points and a parametric component that models the influence of the covariates on the intensity function. The common

**CONTACT** Rasmus Waagepetersen rw@math.aau.dk Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark.
*Hessellund and Xu are joint first authors with equal contributions.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

nonparametric component models background factors such as population density or variation in intensity due to the layout of a city. To fit the model we propose a conditional composite likelihood function that does not depend on the nonparametric factor and is formally equivalent to multinomial logistic regression. We derive the asymptotic joint distribution of the resulting estimators and provide an estimator of the asymptotic covariance matrix. A few articles have considered building full parametric models for clustered multivariate point processes (Jalilian et al. 2015; Waagepetersen et al. 2016; Rajala, Murrell, and Olhede 2018). However, these parametric models impose restrictive assumptions that are difficult to verify in practice and fitting the models can be rather challenging when the number of point types is large.

Our approach is inspired by the case-control methodology introduced in Diggle and Rowlingson (1994) and further considered in Guan, Waagepetersen, and Beale (2008), Zimmerman, Sun, and Fang (2012), and Xu, Waagepetersen, and Guan (2019). However, we do not restrict attention to the bivariate case considered in these references. Our approach also has some resemblance to Diggle, Zheng, and Durr (2005) who considered spatially varying risks of occurrence of one type of bacteria relative to the occurrence of other types. We, however, estimate relative risks using parametric models depending on covariates, where Diggle, Zheng, and Durr (2005) applied nonparametric kernel estimation. Diggle and Rowlingson (1994), Guan, Waagepetersen, and Beale (2008), Zimmerman, Sun, and Fang (2012), and Xu, Waagepetersen, and Guan (2019) further assumed independence between different types of points and that points of at least one type forms a Poisson process while Diggle, Zheng, and Durr (2005) and Zimmerman, Sun, and Fang (2012) assumed that all the different types of points form Poisson processes which are independent. According to the third challenge mentioned above, we do not assume that any of the point processes are Poisson and we do not assume independence between different types of points. This significantly expands the applicability of the proposed methodology to ever-growing multivariate point pattern data collected in the big data era.

Our analysis of the street crime data clearly shows that the different types of street crimes are not distributed as Poisson processes and are also not independent of each other, see Figure 4 for details. Thus, the inferential procedures considered in the existing work cited above, including Diggle and Rowlingson (1994), would not be valid even in the bivariate case. Table 2 in our simulation study demonstrates that ignoring spatial correlations among different point patterns will lead to severe under-coverage of the resulting confidence intervals. Table 4 of our crime data analysis also suggests that failure to take into account spatial correlations within and between types of points may result in misleading interpretations of the effects of some covariates.

Our theoretical investigation reveals that the asymptotic covariance matrix of our proposed estimator depends on the so-called pair correlation functions (PCFs) and cross PCFs of the multivariate point process, neither of which can be consistently estimated due to the common nonparametric component included in the model of the first-order intensity functions. A major novelty of our approach is our discovery that the asymptotic covariance matrix can be consistently estimated by an estimator expressed in terms of ratios of the PCFs and cross PCFs, but not the individual PCFs and cross PCFs themselves. In contrast to the individual PCFs and cross PCFs, it is possible to estimate these ratios consistently under the proposed model. However, the naive use of kernel estimators for PCF/cross-PCF ratios can still lead to serious under-coverage of the resulting confidence intervals. To further improve the quality of statistical inference, we developed a novel regularized nonparametric estimator for these ratios by imposing some mild shape constraints. To the best of our knowledge, no such regularized estimator has been studied in the literature. Consequently, valid statistical inferences can be performed for the estimated regression coefficients without restrictive parametric assumptions.

The proposed semiparametric regression model for the first-order intensity functions allows us to study relative risks given by the ratios of the first-order intensity functions. Our estimators of ratios of PCFs and cross PCFs allow us to generalize this concept to the second-order setting. The application to street crime data in Section 6 shows that practical insights can be gained by studying these PCF and cross PCF ratios. This is another novelty of our work. A final novel feature of our semiparametric model is that we can combine information for all types of points to estimate the nonparametric component and subsequently obtain semiparametric estimates of the intensity function for each type of points. This provides a more precise alternative to the usual nonparametric kernel intensity function estimator that is applied to each type of points separately.

The rest of the article is organized as follows. Section 2 provides an overview of multivariate point processes with a focus on intensity and cross PCFs. The semiparametric model and its inference are introduced in Section 3 and theoretical investigations are given in Section 4. Simulation studies are presented in Section 5 and an application to Washington DC street crime data is given in Section 6. Concluding remarks are given in Section 7 and all technical proofs are collected in the supplementary materials.

## 2. Background on Multivariate Point Processes

Denote by $X = (X_1, \ldots, X_p)$ a multivariate spatial point process, where $X_i$ is a random subset of $\mathbb{R}^d$ with the property that $X_i \cap B$ is of finite cardinality for all bounded $B \subseteq \mathbb{R}^d$ and $i = 1, \ldots, p$. We assume that each $X_i$ is observed in a bounded window $W \subset \mathbb{R}^d$ and $X_i \cap X_j = \emptyset$ for any $i \neq j$. Assume that for each $m \geq 1$ and $i = 1, \ldots, p$, there exists a nonnegative function $\lambda_i^{(m)}(\cdot)$ such that

$$\mathrm{E} \sum_{\mathbf{u}_1,\ldots,\mathbf{u}_m \in X_i}^{\neq} 1\left[\mathbf{u}_1 \in A_1, \ldots, \mathbf{u}_m \in A_m\right] = \int_{\prod_{j=1}^m A_j} \lambda_i^{(m)}(\mathbf{u}_1, \ldots, \mathbf{u}_m) \mathrm{d}\mathbf{u}_1 \cdots \mathrm{d}\mathbf{u}_m,$$

where $A_j \subset \mathbb{R}^d$, and $\sum^{\neq}$ indicates that $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are pairwise distinct. The function $\lambda_i^{(m)}(\cdot)$ is called the $m$th order joint intensity function of $X_i$. When $m = 1$, the function $\lambda_i^{(1)}(\cdot)$ is referred to as the intensity and is denoted $\lambda_i(\cdot)$. Assume further that for each $n, m \geq 1$ and $i, j = 1, \ldots, p$, there exists a nonnegative function $\lambda_{ij}^{(m,n)}(\cdot,\cdot)$ such that

$$\mathrm{E} \sum_{\mathbf{u}_1,\ldots\mathbf{u}_m\in X_i}^{\neq} \sum_{\mathbf{v}_1,\ldots,\mathbf{v}_n\in X_j}^{\neq} 1[\mathbf{u}_1 \in A_1, \ldots, \mathbf{u}_m \in A_m, \mathbf{v}_1 \in B_1, \ldots, \mathbf{v}_n \in B_n] \quad (1)$$

$$= \int_{\prod_{j=1}^m A_j} \int_{\prod_{j=1}^n B_j} \lambda_{ij}^{(m,n)}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n)\, \mathrm{d}\mathbf{u}_1 \cdots \mathrm{d}\mathbf{u}_m \mathrm{d}\mathbf{v}_1 \cdots \mathrm{d}\mathbf{v}_n,$$

where $A_k \subset \mathbb{R}^d$ and $B_l \subset \mathbb{R}^d$ for $k = 1, \ldots, m$ and $l = 1, \ldots, n$. The function $\lambda_{ij}^{(m,n)}(\cdot,\cdot)$ is referred to as the $(m, n)$th order cross intensity function between $X_i$ and $X_j$, $i, j = 1, \ldots, p$. The normalized (cross) joint intensities $g_i^{(m)}(\cdot)$ and $g_{ij}^{(m,n)}(\cdot,\cdot)$ are defined as

$$g_i^{(m)}(\mathbf{u}_1, \ldots \mathbf{u}_m) = \lambda_i^{(m)}(\mathbf{u}_1, \ldots, \mathbf{u}_m) / \prod_{l=1}^m \lambda_i(\mathbf{u}_l), \text{ and}$$

$$g_{ij}^{(m,n)}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n) = \frac{\lambda_{ij}^{(m,n)}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n)}{\prod_{l=1}^m \lambda_i(\mathbf{u}_l) \prod_{k=1}^n \lambda_j(\mathbf{v}_k)}, \quad (2)$$

provided the denominators on the right-hand sides are positive (otherwise we define $g_i^{(m)}(\mathbf{u}_1, \ldots, \mathbf{u}_m) = 0$ and $g_{ij}^{(m,n)}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n) = 0$). For $i \neq j$, $g_{ij}^{(1,1)}(\cdot,\cdot)$ is referred to as the cross PCF and $g_{ii}^{(1,1)}(\cdot,\cdot)$ coincides with $g_i^{(2)}(\cdot,\cdot)$ which is known as the PCF. From now on, we write $g_i(\cdot,\cdot)$ for $g_i^{(2)}(\cdot,\cdot)$ and $g_{ij}(\cdot,\cdot)$ for $g_{ij}^{(1,1)}(\cdot,\cdot)$. The notion of cross joint intensities and their normalized versions can be generalized in an obvious way to joint cross intensities $\lambda_{i_1 i_2 \cdots i_k}^{(n_1,\ldots,n_k)}$ and normalized cross joint intensities $g_{i_1 i_2 \cdots i_k}^{(n_1,\ldots,n_k)}$ for $X_{i_1}, \ldots, X_{i_k}$ for any $k \geq 1$, $\{i_1, \ldots, i_k\} \subseteq \{1, 2, \ldots, p\}$, and integers $n_1, \ldots, n_k \geq 1$.

Suppose that a point from $X_i$ is observed at $\mathbf{u}$. Then $\lambda_j(\mathbf{v}) g_{ij}(\mathbf{u}, \mathbf{v})$ can be interpreted as the conditional intensity of $X_j$ at $\mathbf{v}$ given that $\mathbf{u} \in X_i$. Thus, the cross PCF informs on how presence of a point in $\mathbf{u}$ affects the intensity of further points in $X_j$. In the special case when $X_i$ and $X_j$ are independent, $g_{ij}(\mathbf{u}, \mathbf{v}) \equiv 1$. If $X = (X_1, \ldots, X_p)$ consists of independent Poisson processes, we call $X$ a multivariate Poisson process. Then $\lambda_i^{(m)}(\mathbf{u}_1, \ldots, \mathbf{u}_m) = \prod_{l=1}^m \lambda_i(\mathbf{u}_l)$ and $\lambda_{ij}^{(m,n)}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n) = \prod_{l=1}^m \lambda_i(\mathbf{u}_l) \prod_{k=1}^n \lambda_j(\mathbf{v}_k)$. Consequently, $g_{ij}(\mathbf{u}, \mathbf{v}) = 1$, $i, j = 1, \ldots, p$, for a multivariate Poisson process which is the reference model of complete spatial independence.

Throughout the article, we assume that the multivariate point process is second-order cross-intensity reweighted isotropic meaning that $g_{ij}(\mathbf{u}, \mathbf{v})$ depends only on the distance $||\mathbf{u} - \mathbf{v}||$. For this reason, we abuse notation and denote by $g_{ij}(r)$ the value of $g_{ij}(\mathbf{u}, \mathbf{v})$ when $||\mathbf{u} - \mathbf{v}|| = r$. We often refer to so-called Campbell's formulas. For example, by standard measure theoretical arguments, the definition of $\lambda_i^{(m)}(\cdot)$ implies

$$\mathrm{E} \sum_{\mathbf{u}_1,\ldots,\mathbf{u}_m\in X_i}^{\neq} f(\mathbf{u}_1, \ldots, \mathbf{u}_m)$$
$$= \int_{(\mathbb{R}^d)^m} f(\mathbf{u}_1, \ldots, \mathbf{u}_m) \lambda_i^{(m)}(\mathbf{u}_1, \ldots, \mathbf{u}_m) \mathrm{d}\mathbf{u}_1 \cdots \mathrm{d}\mathbf{u}_m$$

for any nonnegative function $f$ on $(\mathbb{R}^d)^m$. Similar Campbell formulas hold for the cross joint intensities.

## 3. Semiparametric Multinomial Logistic Regression

In this section, we detail the proposed semiparametric model and the multinomial logistic regression approach to statistical inference. Formal asymptotic considerations are deferred to Section 4.

### 3.1. Semiparametric Model

For spatial point pattern data in an environment like a city, the intensity function can be rather complex due to the city layout and variations in population density. To overcome this difficulty, we follow Diggle and Rowlingson (1994) and assume that for each point pattern $X_i$, the intensity function takes the multiplicative form

$$\lambda_i(\mathbf{u}; \boldsymbol{\gamma}_i) = \lambda_0(\mathbf{u}) \exp\left[\boldsymbol{\gamma}_i^{\mathsf{T}} \mathbf{z}(\mathbf{u})\right], \quad i = 1, \ldots, p, \quad (3)$$

where $\lambda_0(\cdot)$ is an unknown background intensity function, $\mathbf{z}(\mathbf{u})$ is a $q$-dimensional vector of spatial covariates at location $\mathbf{u}$, and $\boldsymbol{\gamma}_i \in \mathbb{R}^q$ is the vector of regression parameters. The background intensity $\lambda_0(\cdot)$ can be interpreted as the spatial effects of latent factors such as the urban structure and population density and is assumed to be common for all point types. The model (3) is also closely related to the Cox regression model widely used for the conditional intensity in survival analysis (Cox 1972).

In case of crime, several competing theories regarding causes of crime exist (Weisburd et al. 1993; Haberman 2017). The crime general theory asserts that general factors drive crimes regardless of crime type. Accordingly, the proportions of crime types should be roughly constant across space. The crime specific theory instead asserts that different crimes depend on different factors, including environmental factors, which should lead to a more segregated occurrence of crime types with some crimes being more frequent in some areas than others. Our background intensity accommodates the effects of environmental factors with a common effect for all crimes. Next, based on (3) we can derive conditional probabilities which precisely model the proportions of crime types for each location $u$ and how they depend on spatial covariates, see (5) in Section 3.2.

Following Cohen, Gorr, and Olligschlaeger (2007), crime relevant spatial covariates may be categorized as crime attractors or crime displacements covariates. For example, distances to places like bars, parking lots, and music venues can be viewed as crime attractor covariates. Another example is the indicator of neighborhoods where policing of minor offenses are not strictly enforced. The spatial intensity of policing is an example of a crime displacement covariate since increased police activity in one location merely displaces crime to other locations rather than reducing crime overall (Ratcliffe 2002). In Section 6, we model Washington DC street crime by demographical covariates along with the distance to the nearest police station as a crime displacement covariate. The demographical covariates are not as such crime attractors but can be used to study whether the socio-economic status of a neighborhood has an impact on the occurrence of crimes.

The parameters $\boldsymbol{\gamma}_i$ are not identifiable since subtracting $\mathbf{k}^\mathsf{T}\mathbf{z}(\mathbf{u})$ from the log-linear model for some $\mathbf{k} \in \mathbb{R}^q$ while redefining $\lambda_0(\mathbf{u}) := \lambda_0(\mathbf{u})\exp[\mathbf{k}^\mathsf{T}\mathbf{z}(\mathbf{u})]$ does not change the intensities $\lambda_i$. To address this issue, we pick a baseline process, say $X_p$, and define identifiable parameters $\boldsymbol{\beta}_i = \boldsymbol{\gamma}_i - \boldsymbol{\gamma}_p$ for $i = 1, \ldots, p-1$. Further, without loss of generality, we may assume $\boldsymbol{\gamma}_p = 0$ in which case $\lambda_0(\cdot)$ becomes the intensity of the baseline process. Using the new parameterization, we can evaluate the effects of the covariates $\mathbf{z}(\cdot)$ relative to the baseline process $X_p$ similar to matched case-control studies and Cox regression in survival analysis.

Although estimation of the $\lambda_i(\cdot)$ is not our primary concern, note that given estimates $\hat{\lambda}_0(\cdot)$ and $\widehat{\boldsymbol{\beta}}_i$, we may estimate $\lambda_i(\cdot)$ by

$$\hat{\lambda}_i(\mathbf{u}) = \hat{\lambda}_0(\mathbf{u}) \exp\left[\widehat{\boldsymbol{\beta}}_i^\mathsf{T}\mathbf{z}(\mathbf{u})\right]. \tag{4}$$

If type $i$ points are rare, this estimate may be advantageous compared to an intensity estimate based only on type $i$ points since we can borrow strength by estimating $\lambda_0(\cdot)$ using all types of points, see also Section 6.2.

In terms of criminology research, a solid amount of literature states that crime is clustered in micro-places called hot spots (see Haberman 2017, and the references therein). Identification of hot spots may help police departments to allocate their resources properly (Buerger, Cohn, and Petrosino 1995) and hot spot policing reduces crime (Braga, Papachristos, and Hureau 2014). Numerous nonparametric methods have been developed to identify the hot spots, including kernel density estimation (Ratcliffe 2004; Gorr and Lee 2015). The estimator (4) adds to the existing hot spot detection methods by enhancing nonparametric kernel estimation with additional information from spatial covariates. Forecasting future occurrences of crime is another challenge to police departments. The Broken Windows theory of crime (Wilson and Kelling 1982) states that the tolerance of "soft" crimes in a neighborhood attracts criminals, hence the presence of "soft" crimes can be used to forecast "serious" crimes, see Cohen, Gorr, and Olligschlaeger (2007) and Gorr and Lee (2015). By a straightforward expansion of model (3) to a space-time setup, one could use the estimator (4) to forecast "serious" crimes using an estimate of current soft crime intensity as a covariate.

## 3.2. Multinomial Logistic Regression

We tackle the estimation of model (3) by conditional composite likelihood where we use the reparameterization in terms of the $\boldsymbol{\beta}_i$ from the previous section. Conditioned on that an event is observed at location $\mathbf{u}$, under model (3), the probability that it is from the point process $X_i$ is

$$\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}) = \frac{\lambda_i(\mathbf{u};\boldsymbol{\gamma}_i)}{\sum_{k=1}^p \lambda_k(\mathbf{u};\boldsymbol{\gamma}_k)} = \begin{cases} \dfrac{\exp[\boldsymbol{\beta}_i^\mathsf{T}\mathbf{z}(\mathbf{u})]}{1+\sum_{k=1}^{p-1}\exp[\boldsymbol{\beta}_k^\mathsf{T}\mathbf{z}(\mathbf{u})]}, & i = 1, \ldots, p-1, \\ \dfrac{1}{1+\sum_{k=1}^{p-1}\exp[\boldsymbol{\beta}_k^\mathsf{T}\mathbf{z}(\mathbf{u})]}, & i = p, \end{cases} \tag{5}$$

which does not depend on the background intensity $\lambda_0(\cdot)$. To estimate $\boldsymbol{\beta}$, we define the multinomial conditional composite likelihood as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^p \prod_{\mathbf{u}\in X_i\cap W} \mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}).$$

This is formally equivalent to a multinomial logistic regression likelihood function. It is a composite likelihood function because it ignores possible dependencies between types of points given their locations. The log multinomial conditional composite likelihood function is of the form

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^p \sum_{\mathbf{u}\in X_i\cap W} \left[\boldsymbol{\beta}_i^\mathsf{T}\mathbf{z}(\mathbf{u}) - \log\left(1 + \sum_{k=1}^{p-1}\exp\left[\boldsymbol{\beta}_k^\mathsf{T}\mathbf{z}(\mathbf{u})\right]\right)\right], \tag{6}$$

and the conditional composite likelihood estimator is defined as $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta})$.

## 3.3. Estimation of the Asymptotic Covariance Matrix of $\widehat{\beta}$

In this section, we consider the problem of estimating the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$, which is challenging due to the highly complex between- and within-type correlation structure of the multivariate point process.

We denote by $\mathrm{E}(\cdot)$ and $\mathrm{var}(\cdot)$, expectation and variance with respect to the data generating distribution of $X = (X_1, \ldots, X_p)$, where we assume the intensity function of $X_i$ is of the form (3) with the parameters $\boldsymbol{\gamma}_i$ given by some specific values $\boldsymbol{\gamma}_i^* \in \mathbb{R}^q$ and we let $\boldsymbol{\beta}_i^* = \boldsymbol{\gamma}_i^* - \boldsymbol{\gamma}_p^*$ for $i = 1, \ldots, p-1$. In this section and the rest of the article we will refer to the "pooled" point process $X^{\mathrm{pl}} = \cup_{k=1}^p X_i$, whose intensity function and PCF are

$$\lambda^{\mathrm{pl}}(\mathbf{u};\boldsymbol{\gamma}) = \sum_{k=1}^p \lambda_k(\mathbf{u};\boldsymbol{\gamma}_k) \text{ and } g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta},g) = \sum_{l=1}^p\sum_{l'=1}^p \mathrm{p}_l(\mathbf{u};\boldsymbol{\beta}_l)\mathrm{p}_{l'}(\mathbf{v};\boldsymbol{\beta}_l) g_{ll'}(\mathbf{u},\mathbf{v}). \tag{7}$$

The "$g$" inside $g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta},g)$ signifies the dependence on the $g_{ll'}$. We use in the following the short forms $\lambda_k^*(\cdot)$, $\mathrm{p}_l^*(\cdot)$, $\lambda^{\mathrm{pl}}(\cdot)$, and $g^{\mathrm{pl}}(\cdot,\cdot)$ for $\lambda_k(\cdot;\boldsymbol{\gamma}_i^*)$, $\mathrm{p}_l(\cdot;\boldsymbol{\beta}_l^*)$, $\lambda^{\mathrm{pl}}(\cdot;\boldsymbol{\gamma}^*)$, and $g^{\mathrm{pl}}(\cdot,\cdot;\boldsymbol{\beta}^*,g)$.

It is trivial to see that $\ell(\boldsymbol{\beta})$ in (6) is a concave function of $\boldsymbol{\beta}$ and thus maximizing $\ell(\boldsymbol{\beta})$ is equivalent to solving the estimating equation $\mathbf{e}(\boldsymbol{\beta}) = \mathbf{0}$ where

$$\mathbf{e}(\boldsymbol{\beta}) = \left[\mathbf{e}_1(\boldsymbol{\beta})^\mathsf{T}, \ldots, \mathbf{e}_{p-1}(\boldsymbol{\beta})^\mathsf{T}\right]^\mathsf{T}, \text{ with} \tag{8}$$

$$\mathbf{e}_i(\boldsymbol{\beta}) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}_i}\ell(\boldsymbol{\beta}) = \sum_{\mathbf{u}\in X_i\cap W}\mathbf{z}(\mathbf{u}) - \sum_{l=1}^{p}\sum_{\mathbf{u}\in X_l\cap W}\frac{\mathbf{z}(\mathbf{u})\exp\left[\boldsymbol{\beta}_i^\mathsf{T}\mathbf{z}(\mathbf{u})\right]}{1+\sum_{k=1}^{p-1}\exp\left[\boldsymbol{\beta}_k^\mathsf{T}\mathbf{z}(\mathbf{u})\right]}, \tag{9}$$

for $i = 1,\ldots,p-1$. According to standard estimating equation theory (see, e.g., Crowder 1986) and formally justified by Theorem 2 in Section 4.1, the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ is of the form $\left[\mathbf{S}(\boldsymbol{\beta}^*)\right]^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)\left[\mathbf{S}(\boldsymbol{\beta}^*)\right]^{-1}$ where $\mathbf{S}(\boldsymbol{\beta}^*) = \mathrm{E}\left[-\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\beta}^\mathsf{T}}\mathbf{e}(\boldsymbol{\beta}^*)\right]$ is the so-called sensitivity matrix and $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g) = \mathrm{var}\left[\mathbf{e}(\boldsymbol{\beta}^*)\right]$ is the covariance matrix of $\mathbf{e}(\boldsymbol{\beta}^*)$. The "$g$" inside $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)$ emphasizes that $\mathrm{var}\left[\mathbf{e}(\boldsymbol{\beta}^*)\right]$ depends on the underlying cross PCFs.

The explicit forms of $\mathbf{S}(\boldsymbol{\beta}^*)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)$ are derived in Section 1 of the supplementary materials. The $(i,j)$th block of $\mathbf{S}(\boldsymbol{\beta}^*)$ is of the form

$$\mathbf{S}(\boldsymbol{\beta}^*)_{ij} = \begin{cases} \int_W \mathbf{Z}(\mathbf{u},\mathbf{u})\left[1-\mathrm{p}_i^*(\mathbf{u})\right]\lambda_i^*(\mathbf{u})\mathrm{d}\mathbf{u} & i=j, \\ -\int_W \mathbf{Z}(\mathbf{u},\mathbf{u})\mathrm{p}_j^*(\mathbf{u})\lambda_i^*(\mathbf{u})\mathrm{d}\mathbf{u} & i\neq j, \end{cases} \tag{10}$$

for $i,j = 1,\ldots,p-1$ with $\mathbf{Z}(\mathbf{u},\mathbf{v}) = \mathbf{z}(\mathbf{u})\mathbf{z}(\mathbf{v})^\mathsf{T}$. The $(i,j)$th block of $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)$ corresponding to $\mathrm{cov}\left[\mathbf{e}_i(\boldsymbol{\beta}^*),\mathbf{e}_j(\boldsymbol{\beta}^*)\right]$ takes the form

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)_{ij} = \mathbf{S}(\boldsymbol{\beta}^*)_{ij} + \int_{W^2}\mathbf{Z}(\mathbf{u},\mathbf{v})\lambda_i^*(\mathbf{u})\lambda_j^*(\mathbf{v})g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)\, T_{ij}(\mathbf{u},\mathbf{v};\boldsymbol{\beta},g)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v}, \tag{11}$$

where the function $T_{ij}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)$ is defined as

$$1+\frac{g_{ij}(\mathbf{u},\mathbf{v})}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)} - \sum_{l=1}^{p}\frac{\left[\mathrm{p}_l^*(\mathbf{v})g_{il}(\mathbf{u},\mathbf{v}) + \mathrm{p}_l^*(\mathbf{u})g_{jl}(\mathbf{u},\mathbf{v})\right]}{g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g)}. \tag{12}$$

By Campbell's formulas, we can approximate $\mathbf{S}(\boldsymbol{\beta}^*)$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*,g)$ by $\widehat{\mathbf{S}}(\boldsymbol{\beta}^*)$ and $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}^*,g)$, whose $(i,j)$th blocks are defined as

$$\widehat{\mathbf{S}}(\boldsymbol{\beta}^*)_{ij} = \begin{cases} \sum\limits_{\mathbf{u}\in X^{\mathrm{pl}}}\mathbf{Z}(\mathbf{u},\mathbf{u})\left[1-\mathrm{p}_i^*(\mathbf{u})\right]\mathrm{p}_i^*(\mathbf{u}) & i=j, \\ -\sum\limits_{\mathbf{u}\in X^{\mathrm{pl}}}\mathbf{Z}(\mathbf{u},\mathbf{u})\mathrm{p}_i^*(\mathbf{u})\mathrm{p}_j^*(\mathbf{u}) & i\neq j, \end{cases} \tag{13}$$

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}^*,g)_{ij} = \widehat{\mathbf{S}}(\boldsymbol{\beta}^*)_{ij} + \sum_{\mathbf{u},\mathbf{v}\in X^{\mathrm{pl}}:||\mathbf{u}-\mathbf{v}||\leq R}^{\neq}\mathbf{Z}(\mathbf{u},\mathbf{v})\mathrm{p}_i^*(\mathbf{u})\mathrm{p}_j^*(\mathbf{v})T_{ij}(\mathbf{u},\mathbf{v};\boldsymbol{\beta}^*,g), \tag{14}$$

for $i,j = 1,\ldots,p-1$. Here $R$ denotes a "correlation range" such that $g_{ij}(r)\approx 1$ for $r>R$. In practice, we replace $\boldsymbol{\beta}^*$ by $\widehat{\boldsymbol{\beta}}$ in (13) and (14) and the notion "$g$" emphasizes their dependence on the underlying cross PCFs, which will be replaced by nonparametric estimators discussed in the next sections.

### 3.4. Naive Kernel Estimation of Cross PCF Ratios

The empirical covariance matrix (14) depends critically on cross PCFs which need to be estimated. The definition of a cross PCF in (2) suggests that its estimation requires consistent estimators of the intensity functions which are not available under the model (3), since $\lambda_0(\cdot)$ is unknown. However, a closer look at (12) reveals that for computation of (12) it suffices to estimate the cross PCFs up to a common multiplicative factor, or, equivalently, to estimate ratios of cross PCFs, that is

$$g_{ij,kl}(\mathbf{u},\mathbf{v}) = g_{ij}(\mathbf{u},\mathbf{v})/g_{kl}(\mathbf{u},\mathbf{v}), \quad i,j=1,\ldots,p, \tag{15}$$

for some arbitrary fixed pair of types of points $k$ and $l$. These ratios are also of great interest in their own right as they measure the strength of correlation among two types of points relative to the strength of correlation between two other types of points. Consider the quantity

$$F_{ij}(r;b,\boldsymbol{\beta}) = \sum_{\substack{\mathbf{u}\in X_i\cap W\\ \mathbf{v}\in X_j\cap W}}^{\neq}\frac{k_b(||\mathbf{u}-\mathbf{v}||-r)}{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta})\mathrm{p}_j(\mathbf{v};\boldsymbol{\beta})}, \tag{16}$$

where $k_b(\cdot) = k(\cdot/b)/b$ with $k(\cdot)$ being a kernel function defined on a bounded interval in $\mathbb{R}$ and $b>0$ is a bandwidth. Using Campbell's formula together with Equation (5), it follows that under model (3),

$$\mathrm{E}[F_{ij}(r;b,\boldsymbol{\beta}^*)] = \int_{W^2}\lambda^{\mathrm{pl}}(\mathbf{u})\lambda^{\mathrm{pl}}(\mathbf{v})g_{ij}(\mathbf{u},\mathbf{v})k_b(||\mathbf{u}-\mathbf{v}||-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v},$$

where $\lambda^{\mathrm{pl}}$ was defined in (7). Under suitable conditions and appropriately chosen bandwidth $b$, it is reasonable to expect that $F_{ij}(r;b,\widehat{\boldsymbol{\beta}})\approx c(r)g_{ij}(r)$, where

$$c(r) = \int_{W^2}\lambda^{\mathrm{pl}}(\mathbf{u})\lambda^{\mathrm{pl}}(\mathbf{v})k_b(||\mathbf{u}-\mathbf{v}||-r)\mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v},$$

is a multiplicative factor which, as desired, does not depend on $ij$. Consequently,

$$\hat{g}^{\mathrm{n}}_{ij,kl}(r;b,\widehat{\boldsymbol{\beta}}) = F_{ij}(r;b,\widehat{\boldsymbol{\beta}})/F_{kl}(r;b,\widehat{\boldsymbol{\beta}}) \tag{17}$$

becomes an estimator of (15).

Note that the estimator (17) does not depend on the unknown background intensity $\lambda_0(\cdot)$. The superscript "n" stands for "naive" kernel estimator (a regularized estimator will be introduced in the next section). Our Theorem 3 in Section 4.2 states that under mild conditions, (17) is consistent for $g_{ij,kl}(r)$. The naive plug-in estimator $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})$ is then obtained by replacing $\boldsymbol{\beta}^*$ and the cross PCFs in (12) by $\widehat{\boldsymbol{\beta}}$ and the estimators (17) of cross PCF ratios. For the rest of the article, we use the PCF of the baseline process $X_p$ as the fixed denominator in (15), letting $k=l=p$.

### 3.5. Regularized Cross PCF Ratio Estimators

Even though Theorem 3 in Section 4.2 shows that the naive kernel estimator (17) is consistent under mild conditions, the finite sample performance of the plug-in estimators $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})$ may be unsatisfactory due to high variabilities of the $\hat{g}^{\mathrm{n}}_{ij,pp}(\cdot)$'s. In particular, our numerical experiments suggest that when the number of observed points is small, some diagonal elements of the $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})$ may be negative, resulting in negative estimated variances for some components of $\widehat{\boldsymbol{\beta}}$.

We notice that this phenomenon is mainly caused by the existence of a large number of negative values of $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})$ when $||\mathbf{u}-\mathbf{v}||$ is large, leading to negative values in the diagonal of $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})_{ii}$ as defined in (14). This issue can be resolved or alleviated by imposing constraints on the cross PCFs. In this article, we impose the following constraints

$$g_{ij}(r) \le \sqrt{g_{ii}(r)g_{jj}(r)} \quad \text{for } r \ge R^*, \quad i,j = 1,\ldots,p, \tag{18}$$

for some $R^* \ge 0$. Intuitively, condition (18) means that for lags $r \ge R^*$, the spatial correlation between different point processes is weaker than the (geometric) average of spatial correlation within each individual point process. Condition (18) is not necessarily true for any multivariate point process but is indeed valid with $R^* = 0$ for a large class of multivariate log Gaussian Cox processes (Waagepetersen et al. 2016) (see also Section 5) and for a large subclass of the multivariate shot-noise Cox processes proposed in Jalilian et al. (2015).

To enforce the constraint (18) on the naive kernel estimators, let $\widehat{\mathbf{G}}_r^{\mathrm{n}}$ be a $p \times p$ matrix whose $(i,j)$th element is $\hat{g}^{\mathrm{n}}_{ij,pp}(r; b, \widehat{\boldsymbol{\beta}})$ for some $r > R^*$. The regularized nonparametric estimators, denoted as $\hat{g}^{\mathrm{r}}_{ij,pp}(r; b, \widehat{\boldsymbol{\beta}})$, are collected in the matrix $\widehat{\mathbf{G}}_r^{\mathrm{r}}$ obtained by

$$\widehat{\mathbf{G}}_r^{\mathrm{r}} = \arg\min_{\Theta=[\theta_{ij}]} \left\| \Theta - \widehat{\mathbf{G}}_r^{\mathrm{n}} \right\|_F^2, \quad \text{with} \tag{19}$$

$$\theta_{ij} = \theta_{ji}, \theta_{pp} = 1, \theta_{ij}^2 \le \theta_{ii}\theta_{jj},$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

It can be shown (Section 2 in the supplementary materials) that for $\|\mathbf{u} - \mathbf{v}\| > R^*$, the plug-in estimator with $\hat{g}^{\mathrm{r}}_{ij,pp}(\cdot)$'s satisfies

$$\min_{1\le i\le p} T_{ii}\left(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{r}}\right) \ge 1 - \max_{1\le l\le p} \hat{g}^{\mathrm{r}}_{ll,pp}(\|\mathbf{u}-\mathbf{v}\|; b, \widehat{\boldsymbol{\beta}})/g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{r}}).$$

In contrast, using the naive $\hat{g}^{\mathrm{n}}_{ij,pp}(\cdot)$'s, we can only achieve the lower bound

$$1 - \left[2 \max_{1\le l,l'\le p} \hat{g}^{\mathrm{n}}_{ll',pp}(\|\mathbf{u}-\mathbf{v}\|; b, \widehat{\boldsymbol{\beta}}) - \min_{1\le l\le p} \hat{g}^{\mathrm{n}}_{ll,pp}(\|\mathbf{u}-\mathbf{v}\|; b, \widehat{\boldsymbol{\beta}})\right] / g^{\mathrm{pl}}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}}).$$

Note that the first lower bound above can be much larger than the second lower bound, which partly explains why the regularized cross PCF ratio estimators would produce much fewer large negative $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{r}})$ when $\|\mathbf{u}-\mathbf{v}\| > R^*$, leading to a better covariance matrix estimator. In Section 5 of the supplementary materials, we give a more detailed demonstration through numerical examples.

*Remark 1.* Our numerical investigations suggest that the regularized estimator is quite robust to the choice of $R^*$. The simplest choice is to set $R^* = 0$. Otherwise we recommend to use $R^* = \arg\min_{r\ge 0}\{\max_i P_{ii}(r) > 0.05\}$, where $P_{ii}(r)$ is the percentage of pairs $(\mathbf{u},\mathbf{v})$ that give $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}}) < 0$ within the set $\{(\mathbf{u},\mathbf{v}) : \mathbf{u},\mathbf{v} \in X^{\mathrm{pl}} \text{ and } \|\mathbf{u}-\mathbf{v}\| \in (r-h, r+h)\}$. In other words, when the percentage of negative $T_{ii}(\mathbf{u},\mathbf{v};\widehat{\boldsymbol{\beta}},\hat{g}^{\mathrm{n}})$'s exceeds 5% around the distance $R^*$ for any $i = 1,\ldots,p$, the restriction (18) will be enforced for $r > R^*$.

## 4. Asymptotic Properties

In this section, we study asymptotic properties of $\widehat{\boldsymbol{\beta}}$ when $X$ is observed on a sequence of increasing windows $W_n$. Denote by $\mathbf{e}^{(n)}(\boldsymbol{\beta})$ the multinomial estimating function (8) evaluated on $W_n$ and by $\widehat{\boldsymbol{\beta}}_n$ the sequence of estimators obtained as solutions to $\mathbf{e}^{(n)}(\boldsymbol{\beta}) = \mathbf{0}$. The quantities $\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*, \boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)$ and $\mathbf{S}_n(\boldsymbol{\beta}^*)$ are defined as in Section 3.3 with $W = W_n$ for the last two. We also define "averaged" versions, $\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*, g) = \boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)/|W_n|$ and $\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*) = \mathbf{S}_n(\boldsymbol{\beta}^*)/|W_n|$. Finally, $\|\mathbf{A}\|_{\max} = \max_{ij} a_{ij}$ denotes the maximum norm of $\mathbf{A} = [a_{ij}]_{ij}$.

### 4.1. Consistency and Asymptotic Normality of $\widehat{\boldsymbol{\beta}}_n$

The following conditions are sufficient to establish the consistency of $\widehat{\boldsymbol{\beta}}_n$.

- C1 $W_1 \subset W_2 \subset \cdots$ and $\left|\bigcup_{l=1}^{\infty} W_l\right| = \infty$.
- C2 There exists an $0 < K_1 < \infty$ such that $\|\mathbf{z}(\mathbf{u})\|_{\max}$, $\lambda_i^*(\mathbf{u})$ and $g_{ij}(\mathbf{u},\mathbf{v})$ are bounded above by $K_1$ for all $\mathbf{u},\mathbf{v} \in \cup_{l=1}^{\infty} W_l$ and $i,j = 1,\ldots,p$.
- C3 There exists an $0 < K_2 < \infty$ so that $\int_{\mathbb{R}^d} |g_{ij}(\mathbf{0},\mathbf{u}) - 1|d\mathbf{u} < K_2$ for all $i,j = 1,\ldots,p$.
- C4 $\liminf_{n\to\infty} \lambda_{\min}\left[|W_n|^{-1}\int_{W_n} \mathbf{Z}(\mathbf{u},\mathbf{u})\lambda_i^*(\mathbf{u})\mathrm{p}_p(\mathbf{u};\boldsymbol{\beta}^*)d\mathbf{u}\right] > 0$ for $i = 1,\ldots,p-1$, where $\lambda_{\min}[A]$ denotes the minimal eigenvalue of a matrix $A$.

C1–C3 are mild conditions that have been widely used in the literature. C4 ensures that the averaged sensitivity matrix $\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)$ is invertible for sufficiently large $n$, which is commonly used in the estimating equation literature. Heuristically speaking, C4 requires that sufficient information regarding $\boldsymbol{\beta}^*$ need to be accumulated across space and it could be violated if $\mathbf{z}(\cdot)$ is close to constant.

*Theorem 1.* Under conditions C1–C4, there exists a sequence of solutions $\widehat{\boldsymbol{\beta}}_n$ to the estimating equation $\mathbf{e}_n(\boldsymbol{\beta}) = \mathbf{0}$ for which

$$\widehat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}^*, \text{ as } n \to \infty.$$

The proof of Theorem 1 is given in Section 3.1 of the supplementary materials.

Next, we proceed to establish asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$. Following Biscio and Waagepetersen (2019), we define an $\alpha$-mixing coefficient by regarding $X$ as a marked point process with points in $\mathbb{R}^d$ and marks in $M = \{1,\ldots,p\}$. That is, a point $\mathbf{u}$ in $X_i$ corresponds to a marked point $(\mathbf{u}, i)$. We then for sets $A \subseteq \mathbb{R}^d$ and $B \subseteq M$, define $X_{A,B} = X \cap A \times B$ as the set of marked points in $X$ whose "point parts" fall in $A$ and whose marks fall in $B$.

To define the $\alpha$-mixing coefficient for $X$ we first define an $\alpha$-mixing coefficient for two $\sigma$-algebras $\mathcal{F}$ and $\mathcal{G}$ on a common probability space,

$$\alpha(\mathcal{F},\mathcal{G}) = \sup\{|\mathrm{P}(F\cap G) - \mathrm{P}(F)\mathrm{P}(G)| : F \in \mathcal{F}, G \in \mathcal{G}\}.$$

Define $d(\mathbf{u},\mathbf{v}) = \max\{|u_i - v_i| : 1 \le i \le d\}$ for $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$. The marked point process $\alpha$-mixing coefficient of $X$ is then for $s, c_1, c_2 \ge 0$ given by

$$\alpha^X_{c_1,c_2}(s) = \sup\{\alpha(\sigma(X_{E_1,M}),\sigma(X_{E_2,M})) : E_1 \subset \mathbb{R}^d, E_2 \subset \mathbb{R}^d, |E_1| \le c_1, |E_2| \le c_2, d(E_1,E_2) \ge s\},$$

where $|A|$ is the Lebesgue measure of $A$ and $d(A, B) = \inf\{d(\mathbf{u}, \mathbf{v}) : \mathbf{u} \in A, \mathbf{v} \in B\}$. This coefficient measures the dependence between $X \cap E_1 \times M$ and $X \cap E_2 \times M$, where $E_1$ and $E_2$ are arbitrary Borel subsets of $\mathbb{R}^d$ with volumes less than $c_1$ and $c_2$ and separated by the distance $s$.

The following extra conditions are needed to establish asymptotic normality.

N1 There exists $\epsilon > 0$ such that $\alpha^X_{2,\infty}(s) = O(1/s^{d+\epsilon})$.

N2 There exist an integer $m > 2d/\epsilon + 2$ and $C_g$ such that $g^{(n_1,n_2,\ldots,n_k)}_{i_1 i_2 \cdots i_k}(\cdot, \ldots, \cdot) \le C_g$ for any $\{i_1, \ldots, i_k\} \subseteq \{1, 2, \ldots, p\}$, and integers $n_1 + \cdots + n_k \le m$.

N3 It holds that $\liminf_{n\to\infty} \lambda_{\min}\left[\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*, g)\right] > 0$.

N1 is a standard mixing condition that, for example, holds for multivariate log Gaussian Cox processes with PCFs of bounded range (meaning $g_{ij}(r) = 1$ when $r$ is larger than some $0 \le R < \infty$) or Poisson cluster point processes with sufficiently quickly decaying cluster densities. Condition N2 of bounded normalized joint cross intensities is satisfied for most multivariate point process models. N3 is a standard condition which ensures that the variance of $|W_n|^{-1}\mathbf{e}^{(n)}(\boldsymbol{\beta})$ is not degenerate for sufficiently large $n$.

*Theorem 2.* Under conditions C1–C4 and N1–N3, as $n \to \infty$, we have that

$$|W_n|^{1/2}\bar{\boldsymbol{\Sigma}}_n^{-1/2}(\boldsymbol{\beta}^*, g)\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*) \xrightarrow{d} N\left(0, \mathbf{I}_{(p-1)q}\right).$$

The proof of Theorem 2 is given in Section 3.2 of the supplementary materials.

Theorem 2 implies that the asymptotic variance of $\widehat{\boldsymbol{\beta}}_n$ is of the form

$$|W_n|^{-1}\left[\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\right]^{-1}\bar{\boldsymbol{\Sigma}}_n(\boldsymbol{\beta}^*, g)\left[\bar{\mathbf{S}}_n(\boldsymbol{\beta}^*)\right]^{-1} = \left[\mathbf{S}_n(\boldsymbol{\beta}^*)\right]^{-1}\boldsymbol{\Sigma}_n(\boldsymbol{\beta}^*, g)\left[\mathbf{S}_n(\boldsymbol{\beta}^*)\right]^{-1},$$

where the left-hand side suggests that the variance of $\widehat{\boldsymbol{\beta}}_n$ is of order $|W_n|^{-1}$. Based on Theorem 2, one can make statistical inference regarding $\boldsymbol{\beta}^*$ and other quantities of interest. For example, as in classical multinomial regression models, one may be interested in the probability of a certain event at a given location, that is, $\mathrm{p}^*_i(\mathbf{u})$, or the log-odds $\log\frac{\mathrm{p}^*_i(\mathbf{u})}{\mathrm{p}^*_p(\mathbf{u})} = \mathbf{z}(\mathbf{u})^\top\boldsymbol{\beta}^*_i$ for $i = 1, \ldots, p-1$.

Denote by $\mu(\boldsymbol{\beta}^*)$ a parameter of interest where $\mu : \mathbb{R}^{(p-1)q} \to \mathbb{R}$ is differentiable. A simple application of the Delta method gives for $0 < \alpha < 1$ the $100\alpha$% approximate confidence interval for $\mu(\boldsymbol{\beta}^*)$,

$$\mu(\widehat{\boldsymbol{\beta}}) \pm z_{1-\alpha/2}\sqrt{\left[\boldsymbol{\mu}^{(1)}(\widehat{\boldsymbol{\beta}})\right]^\top\left[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}})\right]^{-1}\widehat{\boldsymbol{\Sigma}}_n(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}}_n)\left[\widehat{\mathbf{S}}_n(\widehat{\boldsymbol{\beta}})\right]^{-1}\boldsymbol{\mu}^{(1)}(\widehat{\boldsymbol{\beta}})}, \quad (20)$$

where $z_\alpha$ is the $100\alpha$th percentile of a standard normal distribution, $\boldsymbol{\mu}^{(1)}(\boldsymbol{\beta}) = \mathrm{d}\mu(\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}$, and estimators of $\boldsymbol{\beta}$ and cross PCFs have been plugged into (13) and (14), see also Sections 3.4, 3.5, and 4.2.

### 4.2. Asymptotic Properties of $\hat{g}^{\mathbf{n}}_{ij,kl}(r; b, \widehat{\beta})$ and $\hat{g}^{\mathbf{r}}_{ij,kl}(r; b, \widehat{\beta})$

Let $W_n$ and $b_n$ be sequences of observation windows and bandwidths, respectively. Denote by $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n)$ a sequence of estimators that is given by

$$\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n) = F_{ij,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n)/F_{kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n),$$

where the $F_{ij,n}$'s are defined as in (16) with $W = W_n$. In this subsection, we show that $\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}_n)$ is a consistent estimator of $g_{ij,kl}(r)$ for any $i, j = 1, \ldots, p$, under the following conditions.

K1 For $i, j = 1, \ldots, p$, the cross joint intensity $g^{(2,2)}_{ij}$ is translation invariant: $g^{(2,2)}_{ij}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2) = g^{(2,2)}_{ij}(\mathbf{0}, \mathbf{u}_2 - \mathbf{u}_1, \mathbf{v}_1 - \mathbf{u}_1, \mathbf{v}_2 - \mathbf{u}_1)$, $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2 \in \cup_{l=1}^{\infty} W_l$, and there exists $K_3 < \infty$ so that $\int_{\mathbb{R}^d} |g^{(2,2)}_{ij}(\mathbf{0}, \mathbf{u}, \mathbf{v}, \mathbf{w} + \mathbf{u}) - g_{ij}(\mathbf{0}, \mathbf{v})g_{ij}(\mathbf{0}, \mathbf{w})|\mathrm{d}\mathbf{u} < K_3$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \cup_{l=1}^{\infty} W_l$.

K2 There exists $K_4 < \infty$ so that $g^{(m,n)}_{ij}(\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{v}_1, \ldots, \mathbf{v}_n) < K_4$ for all $\mathbf{u}_m, \mathbf{v}_n \in \cup_{l=1}^{\infty} W_l$ with $m + n < 4$ and $i, j = 1, \ldots, p$.

K3 The kernel function $k(\cdot)$ has a compact support $[-1, 1]$ and the bandwidth $b_n$ satisfies that (a) $b_n \to 0$; and (b) $|W_n|b_n \to \infty$ as $|W_n| \to \infty$.

*Theorem 3.* Under conditions C2 and K1–K3, one has that

$$\hat{g}^{\mathrm{n}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) \xrightarrow{p} g_{ij,kl}(r), \text{ as } n \to \infty, \text{ for } i, j, k, l = 1, \ldots, p. \quad (21)$$

If we further assume that constraint (18) holds true, then

$$\hat{g}^{\mathrm{r}}_{ij,kl,n}(r; b_n, \widehat{\boldsymbol{\beta}}) \xrightarrow{p} g_{ij,kl}(r), \text{ as } n \to \infty, \text{ for } i, j, k, l = 1, \ldots, p. \quad (22)$$

The proof of Theorem 3 is given in Section 3.3 of the supplementary materials.

## 5. Simulation Studies

In this section, we assess the finite sample performance of the proposed methodology through simulation studies. To evaluate our estimators we need to simulate from a model with known forms of the intensity functions and of the ratios of cross PCFs. This precludes the use of multivariate Gibbs processes as considered, for example, in Rajala, Murrell, and Olhede (2018) and we consider instead a Cox process model. Specifically, the multivariate point patterns are simulated from a multivariate log-Gaussian Cox process where for $i = 1, \ldots, p$, $X_i$ has a random intensity function of the form

$$\Lambda_i(\mathbf{u}) = \lambda_0(\mathbf{u})\exp[\gamma_{i0} + \gamma_{i1}\mathbf{z}(\mathbf{u})] \exp\left[\alpha_i Y(\mathbf{u}) + \sigma_i U_i(\mathbf{u}) - \alpha_i^2/2 - \sigma_i^2/2\right], \quad (23)$$

where $\lambda_0(\cdot)$ is the inhomogeneous background intensity, $\mathbf{z}(\cdot)$ is a spatial covariate, and $Y(\cdot)$ and $U_i(\cdot)$ are independent zero-mean unit variance Gaussian random fields. The spatial correlation functions of $Y(\cdot)$ and $U_i(\cdot)$ are assumed to be exponential $c_Y(\mathbf{u}, \mathbf{v}) = \exp(-||\mathbf{u}-\mathbf{v}||/\xi)$ and $c_{U_i}(\mathbf{u}, \mathbf{v}) = \exp(-||\mathbf{u}-\mathbf{v}||/\phi_i)$ with scale parameters $\xi$ and $\phi_i$. Conditional on the $\Lambda_i$, the $X_i$ are independent Poisson processes. This model has a natural interpretation and can generate both positive and negative correlations between different types of points.

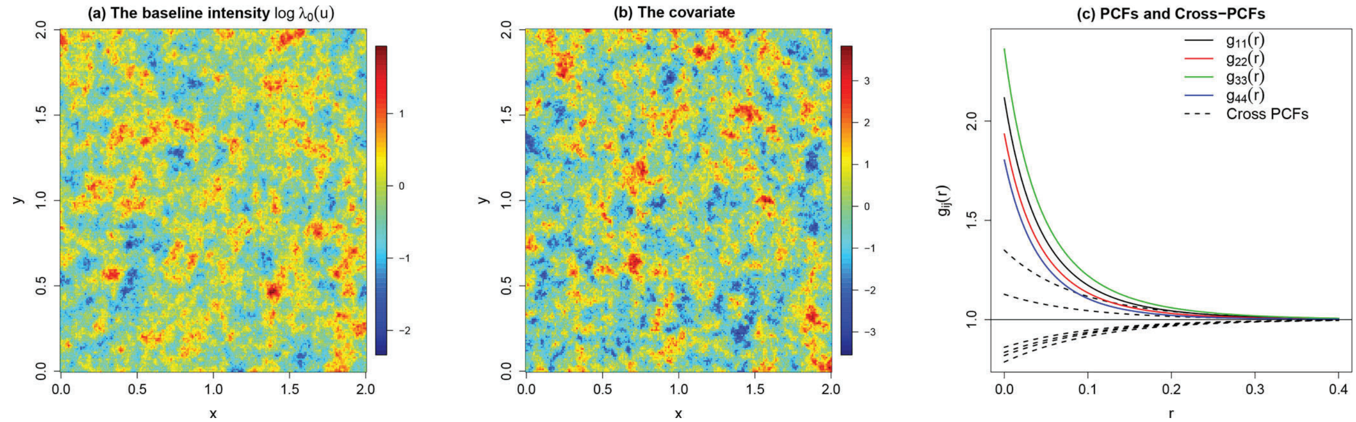


**Figure 1.** The log-background intensity (left panel); the spatial covariate (middle panel); the true PCFs and cross PCFs (right panel).

**Table 1.** The true parameters for the multivariate LGCP.

| $X$ | $\alpha_i$ | $\sigma_i^2$ | $\xi$ | $\phi_i$ | $\gamma_{i0}^*$ | $\gamma_{i1}^*$ | $N_i$ | $X$ | $\alpha_i$ | $\sigma_i^2$ | $\xi$ | $\phi_i$ | $\gamma_{i0}^*$ | $\gamma_{i1}^*$ | $N_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 0.5 | 0.5 | 0.1 | 0.05 | 5.17 | 0 | 150 | $X_2$ | −0.4 | 0.5 | 0.1 | 0.05 | 5.44 | 0.3 | 200 |
| $X_3$ | 0.6 | 0.5 | 0.1 | 0.05 | 5.88 | −0.6 | 300 | $X_4$ | −0.3 | 0.5 | 0.1 | 0.05 | 6.13 | 0.6 | 400 |

The process $Y(\cdot)$ can be viewed as an unobserved factor that affects all types of points and hence induces spatial correlations both within and between different types of points. The latent Gaussian process $U_i(\cdot)$ is a type-specific factor that only affects the $i$th type of points. Conditional on $\lambda_0(\cdot)$ and $\mathbf{z}(\cdot)$, $\mathrm{E}[\Lambda_i(\mathbf{u})] = \lambda_0(\mathbf{u})\exp[\gamma_{i0} + \gamma_{i1}\mathbf{z}(\mathbf{u})]$ and the cross PCF between $X_i$ and $X_j$ is of the form

$$g_{ij}(r;\boldsymbol{\theta}) = \exp\left[\alpha_i\alpha_j\exp\left(-r/\xi\right) + 1[i=j]\sigma_i^2\exp\left(-r/\phi_i\right)\right], \tag{24}$$

where $\boldsymbol{\theta} = (\alpha_1,\ldots,\alpha_p,\xi,\sigma_1^2,\ldots,\sigma_p^2,\phi_1,\ldots,\phi_p) \in \mathbb{R}^{3p+1}$. For $i \neq j$, $\alpha_i\alpha_j > 0$ ($< 0$) implies positive (negative) correlation between points from $X_i$ and $X_j$ whereas $\alpha_i\alpha_j = 0$ implies that $X_i$ and $X_j$ are independent given $\lambda_0(\cdot)$ and $\mathbf{z}(\cdot)$.

## 5.1. Simulation Settings

More specifically, we consider the multivariate log-Gaussian Cox process with $p = 4$ and observed within a sequence of increasing square windows $W_l = [0,l] \times [0,l]$, $1 \leq l \leq 2$. The baseline intensity function in (23) is $\lambda_0(\mathbf{u}) = \exp\left[0.5V(\mathbf{u}) - 0.5^2/2\right]$, where $V(\mathbf{u})$ is a realization of zero-mean unit variance Gaussian random field with the exponential correlation function and a scale parameter 0.05. The spatial covariate $\mathbf{z}(\mathbf{u})$ is chosen as an independent copy of $V(\mathbf{u})$, see Figures 1(a) and (b).

The parameters for the multivariate log-Gaussian Cox process are listed in Table 1, where the intercept parameters $\gamma_{i0}^*$, $i = 1,\ldots,p$, are chosen so that there are on average $N_i$ points in the point pattern $X_i$ in $W_1$ with the $N_i$'s specified in Table 1. We use $X_p$ as the baseline point process and consider three parameters of interest: the intercepts $\beta_{0i}^* = \gamma_{0i}^* - \gamma_{0p}^*$, the slopes $\beta_{1i}^* = \gamma_{1i}^* - \gamma_{1p}^*$, and the log-odds $\theta_i^*(\mathbf{u}) = \log\frac{\mathrm{p}_i(\mathbf{u};\boldsymbol{\beta}^*)}{\mathrm{p}_p(\mathbf{u};\boldsymbol{\beta}^*)} = \beta_{0i}^* + \beta_{1i}^*\mathbf{z}(\mathbf{u})$, for $i = 1,\ldots,p-1$. The log-odds $\theta_i^*(\mathbf{u})$ represent the elevated (or reduced) likelihood of a point in $X_i$ at location $\mathbf{u}$ with an observed covariate $\mathbf{z}(\mathbf{u})$ relative to the probability of a point in $X_p$ at $\mathbf{u}$. For the log odds we consider $\mathbf{z}(\mathbf{u}) = 0.5$. The $\alpha_i$'s are chosen such that there are positive and negative spatial correlations among the $X_i$'s. The resulting PCFs and cross PCFs show (Figure 1(c)) strong between- and within- spatial dependence.

In Section 5.2, we evaluate estimation accuracies for the parameters of interest and the coverage probabilities of their associated confidence intervals. The performances of the nonparametric cross PCF estimators proposed in Sections 3.4 and 3.5 are further considered in Section 5 of the supplementary materials.

## 5.2. Estimation Accuracies and Coverage Probabilities

The log odds $\theta_i^*(\mathbf{u})$ are estimated by replacing the $\boldsymbol{\beta}_i$'s in the definition of the $\theta_i^*(\mathbf{u})$'s by their estimates $\widehat{\boldsymbol{\beta}}_i$. Four types of confidence intervals are investigated, denoted $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$, $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$, $\mathrm{CI}_{g^{\mathrm{Poisson}}}$, and $\mathrm{CI}_{g^{\mathrm{true}}}$. All confidence intervals are constructed using (20) with the sensitivity and the covariance matrices estimated using Equations (13) and (14) with $R = 0.4$ but with different choices of cross PCF estimators. The $\mathrm{CI}_{\hat{g}^{\mathrm{n}}}$ and $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ use, respectively, the "naive" and "regularized" kernel cross PCF ratio estimators (17) and (19). The $R^*$ used for the "regularized" kernel estimators is obtained with the data-driven procedure in Remark 1. The $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ is obtained by assuming $g_{ij}(\cdot) \equiv 1$ for $i,j = 1,\ldots,p$, and $\mathrm{CI}_{g^{\mathrm{true}}}$ is constructed using the true $g_{ij}(\cdot)$'s. The coverage probabilities of $\mathrm{CI}_{g^{\mathrm{true}}}$ serve as bench marks while $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ may reveal potential problems of using multivariate Poisson point process models in presence of spatial correlations. Summary statistics based on 1000 simulations are given in Table 2 and also illustrated in Figure 2.

The "Bias" columns in Table 2 show that the parameter estimates are close to unbiased. Further, as predicted by Theorem 2, the standard errors are approximate halved when the observation window is increased from $W_1$ to the four times larger $W_2$.

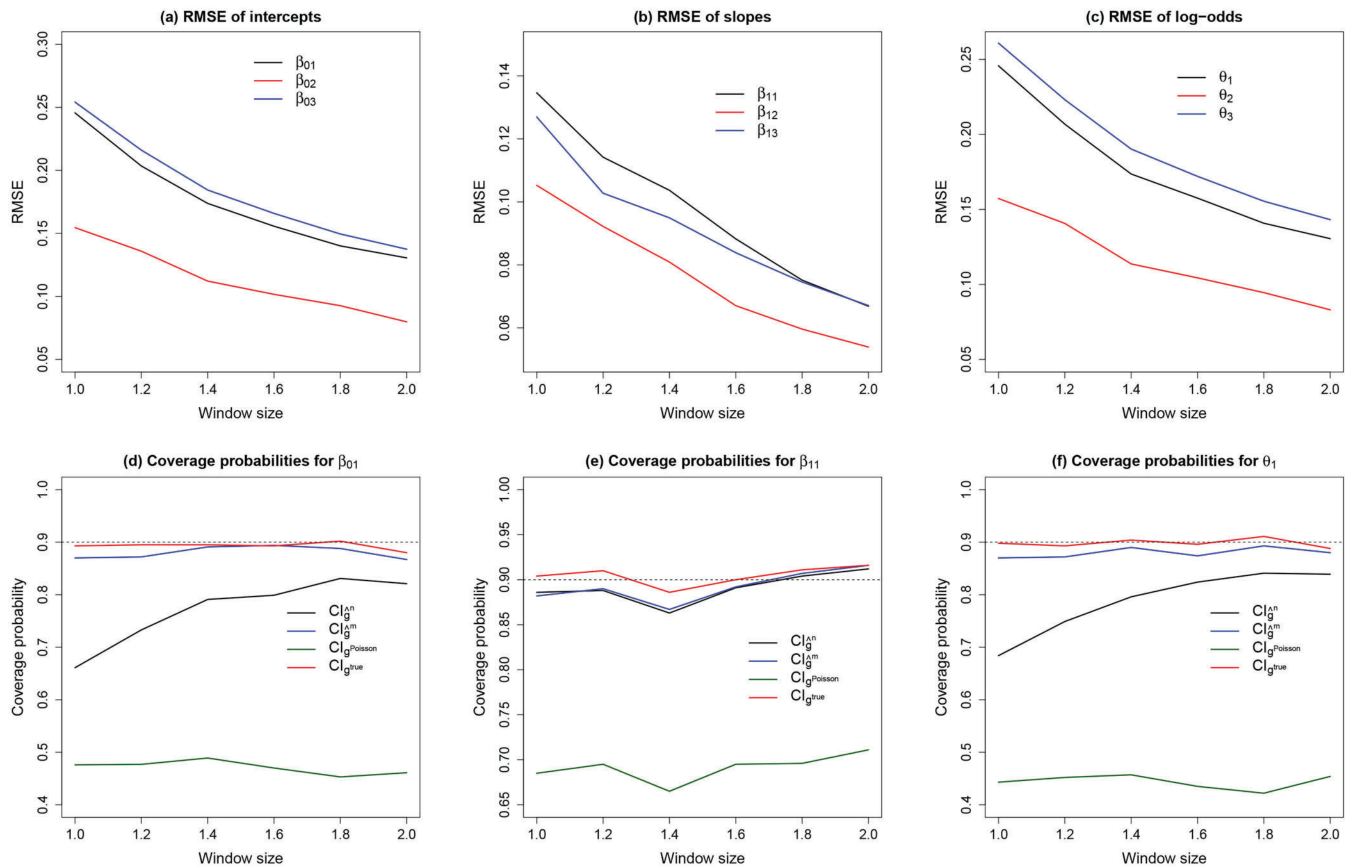


**Figure 2.** Top panels: The root mean squared errors (RMSE) of multinomial composite likelihood estimators; bottom panels: coverage probabilities of various confidence intervals. Observation windows range from $W_1$ to $W_2$.

**Table 2.** Estimation accuracies and coverage probabilities of confidence intervals.

| | | Bias | SE | $\mathrm{CI}_{\hat{g}^n}$ 90% | $\mathrm{CI}_{\hat{g}^n}$ 95% | $\mathrm{CI}_{\hat{g}^r}$ 90% | $\mathrm{CI}_{\hat{g}^r}$ 95% | $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ 90% | $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ 95% | $\mathrm{CI}_{g^{\mathrm{true}}}$ 90% | $\mathrm{CI}_{g^{\mathrm{true}}}$ 95% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $W_1$ | $\widehat{\beta}_{01}$ | −0.002 | 0.246 | 66.1 | 71.8 | 87.0 | 92.6 | 47.6 | 54.6 | 89.3 | 93.8 |
| | $\widehat{\beta}_{02}$ | 0.002 | 0.155 | 66.5 | 72.2 | 93.6 | 97.1 | 62.6 | 70.5 | 90.5 | 94.9 |
| | $\widehat{\beta}_{03}$ | 0.002 | 0.254 | 67.0 | 74.4 | 84.7 | 90.8 | 39.3 | 45.5 | 89.7 | 94.4 |
| | $\widehat{\beta}_{11}$ | −0.001 | 0.135 | 88.6 | 94.4 | 88.2 | 94.4 | 68.5 | 77.6 | 90.4 | 95.9 |
| | $\widehat{\beta}_{12}$ | 0.002 | 0.105 | 89.5 | 94.4 | 89.1 | 94.6 | 75.7 | 83.3 | 90.5 | 95.4 |
| | $\widehat{\beta}_{13}$ | −0.001 | 0.127 | 87.4 | 93.5 | 86.9 | 92.6 | 63.9 | 73.3 | 89.6 | 94.5 |
| | $\hat{\theta}_1$ | −0.003 | 0.246 | 68.4 | 75.4 | 87.0 | 93.0 | 44.3 | 52.2 | 89.8 | 94.5 |
| | $\hat{\theta}_2$ | −0.008 | 0.157 | 70.5 | 77.6 | 92.2 | 96.0 | 61.4 | 70.6 | 89.7 | 95.3 |
| | $\hat{\theta}_3$ | −0.002 | 0.261 | 72.6 | 80.7 | 86.2 | 91.3 | 44.1 | 51.3 | 90.8 | 94.5 |

| | | Bias | SE | $\mathrm{CI}_{\hat{g}^n}$ 90% | $\mathrm{CI}_{\hat{g}^n}$ 95% | $\mathrm{CI}_{\hat{g}^r}$ 90% | $\mathrm{CI}_{\hat{g}^r}$ 95% | $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ 90% | $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ 95% | $\mathrm{CI}_{g^{\mathrm{true}}}$ 90% | $\mathrm{CI}_{g^{\mathrm{true}}}$ 95% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $W_2$ | $\widehat{\beta}_{01}$ | −0.001 | 0.131 | 82.1 | 89.2 | 86.7 | 92.3 | 46.1 | 52.7 | 88.0 | 93.8 |
| | $\widehat{\beta}_{02}$ | −0.006 | 0.080 | 83.3 | 90.4 | 92.3 | 95.7 | 62.0 | 68.5 | 89.6 | 94.7 |
| | $\widehat{\beta}_{03}$ | 0.005 | 0.137 | 81.5 | 88.3 | 86.2 | 92.3 | 34.8 | 42.6 | 87.9 | 94.0 |
| | $\widehat{\beta}_{11}$ | −0.002 | 0.067 | 91.2 | 96.0 | 91.6 | 96.0 | 71.1 | 80.4 | 91.6 | 96.4 |
| | $\widehat{\beta}_{12}$ | −0.001 | 0.054 | 89.7 | 95.5 | 89.7 | 95.5 | 78.1 | 85.8 | 90.5 | 95.6 |
| | $\widehat{\beta}_{13}$ | −0.001 | 0.067 | 88.7 | 95.4 | 88.8 | 95.4 | 63.6 | 72.6 | 89.2 | 95.4 |
| | $\hat{\theta}_1$ | −0.002 | 0.130 | 83.9 | 88.7 | 88.0 | 92.4 | 45.4 | 52.4 | 88.8 | 94.2 |
| | $\hat{\theta}_2$ | −0.006 | 0.083 | 84.0 | 89.7 | 91.2 | 96.0 | 59.4 | 69.0 | 89.2 | 95.2 |
| | $\hat{\theta}_3$ | 0.005 | 0.143 | 83.7 | 88.2 | 86.0 | 91.8 | 40.2 | 47.9 | 88.7 | 93.9 |

The coverage probabilities of $\mathrm{CI}_{g^{\mathrm{true}}}$ are all close to the nominal levels, suggesting that statistical inferences based on Theorem 2 are valid provided all cross PCF functions are correctly specified. On the contrary, in almost all cases, $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ suffers from severe undercoverage that may lead to wrong conclusions in practical applications. Confidence intervals based on the “naive” kernel estimator of cross PCF ratios, that is, $\mathrm{CI}_{\hat{g}^n}$, achieve nominal levels for all slope parameters but suffer from serious undercoverage for intercepts and the log-odds when the observation window is small ($W_1 = [0,1] \times [0,1]$). The undercoverage of $\mathrm{CI}_{\hat{g}^n}$ becomes much less severe when the window expands to $W_2 = [0,2] \times [0,2]$. Finally, confidence intervals based on the “regularized” cross PCF ratio estimators, that is, $\mathrm{CI}_{\hat{g}^r}$, can effectively correct the undercoverage of $\mathrm{CI}_{\hat{g}^n}$ and achieve nominal levels for all parameters of interest. This suggests that it is important to apply the modification proposed in Section 3.5 for practical applications with only limited sample sizes.

Figure 2 paints a more complete picture of how estimation accuracies and coverage probabilities change as $W_l$ expands. The root mean squared error (RMSE) of all estimators decrease as the window size increases, supporting our theoretical findings in Section 4.1. Figure 2 also reveals that while the coverage probabilities of $\mathrm{CI}_{\hat{g}^n}$ for intercepts and log-odds are getting closer to the nominal level as $W_l$ expands, the undercoverage of $\mathrm{CI}_{g^{\mathrm{Poisson}}}$ does not improve at all. This emphasizes the importance of taking into account spatial correlations to make valid statistical inferences. Lastly, the coverage probabilities of $\mathrm{CI}_{\hat{g}^r}$ are close to the nominal level for all parameters and window sizes and only

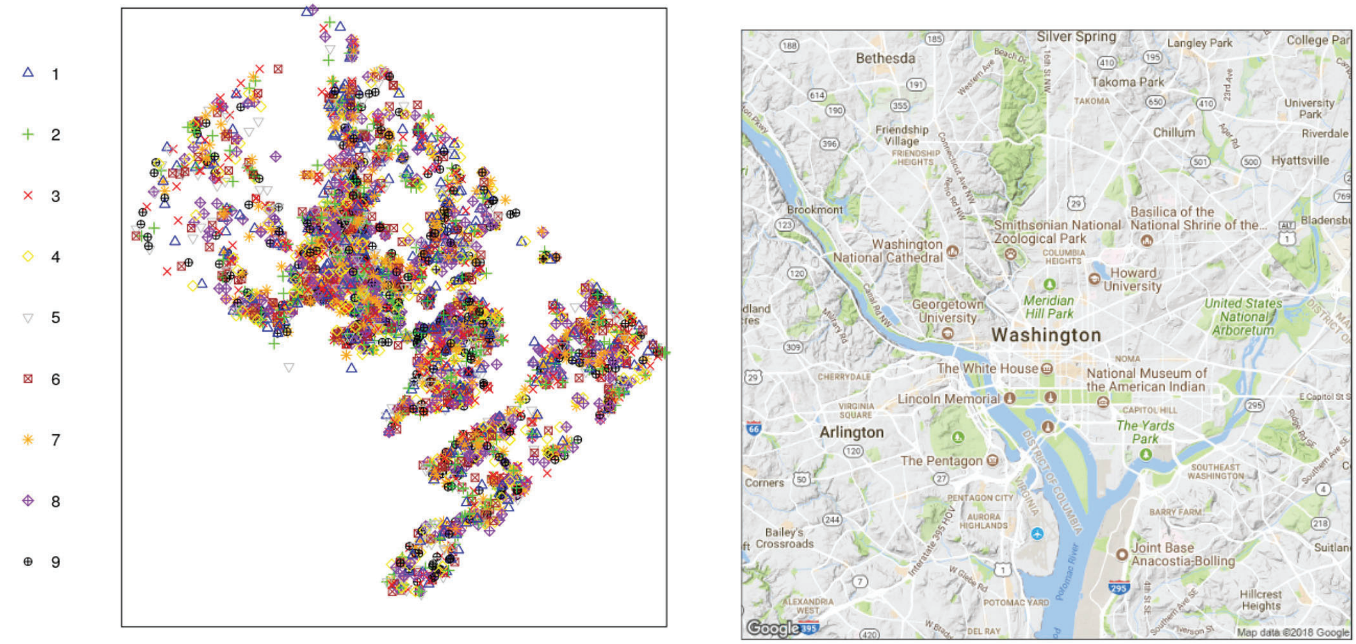


**Figure 3.** Left: Street crimes locations ($n = 5378$); right: a map of Washington DC.

**Table 3.** List of spatial covariates.

| Name | Definition |
|---|---|
| 1. % African | Square root of percentage of African American residents |
| 2. % Hispanic | Square root of percentage of Hispanic residents |
| 3. % Male | Square root of percentage of male residents with age 18-24 |
| 4. % HouseRent | Percentage of housing units occupied by renters |
| 5. % Bachelor | Percentage of residents over age 25 with a bachelor's degree |
| 6. MedIncome | Logarithm of median annual per capita income (in $1000) |
| 7. Pdist | Logarithm of the distance to the nearest police station |

slightly worse than those of $\mathrm{CI}_{g^{\mathrm{true}}}$. Therefore, we recommend $\mathrm{CI}_{\hat{g}^{\mathrm{r}}}$ for practical use.

## 6. Washington DC Street Crime Data

Figure 3 shows spatial locations of nine types of street crimes committed in Washington DC in January and February 2017. The dataset is publicly available from the website *http://opendata.dc.gov/datasets/crime-incidents-in-2017*. Nine types of street crime are included: (1) Other theft, (2) Robbery, (3) Theft from automobile, (4) Motor vehicle theft, (5) Assault with weapon, (6) Sex abuse, (7) Arson, (8) Burglary, and (9) Homicide. The numbers of each crime type are $n_1 = 2254$, $n_2 = 366$, $n_3 = 1832$, $n_4 = 335$, $n_5 = 332$, $n_6 = 44$, $n_7 = 1$, $n_8 = 259$, and $n_9 = 14$. We omit the rare street crimes "Sex abuse," "Arson," and "Homicide." Using spatial covariates similar to those suggested in Reinhart and Greenhouse (2018), the first 6 spatial covariates listed in Table 3 are obtained from US census data and are constant within each of 179 census tracts partitioning Washington DC, see also Section 6.3. We calculated ourselves the last covariate (distance to nearest police station) which varies smoothly across the city. Square root and log transformations have been applied to some covariates to achieve approximate normal distributions.

### 6.1. *Inference Regarding Regression Coefficients and Cross PCFs*

Using model (3), we assume that the intensity of each street crime is given by

$$\lambda_i(\mathbf{u}; \boldsymbol{\gamma}_i) = \lambda_0(\mathbf{u}) \exp\left[\gamma_{i0} + \gamma_{i1} z_1(\mathbf{u}) + \cdots + \gamma_{i7} z_7(\mathbf{u})\right],$$

$$i = 1, \ldots, 5, 8.$$

where the $z_k(\cdot)$'s are listed in Table 3. The common first street crime "Other theft" is used as the baseline. The regression parameters are estimated by maximizing the composite likelihood (6). The asymptotic standard errors and $p$-values are computed with $R = 3$ km and either of two types of cross PCFs: using the "regularized" kernel estimator $\hat{g}^{\mathrm{r}}$ proposed in Section (3.5) with $b = 0.2$ km, or assuming all $g_{ij}(\cdot) \equiv 1$ ( "Poisson") for any $i, j = 1, \ldots, 5, 8$. The $R^*$ used for the "regularized" kernel estimators is obtained through the data-driven procedure outlined in Remark 1. Estimated regression coefficients, standard deviations, and $p$-values are summarized in Table 4, and estimated PCF ratios and cross PCF ratios are illustrated in Figure 4.

Figure 4(a) indicates that within and between clustering for crimes types other than "Other theft" is less strong than for "Other theft" up to around 250 m. After that some crime types appear to be more clustered than "Other theft" but the difference in clustering strength vanishes around 3 km distance. In particular, Figure 4 suggests that a multivariate Poisson model is not appropriate for street crime data.

In Table 4, the Poisson model as expected always gives smaller standard errors for all coefficients. As a result, more regression coefficients appear to be statistically significant at the $\alpha = 0.05$ level (highlighted in blue) compared to those for the proposed method where cross PCFs are estimated from the data. In some cases, the two methods reach contradictory conclusions. For example, the covariate "% HouseRent" is significant

**Table 4.** Estimated coefficients, standard errors, and $p$-values for street crime data.

| Street crime | Covariate | Coef. | Std. err. $\hat{g}^r$ | Std. err. Poisson | $p$-values $\hat{g}^r$ | $p$-values Poisson |
|---|---|---|---|---|---|---|
| Robbery ($n_2 = 366$) | % African | 0.894 | 0.867 | 0.697 | 0.302 | 0.199 |
| | % Hispanic | 0.669 | 0.685 | 0.499 | 0.329 | 0.180 |
| | % Male | 0.141 | 1.183 | 0.962 | 0.905 | 0.884 |
| | % HouseRent | −0.783 | 0.442 | 0.352 | 0.077 | **0.026** |
| | % Bachelor | −1.130 | 0.970 | 0.760 | 0.244 | 0.137 |
| | MedIncome | −0.071 | 0.371 | 0.304 | 0.847 | 0.814 |
| | Pdist | 0.176 | 0.108 | 0.086 | 0.102 | **0.040** |
| Theft from automobile ($n_3 = 1832$) | % African | 2.318 | 0.813 | 0.346 | **0.004** | **<0.0001** |
| | % Hispanic | 2.369 | 0.760 | 0.286 | **0.002** | **<0.0001** |
| | % Male | −2.332 | 1.049 | 0.500 | **0.026** | **<0.0001** |
| | % HouseRent | −0.412 | 0.444 | 0.188 | 0.352 | **0.028** |
| | % Bachelor | 2.936 | 0.891 | 0.417 | **0.001** | **<0.0001** |
| | MedIncome | −0.461 | 0.339 | 0.164 | 0.174 | **0.004** |
| | Pdist | 0.071 | 0.107 | 0.047 | 0.508 | 0.131 |
| Motor vehicle theft ($n_4 = 335$) | % African | −0.451 | 0.872 | 0.702 | 0.605 | 0.520 |
| | % Hispanic | −0.556 | 0.724 | 0.533 | 0.443 | 0.297 |
| | % Male | −0.139 | 1.174 | 0.962 | 0.906 | 0.885 |
| | % HouseRent | −1.295 | 0.443 | 0.355 | **0.003** | **0.0003** |
| | % Bachelor | −1.767 | 0.993 | 0.785 | 0.075 | **0.024** |
| | MedIncome | −0.174 | 0.361 | 0.300 | 0.630 | 0.563 |
| | Pdist | 0.205 | 0.113 | 0.089 | 0.070 | **0.022** |
| Assault with weapon ($n_5 = 332$) | % African | 1.346 | 1.004 | 0.806 | 0.180 | 0.095 |
| | % Hispanic | −0.101 | 0.794 | 0.541 | 0.898 | 0.851 |
| | % Male | −2.76 | 1.358 | 1.132 | **0.042** | **0.0145** |
| | % HouseRent | −1.229 | 0.494 | 0.377 | **0.013** | **0.001** |
| | % Bachelor | −0.619 | 1.124 | 0.839 | 0.582 | 0.461 |
| | MedIncome | −0.798 | 0.391 | 0.314 | **0.041** | **0.011** |
| | Pdist | 0.145 | 0.122 | 0.088 | 0.235 | 0.100 |
| Burglary ($n_8 = 259$) | % African | −2.332 | 1.187 | 0.801 | **0.050** | **0.003** |
| | % Hispanic | −0.029 | 0.983 | 0.583 | 0.977 | 0.961 |
| | % Male | 0.776 | 1.555 | 1.039 | 0.618 | 0.455 |
| | % HouseRent | −1.930 | 0.670 | 0.376 | **0.001** | **<0.0001** |
| | % Bachelor | −3.374 | 1.327 | 0.875 | **0.011** | **0.001** |
| | MedIncome | −0.352 | 0.432 | 0.300 | 0.415 | 0.240 |
| | Pdist | 0.359 | 0.168 | 0.105 | **0.033** | **0.0006** |

NOTE: Highlighted values in Table 4 indicate spatial covariates whose p-values are less than 0.05 under respective models.

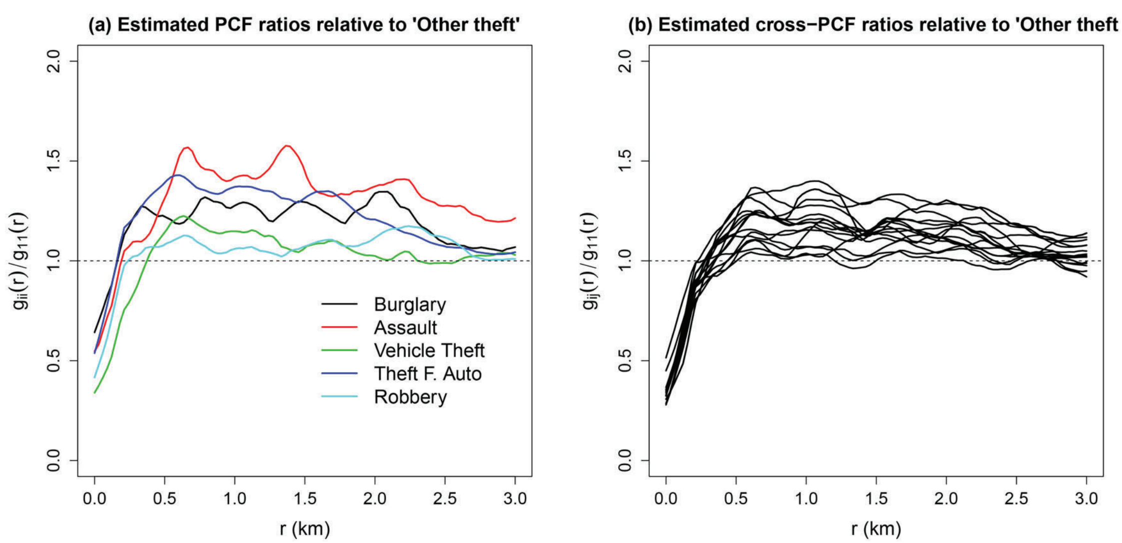


**Figure 4.** (a) Estimated PCF ratios $g_{ii}(r)/g_{11}(r)$ for $i = 2, \dots, 5, 8$; (b) estimated cross PCF ratios $g_{ij}(r)/g_{11}(r)$ for $i, j = 2, \dots, 5, 8$ and $i \neq j$.

under the Poisson model ($p$-value 0.028) when comparing "Theft from auto" to the baseline process "Other theft," while the proposed model asserts otherwise with a $p$-value of 0.352. In such cases, considering the strong spatial correlations displayed in Figure 4, we argue that the proposed method is more reliable.

Based on the proposed method, all estimated coefficients for "% HouseRent" are negative and many of them are significant, suggesting that when "% HouseRent" is large, "Other theft" becomes relatively more frequent compared to all other crime types. Second, no covariate elevates or reduces the relative risk of "Robbery" compared to "Other theft" and no covariate other than "% HouseRent" is significant for the relative risk between "Motor vehicle theft" and "Other theft." Third, "Theft from automobile" tend to occur more often in a neighborhood with more African American/Hispanic population, less young male percentage and residents with relatively low education level, as compared to "Other theft." Fourth, "Assault with weapon" is more likely to occur in a neighborhood with low young male population and low income levels compared to "Other theft." Finally, compared to "Other theft," "Burglary" tends to occur more in areas with low African American population, low education level and larger distance to the police station.

Returning to the discussion regarding crime general and crime specific theories in Section 3.1, our results clearly show that the relative risks of different crime types depend significantly on subsets of the covariates considered. This also means that the conditional probabilities (5) depend significantly on the covariates which results in a clear spatial segregation regarding the relative risks of different crimes, see Figure 5. These results support the crime specific theory.

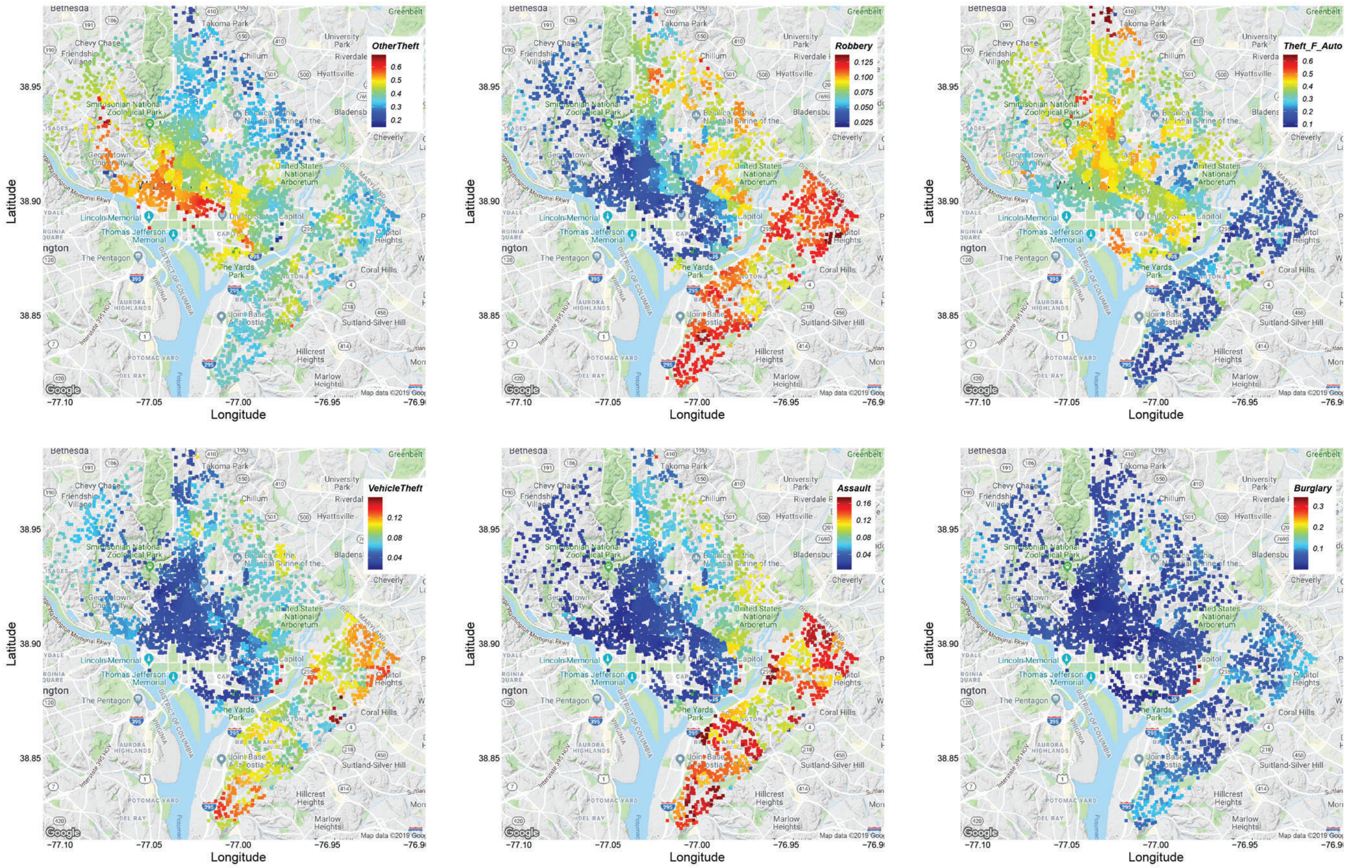


Figure 5. Estimated conditional probability maps for Washington DC.

### 6.2. Conditional Probability Maps and Intensity Estimation

For any location $\mathbf{u}$, using the fitted $\widehat{\boldsymbol{\beta}}$, we can compute $\mathrm{p}_i(\mathbf{u}, \widehat{\boldsymbol{\beta}})$ for $i = 1, \ldots, p$, using (5). This enables us to create the conditional probability maps in Figure 5 which show $\mathrm{p}_i(\mathbf{u}, \widehat{\boldsymbol{\beta}})$, $i = 1, \ldots, 5, 8$ computed at the 5378 observed crime locations. Recall that given a street crime occurs at location $\mathbf{u}$, $\mathrm{p}_i(\mathbf{u}, \widehat{\boldsymbol{\beta}})$ is the fitted probability that the crime is of the $i$th type. The strong spatial patterns in these conditional probabilities are remarkable. For instance, in the southeast part of the city (southeast to the Anacostia River), given a crime occurs, it is much more likely to be of type "Robbery" or "Assault" than in other parts of the city. In contrast, "Theft from automobile" is more likely to be reported in the middle and northern parts of the city while the hot spot for "Other theft" is located in the middle-west part of the city.

Figure 6 shows semiparametric kernel estimates of the six crime intensities using (4) where $\lambda_0$ is estimated using the kernel estimate

$$\hat{\lambda}_0(\mathbf{u}) = \frac{1}{p} \sum_{i=1}^{p} \sum_{\mathbf{v} \in X_i} \exp[-\widehat{\boldsymbol{\beta}}_i^{\mathsf{T}} \mathbf{z}(\mathbf{v})] k[(\mathbf{u} - \mathbf{v})/b]/b^2, \tag{25}$$

where $k$ is a two-dimensional kernel and the bandwidth $b = 3.37$ km is chosen according to the data-driven criterion of Cronie and Van Lieshout (2018). Compared to the conditional probability plots given in Figure 5 that demonstrate relative compositions of different types of crimes at a given location, the marginal intensities provide additional information on how often each type of crime occurs in the same location. Both plots can be useful in practice for the police department to better allocate limited resources to effective fight different types of crimes.

### 6.3. Residual Analysis

In this subsection, we perform a residual analysis for the fitted model. We divide the data according to the 179 census tracts in Washington DC, denoted as $A_1, A_2, \ldots, A_K$, $K = 179$, we define the raw residual for the $i$th type of street crime in $A_k$ as

$$\hat{\varepsilon}_{i,k}(\widehat{\boldsymbol{\beta}}) = \sum_{\mathbf{u} \in X_i} I(\mathbf{u} \in A_k) - \sum_{\mathbf{u} \in X^{\mathrm{pl}} \cap A_k} \mathrm{p}_i(\mathbf{u}; \widehat{\boldsymbol{\beta}}), \tag{26}$$

for $i = 1, \ldots, p$ and $k = 1, \ldots, K$. Equation (26) is essentially a restricted version (within $A_k$) of the intercept component of $\mathbf{e}_i(\widehat{\boldsymbol{\beta}})$ defined in (9). By definition of $\widehat{\boldsymbol{\beta}}$, $\mathbf{e}_i(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$, implying $\sum_{k=1}^{K} \hat{\varepsilon}_{i,k} = 0$ for $i = 1, \ldots, p$. If the model fits the data reasonably well, one should expect most $\hat{\varepsilon}_{i,k}$ to be relatively close to 0.

Use the same arguments leading to (14), the variance of $\hat{\varepsilon}_{i,k}(\boldsymbol{\beta}^*)$ can be estimated by

$$\hat{\sigma}^2_{i,k}(\boldsymbol{\beta}^*, g) = \sum_{\mathbf{u} \in X^{\mathrm{pl}} \cap A_k} \left[1 - \mathrm{p}_i^*(\mathbf{u})\right] \mathrm{p}_i^*(\mathbf{u}) + \sum_{\mathbf{u}, \mathbf{v} \in X^{\mathrm{pl}} \cap A_k}^{\mathbf{u} \neq \mathbf{v}} \sum \mathrm{p}_i^*(\mathbf{u}) \mathrm{p}_i^*(\mathbf{v}) T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g),$$

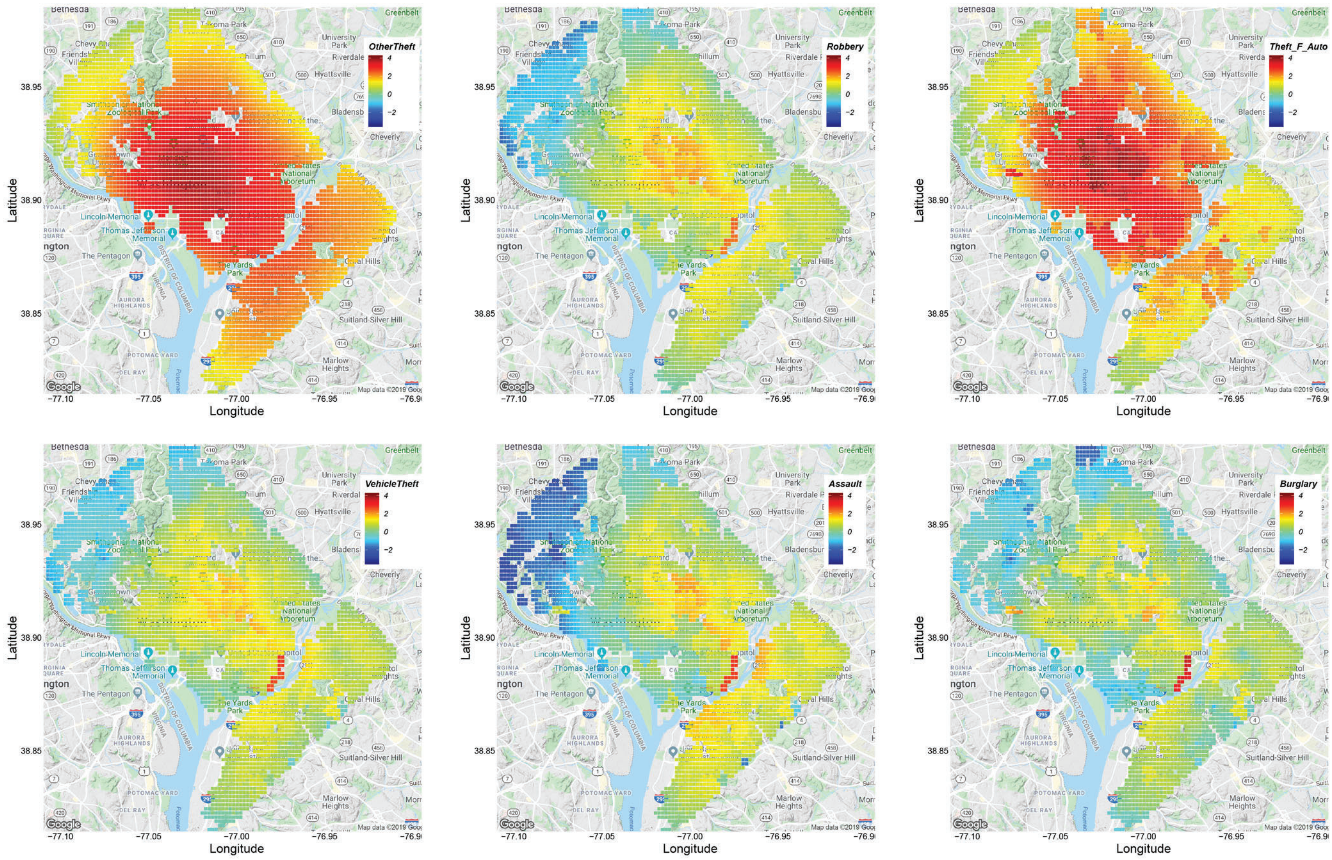


Figure 6. Semiparametric log-intensity (per km$^2$) maps for crime data in Washington DC.

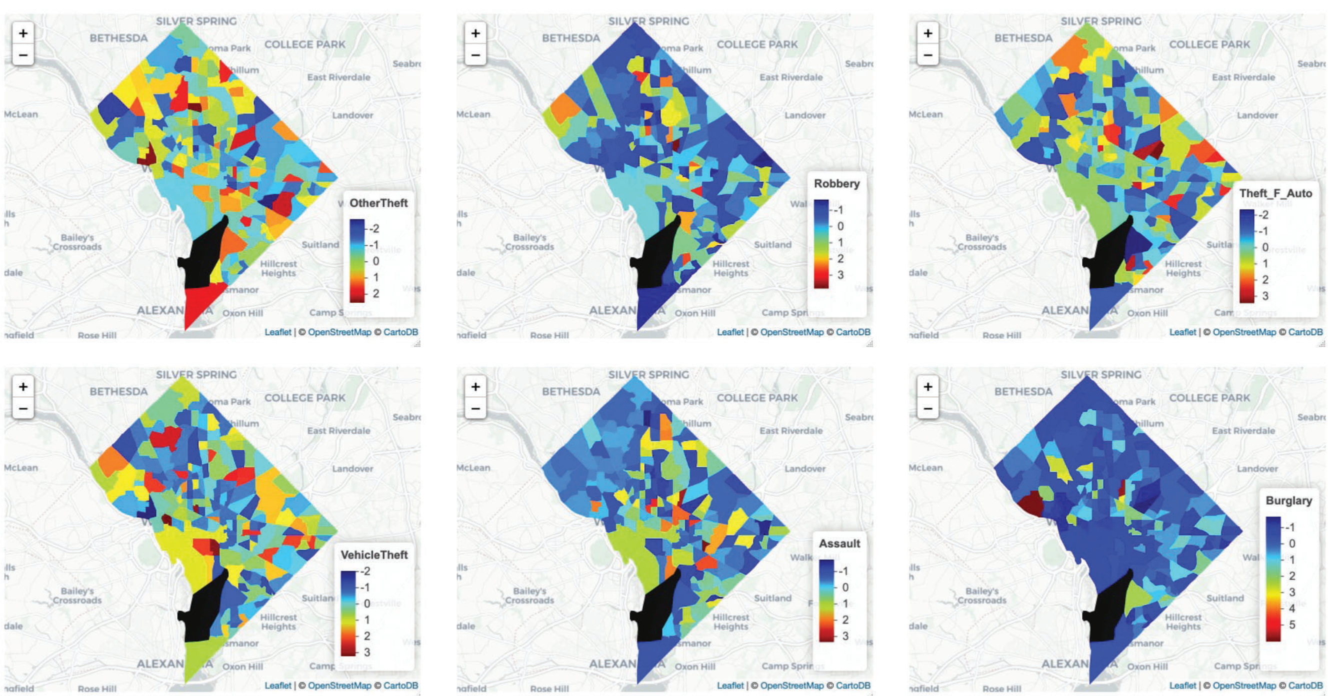


Figure 7. Standardized residuals for 179 census tracts for six types of street crimes.

where $T_{ii}(\mathbf{u}, \mathbf{v}; \boldsymbol{\beta}^*, g)$ is defined in (12). Consequently, by replacing $\boldsymbol{\beta}$ and cross PCFs by their estimates, the standardized residual can be defined as $\hat{\epsilon}_{i,k}(\widehat{\boldsymbol{\beta}}) = \hat{\varepsilon}_{i,k}(\widehat{\boldsymbol{\beta}})/\hat{\sigma}_{i,k}(\widehat{\boldsymbol{\beta}}, \hat{g}^{\mathrm{r}})$, for $i = 1, \ldots, p$ and $k = 1, \ldots, K$.

Standardized residuals for all census tracts in Washington DC are illustrated in Figure 7. One census tract that does not have any reported street crime activities in January and February 2017 is indicated by the black color. Most standardized residuals

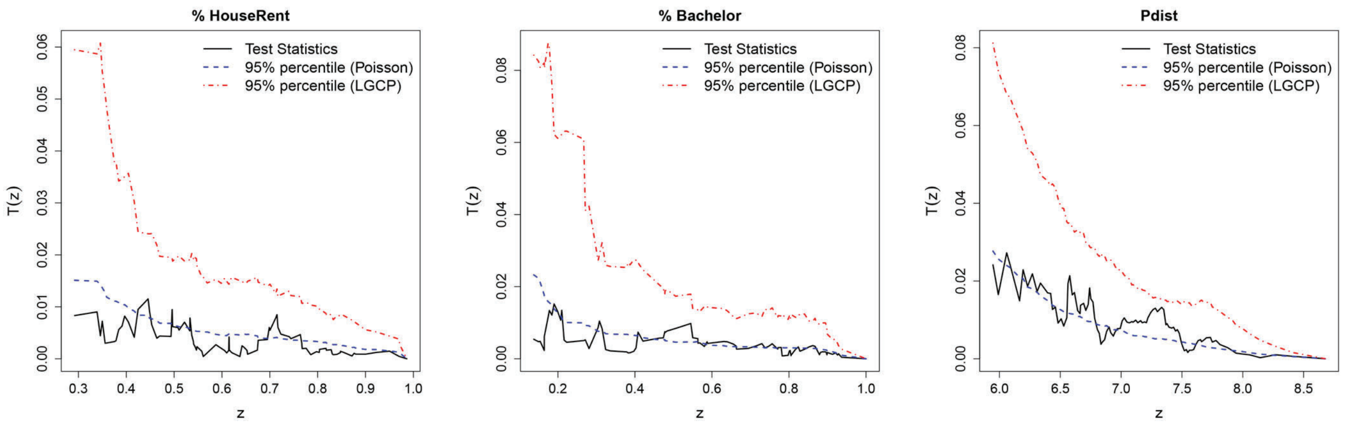


**Figure 8.** Estimated $T_j(z)$, $j = 4, 5, 7$ in (30) for the covariates % HouseRent, % Bachelor, and Pdist, together with 95% percentiles computed from the Poisson model and the LGCP model (plots for the other covariates are given in the supplementary materials).

are inside the range of $[-3, 3]$ for all six types of street crimes, indicating an adequate model fit. Finally, the apparent strong spatial correlations among the residuals further support the use of the proposed method.

### 6.4. Goodness-of-Fit Assessment

In addition to the graphical residual analysis in Section 6.3, it is useful to have a numerical summary of the overall goodness-of-fit of the fitted model. In this section, we propose a Monte Carlo test procedure inspired by the goodness-of-fit tests proposed in Dong and Yu (2021, 2019). To do so, we view the covariate vector $\mathbf{z}(\mathbf{u})$ as a realization of a random vector $\mathbf{Z}(\mathbf{u}) = (Z_1(\mathbf{u}), \dots, Z_q(\mathbf{u}))^\top$ and consider our data as a collection of marked points $(\mathbf{u}, \mathbf{Z}(\mathbf{u}), Y(\mathbf{u}))$ where $\mathbf{u} \in X^{\text{pl}}$ denotes a crime scene and $Y(\mathbf{u}) \in \{1, \dots, p\}$ is the type of crime committed at $\mathbf{u}$. We can define an empirical conditional distribution function as

$$\widehat{F}_{Y|Z_j}(y|z) = \frac{1}{N_j(z)} \sum_{\mathbf{u} \in X^{\text{pl}}} I\left[Y(\mathbf{u}) \le y, z_j(\mathbf{u}) \le z\right], \quad y = 1, \dots, p, z \in \mathbb{R}, \tag{27}$$

where $N_j(z) = \sum_{\mathbf{u} \in X^{\text{pl}}} I\left[z_j(\mathbf{u}) \le z\right]$, $j = 1, \dots, q$. This is an estimate of

$$F_{Y|Z_j}(y|z) = \mathrm{E}\frac{1}{M_j(z)} \sum_{\mathbf{u} \in X^{\text{pl}}} I\left[Y(\mathbf{u}) \le y, Z_j(\mathbf{u}) \le z\right], \tag{28}$$

where $M_j(z) = \mathrm{E}N_j(z)$. Under our model (3), one can show that an alternative estimator of (28) is given by

$$\widehat{F}^*_{Y|Z_j}(y|z) = \frac{1}{N_j(z)} \sum_{\mathbf{u} \in X^{\text{pl}}} \left[\sum_{i=1}^{y} \mathrm{p}_i(\mathbf{u}; \widehat{\boldsymbol{\beta}})\right] I\left[z_j(\mathbf{u}) \le z\right], \quad y = 1, \dots, p, z \in \mathbb{R}, \tag{29}$$

where $\mathrm{p}_i(\mathbf{u}; \widehat{\boldsymbol{\beta}})$ is defined in (5) with $\widehat{\boldsymbol{\beta}}$ obtained from (6).

Following Dong and Yu (2021), if model (3) is appropriate, one would expect $\widehat{F}_{Y|Z_j}(y|z)$ and $\widehat{F}^*_{Y|Z_j}(y|z)$ to be close for any $z$ and $j = 1, \dots, q$. Therefore, we can define for each covariate a test statistic as

$$T_j(z) = \sum_{i=1}^{p} \left|\widehat{F}_{Y|Z_j}(y|z) - \widehat{F}^*_{Y|Z_j}(y|z)\right| \Delta_{j,z}(i), \quad j = 1, \dots, q, \tag{30}$$

where $\Delta_{j,z}(i) = \widehat{F}_{Y|Z_j}(i|z) - \widehat{F}_{Y|Z_j}(i-1|z)$, $i = 1, \dots, p$.

It remains to evaluate the distribution of the $T_j(z)$'s. Dong and Yu (2021) suggested using a bootstrap exploiting that their pairs of covariate vectors and response variables are independent and identically distributed. This is not possible in our situation where the $(\mathbf{Z}(u), Y(u))$'s are not independent. In the following, we pursue some model-based bootstrap alternatives where we replace the unknown background intensity $\lambda_0(\cdot)$ by its nonparametric estimate (25) and try out some simple models for the correlation structure.

The simplest choice is the multivariate Poisson model, where we assume the $X_i$'s are independent inhomogeneous Poisson processes with intensity functions $\lambda_i(\cdot)$'s. Based on $B$ simulations from this model, one can compute $T^{\text{Poisson}}_{j,1}(z), \dots, T^{\text{Poisson}}_{j,B}(z)$ from which point-wise 95% percentiles can be estimated. Figure 8 shows the observed test statistic $T_j(z)$ as a function of $z$ for three covariates and the corresponding 95% percentiles based on simulations from the Poisson model (plots for the remaining covariates are similar and shown in the supplementary materials). The absence of between or within spatial correlation for the Poisson model means that the simulated parameter estimates based on (6) vary too little compared to their variation under the true data generating mechanism where spatial correlation is present as suggested in Figures 4 and 7. It is therefore not surprising that some observed $T_j(z)$'s are above the 95% percentile based on Poisson simulations.

To partially account for spatial correlation we next consider a second special case of model (3) where all point processes are independent log-Gaussian Cox processes (LGCPs), each with an exponential covariance function. Plugging in the kernel estimate of $\lambda_0(\cdot)$, we then estimate the correlation parameters for each LGCP separately using standard minimum contrast methods. Figure 8 shows that all observed $T_j(z)$'s are well below the 95% percentiles based on simulations of the fitted LGCPs. Thus, large values of the observed $T_j(z)$'s can be explained by sampling variation even when we only take into account correlation within each type of points and not between.

Plugging in the nonparametric estimate of $\lambda_0(\cdot)$ is not optimal but seems to be the only alternative at the moment to fit parametric models for the correlation structure. Developing a parametric model for the full correlation structure is beyond the scope of this article.

## 7. Concluding Remarks

We propose a flexible semiparametric model for multivariate point pattern data. The nonparametric component of the model takes into account features of the multivariate intensity function that are difficult to model or specify while the parametric part facilitates a study of effects of covariates on relative risks of occurrence of different types of points. Interesting conditional probability maps can be obtained from the parametric part and the intensity of a specific type of points can be estimated using the full dataset by combining the parametric estimate of the relative risk with an estimate of the nonparametric part.

Our multinomial logistic composite likelihood estimation approach does not require knowledge of the nonparametric model component. It is moreover well founded theoretically since we established the asymptotic properties of the estimation approach in a very general setting that does not require any independence assumptions, neither within nor between the different types of points.

Our nonparametric estimation approach allows us to estimate cross PCFs up to a common multiplicative factor. This is sufficient for estimating the covariance matrix of regression parameter estimates and for inferring ratios of cross PCFs. However, to infer individual cross PCFs, it seems necessary to introduce parametric models for the cross PCFs. We plan to pursue this in future work. There is also room for improving the kernel estimate (25) which can be criticized for ignoring the layout of the city.

Our methodology is applicable in very diverse fields. Our example application is within criminology where the estimated conditional probability maps disclose a remarkable structure in the occurrence of various types of street crimes in Washington DC. Other obvious areas of applications are disease mapping in epidemiology and studies of spatial distributions of plant and animal species in ecology. Our approach can further be extended to space-time multivariate point pattern data, which have attracted much interest in various research areas including criminology, see, for example, the thorough review in the recent article (Reinhart and Greenhouse 2018).

## Supplementary Materials

The supplementary materials for this article contain further simulation studies and plots, proofs, and auxiliary results.

## Funding

Kristian Bjørn Hessellund and Rasmus Waagepetersen gratefully acknowledge The Danish Council for Independent Research—Natural Sciences, grant DFF-7014-00074 "Statistics for point processes in space and beyond," and the "Centre for Stochastic Geometry and Advanced Bioimaging," funded by grant 8721 from the Villum Foundation. Xu's research is supported by NSF grant SES-1902195 and Yongtao Guan is supported by NSF grant DMS-1810591.

## References

Baddeley, A., Jammalamadaka, A., and Nair, G. (2014), "Multitype Point Process Analysis of Spines on the Dendrite Network of a Neuron," *Journal of the Royal Statistical Society*, Series C, 63, 673–694. [1]

Biscio, C. A. N., and Waagepetersen, R. P. (2019), "A General Central Limit Theorem and Subsampling Variance Estimator for $\alpha$-Mixing Point Processes," *Scandinavian Journal of Statistics*, 46, 1168–1190. [6]

Braga, A. A., Papachristos, A. V., and Hureau, D. M. (2014), "The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis," *Justice Quarterly*, 13, 633–663. [4]

Buerger, P. J., Cohn, E. G., and Petrosino, A. J. (1995), "Defining the Hot Spots of Crime: Operationalizing Theoretical Concepts for Field Research," *Crime and Place*, 4, 237–257. [4]

Choiruddin, A., Cuevas-Pacheco, F., Coeurjolly, J.-F., and Waagepetersen, R. (2020), "Regularized Estimation for Highly Multivariate Log Gaussian Cox Processes," *Statistics and Computing*, 30, 649–662. [1]

Cohen, J., Gorr, W. L., and Olligschlaeger, A. M. (2007), "Leading Indicators and Spatial Interactions: A Crime-Forecasting Model for Proactive Police Deployment," *Geographical Analysis*, 29, 105–127. [1,3,4]

Cox, D. R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society*, Series B, 34, 187–220. [3]

Cronie, O., and Van Lieshout, M. N. M. (2018), "A Non-Model-Based Approach to Bandwidth Selection for Kernel Estimators of Spatial Intensity Functions," *Biometrika*, 105, 455–462. [12]

Crowder, M. (1986), "On Consistency and Inconsistency of Estimating Equations," *Econometric Theory*, 2, 305–330. [5]

Diggle, P. J., and Rowlingson, B. S. (1994), "A Conditional Approach to Point Process Modelling of Elevated Risk," *Journal of the Royal Statistical Society*, Series A, 157, 433–440. [2,3]

Diggle, P. J., Zheng, P., and Durr, P. (2005), "Nonparametric Estimation of Spatial Segregation in a Multivariate Point Process: Bovine Tuberculosis in Cornwall, UK," *Journal of the Royal Statistical Society*, Series C, 54, 645–658. [1,2]

Dong, J., and Yu, Q. (2019), "Marginal Distribution Test for Checking Proportional Hazards Model Assumption," *Journal of Statistical Planning and Inference*, 201, 58–70. [14]

——— (2021), "Joint Distribution and Marginal Distribution Methods for Checking Assumptions of Generalized Linear Model," *Communications in Statistics—Theory and Methods*, 50, 1456–1476. [14]

Gorr, W. L., and Lee, Y. (2015), "Early Warning System for Temporary Crime Hot Spots," *Journal of Quantitative Criminology*, 31, 25–47. [4]

Guan, Y., Waagepetersen, R. P., and Beale, C. M. (2008), "Second-Order Analysis of Inhomogeneous Spatial Point Processes With Proportional Intensity Functions," *Journal of the American Statistical Association*, 103, 769–777. [2]

Haberman, C. (2017), "Overlapping Hot Spots? Examination of the Spatial Heterogeneity of Hot Spots of Different Crime Types," *Criminology and Public Policy*, 16, 633–660. [1,3,4]

Jalilian, A., Guan, Y., Mateu, J., and Waagepetersen, R. (2015), "Multivariate Product-Shot-Noise Cox Point Process Models," *Biometrics*, 71, 1022–1033. [1,2,6]

Rajala, T., Murrell, D. J., and Olhede, S. C. (2018), "Detecting Multivariate Interactions in Spatial Point Patterns With Gibbs Models and Variable Selection," *Journal of the Royal Statistical Society*, Series C, 67, 1237–1273. [1,2,7]

Ratcliffe, J. (2002), "Burglary Reduction and the Myth of Displacement," Trends and Issues in Crime and Criminal Justice No. 232. [4]

——— (2004), "Crime Mapping and the Training Needs of Law Enforcement," *European Journal on Criminal Policy and Research*, 10, 65–83. [4]

Reinhart, A., and Greenhouse, J. (2018), "Self-Exciting Point Processes With Spatial Covariates: Modelling the Dynamics of Crime," *Journal of the Royal Statistical Society*, Series C, 67, 1305–1329. [10,15]

Waagepetersen, R. P., Guan, Y., Jalilian, A., and Mateu, J. (2016), "Analysis of Multispecies Point Patterns by Using Multivariate Log-Gaussian Cox Processes," *Journal of the Royal Statistical Society*, 65, 77–96. [1,2,6]

Weisburd, D., Maher, L., Sherman, L., Buerger, M., Cohn, E., and Petrosino, A. (1993), "Contrasting Crime General and Crime Specific Theory: The Case of Hot Spots of Crime," *Advances in Criminological Theory*, 4, 45–70. [1,3]

Wilson, J. Q., and Kelling, G. L. (1982), "Broken Windows: The Police and Neighborhood Safety," *Atlantic Monthly*, 249, 29–38. [4]

Xu, G., Waagepetersen, R., and Guan, Y. (2019), "Stochastic Quasi-Likelihood for Case-Control Point Pattern Data," *Journal of the American Statistical Association*, 114, 631–644. [2]

Zimmerman, D., Sun, P., and Fang, X. (2012), "Likelihood-Based Estimation of Spatial Intensity and Variation in Disease Risk From Locations Observed With Error," *Statistics and Its Interface*, 5, 207–219. [2]