Fall 2014

Spatial marked point processes: Models and inferences

Yen-Ning E Huang
Purdue University

Graduate School Form 30
(Updated 11/20/2014)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yen-Ning E. Huang

Entitled
SPATIAL MARKED POINT PROCESSES: MODELS AND INFERENCES

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Hao Zhang

Tonglin Zhang

Mary Ellen Bock

Bruce Craig

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Hao Zhang

Approved by Major Professor(s): _____

_____

Approved by: Jun Xie 12/02/2014

Head of the Department Graduate Program Date

SPATIAL MARKED POINT PROCESSES:

MODELS AND INFERENCES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Yen-Ning E. Huang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2014

Purdue University

West Lafayette, Indiana

To my family.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation and thanks to my advisors Professor Hao Zhang and Professor Tonglin Zhang, they have been wonderful mentors for me. Professor Hao Zhang has been supportive all the time and he has offered me so much advice. He is always patient to guide me in the right direction. He trained me to seek for research topics that are worth investigating and taught me how to present the important findings in an effective way. Professor Tonglin Zhang has always been enthusiastic about potential research topics. He shared many of his good ideas with me and thus led me to be a more energetic person. Both of my advisors have provided me insightful discussions about the research and their guidance is truly invaluable. I could not have finished my dissertation without their generous help.

I also appreciate Professor Bo Li who is always willing to assist me under any circumstances. I am thankful that Professor Bruce Craig has offered his time to serve in my committee and help me get through all the trouble when scheduling the defense of my dissertation. I am also very grateful to Professor Mary Ellen Bock and Professor Rebecca Doerge for their support over the past few years. They have been kind to me since the first day we met. They are always resourceful and I have learned a lot from them. They have made our department a very unique place where everyone is connected like a big family. I also want to thank Becca Pillion, Marian Duncan and Diane Martin for their various help. It has been my pleasure to be part of this fantastic department.

I would like to acknowledge Dr. Keith Woeste for providing the dataset for me to do case studies. I also want to express my gratitude to my colleagues and friends in our department, including Juan Hu, Cheng Liu, Inkyung Choi, Xian He, Piyas Chakraborty, Whitney Huang, Vitara Pungpapong and Glen DePalma. I am also thankful for the precious friendship from some of my friends outside the department,

including Chun-Hsien Lu, Hsiu-Han Yang, Hsin-Lin Lu, Yi-Cheng Chiu, and many others. They have brought happiness and delight to my life at Purdue.

Special thanks should go to my parents and grandma who support me wholeheartedly. My parents inspired me to start this journey in Statistics and I would not be able to come this far without their encouragement. Words can not express how grateful I am to my parents for all the efforts they have made for me to become who I am, and to keep me optimistic and motivated all the time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Huang, Yen-Ning Ph.D., Purdue University, December 2014. Spatial Marked Point Processes: Models and Inferences. Major Professor: Hao Zhang and Tonglin Zhang.

A spatial marked point process describes the locations of randomly distributed events in a region, with a mark attached to each observed point. Nowadays, the availability of spatiotemporal data is increasing and many spatiotemporal models are studied with applications in a wide range of disciplines. Spatial marked point processes are then extended to spatiotemporal marked point processes if time component is taken into account. In general, the marks can be quantitative or categorical variables. Independence between points and marks is a convenient assumption, but may not be true in practice. Tests for independence between points and marks are proposed previously, though only a few models have been developed to describe dependence between points and marks. In this dissertation, I focus on quantitative marks and the objective is to develop models for both spatial and spatiotemporal marked point processes when points and marks are dependent.

Three approaches to describe dependence between points and marks are studied in this dissertation, while the first two approaches are for spatial marked point processes and the last one is for spatiotemporal marked point processes. First, we derive a covariance function of additive models for marked point processes. This covariance function carries information of dependence between points and marks, which can be used in kriging to make predictions of marks at unknown locations. We expect to obtain better prediction results by using this covariance function when the points and marks are dependent.

The second approach is to consider *intensity-dependent* models. We study both univariate and bivariate intensity marked Log Gaussian Cox processes and apply

an empirical Bayesian estimation procedure with implementation of Markov Chain Monte Carlo methodology for statistical inference. We allow dependence between marks after conditioning on the intensity which is more flexible than conditional independence assumption. The influence of adding cross covariance in modeling bivariate marks is also explored. The first two approaches are applied to model the dependence between points and marks of a white oak data.

The last approach is to consider the partially stationary spatiotemporal marked point process, where the distribution of the spatiotemporal marked process is invariant under parallel shift of time, but may not be invariant under parallel shift of points or marks. It can be classified as a *location-dependent* model. To determine the potential usefulness of this approach, we illustrate through two typical examples in natural hazards: a forest wildfire study and an earthquake study. The results show that the distribution of marks and points is significantly different at local scale. It is expected that the proposed approach will have wide applications in the study of natural hazards.

1. INTRODUCTION AND LITERATURE REVIEW

A point process $N = \{x_i\}$ consists of a set of points in an Euclidean space which represents the times or locations that events occur. If measurements $m(x_i)$ are taken and associated with points x_i, then a marked point process $N_m = \{x_i; m(x_i), i = 1, \ldots, n\}$ appears. Marked point processes are applied to many areas of disciplines including forestry (Stoyan and Penttinen, 2000; Penttinen, Stoyan and Henttonen, 1992), plant pathology (Liu *et al.*, 2007), and epidemiology (Diggle, 2003; Elliott *et al.*, 2001). Often marks and points are assumed to be independent (or separable), which assumes the distribution of marks does not depend on points. Under this assumption, points and marks can be modeled separately. However, the independence assumption may be violated in practice, and the distribution of marks might be affected by point locations. Consequently, the intensity-dependent or location-dependent models may be considered, which can help us learn more about the conditional distribution of marks given points. Some tests for independence between points and marks have been proposed before, for example, Schlather, Ribeiro and Diggle (2004) consider the random-field model to investigate the point pattern and marks separately, and developed exploratory quantities to detect departures from the random-field model. Guan (2006) proposes some formal and graphical tests for independence between marks and points. One of the formal tests uses the mark K function and mark G function, and the other is nonparametric which uses the idea of permutation. We will apply some of their methods to our simulations and real data application.

Although testing methods for independence between marks and points are previously studied, only a few models have been developed for the dependence structure between points and marks. The intensity-dependent marking strategies for the stationary log Gaussian Cox process are discussed in Ho and Stoyan (2008) and Diggle, Menezes and Su (2010). In their settings, the marks are conditionally independent

given the intensity $\Lambda(x_i)$, and $E[m(x_i)|\Lambda(x_i)]$ only depends on $\Lambda(x_i)$. Most of the work assumes only the mean but not variance depends on the intensity, but Myllymäki and Penttinen (2009) extends the setting where the variance depends on the intensity as well.

Other than models for spatial point processes, spatiotemporal point processes are also of interest. It can be extended from spatial point processes by adding the time component into a model. It has wide applications in many disciplines such as environment monitoring, disease mapping and functional MRI. We will consider spatiotemporal marked point processes in this dissertation with applications to natural hazards data.

This dissertation is organized as follows. Later in this chapter, we will give a brief review of point processes, marked point processes as well as spatiotemporal marked point processes. We will introduce our study for marked point processes in Chapter 2. In Chapter 3 we will describe the empirical Bayesian modeling of bivariate marked point processes with simulation studies and application to a white oak dataset. We propose a location-dependent partial stationarity model in Chapter 4, which includes simulation studies and applications to forest wildfire data and earthquake data. We conclude this work by a brief summary of our approaches and provide a discussion with potential direction of future work.

1.1 Poisson Process and Poisson Cluster Process

A point process in one-dimension can be used to model the arrival times or occurring times for a particular event of interest. When we consider a random pattern of points in d-dimension (in most applications $d = 2$ or 3), it is called a spatial point process. The process can be used to describe occurrences of events in space, e.g., locations of trees, earthquakes, the positions of stars in the galaxy, or the locations of patients diagnosed with a specific disease, etc. The analysis of spatial point processes

is useful in seismology, forestry, astrostatistics, spatial epidemiology and many other areas.

Among different models, the Poisson process plays an important role in the study of point processes. A Poisson process is often used because it is one of the simplest stochastic processes. In one-dimensional case, one usually studies the arrival time of the i^{th} event $T_i, i = 1, 2, \ldots$ where $T_i < T_j$ for $i < j$, and also the inter-arrival time $T_{i+1} - T_i$. One of the properties for Poisson processes is that the inter-arrival times are independent and exponentially distributed. Furthermore, the number of arrivals in disjoint time intervals are also independent. If we use $N(a, b]$ to denote the numbers of arrivals in the time interval $(a, b]$ for $0 < a \leq b$, then we have $N(a, b] \sim \text{Poisson}(\lambda(b - a))$ where λ is the intensity of the process representing the expected number of arrivals per unit time.

We can extend the one-dimensional process to a two-dimensional process, i.e. the spatial point process. The occurrence of events in one-dimension has an automatic order (in time), but there is no order in higher dimensional spaces. Instead of studying the inter-arrival times or counting the number of events in specific time intervals, for spatial point processes we consider the number of points in a bounded region D, denoted as $N(D)$. Below we review the definition of the spatial point process based on the method of the counting measure (e.g. p. 42 in Daley and Vera-Jones, 2003).

Let N be the point process defined on a complete separable metric space $\mathcal{S}$ and $N(D)$ be the number of points in a Borel set $D \subseteq \mathcal{S}$. If D is bounded, i.e., $|D|$ is finite, then $N(D) < \infty$ where $|D|$ denotes the Lebesgue measure of D. A spatial point process N is derived if $\mathcal{S}$ is specified as the Euclidean space $\mathbb{R}^d$. Based on the approach of the counting measure, the distribution of N is defined as

$$P[N(D_1) = n_1, \cdots, N(D_k) = n_k], \tag{1.1}$$

where $\{D_1, \cdots, D_k\}$ is a collection of bounded Borel subsets of $\mathcal{S}$. According to the distribution given by Equation (1.1), the *kth order intensity function* of N is defined as

$$\lambda_k(\mathbf{s}_1, \cdots, \mathbf{s}_k) = \lim_{|d\mathbf{s}_i| \to 0, i=1,\cdots,k} \left\{ \frac{\mathbb{E}[\prod_{i=1}^{k} N(d\mathbf{s}_i)]}{\prod_{i=1}^{k} |d\mathbf{s}_i|} \right\}, \tag{1.2}$$

where $(\mathbf{s}_1, \cdots, \mathbf{s}_k)$ are distinct points and $d\mathbf{s}_i$ is an infinitesimal region containing $\mathbf{s}_i$. The first-order intensity $\lambda(\mathbf{s}) = \lambda_1(\mathbf{s})$ can be derived when $k = 1$. If the first-order intensity $\lambda(\mathbf{s})$ is constant, then N is homogeneous; otherwise it is inhomogeneous. The mean structure of N can be expressed as

$$\mu(D) = \mathbb{E}[N(D)] = \int_D \lambda(\mathbf{s}) d\mathbf{s},$$

which represents the expected number of points within D.

Based on the second-order intensity function $\lambda_2(\mathbf{s}_1, \mathbf{s}_2)$, the covariance structure of N can be expressed as

$$\begin{aligned} \mathbb{C}\text{ov}[N(D_1), N(D_2)] &= \int_{D_1} \int_{D_2} [\lambda_2(\mathbf{s}_1, \mathbf{s}_2) - \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)] d\mathbf{s}_2 d\mathbf{s}_1 + \int_{D_1 \cap D_2} \lambda(\mathbf{s}) d\mathbf{s} \\ &= \int_{D_1} \int_{D_2} [g(\mathbf{s}_1, \mathbf{s}_2) - 1] \lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2) d\mathbf{s}_2 d\mathbf{s}_1 + \int_{D_1 \cap D_2} \lambda(\mathbf{s}) d\mathbf{s}, \end{aligned}$$

where $g(\mathbf{s}_1, \mathbf{s}_2) = \lambda_2(\mathbf{s}_1, \mathbf{s}_2)/[\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)]$ is the pair correlation function.

The spatial Poisson process can be defined if $N(D_i)$ follows a Poisson distribution with mean $\mu(D_i)$ for every bounded region D_i, and $N(D_1), \ldots, N(D_k)$ are independent for any disjoint bounded regions. That is,

$$P_k[N(D_1) = n_1, \cdots, N(D_k) = n_k] = \prod_{i=1}^{k} \frac{\mu^{n_i}(D_i)}{n_i!} e^{-\mu(D_i)},$$

where $D_1, \cdots, D_k$ are disjoint Borel subsets in $\mathbb{R}^d$. If N is Poisson, then $g(\mathbf{s}_1, \mathbf{s}_2) = 1$ for any $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$.

We can transform a point process to construct a new one. Common constructions include clustering, thinning, mapping and superposition (Stoyan *et al.*, 1995). We will give more details on clustering here since the cluster point process are considered

in our simulation study in Chapter 4. To generate a cluster process, we replace each point x in the existing point process N by a random set of finite points N_x, and the collection of all the clusters forms a cluster point process $N_c = \cup_{x \in N} N_x$. The original point process N is called the parent process and the cluster N_x for x is called the offspring process. The clusters for different points are often assumed to be independent.

Realizations of many spatial point processes can be simulated by R. The plots for Poisson processes and Poisson cluster processes are displayed on the next page. In Figure 1.1 and Figure 1.2 there are a few realizations for these two point processes with different structures and parameters. Figure 1.1 shows three different realizations of Poisson processes with uniform intensity 50 in the unit square (upper panel), and three different realizations of Poisson processes with independent bivariate Normal intensity with center $(0, 0)$ and standard deviation 0.5. In Figure 1.2 there are four realizations of Poisson cluster processes with uniform intensity 50 for the parent process and independent bivariate Normal intensity for the offspring process of mean 4 and different standard deviations 0.02, 0.04, 0.06 and 0.08. It is easier to observe the clusters if we specify smaller standard deviation for the intensity of offspring points.

1.2 Log Gaussian Cox Process (LGCP)

Sometimes the Poisson process is not enough to model point patterns, but it can be used to construct more complex models. The Cox process is one of those that is constructed from a Poisson process, which can be used to model clustered or aggregated point patterns. If the intensity function of a Poisson process is itself a realization of a random field, a Cox process is derived. It is an extension of Poisson process which is first introduced by Cox (1955) under the name *doubly stochastic Poisson process*, but nowadays is usually called the Cox process. Let $\lambda = \{\lambda(x) : x \in \mathbb{R}^2\}$ be the realization of the underlying random field Λ. The point process is called a Cox process with intensity surface Λ if given $\Lambda = \lambda$ it is an inhomogeneous Poisson

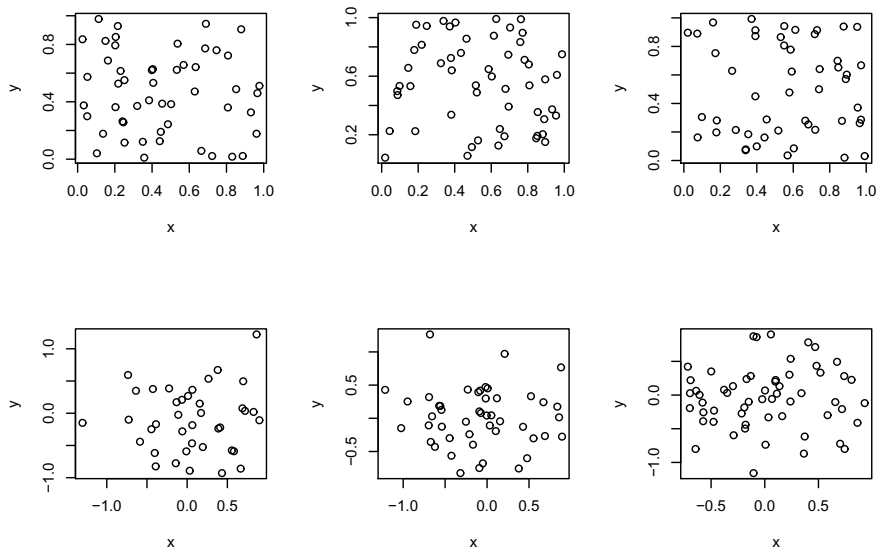

Figure 1.1. Realizations of Poisson processes with uniform intensity 50 in the unit square (upper panel) and Normal intensity centered at (0,0) with standard deviation 0.5 (lower panel).

process with intensity function λ. It is expected more points to occur in the areas with larger Λ. If different processes are chosen for Λ, we can construct different point processes which makes Cox processes a very flexible tool.

There are a few parametric models for Cox processes. For example, the shot-noise G Cox process (SNGCP) (Brix, 1999), and the log Gaussian Cox process (LGCP) which has been introduced in astronomy by Coles and Jones (1991) and in statistics by Møller *et al.* (1998). Among the above different models, the LGCP is our main focus here.

If $\log \Lambda = Z$ is a Gaussian process, then the point process X is called a log Gaussian Cox process (LGCP). Møller *et al.* (1998) extend the definition of an LGCP to multivariate LGCPs, and Brix and Diggle (2001), Brix and Møller (2001) give the definition of multivariate spatiotemporal LGCPs. In general, we can consider a Cox process where the intensity function $\Lambda = h(Z)$ for other non-negative function h,

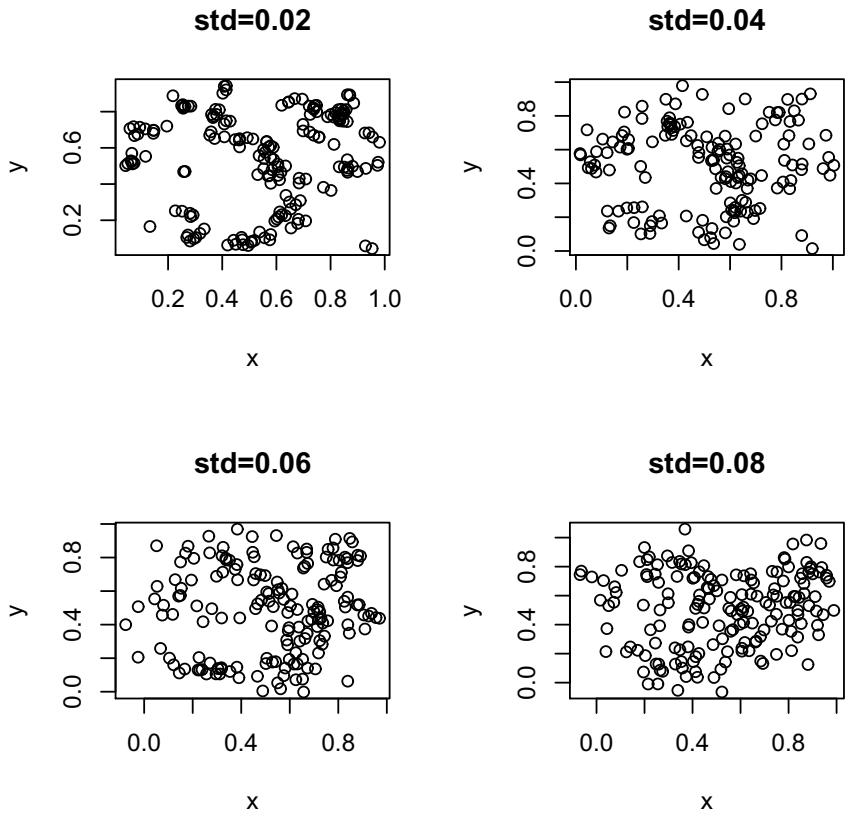

Figure 1.2. Realizations of Poisson cluster processes with uniform intensity 50 in the unit square and Normal intensity for the offspring points with mean 4 and different standard deviations 0.02, 0.04, 0.06 and 0.08.

but in particular the LGCP where $h(Z) = \exp(Z)$ is more convenient. Let μ_Z and C_Z denotes the mean and covariance of the Gaussian process Z. Figure 1.3 shows four realizations of log Gaussian Cox processes. Two realizations of homogeneous LGCP with exponential covariance function are given in the upper panel, and two realizations of inhomogeneous LGCP with Gaussian and Matérn covariance are given in the lower panel.

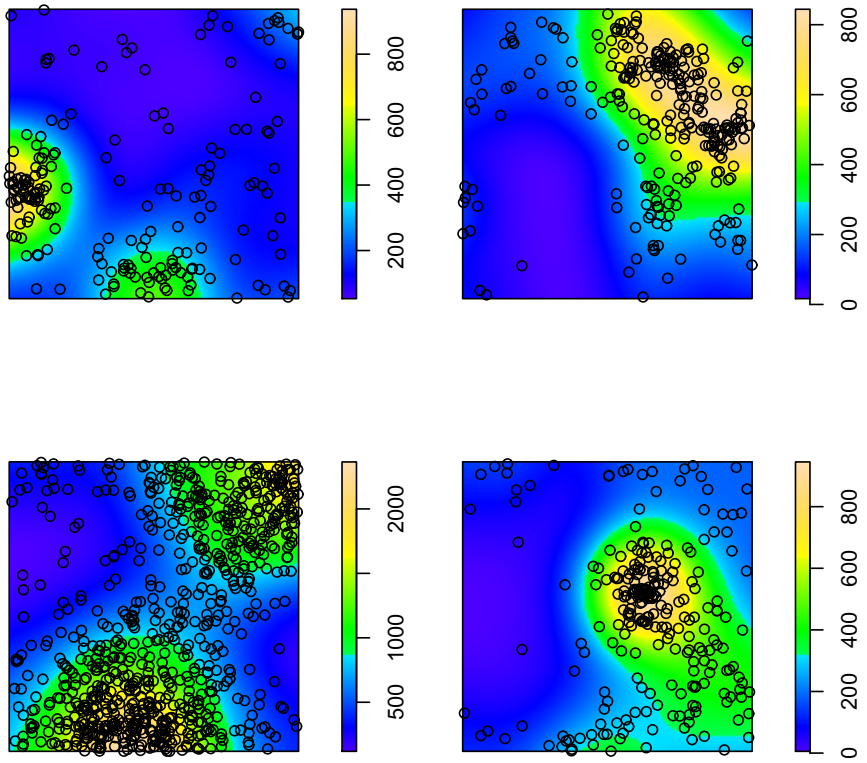

Figure 1.3. Realizations of LGCPs. *Upper left* : homogeneous LGCP with exponential covariance ($\sigma^2 = 1$, $\phi = 0.2$); *upper right* : homogeneous LGCP with exponential covariance ($\sigma^2 = 2$, $\phi = 0.2$); *lower left* : inhomogeneous LGCP with Gaussian covariance ($\sigma^2 = 2$, $\phi = 0.2$); *lower right* : inhomogeneous LGCP with Matérn covariance ($\sigma^2 = 2$, $\phi = 0.2$, $\nu = 0.6$) .

1.3 Marked Point Process (MPP)

For an unmarked point process $N = \{x_i, i = 1, \ldots, n\}$, if we attach additional information called marks $m(x_i)$ to the points, then collectively a marked point process $N_m = \{x_i; m(x_i), i = 1, \ldots, n\}$ is derived. The process is defined on the product space of points and marks. Figure 1.4 shows an example of a marked point pattern with the locations of 126 pine saplings in a Finnish forest marked by their heights. This

dataset is from R-library `spatstat` (Baddeley and Turner, 2005). On the left panel each circle represents the location of a pine sapling where the radius of circle are proportional to its height. The right panel shows density estimates of the point pattern using Gaussian kernel.

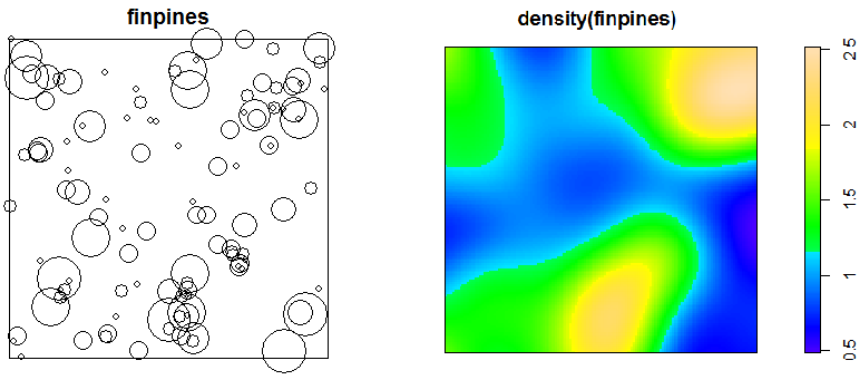

Figure 1.4. Finnish Pines marked by their heights

The simplest assumption for a marked point process is that the marks and points are independent. Under this assumption we can model the point pattern and the marks separately. Different concepts of independence between marks and points are discussed before. For an independent marked point process, the marks are independent and identically distributed and also independent of the points, or we have a weakly independent stationary marked point process if the points and marks are both given by a stationary sequences but the two sequences are independent of each other. Also the random-field model (Takahata, 1994; Mase, 1996) for unbiased sampling in geostatistics, which assumes the marks are generated by a random field independent of the points. Under this definition the independent marked point process is a special case of random-field model, but the weakly independent stationary marked point process may not be a random-field model.

Although different assumptions of independence are studied before, most of the time the independence assumption between points and marks does not hold in prac-

tice. There is a natural ordering of the points for a one-dimensional marked point process, and both the history of point process and history of mark sequence can be discussed. Cox and Isham (1980) summarize different types of dependence between points and marks given the history of the marked point process at time t, and also give comments to a few possibilities they list. Here we focus on spatial marked point processes so that the dependence structure between points and marks may be different from one-dimensional marked point process. In some cases where the mark values might be affected by the local point density, the intensity-dependent models for the marks should be considered and we expect to learn more about the mark distribution given the points.

1.3.1 Test for Independence Between Points and Marks of MPPs

Similar to classical geostatistics, some characteristics such as the mark variogram (Cressie, 1993) or the mark correlation (Isham, 1985) for the marks in MPPs have been developed.

So far not many models have been developed for describing the dependence structure between points and marks, but there are some exploratory analysis and tests for independence between points and marks proposed before. For example, Schlather, Ribeiro and Diggle (2004) introduce exploratory quantities $E(h)$ and $V(h)$ for stationary and isotropic marked point processes to describe the strength and range of interactions between points and marks. The method is based on

$$E(h) = E\{m(\mathbf{o})|\mathbf{o}, \mathbf{h} \in N, \|\mathbf{h}\| = h)\}$$

and

$$V(h) = E[\{m(\mathbf{o}) - E(h)\}^2|\mathbf{o}, \mathbf{h} \in N, \|\mathbf{h}\| = h)].$$

These two functions are constant if the marked point process follows the stationary random-field model, where their values are the mean and variance of the mark,

respectively. If they are quite different from constant values it would be a sign indicating that the marked point process deviates from the random-field model. In R-library `spatstat` there are functions Emark and Vmark which can generate plots of E and V for exploratory analysis about mark independence diagnostics. We will apply these two functions for diagnostics in our simulation and case study in Chapter 3.

There are also some other formal and graphical tests proposed to check for independence between marks and points. Guan (2006) uses the mark K function and the mark G function to test for the independence assumption, as well as a nonparametric approach based on the idea of permutation. And Guan (2007) proposes another test for independence between marks and points using a subsampling approach. Zhang and Zhuang (2014) propose a local odds ratio approach to estimate the localized dependence structure between marks and points, which can be applied to natural hazards studies.

In chapter 3, we apply the test using mark K function and the method is briefly introduced below. Let $\mu = E\{m(x)\}$. Under the null hypothesis of independence between marks and points, note that $K_m(r) = \mu \times K_p(r)$, we estimate μ by a slope estimator obtained from regressing $\hat{K}_m(r)$ on $\hat{K}_p(r)$:

$$\hat{\mu} = \frac{\int_0^R \hat{K}_m(r)dr}{\int_0^R \hat{K}_p(r)dr},$$

where R is an upper bound on the lag vector of observations since the dependence between marks and points may exist only at a small physical scale. Under the null hypothesis of independence, the slope estimator $\hat{\mu}$ should be close to the average of the observed marks, $\bar{M}$. We can establish the asymptotic normality of $\hat{\mu} - \bar{M}$ under some conditions on the underlying process, and therefore come up with a statistic of the form $(\hat{\mu} - \bar{M})^2 / \widehat{\mathrm{Var}}(\hat{\mu} - \bar{M})$, where the denominator is a consistent estimator of the true variance. Then we have an approximate size α test that the null hypothesis of independence will be rejected if the test statistic is greater than the critical value of Chi-squared distribution with one degree of freedom, given the significance level α.

1.3.2 Additive Model for MPPs

The additive model for marked point processes can be written as follows (Schlather *et al.*, 2004):

$$m(x) = pZ(x) + (1-p)M(x) \tag{1.3}$$

where $Z(x)$ is a stationary and isotropic Gaussian random field with covariance $C(r)$. The function $M(x)$ can be some characteristics to describe the intensity of point processes in different ways. For instance, the number of points of N that fall within the circle of a given radius centered at x, or the distance from x to its nearest neighbor. Different values of p correspond to different strength of dependence between marks and points. We will consider the additive model and derive a covariance function for the marks in Chapter 2.

1.3.3 Intensity Marked LGCP

Let $\boldsymbol{M} = (m(x_1), \ldots, m(x_n))'$. A marked point process $[X, \boldsymbol{M}]$ is called an intensity-marked log Gaussian Cox process if the marks depend on the underlying point process through its (random) intensity $\{\Lambda(s)\} = \exp(Z(s))$, where $\{Z(s)\}$ is a Gaussian random field with mean μ_Z and covariance function $C_z(r)$. The intensity-dependent marking strategies for the stationary log Gaussian Cox process are discussed by Ho and Stoyan (2008); Diggle, Menezes and Su (2010). In their settings, the marks are conditionally independent given the intensity, and $E[m(x_i)|\Lambda(x_i)]$ is a function of $\Lambda(x_i)$. Myllymäki and Penttinen (2009) further consider the markings that allow variance to depend on the intensity, while in most of other work they assume only the mean but not variance depends on the intensity.

One of the cases considered in Myllymäki and Penttinen (2009) is the Gaussian intensity-marked Cox process

$$m(x_i)|\Lambda(x_i) \sim N(a + b\Lambda(x_i), c^2\Lambda(x_i) + d^2).$$

With the sign of parameter b, this model can be used to describe positive or negative dependence between the marks and intensity. They also propose Gamma intensity-marked Cox process and some mark characteristics are derived.

1.4 Spatiotemporal Process Modeling

The spatial point process can be extended to the spatiotemporal case. Instead of only locations, the data also include the time component in the spatiotemporal point process. If a two-dimensional spatial coordinate is used, then it is often represented by the longitude and latitude. If we use a three-dimensional spatiotemporal coordinate, then it is often represented by the longitude, latitude, and time. Many theories of spatial-temporal point processes can be generalized from spatial point processes. However, the temporal component enables a natural ordering of the points that does not exist in spatial processes in general.

Now the point process becomes $N = \{x_i = (s_i, t_i), i = 1, \ldots, n\}$ where s_i denotes the location and t_i denotes the time, and N is within a bounded region $S \times T \in \mathbb{R}^2 \times \mathbb{R}$. We still use $N(D)$ to denote the number of points in a bounded region $D \in \mathbb{R}^2 \times \mathbb{R}$. The spatiotemporal Poisson process can be defined similarly as in Section 1.1, but $|D|$ becomes the volume of D, the intensity function becomes $\lambda(s, t)$ for inhomogeneous spatiotemporal point process, and the mean of $N(D)$ becomes $\int_D \lambda(s, t) dsdt$.

1.5 Spatiotemporal Marked Point Process (SMPP)

Spatiotemporal marked point processes (SMPPs), which can be extended from spatiotemporal point processes, have been widely considered in data involving both the spatiotemporal coordinates of events and the corresponding measurements attached to each point. Methods of MPPs are often used to model different types of natural hazards located in space and time. Many successful applications of MPPs can be found in literature, including MPP modeling and prediction of earthquakes (Holden *et al.*, 2003; Ogata, 1988; Ogata and Katsura, 1993; Vere-Jones, 1995), where

each earthquake is recorded by a magnitude and a space-time coordinate. MPPs for forest wildfires have been discussed by Zhang and Zhuang (2014), where each wildfire is represented by its area burned and space-time coordinate. An SMPP is derived if time is considered.

The definition of MPPs is well-established and can be found in many textbooks (e.g. in Daley and Vere-Jones, 2003; Karr, 1991). We also introduced the MPPs in Section 1.3. The definition for MPPs can be easily generalized to SMPPs. Overall, an SMPP is a point process defined on the product of space, time, and marks, but the concept has its own life in applications. Let $\mathcal{S}$ be a measurable subset of $\mathbb{R}^2$ (or a two-dimensional sphere $\mathcal{S}^2$) and $\mathcal{M}$ be a measurable subset of $\mathbb{R}^q$. Then, an SMPP N with points in $\mathcal{S}\times\mathbb{R}$ and marks in $\mathcal{M}$ is a point process on $\mathcal{S}\times\mathbb{R}\times\mathcal{M}$ with the additional property that the underlying point process (which is derived by ignoring the mark) is itself a spatiotemporal point process and for any bounded $A\times B\in\mathscr{B}(\mathcal{S}\times\mathbb{R})$ there is $N(A\times B\times\mathcal{M})<\infty$, where $N(A\times B\times C)$ is the number of points in $A\times B\times C$ for $A\in\mathscr{B}(\mathcal{S})$, $B\in\mathscr{B}(\mathbb{R})$, and $C\in\mathscr{B}(\mathcal{M})$, respectively.

Based on the distribution of N, the k-th order intensity function of N (if it exists) can be defined as

$$\lambda_k[(\mathbf{s}_1,t_1,m_1),\cdots,(\mathbf{s}_k,t_k,m_k)] \tag{1.4}$$

$$=\lim_{|d\mathbf{s}_i\times dt_i\times dm_i|\to 0,i=1,\cdots,k}\left\{\frac{E[\prod_{i=1}^k N(d\mathbf{s}_i\times dt_i\times dm_i)]}{\prod_{i=1}^k|d\mathbf{s}_i\times dt_i\times dm_i|}\right\} \tag{1.5}$$

where $(\mathbf{s}_i,t_i,m_i)$ are distinct, $d\mathbf{s}_i\times dt_i\times dm_i$ is an infinitesimal region containing $(\mathbf{s}_i,t_i,m_i)$, and $|d\mathbf{s}_i\times dt_i\times dm_i|$ is the Lebesgue measure of $d\mathbf{s}_i\times dt_i\times dm_i$.

In most of applications, main objectives are often to assess the strength of the mean and interaction between events, which are usually described by the first-order intensity function and the pair correlation function. Using the first-order intensity function denoted as $\lambda(\mathbf{s},t,m)=\lambda_1(\mathbf{s},t,m)$, the mean measure of N can be expressed as

$$\mu(A\times B\times C)=\int_A\int_B\int_C\lambda(\mathbf{s},t,m)dmdtd\mathbf{s},$$

for $A \in \mathscr{B}(\mathcal{S})$, $B \in \mathscr{B}(\mathbb{R})$, and $C \in \mathscr{B}(\mathcal{M})$. Using the pair correlation

$$g[(\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)] = \frac{\lambda_2[(\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)]}{\lambda(\mathbf{s}_1, t_1, m_1)\lambda(\mathbf{s}_2, t_2, m_2)} \tag{1.6}$$

we can express the covariance structure of N as

$$\mathrm{Cov}[N(A_1 \times B_1 \times C_1), N(A_2 \times B_2 \times C_2)] \tag{1.7}$$

$$= \int_{A_1}\int_{B_1}\int_{C_1}\int_{A_2}\int_{B_2}\int_{C_2} \{g[(\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)] - 1\}\lambda(\mathbf{s}_1, t_1, m_1) \tag{1.8}$$

$$\times \lambda(\mathbf{s}_2, t_2, m_2) dm_2 dt_2 d\mathbf{s}_2 dm_1 dt_1 d\mathbf{s}_1 + \mu[(A_1 \cap A_2) \times (B_1 \cap B_2) \times (C_1 \cap C_2)] \tag{1.9}$$

$$\tag{1.10}$$

for any $A_1, A_2 \in \mathscr{B}(\mathcal{S})$, $B_1, B_2 \in \mathscr{B}(\mathbb{R})$, and $C_1, C_2 \in \mathscr{B}(\mathcal{M})$. If N is a Poisson SMPP, then $g[(\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)] = 1$ for any $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{S}$, $t_1, t_2 \in \mathbb{R}$, and $m_1, m_2 \in \mathcal{M}$. Otherwise, there is either attractions or repulsions among events contained in N, which indicates that we cannot ignore the interaction between points.

1.6 Estimation of Intensity Functions

After the introduction of different types of point process, it is of interest to estimate the first-order intensity of a point process. Below we will review methods of estimation from nonparametric and parametric approaches.

1.6.1 Nonparametric Approach: Kernel Methods

For a homogeneous point process where the intensity $\lambda(\mathbf{s}) = \lambda$ is constant, a natural estimate of the intensity is $\hat{\lambda} = N(\mathcal{S})/|\mathcal{S}|$ where $\mathcal{S}$ denotes the whole study region. This is actually the maximum likelihood estimate of λ when N is a homogeneous Poisson process. In the case of inhomogeneous point processes where the intensity is not constant, the Berman-Diggle estimator (Berman and Diggle, 1989) are widely used and can be defined as follows:

$$\widehat{\lambda_{BD}(\mathbf{s})} = \frac{N(b(\mathbf{s}, h)) \cap A}{|b(\mathbf{s}, h) \cap A|}$$

for $\mathbf{s}_0 \in A \subseteq \mathcal{S}$, where $b(\mathbf{s}, h)$ denotes the open ball centered at $\mathbf{s}$ with radius $h > 0$. This can be considered as a kernel estimate of the intensity function using uniform kernel.

A a nonparametric kernel estimate of the intensity function $\lambda(\mathbf{s})$ is defined as

$$\hat{\lambda}_h(\mathbf{s}) = \sum_{i=1}^{n} K_h(\mathbf{s} - \mathbf{s}_i)/p_h(\mathbf{s}). \tag{1.11}$$

(Diggle, 1985). Here K_h is a kernel function with bandwidth $h > 0$, $K_h(\mathbf{s}) = K(\mathbf{s}/h)/h^d$ for d-dimensional point process. The kernel function K can be a full symmetric density function on $\mathbb{R}^d$. The denominator $p_h(\mathbf{s}) = \int_{\mathcal{S}} K_h(\mathbf{s} - \mathbf{s}')d\mathbf{s}'$ is an edge correction. Usually this estimator will be sensitive to the choice of bandwidth h, whereas the choice of kernel function K is of less importance. We need the edge correction because the points outside the boundary of $\mathcal{S}$ are not included in the numerator even if they are within distance h of $\mathbf{s}$. We might get a biased estimator if we ignore the edge effects especially when h is large.

1.6.2 Parametric Approach: Composite Likelihood Methods

The idea of composite likelihood is developed by Lindsay (1988), which has been used in different settings when a full maximum likelihood approach is not available or computationally intensive. Notice that composite likelihoods are referred to as various names like psudolikelihood (Besag, 1975; Molenberghs and Verbeke, 2005; Baddeley and Turner, 2000), approximate likelihood (Stein, Chi and Welty, 2004) or partial likelihood (Cox, 1975). Varin, Reid and Firth (2011) provide a survey on developments of composite likelihood methods.

Composite likelihood is a product of likelihood components where the components do not necessarily need to be independent. Each component can be a conditional or marginal density depending on the context, and therefore the derivative of the composite log-likelihood is an unbiased estimating equation. When the full maximum likelihood is not feasible, composite likelihood is a substitution for estimation. The composite likelihood estimator can be consistent even if the full maximum likelihood

estimator is not. Composite likelihoods may be viewed as misspecified likelihoods. They sometimes represent the part of model that we are more confident to model with. The full joint distribution may not be specified but we can still get a robust estimator from composite likelihood approach. Some investigations about the robustness of composite likelihood inference are provided in Lele and Taper (2002), and Wang and Williamson (2005) through simulation studies. For analysis of longitudinal data in which missing values are common, it is found that some versions of composite likelihood are robust to the specification of the missing data structure (Parzen *et al.*, 2007; Yi, Zeng and Cook, 2009).

The definition of composite likelihood is as follows. Consider n-dimensional random variable Y with probability density function $f(y;\theta)$, where θ is a p-dimensional parameter vector. Let $\mathcal{L}_i(\theta;y) \propto f(y \in A_i;\theta)$ be the likelihoods of a set of marginal or conditional events $\{A_i, i = 1,\ldots,n\}$. The composite likelihood can be written as the weighted product $\mathcal{L}_C(\theta;y) = \prod_{i=1}^{n} \mathcal{L}_i(\theta;y)^{w_i}$, where w_i are non-negative weights to be selected. Equivalently, we can also write down the composite log-likelihood

$$\ell_C(\theta;y) = \sum_{i=1}^{n} w_i \log \mathcal{L}_i(\theta;y).$$

And the maximum composite likelihood estimator can be found by maximizing $\ell_C(\theta;y)$ as $\hat{\theta} = \arg\max_\theta \ell_C(\theta;y)$, which is usually solved by $\dot{\ell}_C(\theta;y) = 0$ where $\dot{\ell}_C$ is the gradient of ℓ_C.

If the weights are chosen to be all equal then they can be ignored. The unweighted composite likelihood approach has been used to estimate parameters of point process data, see Schoenberg (2005) and Waagepetersen (2007). Guan and Shen (2010) introduced a weighted estimating equation approach for inhomogeneous spatial point process, where they maximize the weighted likelihood

$$\mathcal{L}(\theta) = \left\{ \prod_{x \in N \cap \mathcal{S}} \lambda(x;\theta)^{w(x)} \right\} \exp\left\{ -\int_{\mathcal{S}} w(s)\lambda(s;\theta)ds \right\}$$

for estimating the parameters θ we assume in the intensity function $\lambda(s;\theta)$. This can be applied to both Poisson and non-Poisson processes.

In particular applications, choosing unequal weights can improve efficiency and the selection of weights is also worth investigating. We will consider a kernel weighted composite likelihood approach for analyzing point process data in Chapter 4.

2. COVARIANCE FUNCTION OF AN ADDITIVE MODEL FOR MPPS

In general, spatial data are correlated and modeling the covariance structure is a crucial step in analyzing spatial data. Usually the goal is to make predictions for the spatial variables of interest at unobserved locations. In geostatistics, kriging is a commonly used method which is the best linear unbiased predictor that minimizes the mean squared error. It is proper to make kriging predictions for geostatistical data where the locations of observations are assumed to be fixed. The objective of this chapter is to derive a covariance function that can be used in kriging prediction for MPPs. Here we treat the marks as a realization of an underlying spatial process without considering the distribution of the point process. This covariance function contains information of dependence between points and marks, and it has potential to result in better predictions with a more appropriate interpretation. Using a covariance function that carries dependence between points and marks to make kriging prediction can be considered as an intermediate method between geostatistics and marked point process approach.

The plan of this chapter is as follows. We derive a covariance function for an additive model for MPPs in Section 2.1, and provide a simulation study in Section 2.2. We apply univariate additive models to a white oak data in Section 2.3. Section 2.4 provides a discussion.

2.1 Additive Model for MPPs

To capture the dependence between points and marks of a marked point process $N = \{x_i; m(x_i), i = 1, \ldots, n\}$, we consider the following additive model for marks (Schlather, Ribeiro and Diggle, 2004)

$$m(x) = pZ(x) + (1-p)M(x) \tag{2.1}$$

where $Z(x)$ is a stationary and isotropic Gaussian random field with covariance $C_Z(h) = \text{Cov}(Z(s+h), Z(s)) = \sigma^2 \text{exp}(-h/\phi)$. Here we let $M(x)$ to be the number of points of N that fall in the ball of a given radius r centered at x. A smaller value of p indicates stronger dependence between points and marks.

By Law of total covariance, we can write the covariance function between marks at locations x_1, x_2 as below:

$$\begin{aligned}
&\text{Cov}(m(x_1), m(x_2)) \\
&= E(\text{Cov}(m(x_1), m(x_2)|N(A)) + \text{Cov}(E(m(x_1)|N(A)), E(m(x_2)|N(A))) \\
&= p^2 \text{Cov}(Z(x_1), Z(x_2)) + \text{Cov}(\mu_Z + (1-p)M(x_1), \mu_Z + (1-p)M(x_2)) \\
&= p^2 \text{Cov}(Z(x_1), Z(x_2)) + (1-p)^2 \text{Cov}(M(x_1), M(x_2)).
\end{aligned} \tag{2.2}$$

where μ_Z is the mean of $Z(x)$. Some existing covariance function can be specified to model $\text{Cov}(Z(x_1), Z(x_2))$ (e.g., exponential covariance), then $\text{Cov}(m(x_1), m(x_2))$ can be derived explicitly for homogeneous Poisson processes with intensity λ. When $d = ||x_1 - x_2|| \leq 2r$, we can write $M(x_1)$ and $M(x_2)$ as the number of points in disjoint areas A_1, A_2 and A_3, where A_2 is the intersection of two balls centered at x_1 and x_2 with radius r:

$$M(x_1) = N(A_1) + N(A_2)$$

$$M(x_2) = N(A_2) + N(A_3)$$

The covariance between $M(x_1)$ and $M(x_2)$ can be written as

$$\begin{aligned}
&\mathrm{Cov}(M(x_1), M(x_2)) \\
&= E(M(x_1)M(x_2)) - E(M(x_1))E(M(x_2)) \\
&= E(N(A_2)^2) - E(N(A_2))^2 = \mathrm{Var}(N(A_2)) \\
&= \lambda(2r^2\arccos(d/2r) - d/2\sqrt{4r^2 - d^2}),
\end{aligned}$$

and the correlation function between $M(x_1)$ and $M(x_2)$ is

$$\begin{aligned}
&\rho(M(x_1), M(x_2)) \\
&= \frac{\mathrm{Cov}(M(x_1), M(x_2))}{\sqrt{\mathrm{Var}(M(x_1))\mathrm{Var}(M(x_2))}} \\
&= \frac{\lambda(2r^2\arccos(d/2r) - d/2\sqrt{4r^2 - d^2})}{\lambda\pi r^2} \\
&= \frac{1}{\pi r}(2r\arccos(d/2r) - d\sqrt{1 - (d/2r)^2}).
\end{aligned}$$

If $d > 2r$, then the intersection is empty and $\mathrm{Cov}(m(x_1), m(x_2)) = p^2\mathrm{Cov}(Z_1(x_1), Z_2(x_2))$ because $\mathrm{Cov}(M(x_1), M(x_2)) = 0$. Figure 2.1 shows plots for this correlation function with the parameter r set to be 5, 10, 15 and 20.

2.2 Simulation Study

We studied the prediction performance from univariate ordinary kriging, i.e., the best linear unbiased prediction (BLUP) when the mean is unknown. The ordinary kriging was considered for univariate variable using the covariance function we proposed in the previous section, and the results were compared to some existing and commonly used covariance functions like exponential and spherical covariance functions. The marks $m(x_i), i = 1, 2, \ldots, n$ were considered as the underlying process which we made predictions at some location x_0 using ordinary kriging predictor $\hat{m}(x_0) = \sum_{i=1}^{n} a_i m(x_i)$, where the BLUP minimized the mean squared error $E\left(\sum_{i=1}^{n} a_i m(x_i) - m(x_0)\right)^2$ under the constraint $\sum_{i=1}^{n} a_i = 1$. Let $V(\theta) = \mathrm{Var}(\boldsymbol{M})$ be the covariance matrix of the marks with parameters θ, where the parameters

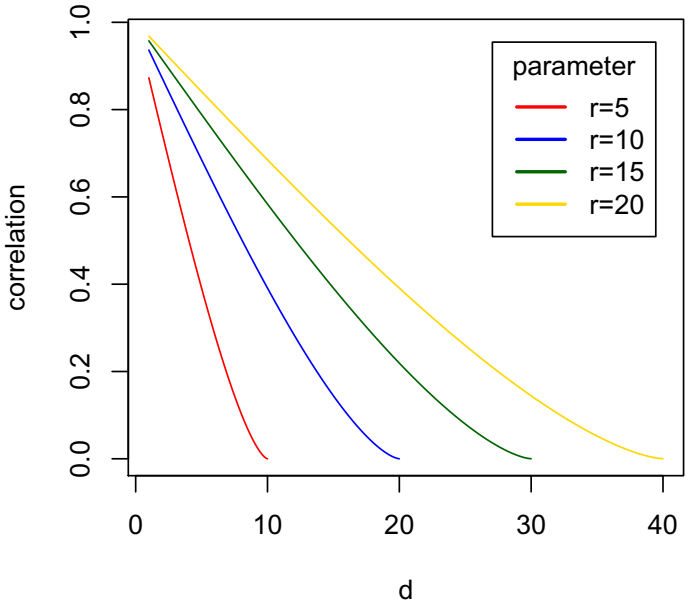

Figure 2.1. Plots for New correlations with parameter $r = 5, 10, 15, 20$.

are to be estimated from data. Existing methods for estimating θ including least squares based methods, composite likelihood estimation (Curriero and Lele, 1999), maximum likelihood estimation (Mardia and Marshall, 1984) or Bayesian estimation (Agarwal and Gelfand, 2005). Here we used the weighted least squares estimator (Cressie, 1985) to estimate θ. After we obtained $\hat{\theta}$, the ordinary kriging predictor is $\hat{m}(x_0) = \hat{\mu} + V(\hat{\theta})^{-1}\hat{k}(\boldsymbol{M} - 1\hat{\mu})$, where $\hat{\mu} = 1'V(\hat{\theta})^{-1}\boldsymbol{M}/1'V(\hat{\theta})^{-1}1$ and $k = \text{Cov}(\boldsymbol{M}, m(x_0)) = (\text{Cov}[m(x_1), m(x_0)], \ldots, \text{Cov}[m(x_n), m(x_0)])'$. We specified different covariance functions to the mark process and obtained the ordinary kriging predictors. The comparison between the use of different covariance functions was based on five-fold cross-validation results, mainly the mean squared error (MSE).

Locations of points were generated from homogeneous Poisson process with intensity $\lambda = 100$ and 500 within $[0, 100] \times [0, 100]$ square. For univariate case from (2.1), we set $\phi = 25, 50, 75$, $\sigma^2 = 50, 100$ and $p = 0.2, 0.4, 0.6, 0.8$. The radius r defined in $M(x)$ was specified to be 3 and 5. The marks were generated from (2.1) given different combinations of parameters. Exponential and spherical covariances were selected to compare with the newly derived covariance function, while exponential covariance corresponds to the case when $p = 1$ in (2.2). We simulated 1000 realizations for each combination of parameters. The MSE of five-fold cross-validation was calculated for each of 1000 realizations, and the average of 1000 MSEs for each combination of parameters was reported in Table 2.1 and Table 2.2. We only displayed the results from exponential and the new covariance since spherical covariance was outperformed by the other two. In most of cases, the MSE was smaller when σ^2 and r were smaller for both covariance functions. In general, we got smaller MSE from newly derived covariance than exponential covariance for $p = 0.2, 0.4$. When $p = 0.6, 0.8$, it was possible to get smaller MSE from either of the two covariance functions. We expected the newly derived covariance to perform better when p is small, because smaller p represented higher dependence between the points and marks. From our simulation results, it was still likely that the newly derived covariance function performed better in prediction even when $p > 0.5$.

2.3 Case Study: White Oak Data

2.3.1 Dataset Description

In this section, we analyzed a data set collected from a 23-year-old limited range white oak progeny trial in southern Indiana (Harrison-Crawford State Forest). We considered the point pattern of 334 trees with significant forking or severe defects. The range of tree height was from 3.5 m to 14.2 m with a mean value of 10.6 m. This data set actually motivated this work.

2.3.2 Comparison Between Different Covariance Functions

The white oak data should be considered as a point process data, but we treated the mark itself as the underlying process in geostatistics and do not include the distribution of points here. Nevertheless, the covariance function derived in Section 2.1 is aiming to capture the dependence between points and marks to some degree, and we used it to do kriging prediction on the height of white oak data. Five-fold cross-validation was used to compare the prediction performance of different covariance structures specified in kriging.

Table 2.3 shows the mean squared error (MSE) from kriging using four different covariance functions. Exponential covariance corresponds to the case when $p = 1$ in (2.2), the mark related covariance (Mark) corresponds to the case when $p = 0$, and "New" represents the newly derived covariance where p was a parameter to be estimated and $0 < p < 1$. For this dataset, we got smaller MSE when we used Mark or New covariance than exponential or spherical covariance. Spherical covariance resulted in worst prediction performance while New covariance did the best job for this dataset.

2.4 Discussion

In this chapter, we have derived a covariance function of additive model for mark point processes. If we use kriging method instead of point process model to do analysis, we expect the newly derived covariance function can help to get better prediction results when the points and marks are dependent. We specify exponential covariance for $C_Z(h)$, which is the first component of this newly derived covariance, but other choices like Matérn covariance can also be considered. We use weighted least squares estimator to estimate the parameters in the covariance function, but other estimation methods as mentioned in Section 2.2 can also be considered.

So far we only consider the univariate additive model for marked point processes. It is also possible to further consider the bivariate case of additive model:

$$m_i(x) = p_i Z_i(x) + (1 - p_i)M(x), i = 1, 2.$$

For a simple case where $p_1 = p_2 = p$, we have

$$\begin{aligned}
&\mathrm{Cov}(m_1(x_1), m_2(x_2)) \\
&= E(\mathrm{Cov}(m_1(x_1), m_2(x_2)|N(A)) + \mathrm{Cov}(E(m_1(x_1)|N(A)), E(m_2(x_2)|N(A))) \\
&= p^2\mathrm{Cov}(Z_1(x_1), Z_2(x_2)) + \mathrm{Cov}(\mu_1 + (1-p)M(x_1), \mu_2 + (1-p)M(x_2)) \\
&= p^2\mathrm{Cov}(Z_1(x_1), Z_2(x_2)) + (1-p)^2\mathrm{Cov}(M(x_1), M(x_2)).
\end{aligned}$$

Some existing bivariate cross covariance function can be specified to model $\mathrm{Cov}(Z_1(x_1), Z_2(x_2))$ (e.g., bivariate Matérn cross covariance function), and then we can derive $\mathrm{Cov}(m_1(x_1), m_2(x_2))$ explicitly for homogeneous Poisson processes with intensity λ. It is of interest to compare the bivariate version of this covariance function with other existing bivariate covariance, say bivariate Matérn covariance and see how much difference they will produce.

Table 2.1

Five-fold cross-validation results from exponential covariance (Exp) and newly derived covariance (New) for simulated homogeneous Poisson process with intensity $\lambda = 100$, and other parameters $\sigma^2 = 50, 100$, $\phi = 25, 50, 75$, $r = 3, 5$ and $p = 0.2, 0.4, 0.6, 0.8$.

σ^2	ϕ	r	Cov	$p = 0.2$	$p = 0.4$	$p = 0.6$	$p = 0.8$
50	25	3	Exp	22.64	29.13	25.77	49.09
			New	21.33	28.15	26.89	49.67
		5	Exp	46.13	45.89	40.88	61.68
			New	45.39	45.07	41.43	61.94
	50	3	Exp	28.50	40.54	48.20	55.40
			New	26.07	36.85	49.54	49.60
		5	Exp	58.48	49.17	57.28	60.11
			New	44.43	42.79	54.85	51.14
	75	3	Exp	30.37	30.81	33.70	70.83
			New	28.63	30.46	28.65	66.81
		5	Exp	78.85	58.50	42.40	72.50
			New	76.67	54.67	43.80	70.11
100	25	3	Exp	25.71	50.62	47.92	65.43
			New	23.55	47.26	48.66	61.30
		5	Exp	70.63	59.46	50.16	80.51
			New	67.01	56.20	51.99	81.04
	50	3	Exp	30.21	43.10	55.11	68.65
			New	25.13	37.20	51.22	70.13
		5	Exp	63.56	51.35	50.92	73.45
			New	60.22	47.25	48.84	75.60
	75	3	Exp	33.88	33.95	34.32	79.54
			New	28.15	29.37	35.67	80.21
		5	Exp	72.59	65.21	50.46	75.79
			New	67.65	62.35	51.56	77.85

Table 2.2

Five-fold cross-validation results from exponential covariance (Exp) and newly derived covariance (New) for simulated homogeneous Poisson process with intensity $\lambda = 500$, and other parameters $\sigma^2 = 50, 100$, $\phi = 25, 50, 75$, $r = 3, 5$ and $p = 0.2, 0.4, 0.6, 0.8$.

σ^2	ϕ	r	Cov	$p = 0.2$	$p = 0.4$	$p = 0.6$	$p = 0.8$
50	25	3	Exp	23.65	26.82	37.25	59.61
			New	22.78	25.97	34.39	58.00
		5	Exp	65.94	48.80	46.15	56.72
			New	59.94	46.99	41.91	53.09
	50	3	Exp	28.77	26.26	39.36	44.45
			New	24.54	24.29	42.02	42.64
		5	Exp	61.62	48.40	45.59	54.50
			New	54.94	42.55	44.71	57.53
	75	3	Exp	23.15	24.92	30.77	51.17
			New	20.49	22.29	33.48	52.80
		5	Exp	63.73	41.51	42.45	49.64
			New	55.41	39.44	39.16	45.16
100	25	3	Exp	24.98	33.58	57.50	83.10
			New	22.63	28.85	60.22	85.24
		5	Exp	63.67	53.97	62.42	78.21
			New	54.32	48.13	63.25	80.26
	50	3	Exp	25.83	33.12	40.89	64.39
			New	21.33	32.62	39.56	61.69
		5	Exp	64.83	52.42	45.60	69.32
			New	63.41	48.81	42.16	74.45
	75	3	Exp	24.87	30.51	39.96	75.90
			New	22.78	27.09	35.68	76.30
		5	Exp	69.40	56.90	43.89	78.37
			New	66.17	52.01	45.19	77.29

Table 2.3

Five-fold cross-validation results of four covariance functions: exponential (Exp), mark related covariance (Mark), the summation of Exp and Mark (New), and spherical covariance (Sph).

	C_1 (Exp)	C_2 (Mark)	C_3 (New)	C_4(Sph)
MSE	290.48	285.22	285.10	293.65
MSE (with nugget)	287.21	284.59	284.34	300.18

3. EMPIRICAL BAYESIAN MODELING OF BIVARIATE MPPS

In this chapter, we consider *intensity-dependent* models for both univariate and bivariate intensity marked log Gaussian Cox processes. An empirical Bayesian estimation procedure with implementation of Markov Chain Monte Carlo methodology is applied for statistical inference.

The structure of this chapter is as follows. In Section 3.1 we describe a more general case for univariate density-dependent markings for log Gaussian Cox processes, and further consider the case of bivariate marks. Details for empirical Bayesian modeling are given in Section 3.2, and simulation results are presented in Section 3.3. We apply both univariate and bivariate marking models to a white oak data in Section 3.4. Section 3.5 provides a discussion.

3.1 Intensity-Dependent Marked Point Processes

3.1.1 Univariate Marks

We assume the observed point patterns and marks $\{(x_i, m_i)\}$ are a partial realization of a marked point process $[X, M]$ where the marks possibly depend on the underlying point process through its intensity. We assume that the point process is a Cox point process so that the random intensity function, denoted by $\Lambda(x)$, is an exponential function of a stationary Gaussian process $Z(x)$, namely, $\Lambda(x) = \exp(Z(x))$. Furthermore, conditional on the point process, the mark is a Gaussian process whose mean function is a linear function of the intensity of the the point process while the covariance function, denoted by $C(x)$, is stationary and has a parametric form. These assumptions on the covariance function are simply meant to make the inferences man-

ageable as in geostatistics, and could be weakened (e.g., extending to non-stationary cases).

Given data (x_i, m_i), $i = 1, \ldots, n$, write $M = (m(x_1), \ldots, m(x_n))'$, $Z = (Z(x_1), \ldots, Z(x_n))'$. Under the assumptions, Z is multivariate normal with a constant mean and covariance matrix whose (i, j)th element is $C_Z(|x_i - x_j|)$ where C_Z is the covariance function. Given Z, M is multivariate normal with a mean vector

$$\boldsymbol{\mu} = (a + b\Lambda(x_1), \ldots, a + b\Lambda(x_n))' \tag{3.1}$$

where $\Lambda(x) = \exp(Z(x))$ denotes the intensity function, and the conditional covariance matrix is given by another stationary covariance function. This formulation of model allows explicit expression of the likelihood function, which is discussed in the next section.

A similar but different intensity-dependent marked point process is considered in Myllymäki and Penttinen (2009) where the marks are independent conditional on the intensity. Specifically,

$$m(x_i)|\Lambda(x_i) \sim N(a + b\Lambda(x_i), c^2\Lambda(x_i) + d^2). \tag{3.2}$$

We note this model implies that the correlation between $m(x_i)$ and $m(x_j)$ is bounded by the correlation between $\Lambda(x_i)$ and $\Lambda(x_j)$, which is an unnecessary constraint imposed on the model. Indeed, applying the following formula

$$\mathrm{Cov}(X, Y) = E\left(\mathrm{Cov}(X, Y|W)\right) + \mathrm{Cov}(E(X|W), E(Y|W)),$$

we get

$$\begin{aligned} \mathrm{Cov}(m(x_i), m(x_j)) &= b^2\mathrm{Cov}(\Lambda(x_i), \Lambda(x_j)), i \neq j \\ \mathrm{Var}(m(x_i)) &= c^2 E\Lambda^2(x_i) + b^2\mathrm{Var}(\Lambda(x_i)) \end{aligned} \tag{3.3}$$

Then it follows that $\mathrm{Corr}(m(x_i), m(x_j)) \leq \mathrm{Corr}(\Lambda(x_i), \Lambda(x_j))$ if $c \neq 0$.

It is not hard to see that our model does not have this constraint. Our model also extends to the multivariate case straightforwardly, as seen in the next section.

Finally, we make some comparisons between the marked point process and the process in geostatistics. The mark $m(x)$ marginally is a stationary process with a constant mean $\mu_m = a + b\exp\{\mu_Z + (1/2)C_Z(0)\}$ and covariance matrix

$$C_m(x) = C(x) + b^2\mu_m^2\{\exp(C_Z(x)) - 1\}$$

where μ_Z and C_Z denotes the mean and covariance function of the process $Z(x)$, and $C(x)$ is the conditional covariance function of $m(x)$. Here we used the covariance function of a lognormal process (Aitchison and Brown, 1957). Given the observed marks, it is also possible to make prediction for the mark using kriging. We note some differences between this kriging prediction and those given by our model for the marked point process.

First of all, in kriging prediction the randomness of locations of events $x_1, \ldots, x_n$ is not considered and this could lead to inferior result. In the marked point process, prediction of $m(x_0)$ based on the data $(x_0, m_i), i = 1, \ldots, n$, is given by the conditional mean

$$E(m(x_0)|x_i, m_i, i = 1, \ldots, n),$$

which has a smaller mean square error than $E(m(x_0)|m_i, i = 1, \ldots, n)$. The latter is the kind of prediction used in goestatistics.

3.1.2 Bivariate Marks

In this section, we consider a bivariate marked point process where there are two marks attached to each point. The two marks both depend on the intensity of the point process and may be correlated conditional on the intensity. The setting is similar to univariate case where the point process is assumed to be a Cox point process. Conditional on intensity, assume the marked point process is bivariate Gaussian with mean

$$\mu(x_i) = \begin{pmatrix} a_1 + b_1\Lambda(x_i) \\ a_2 + b_2\Lambda(x_i) \end{pmatrix}$$

and covariance function is bivariate Matérn

$$\mathbf{C}(\mathbf{h}) = \begin{pmatrix} C_{11}(h) & C_{12}(h) \\ C_{21}(h) & C_{22}(h) \end{pmatrix}$$

where each $C_{ii}(h) = \text{Cov}(m_i(s+h), m_i(s)) = \sigma_i^2 M(h|\nu_i, \alpha_i)$ for $i = 1, 2$ is a marginal Matérn covariance function (Guttorp and Gneiting ,2006; Matern, 1986) with spatial correlation at distance h

$$M(h|\nu, \alpha) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\alpha h)^{\nu} K_{\nu}(\alpha h). \tag{3.4}$$

Here $\sigma^2 > 0$ is the variance parameter, $\alpha > 0$ is the scale parameter, $v > 0$ is the smoothness parameter and K_{ν} is a modified Bessel function. And each cross covariance function is also a Matérn function

$$C_{12}(h) = C_{21}(h) = \rho_{12}\sigma_1\sigma_2 M(h|\nu_{12}, \alpha_{12})$$

with correlation coefficient ρ_{12} (Gneiting, Kleiber and Schlather, 2010). Details for estimating the covariance parameters and mark-related parameters are given in the next section.

3.2 Empirical Bayesian Modeling

Parameter estimation for log Gaussian Cox processes is known to be complex. Møller and Waagepetersen (2004) reviewed some inferential methods and proposed an empirical Bayesian method. In this section, we extend the empirical Bayesian method to the intensity-marked log Gaussian Cox process. We will first discuss the univariate case.

Conditional on the process $\{Z(s)\}$ (or $\{\Lambda(s)\}$), the joint probability density of the marks and points can be given explicitly

$$f(X, M|\theta_M, Z) = \left(\prod_{i=1}^{n} e^{Z(x_i)}\right) f_{M|Z}(m|z) e^{-\int_W \Lambda(s)ds} \tag{3.5}$$

where $\exp(Z(x_i)) = \Lambda(x_i)$ denotes intensity at location x_i and $f_{M|Z}(m|z)$ denotes the joint parametric density of the conditional mark distribution. For example, for our

model in the precious section, $M|Z \sim N(\boldsymbol{\mu_1}, \Sigma_1)$ where $\boldsymbol{\mu_1} = (a + b\Lambda(x_1), ..., a + b\Lambda(x_n))'$ and $(\Sigma_1)_{i,j} = \text{Cov}(m(x_i), m(x_j)) = \sigma^2 e^{-h/\phi}$ with $h = ||x_i - x_j||$.

For a given prior $f(\theta_M)$, the posterior distribution is

$$f(\theta_M, Z|X, M) \propto f(\theta_M) f(Z) f(X, M|\theta_M, Z).$$

The full Bayesian approach will be challenging to implement mainly because that some parameters in the covariance function are hard to estimate well (Zhang, 2004). Here we adopted the approach in Møller, Syversveen and Waagepetersen (1998) and estimate parameters in C_Z by a minimum contrast method, that minimizes

$$\int_{\epsilon}^{a_0} \{\hat{C}_Z(r)^\eta - C_Z(r)^\eta\}^2 dr \tag{3.6}$$

with respect to the parameters of $C_Z(r)$ using user-specified values $0 \leq \epsilon < a_0$ and $\eta > 0$. If we choose Matérn covariance function for $C_Z(r)$, then the parameters will be μ_Z, ϕ_Z and σ_Z^2. $\hat{C}_Z(r)$ in (3.6) is a non-parametric estimate of covariance function C_Z that is given through a kernel function. We refer to Møller, Syversveen and Waagepetersen (1998) for details.

In the next two sections, we discuss in more details the choice the prior distribution and the implementation of MCMC sampling.

For the bivariate marked point process defined previously, the estimation method extends easily. For example, the conditional density in (3.5) now becomes

$$f(X, M_1, M_2|\theta, Z) = \left(\prod_{i=1}^{n} e^{Z(x_i)} \right) f_{M_1, M_2|Z}(m_1, m_2|z) e^{-\int_W \Lambda(s) ds}$$

where $f_{M_1,M_2|Z}$ is the conditional density function of (M_1, M_2) given Z, which is $\sim N(\boldsymbol{\mu_2}, \Sigma_2)$. The conditional covariance matrix Σ_2 is given by the bivariate covariance function, which we assumed to be exponential. Therefore, Σ_2 can be written as

$$\Sigma_2 = \begin{pmatrix} \sigma_1^2 \exp(-\mathbf{d}/\phi_1) & \rho\sigma_1\sigma_2 \exp(-\mathbf{d}/\phi_{12}) \\ \rho\sigma_1\sigma_2 \exp(-\mathbf{d}/\phi_{12}) & \sigma_2^2 \exp(-\mathbf{d}/\phi_2) \end{pmatrix}$$

where $\mathbf{d}$ is the distance matrix between the observed points $\{x_1, ..., x_n\}$. To ensure Σ_2 is a valid covariance matrix, we need the parameters $\{\phi_1, \phi_2, \phi_{12}, \rho\}$ to satisfy some

constraints (Gneiting, Kleiber and Schlather, 2010). The Bayesian estimation of parameters related to bivariate Matérn covariance matrix is similar to the univariate model, which is briefly discussed in Section 3.2.2.

3.2.1 Prior Specification

Parameters of prior distribution of the Gaussian random field $\{Z(s)\}$ are estimated from data using minimum contrast method as mentioned before, and the prior distributions of the mean parameters a and b are chosen to be normal. The parameters σ^2 and ϕ are known to be estimated with relatively large variances, which can be justified theoretically under the fixed domain asymptotics (Zhang, 2004). For this reason, we reparametrize $(\Sigma_1)_{i,j} = \mathrm{Cov}(m(x_i), m(x_j)) = \sigma^2 \exp(-h/\phi) = \tau^2 \phi \exp(-h/\phi)$ where $\tau^2 = \sigma^2/\phi$, and the estimates for τ^2 will be estimated instead of σ^2. We use Inverse-Gamma prior for τ^2 and fix ϕ at a few different values.

Similarly, for bivariate marks we set normal priors for a_1, b_1, a_2, b_2, Inverse-Gamma priors for σ_1^2 and σ_2^2 and uniform priors for the rest four parameters $\phi_1, \phi_2, \phi_{12}, \rho$. As we mention before, there are some constraints for these parameters in order to make sure Σ_2 is a valid covariance matrix. Therefore we consider $\{\phi_1, \phi_2, \phi_{12}, \rho\}$ as a set of parameters and update them simultaneously. Initial values for the four parameters will be generated following the constraints as well as the proposed new values.

3.2.2 Implementation

For the univariate mark, the parameter vector to be estimated is $\theta_M = (a, b, \tau^2, \phi)$. We update the parameters θ_M and Gaussian random field Z one by one using Metropolis-Hasting algorithm in the MCMC procedure.

To update a, draw a candidate value a^* from proposal $q_1(a^*|a^{(t)})$. Accept the proposal, i.e., let $a^{(t+1)} = a^*$, with probability

$$R = \min\left\{1, \frac{f(\theta^*|X, M) q_1(a^{(t)}|a^*)}{f(\theta^{(t)}|X, M) q_1(a^*|a^{(t)})}\right\}$$

where $\theta^* = \{a^*, b^{(t)}, \tau^{2(t)}, \phi^{(t)}, Z^{(t)}\}$, and b, τ^2, ϕ and Z can be updated similarly. Fixing ϕ at some values is the same as choosing a discrete distribution as the proposal distribution. For parameters other than ϕ, we consider the proposal distributions of following forms:

$$\begin{aligned} a^* &\sim N(a^{(t)}, \sigma_a^2), \\ b^* &\sim N(b^{(t)}, \sigma_b^2), \\ \tau^{2*} &\sim \text{Unif}(\max(0, \tau^{2(t)} - \delta_{\tau^2}), \tau^{2(t)} + \delta_{\tau^2}), \\ Z^* &\sim N(Z^{(t)}, \sigma_{qz}^2) \end{aligned} \tag{3.7}$$

The values of $\sigma_a^2, \sigma_b^2, \delta_{\tau^2}$ and σ_{qz}^2 are specified in simulation study and application.

When applying the Metropolis-Hasting algorithm to the bivariate model, we need to be more careful when updating the parameters for bivariate marks due to the constraints the parameters have to satisfy. The parameters $a_1, a_2, b_1, b_2, \sigma_1^2$ and σ_2^2 can be updated one by one similar to univariate case, with Normal distributions as proposal for a_1, a_2, b_1, b_2 and uniform proposal for σ_1^2 and σ_2^2 as follows: $a_i^* \sim N(a_i^{(t)}, \sigma_{a_i}^2)$, $b_i^* \sim N(b_i^{(t)}, \sigma_{b_i}^2)$, $\sigma_i^{2*} \sim \text{Unif}(\max(0, \sigma_i^{2(t)} - \delta_{\sigma_i^2}), \sigma_i^{2(t)} + \delta_{\sigma_i^2})$, for $i = 1, 2$. However, for the rest four parameters $\{\phi_1, \phi_2, \phi_{12}, \rho\}$ we decide whether to accept the set of proposal $\{\phi_1^*, \phi_2^*, \phi_{12}^*, \rho^*\}$ or not at the same time. We generate ϕ_1^*, ϕ_2^* from uniform proposals first: $\phi_i^* \sim \text{Unif}(\max(0, {\phi_i}^{(t)} - \delta_{\phi_i}), {\phi_i}^{(t)} + \delta_{\phi_i})$ for $i = 1, 2$, and then generate ρ^*, ϕ_{12}^* from uniform proposals satisfying constraints of these four parameters. By updating the parameters this way we can make sure the covariance matrix Σ_2 is positive definite.

For approximating the integral $\int_W \Lambda(s)ds$ in (3.5), we follow the methods in Myllymäki and Penttinen (2009) which defined a partition of W with equal size rectangular sets and the center points forming a grid are denoted as $s_1, ..., s_k$. The integral is approximated as follows

$$\int_W \Lambda(s)ds \approx A \sum_{j=1}^{k} e^{Z(s_j)}.$$

where A is the size for each rectangular set. The values of $\{Z(s_j)\}$ for $j = 1, ..., k$ are simulated during MCMC iteration. We get better results if the grid is denser. In the next section we show the simulation results using 40×40 grid.

3.3 Simulation Study

The objectives of this section are to demonstrate exploratory mark independence diagnostics for marked point processes using Emark function in R-library `spatstat`, and assess how well our MCMC procedure can estimate the parameters for univariate and bivariate models. We simulate realizations of log Gaussian Cox process in a $[0, 100] \times [0, 100]$ window W using different combinations of parameters. Consider Gaussian random field $\{Z(s)\}$ with mean $\mu_Z = -3$ and the exponential covariance function with variance $\sigma_Z^2 = 0.5, 1$ and the range parameter $\phi_Z = 10$. Given these point processes, the marks were generated from (3.1) where the marks were auto-correlated conditioning on $\{Z(s)\}$, and the covariance function was chosen to be the exponential $\text{Cov}(m(x_i), m(x_j)) = \tau^2\phi \exp(-h/\phi)$, where we chose two combinations $(\tau^2, \phi) = (4, 2.5)$ and $(2, 5)$. In addition, we chose $a = 80$ and $b = -10, -5, 5, 10$ for the mean values of marks. Figure 3.1 and Figure 3.2 show the mark independence diagnostics generated from Emark and Vmark function in R-library `spatstat`. These two functions estimate $E(r)$ and $V(r)$, the conditional mean and variance of the mark attached to a point, given that there exists another point at distance r away. If the marks are independent with the points then we expect both $E(r)$ and $V(r)$ to be constant as the horizontal line $E^{\text{iid}}(r)$ and $V^{\text{iid}}(r)$, respectively. The other two lines $\hat{E}^{\text{iso}}(r)$ and $\hat{E}^{\text{trans}}(r)$ (or $\hat{V}^{\text{iso}}(r)$ and $\hat{V}^{\text{trans}}(r)$) are the estimates of $E(r)$ (or $V(r)$) for the marked point pattern proposed by Schlather, Ribeiro and Diggle (2004) as diagnostics for dependence between the marks and the points, with Ripley's isotropic correction (Ripley, 1976) and translation correction (Ohser and Stoyan, 1981). The two functions $E(r)$ and $V(r)$ seem to be non-constant particularly for small r values, which suggest there exists dependence between the marks and points.

We set normal priors for the parameters $a \sim N(90, 100)$ and $b \sim N(3, 100)$, and $\tau^2 \sim$ Inverse-Gamma$(3, 60)$. The proposal distributions for a, b and τ^2 were chosen as in (3.7) with $(\sigma_a^2, \sigma_b^2, \delta_{\tau^2}, \sigma_{qz}^2) = (5, 3, 2, 1)$. The range parameter ϕ was not estimated but rather fixed at twice the true values of ϕ. The main reason for this is to simplify the computations. However, we will estimate the range parameter in the bivariate model.

We ran 120,000 iterations of MCMC algorithm and omit the first 20,000 iterations as burn-in. The results for parameter estimations when $\sigma_Z^2 = 0.5$ using 40×40 grid points are given in Table 3.1. We can see when b is positive/negative we tend to overestimate/underestimate a. Estimates for parameter τ^2 have smaller variances particularly when the spatial correlation is stronger (i.e., a larger ϕ). Since the mean of marks is written as $a+b\Lambda(x)$ and estimates of $\{\Lambda(x)\}$ (or $\{Z(x)\}$) is usually very small, we found that the estimates of b have larger Monte Carlo standard error (MCSE). It took longer for b to converge than the other parameters. We used mcse.mat function in R-library `mcmcse` to calculate MCSE for all parameters and determine when to stop the MCMC iterations (Flegal and Jones, 2010).

For bivariate marks, we used the same point process simulated in the $[0, 100] \times [0, 100]$ window. The marks had parameters $(a_1, a_2, \sigma_1^2, \sigma_2^2, \phi_1, \phi_2, \phi_{12}, \rho) = (80, 20, 400, 20, 10, 10, 10, 0.8)$ and $b_1 = -5, 5$, $b_2 = -2, 2$. Similar to univariate case, we chose normal priors for parameters a_1, b_1, a_2, b_2 as follows: $a_1 \sim N(90, 100)$, $b_1 \sim N(3, 100)$, $a_2 \sim N(30, 100)$ and $b_2 \sim N(1, 100)$. The priors for σ_1^2 and σ_2^2 were Inverse-Gamma$(3, 60)$ and Inverse-Gamma$(3, 40)$ respectively. For ϕ_1, ϕ_2 and ϕ_{12} the priors were all Uniform$(5, 30)$ and the prior for ρ was Uniform$(0, 1)$. The proposals were described in Section 3.2.2, with $(\sigma_{a_1}^2, \sigma_{a_2}^2, \sigma_{b_1}^2, \sigma_{b_2}^2, \delta_{\sigma_1^2}, \delta_{\sigma_2^2}, \delta_{\phi_1}, \delta_{\phi_2}) = (5, 3, 2, 1, 10, 10, 2, 2)$.

We ran 160000 iterations of MCMC algorithm and omitted the first 20000 iterations as burn-in. The results for parameter estimations using 40×40 grid points are given in Table 3.2. We still observed that a_1, a_2 are overestimated when b_1, b_2 were positive and underestimated when b_1, b_2 were negative.

Estimates for $\sigma_1^2, \sigma_2^2, \phi_1, \phi_2$ and ϕ_{12} were higher than the true values, and the estimates of ρ were generally smaller than the true value, or even smaller when b_1 and b_2 have opposite signs. For the range parameters we had better estimates for ϕ_1 than ϕ_2 and ϕ_{12}. It was difficult to estimate these parameters related to covariance structure of bivariate marks. However, the estimates of ρ still indicated moderate correlation between the bivariate marks.

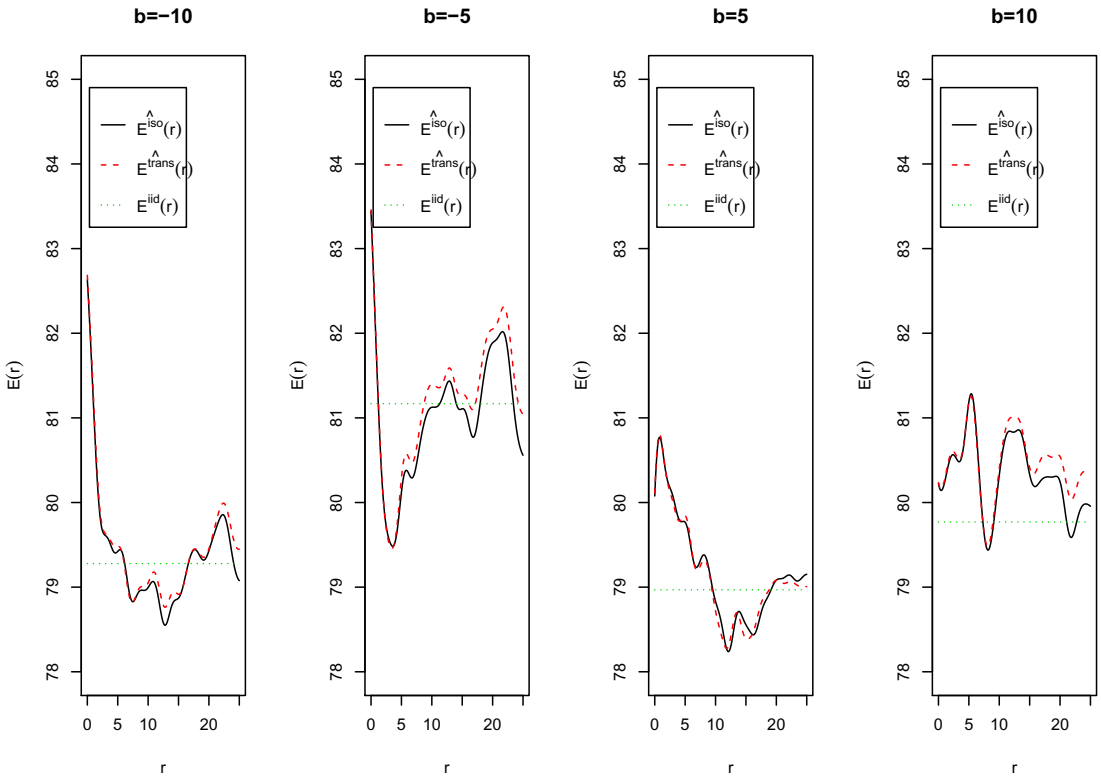

Figure 3.1. Mark independence diagnostics using Emark function in R-library `spatstat` for four intensity marked Log Gaussian Cox processes generated from model (3.1) with parameters $a = 80, b = -10, -5, 5, 10, (\phi, \sigma^2) = (5, 10)$ and $(\mu_Z, \phi_Z, \sigma_Z^2) = (-3, 0.5, 10)$.

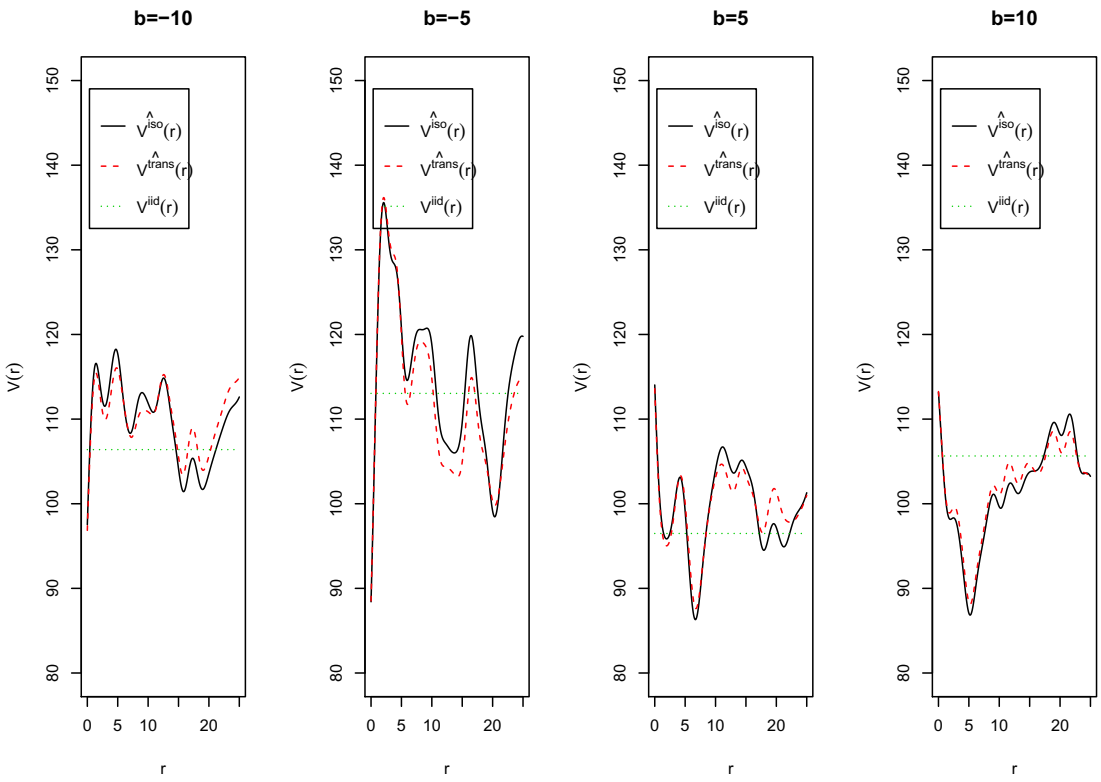

Figure 3.2. Mark independence diagnostics using Vmark function in R-library `spatstat` for four intensity marked Log Gaussian Cox processes generated from model (3.1) with parameters $a = 80, b = -10, -5, 5, 10, (\phi, \sigma^2) = (5, 10)$ and $(\mu_Z, \phi_Z, \sigma_Z^2) = (-3, 0.5, 10)$.

3.4 Case study: White Oak Data

3.4.1 Dataset Description

In this section, we analyzed the same data as in Chapter 2. We only used height in Chapter 2, here we considered both height and diameter. The range for diameter is from 2.6 cm to 26.7 cm with mean 12.8 cm. More details about this data can be found in Section 2.3.

Table 3.1

Simulation results for univariate mark: posterior mean (MCSE) for 40×40 grid points with $a = 80$ and $\sigma^2 = 10$.

.

		$\hat{\mu}_Z, \hat{\sigma}_Z^2, \hat{\phi}_Z$	$\hat{a}$	$\hat{b}$	$\hat{\tau}^2$
$\phi = 2.5$	$b = 5$		81.07 (0.06)	7.89 (0.93)	4.45 (0.12)
$(\tau^2 = 4)$	$b = 10$		82.55 (0.03)	8.66 (0.95)	4.94 (0.07)
	$b = -5$		78.29 (0.04)	-7.67 (0.91)	5.02 (0.25)
	$b = -10$		77.12 (0.04)	-8.61 (0.84)	5.26 (0.23)
		-3.54, 0.46, 6.57			
		(-3.00, 0.50, 10.00)			
$\phi = 5$	$b = 5$		82.08 (0.03)	5.55 (0.79)	1.88 (0.09)
$(\tau^2 = 2)$	$b = 10$		84.33 (0.02)	8.88 (0.61)	1.98 (0.04)
	$b = -5$		78.75 (0.07)	-6.47 (0.89)	1.87 (0.07)
	$b = -10$		77.83 (0.06)	-8.05 (0.70)	1.95 (0.05)

3.4.2 Check for Independence Between Points and Marks

For exploratory analysis, we first did mark independence diagnostics using the Emark and Vmark functions in R-library `spatstat` and the plots for height are given in Figure 3.3. These functions as well as the edge corrected estimates $\hat{E}^{\text{iso}}(r)$ and $\hat{E}^{\text{trans}}(r)$ were mentioned in Section 3.3. We observed non-constant mean values and variances within short range which indicates dependence between points and height. However, the mean values and variances for diameter did not vary a lot which may suggest independence between points and diameter. To test the independence between marks and points more formally, we applied the test using the mark K function $K_m(r)$ proposed by Guan (2006). The mark K function $K_m(r)$ is defined as $\lambda^{-1}E\{m(x) \times N_r(x)\}$ where λ is the first-order intensity of the point process and $N_r(x)$ is the number of points within distance r from a randomly chosen point x. Let

Table 3.2

Simulation results for bivariate marks: posterior mean (MCSE) for 40 × 40 grid points with $(a_1, a_2, \sigma_1^2, \sigma_2^2, \phi_1, \phi_2, \phi_{12}, \rho) = (80, 20, 400, 20, 10, 10, 10, 0.8)$ and $b_1 = -5, 5$, $b_2 = -2, 2$.

$\hat{\mu}_Z, \hat{\sigma}_Z^2, \hat{\phi}_Z = -3.54, 0.46, 6.57$ (-3.00, 0.50, 10.00)				
(b_1, b_2)	$(5, 2)$	$(5, -2)$	$(-5, 2)$	$(-5, -2)$
$\hat{a}_1$	84.98 (0.17)	83.75 (0.16)	78.65 (0.15)	77.97 (0.15)
$\hat{b}_1$	5.98 (0.65)	4.65 (0.56)	-4.75 (0.34)	-6.75 (0.47)
$\hat{a}_2$	22.23 (0.18)	17.90 (0.19)	21.78 (0.14)	16.98 (0.20)
$\hat{b}_2$	2.01 (0.56)	-3.24 (0.70)	2.75 (0.43)	-3.35 (0.37)
$\hat{\sigma}_1^2$	502.01 (0.86)	531.98 (0.70)	498.20 (0.79)	485.12 (0.87)
$\hat{\sigma}_2^2$	29.55 (0.77)	34.31 (0.80)	27.34 (0.82)	28.54 (0.88)
$\hat{\phi}_1$	10.39 (0.95)	10.87 (0.98)	12.01 (0.94)	10.77 (0.97)
$\hat{\phi}_2$	16.97 (0.99)	14.02 (0.97)	9.11 (0.95)	17.13 (0.98)
$\hat{\phi}_{12}$	15.35 (0.99)	14.57 (0.98)	13.35 (0.97)	16.35(0.99)
$\hat{\rho}$	0.72 (0.03)	0.60 (0.04)	0.66 (0.02)	0.73 (0.05)

$\mu = E\{m(x)\}$, then under the null hypothesis of independence between marks and points, we should have $K_m(r) = \mu \times K_p(r)$ where $K_p(r) = \lambda^{-1} E\{N_r(x)\}$ is the usual K function for spatial point process. We estimated μ by a slope estimator obtained from regressing $\hat{K}_m(r)$ on $\hat{K}_p(r)$:

$$\hat{\mu} = \frac{\int_0^R \hat{K}_m(r)dr}{\int_0^R \hat{K}_p(r)dr}$$

where R is an upper bound on the lag vector of observations since the dependence between points and marks may exist only at a small physical scale. Under the null hypothesis of independence, the slope estimator $\hat{\mu}$ should be close to the average of the observed marks, $\bar{M}$. The asymptotic normality of $\hat{\mu} - \bar{M}$ can be established under some conditions on the underlying process, and therefore come up with a statistic of

the form $(\hat{\mu} - \bar{M})^2 / \widehat{\mathrm{Var}}(\hat{\mu} - \bar{M})$. For our dataset we rejected the null hypothesis of independence between height and points (p-value < 0.001), but failed to reject independence between diameter and points (p-value=0.154) which was consistent with our exploratory analysis. Therefore we only applied our univariate model to the height. Nevertheless, the correlation between height and diameter is high, it is still reasonable to consider the two variables as dependent bivariate marks.

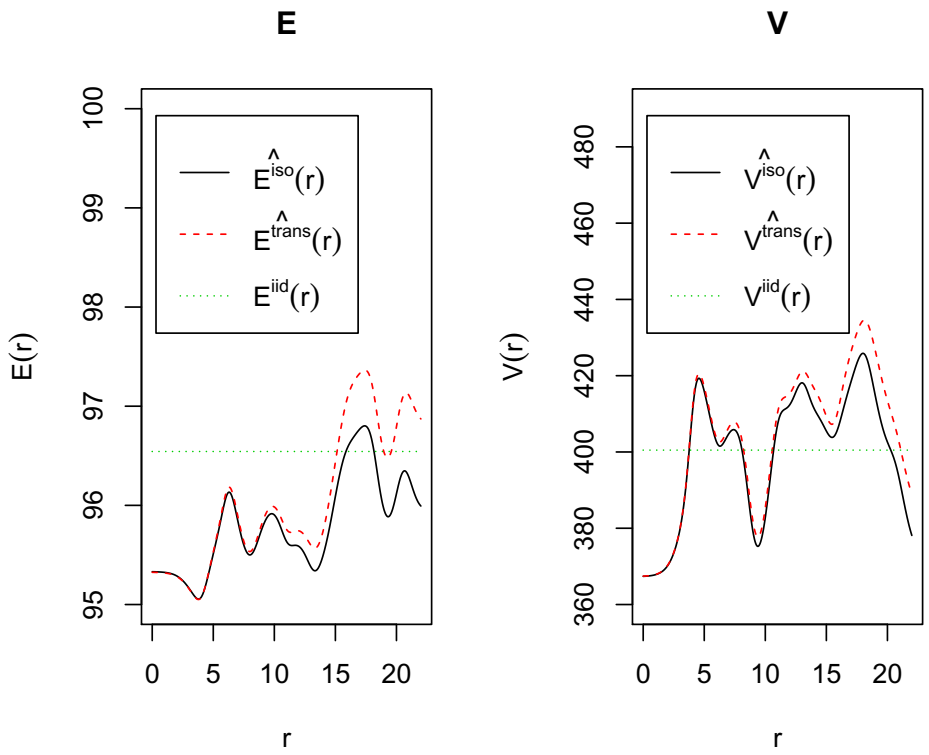

Figure 3.3. Mark independence diagnostics using Emark and Vmark function in R-library `spatstat` for the height of white oaks. The unit for height measurement is meter.

3.4.3 Empirical Bayesian Estimation

Univariate Marks

Based on the results of the previous section on the diagnostics of mark-points dependence, we applied the mark-dependent model to the height of white oaks. The Gaussian random field $\{Z(s)\}$ was assumed to have mean μ_Z and a Matérn covariance function $\sigma_Z^2 M(r|\nu, 1/\phi_Z)$ with parameters $\{\phi_Z, \sigma_Z^2, \nu\}$, where $M(h|\nu, \alpha)$ was defined in (3.4). The parameters related to $\{Z(s)\}$ are estimated using the minimum contrast (3.6) where η was set to be 1, $\epsilon = 12, a_0 = 72$. We found the estimated values to be $(\hat{\mu}_Z, \hat{\sigma}_Z^2, \hat{\phi}_Z, \hat{\nu}) = (-2.61, 0.63, 31.40, 0.5)$. The estimate of range parameter ϕ_Z indicated the spatial dependence occurs within distance around 31.4 m. These estimates were used in the prior distribution of Z.

We fixed ϕ at four values 12, 24, 36, 48, and applied Bayesian model to update parameters a, b, τ^2 and the Gaussian random field Z. We set normal priors for the parameters $a \sim N(120, 100)$ and $b \sim N(3, 100)$, and $\tau^2 \sim$ Inverse-Gamma$(3, 60)$. The proposal distributions for a, b and τ^2 were chosen as in (3.7) with $(\sigma_a^2, \sigma_b^2, \delta_{\tau^2}, \sigma_{qz}^2) = (4, 2, 5, 1)$. We ran 120000 iterations of MCMC and omitted the first 20000 iterations as burn-in. The posterior means and Monte Carlo standard errors (MCSE) are given in Table 3.3.

Table 3.3
Posterior mean (MCSE) for univariate mark (height).

$(\hat{\mu}_Z, \hat{\sigma}_Z^2, \hat{\phi}_Z)$=(-2.61, 0.63, 31.40)			
	$\hat{a}$	$\hat{b}$	$\hat{\tau}^2$
$\phi = 12$	75.62 (0.03)	35.03 (0.06)	10.01 (0.10)
$\phi = 24$	75.40 (0.08)	33.78 (0.03)	9.92 (0.30)
$\phi = 36$	77.49 (0.18)	34.00 (0.05)	10.43 (0.94)
$\phi = 48$	77.64 (0.27)	35.00 (0.10)	10.55 (0.92)

We used Bayes Factor (Kass, 1995) to choose the best combination of parameters $\{a, b, \phi, \tau^2\}$. Let $\theta_k = \{a_k, b_k, \phi_k, \tau_k^2\}, k = 1, \ldots, 4$ and each θ_k corresponds to the combination where $\phi_1 = 12, \phi_2 = 24, \phi_3 = 36$ and $\phi_4 = 48$. We used the notation B_{ij} to denote the Bayes Factor for θ_i against θ_j when we set equal priori on them, that is, $B_{ij} = f([X, M]|\theta_i, Z_i)/f([X, M]|\theta_j, Z_j)$. Simple Monte Carlo integration $\frac{1}{m}\sum_{i=1}^{m} f([X, M]|\theta_k^{(i)}, Z_k^{(i)})$ was used for estimating $\hat{f}([X, M]|\theta_k, Z_k)$, where $\theta_k^{(i)} = \{a_k^{(i)}, b_k^{(i)}, \phi_k^{(i)}, \tau_k^{2(i)}\}$ were samples from the prior distributions of parameters we obtained from MCMC iterations.

The Bayes Factor provided information about which statistical model was preferred given the observed data. If $B_{ij} > 1$ then it means the data in favors of θ_i instead of θ_j. Moreover, follow the interpretation suggested by Jeffreys (1961), we had decisive evidence against θ_j if B_{ij} was over 100.

From the summary of Bayes Factors for $i < j$ (Table 3.4), we found evidence that the data strongly support $\theta_1 = \{a_1, b_1, \phi_1, \tau_1^2\} = \{75.62, 35.03, 12.00, 10.01\}$ instead of others. As a result, θ_1 was chosen to be the preferred parameter combination. The plots of parameter estimates (when ϕ was set to be 12) verse sample size were created by estvssamp function in R-library `mcmcse` in Figure 3.4.

Table 3.4
Bayes Factors for choosing the best combination of parameters $\{a, b, \phi, \tau^2\}$.

k={1,2,3,4}		
$\phi = \{12, 24, 36, 48\}$		
$B_{12} = 369$	$B_{13} > 400$	$B_{14} > 400$
	$B_{23} > 400$	$B_{24} > 400$
		$B_{34} = 305$

The estimated parameters were then used to perform five-fold cross-validation to compare our approach with kriging prediction. We ignored the point process and make prediction for the marks using kriging, where the covariance matrix for marginal

distribution of marks were specified in Section 3.1. The mean absolute error (MAE) for prediction were used to compare the performance of two approaches. We found smaller MAE from the marked point process approach (MAE=1.99 meter) than the kriging prediction (MAE=2.05 meter) for height in our white oak data, indicating the marked point process approach has better performance on prediction. If the points and marks are independent, it is possible that kriging will do a fairly good job on prediction even if we ignore the point process. On the other hand, if the points and marks are dependent, then using kriging without considering the randomness of locations in the point process will not be the best way to make prediction.

Bivariate Marks

We applied the bivariate mark-intensity dependent model to the two marks, height and diameter. The latent Gaussian process $Z(x)$ was the same as in the previous section and its parameters were estimated in the same way. The prior distribution of $Z(x)$ was specified empirically as before. We considered two estimation approaches for bivariate marks. One is to first estimate the parameters of bivariate marks separately, which means we applied the method in the previous section to each of the two marks. We then estimated the parameters in the cross covariance function using bivariate data. The other one is to estimate their parameters in the bivariate model simultaneously. For the second approach we chose normal priors for parameters a_1, b_1, a_2, b_2 as follows: $\mathrm{N}(120, 100), \mathrm{N}(3, 100), \mathrm{N}(30, 100)$ and $\mathrm{N}(1, 100)$ respectively. The priors for σ_1^2 and σ_2^2 were Inverse-Gamma$(3, 60)$ and Inverse-Gamma$(3, 40)$ respectively. For ϕ_1, ϕ_2 and ϕ_{12} the priors were all Uniform$(12, 72)$ and the prior for ρ was Uniform$(0, 1)$. The proposals were described in Section 3.2.2, with $(\sigma^2_{a_1}, \sigma^2_{a_2}, \sigma^2_{b_1}, \sigma^2_{b_2}, \delta_{\sigma_1^2}, \delta_{\sigma_2^2}, \delta_{\phi_1}, \delta_{\phi_2}) = (4, 2, 2, 1, 8, 8, 5, 5)$. Initial vales of $\phi_1, \phi_2, \phi_{12}$ and ρ need to be generated following the constraints of bivariate Matérn covariance function.

Estimates from the two approaches are summarized in Table 3.5. They are compared through five-fold cross-validation. The results show that the dependent model

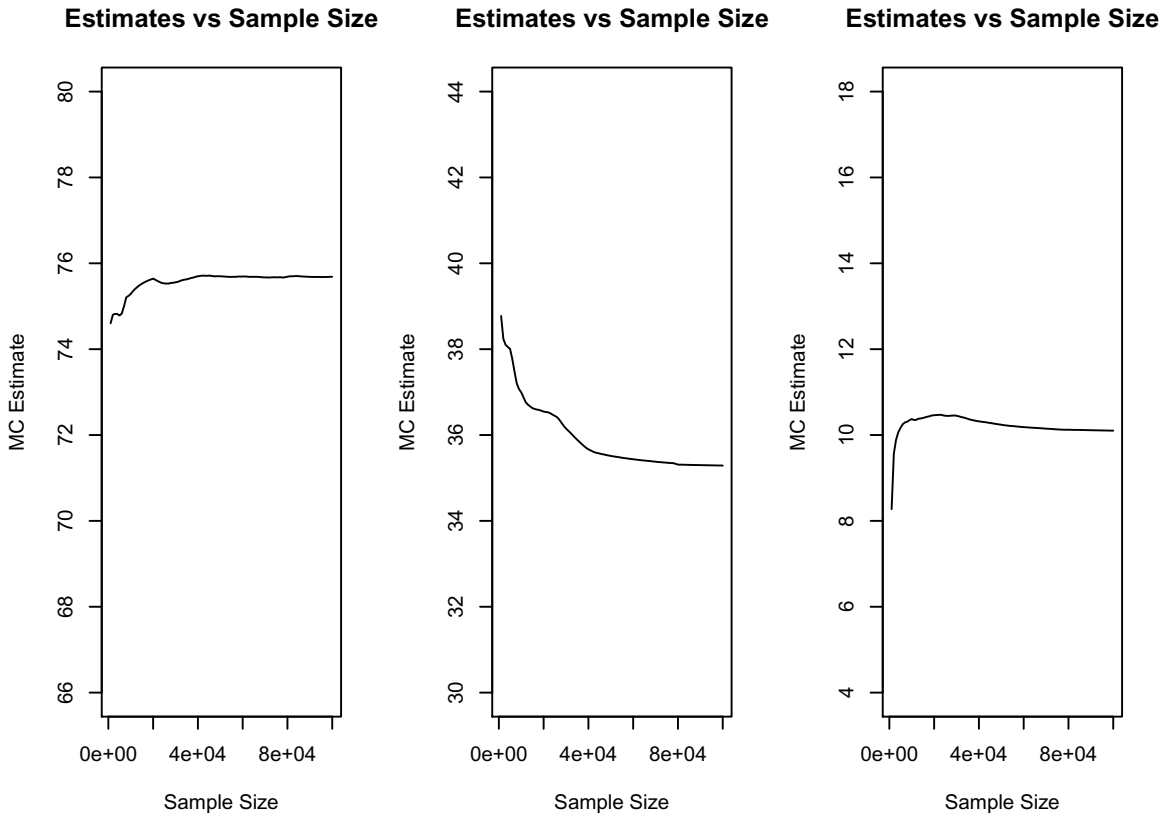

Figure 3.4. Plots to show how Monte Carlo estimates for a, b and τ^2 (from left to right) change when sample size increases.

results in a smaller mean absolute error (MAE). Hence the dependent model is preferred. All the estimates for b are positive which indicates positive dependence between the intensity and marks regardless of the model choices.

3.5 Discussion

In this chapter we consider the model of univariate and bivariate intensity-dependent marked LGCP. Here we only use normal for the mark distribution, but other choices like exponential or gamma distribution introduced by Myllymäki and Penttinen (2009) are also possible. We update both the intensity and other mark-related parameters in

Table 3.5

Posterior mean (MCSE) for bivariate marks (height and diameter) from two approaches. Independent Marks: To assume the two marks to be independent conditional on the intensity, and estimate their parameters separately; Dependent Marks: To include cross covariance between height and diameter and estimate their parameters simultaneously.

Height (m)	$\hat{a}_1$	$\hat{b}_1$	MAE
Independent Marks:	75.62 (0.03)	35.03 (0.06)	1.99
Dependent Marks:	78.43 (0.04)	24.47 (0.32)	1.72
Diameter (cm)	$\hat{a}_2$	$\hat{b}_2$	MAE
Independent Marks:	10.37 (0.03)	10.03 (1.21)	4.17
Dependent Marks:	11.27 (0.01)	3.20 (0.23)	3.77

our MCMC algorithm to avoid the kernel estimates of intensity function, since kernel methods are sensitive to the choice of bandwidth. But the estimates of intensity still affects parameter b and makes the convergence slower.

An alternative and easier way is to consider marks as covariates to estimate intensity $\Lambda(x)$ using common LGCP modeling methodology. It may also help to explain the dependence relationship between marks and points, but it makes more sense to let the variables like height and diameter to depend on the point process instead of the other way around since it is reasonable that how the trees grow will be affected by how dense the nearby trees are.

If we have univariate mark, adding the dependence structure between marks after conditioning on intensity is more flexible than the conditional independence assumption. Another possible improvement on this model may be to make use of the nearest neighbor concept and add trees close by to help modeling the marks. However, this way we have to determine the number of neighbor trees, and we need to estimate more parameters which will result in longer time to run MCMC iterations.

For bivariate marks we have height and diameter of trees in our study. These two variables are highly correlated and consequently, we get better results not to ignore the relationship between them. Nevertheless, we need to estimate more parameters from bivariate covariance function than the univariate case, and have to pay attention to the constraints between parameters to make sure the covariance matrix is positive definite.

Currently the marks we deal with are quantitative variables, it is also of interest to consider categorical marks like the species of trees and try to model how the different species can be dependent on each others. We are also curious about the difference between using point process and geostatistical methodology. In practice, from the nature of forest data we should consider it as point process since the sampling locations x_i are random, and proceed the analysis using point process methodology. Using geostatistical methodology we will not be able to describe the dependence between intensity and marks, but in general we can still apply geostatistical methods to model and predict the marks regardless of the point process structure. How much difference the two approaches will generate is still under study. This is worth investigating both empirically and theoretically as to extend our understanding of the dissimilarity between point process and geostatistics.

4. A PARTIALLY STATIONARY MODEL FOR SPATIOTEMPORAL MARKED POINT PROCESSES

Our objective in this chapter is to develop a model between the stationary and non-stationary marked point processes that can be used to model dependence between points and marks. The partially stationary spatiotemporal marked point process is our focus in this chapter. Before we give definition for partially stationary, we first introduce definitions for different types of stationary.

A spatiotemporal point process is said to be strong stationary if the joint distribution of $N(A_1 \times B_1), \cdots, N(A_k \times B_k)$ is equal to the joint distribution of $N((A_1 + t) \times (B_1 + t)), \cdots, N((A_k + t) \times (B_k + t))$ for any $k \in \mathbb{N}^+$ and $t \in \mathbb{R}$, where $A_1, \cdots, A_k \in \mathscr{B}(\mathcal{S})$ and $B_1, \cdots, B_k \in \mathscr{B}(\mathbb{R})$ are defined in Section 1.5.

Another type of stationary is based on intensity function. A spatiotemporal point process is said to be first-order stationary in both space and time if its first-order properties are invariant under translation:

$$\lambda(\mathbf{s}, t) = \lambda_c,$$

and second-order stationary if its second-order properties are invariant under translation:

$$\lambda_2[(\mathbf{s}, t), (\mathbf{s}', t')] = \lambda_2(\mathbf{s} - \mathbf{s}', t - t'),$$

where $\lambda(\mathbf{s}, t)$ and $\lambda_2[(\mathbf{s}, t), (\mathbf{s}', t')]$ denote the first-order intensity and second-order intensity of the point process, respectively. They are defined as

$$\lambda(\mathbf{s}, t) = \lim_{|d\mathbf{s} \times dt| \to 0} \left\{ \frac{E[N(d\mathbf{s} \times dt)]}{|d\mathbf{s} \times dt|} \right\},$$
$$\lambda_2[(\mathbf{s}, t), (\mathbf{s}', t')] = \lim_{|d\mathbf{s} \times dt|, |d\mathbf{s}' \times dt'| \to 0} \left\{ \frac{E[N(d\mathbf{s} \times dt) N(d\mathbf{s}' \times dt')]}{|d\mathbf{s} \times dt||d\mathbf{s}' \times dt'|} \right\}$$

A first-order and second-order stationary spatiotemporal marked point process can be defined similarly as

$$\lambda(\mathbf{s}, t, m) = \lambda_c$$

and

$$\lambda_2[(\mathbf{s}, t, m), (\mathbf{s}', t', m')] = \lambda_2(\mathbf{s} - \mathbf{s}', t - t', m - m'),$$

which is first-order and second-order stationary in space, time and mark.

Our definition for partially stationary spatiotemporal marked point process is given as follows. We define a spatiotemporal marked point process N to be *partial stationary* if the joint distribution of $N(A_1 \times B_1 \times C_1), \cdots, N(A_k \times B_k \times C_k)$ is equal to the joint distribution of $N(A_1 \times (B_1 + t) \times C_1), \cdots, N(A_k \times (B_k + t) \times C_k)$ for any $k \in \mathbb{N}^+$ and $t \in \mathbb{R}$, where $A_1, \cdots, A_k \in \mathscr{B}(\mathcal{S})$, $B_1, \cdots, B_k \in \mathscr{B}(\mathbb{R})$, and $C_1, \cdots, C_k \in \mathscr{B}(\mathcal{M})$ are defined in Section 1.5. Under the assumption of partial stationarity, it is nonstationary in first-order and partially stationary in second-order. The distribution of N is only invariant under the parallel shift of the time. If the distribution of N is invariant under parallel shift in space, time and mark, or only invariant under parallel shift of space and time, different types of stationarity can also be defined accordingly.

One of the applications for partially stationary spatiotemporal marked point process is to model natural hazards since we expect many natural events to satisfy this assumption. Natural hazard forecasting with time, location and magnitude has always been a difficult task. Spatiotemporal marked point processes (SMPPs) are widely used stochastic models which are appropriate for modeling the occurrence and magnitude of natural hazards. We consider SMPPs because the location where an event occurs and the magnitude of the event can be represented by points and marks in SMPPs. In this chapter, our objective is to develop models for SMPPs that can describe the patterns of natural hazards and try to identify regions with higher risks where severe natural hazards are more likely to occur.

We develop an approach to model the partially stationary SMPPs, and apply our method to model the patterns of natural hazards. The plan of this chapter is as follows. The statistical model for partially stationary SMPPs is introduced in Section 4.1. A kernel weighted composite likelihood approach for modeling partially stationary SMPPs is given in Section 4.2, and the estimators of parameters in the model are derived. The asymptotic properties for estimators and selection of bandwidth for the kernel function are provided in Section 4.3 and 4.4, respectively. Section 4.5 provides a simulation study for the performance of our approach. The proposed method is applied to a forest wildfire data and an earthquake data in Section 4.6. A discussion is given in Section 4.7.

4.1 Statistical Model

In the analysis of SMPP data, the main interest is to model the first-order intensity function that has generated the SMPP pattern. Let N be a partially stationary SMPP, then its first-order intensity function does not vary in time indicating that the first-order intensity function can be expressed as $\lambda(\mathbf{s}, t, m) = \lambda(\mathbf{s}, m)$. Therefore, in general the first-order intensity function can be written as

$$\lambda(\mathbf{s}, m) = \lambda(\mathbf{s}, t, m) = f(m|\mathbf{s})\lambda_s(\mathbf{s}), \tag{4.1}$$

for all $\mathbf{s} \in \mathcal{S}$, $t \in \mathbb{R}$, and $m \in \mathcal{M}$, where $\lambda_s(\mathbf{s}) = \int_{\mathcal{M}} \lambda(\mathbf{s}, m)dm$ is the marginal first-order intensity function of points and $f(m|\mathbf{s})$ is the conditional density function of marks.

According to how we formulate $\lambda(\mathbf{s}, t, m)$ as in (4.1), we consider a way to model $f(m|\mathbf{s})$ and $\lambda_s(\mathbf{s})$ separately. A local parametric model is considered for $f(m|\mathbf{s})$ and a nonparametric model for $\lambda_s(\mathbf{s})$. In the local parametric model, we assume that $f(m|\mathbf{s}) \in \mathcal{F}_\Theta = \{f_\theta(m) : \theta \in \Theta \in \mathscr{B}(\mathbb{R}^p)\}$, where $f_\theta(m)$ is a parametric density function which is completely determined by the unknown parameter θ. We assume that the parameter contained in $f(m|\mathbf{s})$ may depend on $\mathbf{s}$ such that $f(m|\mathbf{s})$ can be expressed as $f_{\theta(\mathbf{s})}(m)$ in general, where $\theta(\mathbf{s})$ is a p-dimensional multivariate smooth

function taking values on Θ for any $\mathbf{s} \in \mathcal{S}$. In the nonparametric model for $\lambda_s(\mathbf{s})$, we assume that $\lambda_s(\mathbf{s})$ is positive and smooth for every $\mathbf{s} \in \mathcal{S}$. Therefore, the first-order intensity function of N can be expressed as

$$\lambda(\mathbf{s}, t, m) = \lambda_s(\mathbf{s}) f_{\theta(\mathbf{s})}(m), \tag{4.2}$$

for any $\mathbf{s} \in \mathcal{S}$, $t \in \mathbb{R}$, and $m \in \mathcal{M}$. In this study, the term $f_{\theta(\mathbf{s})}(m)$ can be interpreted as the local parametric component and $\lambda_s(\mathbf{s})$ as the nonparametric component.

A special case of the model given by Equation (4.2) is the separable model in which $\theta(\mathbf{s})$ is treated as a constant. Such a model has been investigated by Schoenberg (2004) for Poisson SMPPs in a hypotheses testing problem. The assumption with a constant $\theta(\mathbf{s})$ is usually called the independence (or separability) assumption between points and marks in literature. As we mentioned in Section 1.3, it becomes quite convenient in modeling, estimation, and prediction in an SMPP if marks and points are independent. Many commonly used Hawkes models, such as the epidemic-type aftershock sequences (ETAS) model (Ogata, 1998), may exhibit independence between marks and points (Schoenberg, 2004). However, the independence assumption is often unrealistic in applications. For instance, the relative positions of trees in a forest have repercussions on their size owing to their competition for light or nutrient (Schlather, Ribeiro, and Diggle, 2004), indicating that tree sizes and locations of trees may not be independent. Forest wildfire activities exhibit power-law relationships between frequency and burned area (Malamud, Millington, and Perry, 2005), which means the burned area and the locations of forest wildfires may not be independent either. If the independence assumption is violated, then intensity-dependent models may be used (Ho and Stoyan, 2008; Malinowski, Schlather and Zhang, 2012; Myllymäki and Penttinen, 2009). An intensity-dependent model attempts to model the expected value of marks by treating $\lambda_s(\mathbf{s})$ as an explanatory variable. Our study in Chapter 3 also focused on intensity-dependent models. However, the model we consider in this chapter is different from intensity-dependent assumption, because Model (4.2) assumes that the parameters of marks depend on their locations but

not only their intensity functions. Therefore, Model (4.2) should be classified as a *location-dependent model* instead.

The behavior of N can be completely determined by Equation (4.2) if N is a Poisson SMPP. For any $A \in \mathscr{B}(\mathcal{S})$ and $C \in \mathscr{B}(\mathcal{M})$, if $B_1, \cdots, B_k \in \mathscr{B}(\mathbb{R})$ are disjoint, then $N(A \times B_1 \times C), \cdots, N(A \times B_k \times C)$ are independent Poisson random variables with mean $|B_1|\mu(A \times C), \cdots, |B_k|\mu(A \times C)$, respectively, where

$$\mu(A \times C) = \mu_0(A \times [t, t+1] \times C) = \int_A \int_C \lambda(\mathbf{s}) f_{\theta(\mathbf{s})}(m) dm d\mathbf{s}$$

for any $t \in \mathbb{R}$. Let

$$M_\eta(t; A, C) = \frac{N(A \times \eta[0,t] \times C) - E[N(A \times \eta[0,t] \times C)]}{\sqrt{\eta}}. \tag{4.3}$$

Using the theory of empirical distributions (e.g. Chapter 19 of van der Vaart, 1998), we have

$$\mu^{-1/2}(A \times C) M_\eta(\cdot; A, C) \xrightarrow{D} \mathbb{B}([0, \infty)) \tag{4.4}$$

as $\eta \to \infty$, where $\eta B = \{\eta t : t \in B\}$ for any $B \in \mathscr{B}(\mathbb{R})$ and $\mathbb{B}([0, \infty))$ is the standard Brownian motion on $[0, \infty)$. However, if N is not Poisson, then the asymptotic distribution given by Equation (4.4) does not hold. Fortunately, a similar conclusion can be drawn with a few regularity conditions.

The most important regularity condition is the mixing condition, which was first proposed by Rosenblatt (1956) and then followed by many other authors (e.g. by Ibragimov, 1962). The mixing condition has later been considered for asymptotic properties of methods in spatial point processes (Ivanoff, 1982). Different mixing conditions can be defined, such as *strong mixing, uniform strong mixing*, *φ-mixing* and B-*mixing*. Among various types of mixing conditions, *strong mixing* is the simplest to verify. The definition is given below. A homogeneous spatial point process on $\mathbb{R}^n$ is said to be *strong mixing* if $\alpha(\eta u, \eta v) \to 0$ as $\eta \to 0$, where $\alpha(u, v)$ is defined as

$$\alpha(u, v) = \sup_{\substack{\rho(E_1, E_2) \geq u \\ d(E_1) \leq v, d(E_2) \leq v}} \sup_{\substack{U_1 \in \mathscr{F}(E_1) \\ U_2 \in \mathscr{F}(E_2)}} |P(U_1 \cap U_2) - P(U_1)P(U_2)|,$$

$d(E) = \sup_{t,t' \in E} |t - t'|$ is the diameter of E, and $\rho(E_1, E_2) = \min_{t \in E_1, t' \in E_2} |t - t'|$ is the minimum distance between disjoint sets $E_1, E_2 \in \mathscr{B}(\mathbb{R}^n)$. Here, we modify this definition for partially stationary SMPPs.

Let $P_{A,C}(\cdot)$ be the distribution of $N(A \times \cdot \times C)$ for given $A \in \mathscr{B}(\mathcal{S})$ and $C \in \mathscr{B}(\mathcal{M})$. We say N satisfies the *partially strong mixing condition* (i.e. in time) if

$$\lim_{\eta \to \infty} \sup_{A \in \mathscr{B}(\mathcal{S}), C \in \mathscr{B}(\mathcal{M})} \alpha_{A,C}(\eta u, \eta v) = 0 \tag{4.5}$$

for any distinct positive u and v, where

$$\alpha_{A,C}(u, v) = \sup_{\substack{\rho(E_1, E_2) \geq u \\ d(E_1) \leq v, d(E_2) \leq v}} \sup_{\substack{U_1 \in \mathscr{F}(E_1) \\ U_2 \in \mathscr{F}(E_2)}} |P_{A,C}(U_1 \cap U_2) - P_{A,C}(U_1) P_{A,C}(U_2)|,$$

Besides the partially strong mixing condition, we also need to consider higher-order properties in our regularity conditions, which can be summarized by the covariance function and the cumulant density function. The covariance function of N is well defined if the covariance structure given by Equation (1.7) is absolutely continuous with respect to the Lebesgue measure on $\mathcal{S} \times \mathbb{R} \times \mathcal{M}$ as

$$\begin{aligned} &\Gamma((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)) \\ =&\{g[(\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2)] - 1\} \lambda(\mathbf{s}_1, t_1, m_1) \lambda(\mathbf{s}_2, t_2, m_2) \\ &+ \lambda(\mathbf{s}_1, t_1, m_1) \delta_{(\mathbf{s}_1, t_1, m_1)}(\mathbf{s}_2, t_2, m_2), \end{aligned} \tag{4.6}$$

where $\delta_{(\mathbf{s},t,m)}$ is the point measure at $(\mathbf{s}, t, m)$. If N is partially stationary, then $\Gamma(\cdot, \cdot)$ can be expressed as

$$\begin{aligned} &\Gamma((\mathbf{s}_1, t_1, m_1), (\mathbf{s}, t_2, m_2)) = \Gamma((\mathbf{s}_1, 0, m_1), (\mathbf{s}_2, t_2 - t_1, m_2)) \\ =&[g((\mathbf{s}_1, 0, m_1), (\mathbf{s}_2, t_2 - t_1, m_2)) - 1] \lambda(\mathbf{s}_1, m_1) \lambda(\mathbf{s}_2, m_2) \\ &+ \lambda(\mathbf{s}_1, m_1) \delta_{\mathbf{s}_1, t_1, m_1}(\mathbf{s}_2, t_2, m_2). \end{aligned} \tag{4.7}$$

Let

$$\psi(\varphi) = \log \left[E e^{\int_{\mathbb{R}^d} \log \varphi(\mathbf{s}, t, m) N(d\mathbf{s} \times dt \times dm)} \right]$$

be the logarithm probability generating function (PGF) of N, where φ, $0 \leq \varphi \leq 1$, is a function with compact support on $\mathcal{S} \times \mathbb{R} \times \mathcal{M}$. The kth-order factorial cumulant of N is defined by

$$C_{(k)}(D_1, \cdots, D_k) = \lim_{\eta \uparrow 1} \left[\frac{\partial^k}{\partial a_1 \cdots \partial a_k} \psi \left(\eta + \sum_{i=1}^{k} a_i I_{D_i} \right) \right]_{a_1 = a_2 = \cdots = a_k = 0}, \tag{4.8}$$

where I_D is the indicator function of $D \in \mathscr{B}(\mathcal{S} \times \mathbb{R} \times \mathcal{M})$. If $C_{(k)}$ is absolutely continuous, then its kth-order density function, denoted by $Q_k((\mathbf{s}_1, t_1, m_1), \cdots, (\mathbf{s}_k, t_k, m_k))$ for $\mathbf{s}_1, \cdots, \mathbf{s}_k \in \mathcal{S}$, $t_1, \cdots, t_k \in \mathbb{R}$, and $m_1, \cdots, m_k \in \mathcal{M}$, is called the kth-order factorial cumulant density of N.

Regularity Conditions:

(C1) The SMPP N is partially stationary.

(C2) The SMPP N satisfies the partially strong mixing condition given by (4.5).

(C3) The covariance function of N is continuous and satisfies

$$0 < \omega((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2)) = \int_{\mathbb{R}} \Gamma((\mathbf{s}_1, 0, m_1), (\mathbf{s}_2, t, m_2)) dt < \infty$$

with

$$\int_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S} \times \mathcal{M}} G(\mathbf{x}_1, \mathbf{x}_2) \omega(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 < \infty$$

for any bounded integrable function $G(\mathbf{x}_1, \mathbf{x}_2)$ on $(\mathcal{S} \times \mathcal{M})^2$.

(C4) The cumulant Q_k satisfies

$$\int_{\mathbf{x}_1, \cdots, \mathbf{x}_k \in \mathcal{S} \times \mathbb{R} \times \mathcal{M}} |Q_k(\mathbf{x}_1, \cdots, \mathbf{x}_k)| d\mathbf{x}_1 \cdots d\mathbf{x}_{k-2} < c_1, k = 2, 3, 4,$$

and

$$\int_{\mathbf{x}_1, \cdots, \mathbf{x}_k \in \mathcal{S} \times \mathbb{R} \times \mathcal{M}} |Q_k(\mathbf{x}_1, \cdots, \mathbf{x}_k)| d\mathbf{x}_1 \cdots d\mathbf{x}_{k-1} < c_2, k = 2, 3, 4$$

for some constants $c_1, c_2 \in \mathbb{R}$.

Theorem 1 *If Conditions (C1)–(C4) hold, then* $\tau^{-1}(A,C)M_\eta(\cdot;A,C) \xrightarrow{D} \mathbb{B}([0,\infty))$ *as* $\eta \to \infty$ *for any* $A \in \mathscr{B}(\mathcal{S})$ *and* $C \in \mathscr{B}(\mathcal{M})$*, where*

$$\tau^2(A,C) = \int_{\mathbf{x}_1,\mathbf{x}_2 \in A\times C} \omega(\mathbf{x}_1,\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{4.9}$$

is positive and finite.

Proof: Let $\tilde{N}_{A,C}$ be the process defined as $\tilde{N}_{A,C}(B) = N(A\times B\times C)$ for $A \in \mathscr{B}(\mathcal{S})$ and $C \in \mathscr{B}(\mathcal{M})$. Then, $\tilde{N}_{A,C}$ is a strong stationary point process on $\mathbb{R}$ with a constant first-order intensity function equal to $\tilde{\lambda}_{A,C} = \int_A \int_C \lambda(\mathbf{s},m) dm d\mathbf{s}$ and covariance function equal to

$$\tilde{\Gamma}_{A,C}(t) = \int_{\mathbf{s}_1,\mathbf{s}_2 \in A, m_1, m_2 \in C} \Gamma((\mathbf{s}_1,0,m_1),(\mathbf{s}_2,t,m_2)) dm_2 dm_1 d\mathbf{s}_2 d\mathbf{s}_1.$$

Then, $\tau^2(A,C) = \int_{\mathbb{R}} \tilde{\Gamma}_{A,C}(t)dt$, which is (4.9). Let $\tilde{Q}_k$ be the k-th cumulant of $\tilde{N}_{A,C}$. Then, $\tilde{Q}_k(t_1,\cdots,t_k) = \tilde{Q}_k(0,t_2-t_1,\cdots,t_k-t_1)$. If Condition (C4) holds, then there exist finite c_1 and c_2 (which may depend on A and C) such that

$$\int_{t_1,\cdots,t_{k-2}\in\mathbb{R}} |\tilde{Q}_k(t_1,\cdots,t_k)| dt_1 \cdots dt_{k-2} < c_1$$

and

$$\int_{t_1,\cdots,t_{k-1}\in\mathbb{R}} |\tilde{Q}_k(t_1,\cdots,t_k)| dt_1 \cdots dt_{k-1} < c_2$$

for $k = 2,3,4$. Using Corollary 7.2 of Ivanoff (1982), the final conclusion is drawn. $\diamondsuit$

Corollary 1 *Assume Conditions (C1)–(C4) hold. For a continuous bounded non-negative function* $G(\mathbf{s},m)$ *on* $\mathcal{S}\times\mathcal{M}$*, let*

$$Z_{G,\eta}(t) = \frac{X_G(\eta[0,t]) - \eta t \mu_G}{\tau_G \sqrt{\eta}}$$

where

$$X_G(B) = \int_{\mathcal{S}\times\mathcal{M}} G(\mathbf{s},m) N(d\mathbf{s}\times B\times dm),$$

$$\mu_G = \int_{\mathcal{S}\times\mathcal{M}} G(\mathbf{s},m)\lambda(\mathbf{s},m) d\mathbf{s} dm,$$

and

$$\tau_G^2 = \int_{\mathbf{s}_1,\mathbf{s}_2\in\mathcal{S},t\in\mathbb{R},m_1,m_2\in\mathcal{M}} \omega((\mathbf{s}_1,m_1),(\mathbf{s}_2,m_2))G(\mathbf{s}_1,m_1)G(\mathbf{s}_2,m_2)d\mathbf{s_1}dm_1d\mathbf{s}_2dm_2dt.$$

Then, $Z_{G,\eta}(\cdot) \xrightarrow{D} \mathbb{B}([0,\infty))$ *as* $\eta \to \infty$.

Proof: Let $A = d\mathbf{s}$ and $C = dm$ in Theorem 1. Then, $X_G(\eta[0,t]) = N(d\mathbf{s}\times\eta[0,t]\times dm)$, $E[N(d\mathbf{s}\times\eta[0,t]\times dm)] = t\eta|d\mathbf{s}\times dm|$, and $V[N(d\mathbf{s}\times\eta[0,t]\times dm)] = t\eta\tau^2(d\mathbf{s},dm)$. This expression implies that the classical approach to defining an integral on an Euclidean space using the simple function can be applied (e.g. p. 199 in Billingsley, 1995). Using $V[X_G(\eta[0,t])] = t\eta\tau_G^2$, the conclusion is drawn. ♢

Theorem 1 and Corollary 1 motivate a way to derive estimating equations of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ based on Equation (4.2), which is introduced in the next section.

4.2 The Kernel Weighted Composite Likelihood (KWCL) Approach

We use the kernel weighted composite likelihood (KWCL) approach to estimate $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ in Model (4.2). The KWCL approach can be treated as an extension of the local likelihood approach (Loader, 1996; Tibshirani and Hastie, 1987), which provides a flexible method to estimate local varied parameters. The local likelihood approach estimates the local varied parameters by a localized version of the loglikelihood function. It gives more weights to observations closer to a particular location. We propose the KWCL approach based on this idea. Here, we use a kernel function to determine the observations. The bandwidth controls the bias and the variance of the resulting estimator, which is balanced by its mean square error (MSE) values. The strategy to select the best bandwidth is discussed in Section 4.4.

As we reviewed in Section 1.6.2, composite likelihood is an inference function derived by multiplying a collection of component of likelihood functions. Because each individual component is a conditional or a marginal probability density (or mass) function, the resulting estimating equation is often unbiased (Varin, Reid, and

Firth, 2011). Recently, the unweighted composite likelihood approach has been used in estimation of parameters of point process data (Schoenberg, 2005; Waagepetersen, 2007). Our approach may be considered as an extension of the previous unweighted composite likelihood approach.

Assume that the first-order intensity function $\lambda(\mathbf{s}, t, m)$ is given in terms of a local parametric model displayed in (4.2), where $\theta(\mathbf{s})$ is a vector of unknown parameters. Let $D \in \mathscr{B}(\mathcal{S}, \mathbb{R}, \mathcal{M})$ be the region in which a realization of N has been observed. As before, assume $\mathcal{S} \in \mathscr{B}(\mathbb{R}^2)$ and $\mathcal{M} \in \mathscr{B}(\mathbb{R}^q)$, where q is a certain positive integer. Let the observations of N be denoted as $(\mathbf{s}_i, t_i, m_i) \in D$ for $i = 1, \cdots, n$. Based on Equation (4.2), we can write down the composite likelihood function

$$\begin{aligned}
&\mathcal{L}_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) \\
&= \prod_{i=1}^{n} \lambda(\mathbf{s}, t_i, m_i) \exp\left(- \int_{(\mathbf{s}', t, m) \in D} \lambda(\mathbf{s}, t, m) dm dt d\mathbf{s}' \right) \\
&= \prod_{i=1}^{n} [\lambda_s(\mathbf{s}) f_{\theta(\mathbf{s})}(m_i)] \exp\left(- \int_{(\mathbf{s}', t, m) \in D} \lambda_s(\mathbf{s}) f_{\theta(\mathbf{s})}(m) dm dt d\mathbf{s}' \right)
\end{aligned} \tag{4.10}$$

Using kernel $K(\mathbf{s})$ as the weights and then take log, we can also write down the kernel weighted composite loglikelihood function

$$\begin{aligned}
\ell_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) = &\sum_{i=1}^{n} K_h(\mathbf{s}_i - \mathbf{s}) \{ \log[\lambda_s(\mathbf{s})] + \log[f_{\theta(\mathbf{s})}(m_i)] \} \\
&- \int_{(\mathbf{s}', t, m) \in D} K_h(\mathbf{s}' - \mathbf{s}) \lambda_s(\mathbf{s}) f_{\theta(\mathbf{s})}(m) dm dt d\mathbf{s}',
\end{aligned} \tag{4.11}$$

where $K_h(\mathbf{s}) = K(\mathbf{s}/h)/h^2$ and $K(\mathbf{s})$ is a kernel function on $\mathbb{R}^2$. Here we choose K as a full symmetric density function on $\mathbb{R}^2$ satisfying

$$0 < \xi = \int_{\mathbb{R}^2} K^2(\mathbf{s}) d\mathbf{s} < \infty.$$

As long as the kernel function $K(\mathbf{s})$ and the bandwidth h have been chosen, $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ can be estimated by maximizing $\ell_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))$.

The function $\ell_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))$ can be viewed as a limit of the kernel weighted composite loglikelihood for binary variables $Y_i = I[N(D_i) > 0]$, $i = 1, \cdots, k$, where D is

partitioned into L small pixels and $I(\cdot)$ is an indicator function. This limit is obtained as $k \to \infty$ with $\sup_{i \leq k} \rho(D_i) \to 0$. In the case of a Poisson SMPP, the composite loglikelihood function coincides with the loglikelihood function, which indicates that the KWCL approach reduces to the kernel weighted likelihood (KWL) approach; otherwise, the two approaches are different.

Observing the right side of (4.11), we find that $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ can be treated as constants after $\mathbf{s}$ is chosen. We treat both as unknown parameters in our KWCL approach. Using the basic theory of the composite likelihood approach, the kernel weighted composite likelihood estimators (KWCLE) of $\lambda_s(\mathbf{s})$ and $\theta(\mathbf{s})$ can be derived by maximizing $\ell_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))$ as

$$(\hat{\theta}(\mathbf{s}), \hat{\lambda}_s(\mathbf{s})) = \underset{\theta(\mathbf{s}), \lambda_s(\mathbf{s})}{\arg\max}\, \ell_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s})), \tag{4.12}$$

which are usually solved by

$$\dot{\ell}_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) = 0, \tag{4.13}$$

where $\dot{\ell}_h$ is the gradient of ℓ_h. In the following of this paper, we use $\dot{\ell}$ and $\ddot{\ell}$ to represent the gradient vector and the Hessian matrix of ℓ.

A particular interest is the case when $f_{\theta(\mathbf{s})}(m)$ is the density of a truncated exponential distribution given by

$$f_{\theta(\mathbf{s})}(m) = \theta(\mathbf{s}) e^{-\theta(\mathbf{s})(m - \underline{m})}, m \geq \underline{m}, \theta(\mathbf{s}) \in \mathbb{R}^+, \tag{4.14}$$

as this model is summarized from the famous power-law frequency-size relationship of natural hazards. For example, in an effort to examine forest fires in the conterminous USA, Malamud, Millington, and Perry (2005) found that despite the complexities concerning their initiation and propagation, wildfires exhibit power-law frequency-area statistics over many orders of magnitude. In summary, the power-law states a statistically robust fit as $y = \alpha x^{-\beta}$, $\alpha > 0$, where x represents the area burned and y represents the frequency density. Let $m = \log(x)$ and assume the data only contains those with $x \geq e^{\underline{m}}$. Then, the power-law relationship can be summarized by a

probability density function as displayed in (4.14). In addition, the usage of the exponential distribution is consistent with the well-known Gutenberg-Richter relationship between magnitudes and frequencies for earthquakes (Gutenberg and Richter, 1954).

The power-law relationship has been examined by many similar studies, which include McCarthy and Gill (1997) for Australia, Ricotta, Avena and Marchetti (1999); Ricotta *et. al* (2001) for Italy and Spain, Niklasson and Granstrom (2000) for Northern Sweden, Song *et. al* (2001) for China, Zhang *et. al* (2003) for Russia, and Fiorucci, Gaetani and Minciardi (2008) for Italy. These studies suggest that wildfires exhibit robust power-law relationships between fire frequency and burned area at regional scales. Besides forest wildfire studies, the power-law relationship has also been confirmed by many authors in earthquake studies (Ogata, 1988).

Let $f_{\theta(\mathbf{s})}(m)$ be given by (4.14) and assume $D = \mathcal{S} \times [0, T] \times \mathcal{M}$ with $\mathcal{M} = [\underline{m}, \infty)$. Then, Equation (4.13) becomes

$$\sum_{i=1}^{n} K_h(\mathbf{s}_i - \mathbf{s})[\frac{1}{\theta(\mathbf{s})} - (m_i - \underline{m})] = 0$$

and

$$\sum_{i=1}^{n} \frac{K_h(\mathbf{s}_i - \mathbf{s})}{\lambda_s(\mathbf{s})} - T \int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) d\mathbf{s}' = 0.$$

The KWCLE of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ are

$$\begin{aligned} \hat{\theta}(\mathbf{s}) =& \frac{\sum_{i=1}^{n} K_h(\mathbf{s}_i - \mathbf{s})}{\sum_{i=1}^{n} K_h(\mathbf{s}_i - \mathbf{s})(m_i - \underline{m})}, \\ \hat{\lambda}_s(\mathbf{s}) =& \frac{\sum_{i=1}^{n} K_h(\mathbf{s}_i - \mathbf{s})}{T \int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) d\mathbf{s}'}. \end{aligned} \tag{4.15}$$

The estimators $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ provided by Equation (4.15) are considered as a special case of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ in Equation (4.13). Both of them are considered in this article, where the former are used in our simulations and applications displayed in Sections 4.5 and 4.6, respectively, but the latter are considered more generally in our asymptotics in Section 4.3.

4.3 Asymptotics

Denote

$$\dot{\ell}_{h,0}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) = \partial \ell(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))/\partial \lambda_s(\mathbf{s})$$

and

$$\dot{\ell}_{h,j}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) = \partial \ell(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))/\partial \theta_j(\mathbf{s})$$

for $\theta(\mathbf{s}) = (\theta_1(\mathbf{s}), \cdots, \theta_p(\mathbf{s}))$ with $j = 1, \cdots, p$. Using the expression of stochastic integral, the left side of Equation (4.13) can be expressed as

$$\begin{aligned} &\frac{1}{T}\dot{\ell}_{h,0}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) \\ =&\frac{1}{T}\int_{\mathbf{s}' \in \mathcal{S}, t \in [0,T], m \in \mathcal{M}} \frac{K_h(\mathbf{s}' - \mathbf{s})}{\lambda_s(\mathbf{s})} N(d\mathbf{s}' \times dt \times dm) - \int_{\mathbf{s}' \in \mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) d\mathbf{s}' \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} &\frac{1}{T}\dot{\ell}_{h,j}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) \\ =&\frac{1}{T}\int_{\mathbf{s}' \in \mathcal{S}, t \in [0,T], m \in \mathcal{M}} K_h(\mathbf{s}' - \mathbf{s}) \frac{\dot{f}_{j,\theta(\mathbf{s})}(m)}{f_{\theta(\mathbf{s})}(m)} N(d\mathbf{s}' \times dt \times dm),\ j = 1, \cdots, p, \end{aligned} \tag{4.17}$$

where $\dot{f}_{j,\theta(\mathbf{s})}(m) = \partial f_{\theta(\mathbf{s})}(m)/\partial \theta_j(\mathbf{s})$, $j = 1, \cdots, p$. Then, $\hat{\lambda}_s(\mathbf{s})$ can be solved by $\dot{\ell}_{h,0}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) = 0$ and $\hat{\theta}(\mathbf{s})$ can be solved by $\dot{\ell}_{h,j}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})) = 0$, $j = 1, \cdots, p$, which implies that (4.16) and (4.17) can be treated as equations of Z-estimation (e.g. Chapter 5 of van der Vaart (1998)). However, we cannot directly use the conclusions since dependence is presented on right sides of (4.16) and (4.17). Therefore, it is necessary to provide rigorous proofs of the asymptotic properties of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$.

Denote $\theta_0(\mathbf{s}) = (\theta_{0,1}(\mathbf{s}), \cdots, \theta_{0,p}(\mathbf{s}))^T$ and $\theta(\mathbf{s}) = (\theta_1(\mathbf{s}), \cdots, \theta_p(\mathbf{s}))^T$. In the following, we always use $\theta_0(\mathbf{s})$ and $\lambda_{s0}(\mathbf{s})$ to represent the true functions of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$, $\Gamma_0((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2))$, $g_0((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2))$, and $\omega_0((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2))$ to represent the true functions of $\Gamma((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2))$, $g((\mathbf{s}_1, t_1, m_1), (\mathbf{s}_2, t_2, m_2))$, and $\omega((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2))$, respectively. We use E_λ and V_λ to denote the expected value and variance of a certain expression, if the first-order intensity function of N is $\lambda(\mathbf{s}, t, m) = \lambda_s(\mathbf{s}) f_{\theta(\mathbf{s})}(m)$. Then, E_{λ_0} and V_{λ_0} are the true expected value and true variance under $\lambda_0(\mathbf{s}, t, m) = \lambda_{s0}(\mathbf{s}) f_{\theta_0(\mathbf{s})}(m)$, respectively.

The asymptotic properties of $\hat{\lambda}_s(\mathbf{s})$ and $\hat{\theta}(\mathbf{s})$ are considered under the case when the domain increases. We assume $\mathcal{S}$ and $\mathcal{M}$ are bounded such that $N(D)$ is finite if T is finite, where $D = \mathcal{S} \times [0, T] \times \mathcal{M}$. The consistency and asymptotic normality of $\hat{\lambda}_s(\mathbf{s})$ and $\hat{\theta}(\mathbf{s})$ are considered under $T \to \infty$. We also need the following regularity conditions.

Additional Regularity Conditions:

(C5) Both $\mathcal{S}$ and $\mathcal{M}$ are connected measurable subset with $|\partial\mathcal{S}| = |\partial\mathcal{M}| = 0$.

(C6) Both $\lambda_s(\mathbf{s})$ and $\theta(\mathbf{s})$ are twice continuously differentiable in any $\mathbf{s} \in \mathcal{S}$.

(C7) $f_{\theta'}(m)$ is twice continuously differentiable in $\theta' \in \Theta$ for any $m \in \mathcal{M}$.

(C8) If $f_{\theta'}(m) = f_{\theta''}(m)$ almost surely for all $\mathbf{s} \in \mathcal{S}$, then $\theta' = \theta''$, and the inverse of the Fisher Information matrix $I(\theta') = (I_{ij}(\theta'))_{p\times p}$ exists for all $\theta' \in \Theta$.

(C9) Θ is compact.

(C10) There exists a function $H(m)$ such that $E_{\theta'}|H(m)| < \infty$ and $|\log f_{\theta'}(m) - \log f_{\theta''}(m)| < H(m)$ for all $m \in \mathcal{M}$ and $\theta', \theta'' \in \Theta$.

(C11) The solutions of $\lambda_s(\mathbf{s})$ and $\theta(\mathbf{s})$ in Equation (4.13) are unique.

(C12) The true function satisfies $\lambda_{s0}(\mathbf{s}) > 0$ at any interior point $\mathbf{s} \in \mathcal{S}$.

(C13) The true pair correlation function $g_0((\mathbf{s}_1, 0, m_1), (\mathbf{s}_2, t, m_2))$ is continuous on $(\mathcal{S} \times \mathbb{R} \times \mathcal{M})^2$.

Lemma 1 *Assume Conditions (C1)–(C8) hold. For any interior point* $\mathbf{s} \in \mathcal{S}$ *with* $\lambda_s(\mathbf{s}) > 0$, *let* $\mu_h(\mathbf{s}) = (\mu_{h,0}(\mathbf{s}), \cdots, \mu_{h,p}(\mathbf{s}))^T$ *with*

$$\begin{aligned} \mu_{h,0}(\mathbf{s}) =& \frac{1}{\lambda_s(\mathbf{s})} \int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S}} K(\mathbf{u})[\lambda_{s0}(\mathbf{s}+h\mathbf{u}) - \lambda_s(\mathbf{s})]d\mathbf{u} \\ \mu_{h,j}(\mathbf{s}) =& \int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S}, m\in\mathcal{M}} K(\mathbf{u})\lambda_{s0}(\mathbf{s}+h\mathbf{u}) \frac{\partial \log f_{\theta(\mathbf{s})}(m)}{\partial \theta_j(\mathbf{s})} \\ & \times f_{\theta_0(\mathbf{s}+h\mathbf{u})}(m) d\mathbf{u}dm, \ j = 1, \cdots, p, \end{aligned} \tag{4.18}$$

and $\Sigma_h(\mathbf{s}) = (\sigma_{h,ij}(\mathbf{s}))_{(p+1)\times(p+1)}$ *with*

$$\begin{aligned}
\sigma_{h,00}(\mathbf{s}) =& \frac{1}{\lambda_s^2(\mathbf{s})} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s}) \\
& \times \omega_0((\mathbf{s}',m'),(\mathbf{s}'',m''))dm''d\mathbf{s}''dm'd\mathbf{s}', \\
\sigma_{h,0i}(\mathbf{s}) =& \frac{1}{\lambda_s(\mathbf{s})} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s})\omega_0((\mathbf{s}',m'),(\mathbf{s}'',m'')) \\
& \times \frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_i(\mathbf{s})} d\mathbf{u}dm''d\mathbf{s}'dm', \\
\sigma_{h,ij}(\mathbf{s}) =& \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s})\omega_0((\mathbf{s}',m'),(\mathbf{s}'',m'')) \\
& \times \frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_i(\mathbf{s})} \frac{\partial \log f_{\theta(\mathbf{s})}(m'')}{\partial\theta_j(\mathbf{s})} d\mathbf{s}''dm''d\mathbf{s}'dm',
\end{aligned} \tag{4.19}$$

for $i,j = 1,\cdots,p$. *If* h *is fixed, then*

$$\sqrt{T}[\frac{1}{T}\dot{\ell}_h(\theta(\mathbf{s}),\lambda_s(\mathbf{s})) - \mu_{h,0}(\mathbf{s})] \xrightarrow{D} N(0,\Sigma_h(\mathbf{s})) \tag{4.20}$$

as $T\to\infty$.

Proof: According to Corollary 1, there is

$$V^{-\frac{1}{2}}[\dot{\ell}_{h,0}(\theta(\mathbf{s}),\lambda_s(\mathbf{s}))]\{\dot{\ell}_{h,0}(\theta(\mathbf{s}),\lambda_s(\mathbf{s})) - E[\dot{\ell}_{h,0}(\theta(\mathbf{s}),\lambda_s(\mathbf{s}))]\} \xrightarrow{D} N(0,1)$$

as $T\to\infty$. Then, it is enough to compute the expected value and variance of $\dot{\ell}_{h,0}(\theta(\mathbf{s}),\lambda_s(\mathbf{s}))$. First, we consider the expected value of Equations (4.16) and (4.17) for a given $\mathbf{s}\in\mathcal{M}$. Straightforwardly,

$$\begin{aligned}
& E_{\lambda_0}[\frac{1}{T}\dot{\ell}_{h,0}(\theta(\mathbf{s}),\lambda_s(\mathbf{s}))] \\
=& \int_{\mathbf{s}'\in\mathcal{S},m\in\mathcal{M}} \frac{K_h(\mathbf{s}'-\mathbf{s})}{\lambda_s(\mathbf{s})}[\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m) - \lambda_s(\mathbf{s})f_{\theta(\mathbf{s})}(m)]d\mathbf{s}'dm \\
=& \int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S},m\in\mathcal{M}} \frac{K(\mathbf{u})}{\lambda_s(\mathbf{s})}[\lambda_{s0}(\mathbf{s}+h\mathbf{u})f_{\theta_0(\mathbf{s}+h\mathbf{u})}(m) - \lambda_s(\mathbf{s})f_{\theta(\mathbf{s})}(m)]d\mathbf{u}dm.
\end{aligned}$$

and

$$\begin{aligned}
& E_{\lambda_0}[\frac{1}{T}\dot{\ell}_{h,j}(\theta(\mathbf{s}),\lambda_s(\mathbf{s}))] \\
=& \int_{\mathbf{s}'\in\mathcal{S},m\in\mathcal{M}} \frac{K_h(\mathbf{s}'-\mathbf{s})\frac{\partial f_{\theta(\mathbf{s})}(m)}{\partial\theta_j(\mathbf{s})}}{f_{\theta(\mathbf{s})}(m)}\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m)d\mathbf{s}'dm \\
=& \int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S},m\in\mathcal{M}} \frac{K(\mathbf{u})\frac{\partial f_{\theta(\mathbf{s})}(m)}{\partial\theta_j(\mathbf{s})}}{f_{\theta(\mathbf{s})}(m)}\lambda_{s0}(\mathbf{s}+h\mathbf{u})f_{\theta_0(\mathbf{s}+h\mathbf{u})}(m)d\mathbf{s}'dm.
\end{aligned}$$

Then, $\mu_h(\mathbf{s}) = E_{\lambda_0}[T^{-1}\dot{\ell}_{h,j}(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))]$. Next, we consider the variance-covariance matrix of Equations (4.16) and (4.17). We have

$$
\begin{aligned}
& V_{\lambda_0}[\frac{1}{T}\dot{\ell}_{h,0}(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))] \\
=& \frac{1}{T^2\lambda_s^2(\mathbf{s})} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},t',t''\in[0,T],m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s}) \\
& \times \Gamma((\mathbf{s}',0,m'),(\mathbf{s}'',t''-t',m''))d\mathbf{s}''dt''dm''d\mathbf{s}'dt'dm' \\
=& \frac{1}{T^2\lambda_s^2(\mathbf{s})} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s}) \\
& \times \left[\int_{t'\in[0,T],v\in[-t',T-t']} \Gamma_0((\mathbf{s}',0,m'),(\mathbf{s}'',v,m''))dvdt'\right] d\mathbf{u}dm''d\mathbf{s}'dm', \\
& \mathrm{Cov}_{\lambda_0}[\frac{1}{T}\dot{\ell}_{h,0}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})), \frac{1}{T}\dot{\ell}_{h,i}(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))] \\
=& \frac{1}{T^2\lambda_s(\mathbf{s})} \int_{\mathbf{s}',\mathbf{s}+h\mathbf{u}\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s}+h\mathbf{u})K(\mathbf{u})\frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_i(\mathbf{s})} \\
& \times \Gamma((\mathbf{s}',0,m'),(\mathbf{s}'',t''-t',m''))d\mathbf{s}''dt''dm''d\mathbf{s}'dt'dm' \\
=& \frac{1}{T^2\lambda_s(\mathbf{s})} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s})\frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_i(\mathbf{s})} \\
& \times \left[\int_{t'\in[0,T],v\in[-t',T-t']} \Gamma_0((\mathbf{s}',0,m'),(\mathbf{s}'',v,m''))dt'dv\right] d\mathbf{s}''dm''d\mathbf{s}'dm',
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathrm{Cov}_{\lambda_0}[\frac{1}{T}\dot{\ell}_{h,i}(\theta(\mathbf{s}), \lambda_s(\mathbf{s})), \frac{1}{T}\dot{\ell}_{h,j}(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))] \\
=& \frac{1}{T^2} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},t',t''\in[0,T],m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s})\frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_i(\mathbf{s})}\frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_j(\mathbf{s})} \\
& \times \Gamma((\mathbf{s}',0,m'),(\mathbf{s}'',t''-t',m''))d\mathbf{s}''dt''dm''d\mathbf{s}'dt'dm' \\
=& \frac{1}{T^2} \int_{\mathbf{s}',\mathbf{s}''\in\mathcal{S},m',m''\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})K_h(\mathbf{s}''-\mathbf{s})\frac{\partial \log f_{\theta(\mathbf{s})}(m')}{\partial\theta_i(\mathbf{s})}\frac{\partial \log f_{\theta(\mathbf{s})}(m'')}{\partial\theta_j(\mathbf{s})} \\
& \times \left[\int_{t'\in[0,T],v\in[-t',T-t']} \Gamma_0((\mathbf{s}',0,m'),(\mathbf{s}'',v,m''))dt'dv\right] d\mathbf{s}''dm''d\mathbf{s}'dm'
\end{aligned}
$$

for $i, j = 1, \cdots, p$. Then, $\Sigma_h(\mathbf{s}) = \lim_{T\to\infty} V[T^{-1/2}\dot{\ell}_h(\theta(\mathbf{s}), \lambda_s(\mathbf{s}))]$. The final conclusion is drawn using the Continuous Mapping Theorem. ♢

Theorem 2 *(Consistency) Assume Conditions (C1)–(C13) hold. If* $h \to 0$ *and* $Th \to \infty$ *as* $T \to \infty$*, then* $\hat{\theta}(\mathbf{s}) \xrightarrow{P} \theta_0(\mathbf{s})$ *and* $\hat{\lambda}_s(\mathbf{s}) \xrightarrow{P} \lambda_{s0}(\mathbf{s})$ *for every interior point* $\mathbf{s}$ *of* $\mathcal{S}$*.*

Proof: Note that

$$\lim_{h\to 0}\mu_{h,0}(\mathbf{s}) = \frac{1}{\lambda_s(\mathbf{s})}\int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S},m\in\mathcal{M}} K(\mathbf{u})[\lambda_{s0}(\mathbf{s}+h\mathbf{u}) - \lambda_s(\mathbf{s})]d\mathbf{u}dm = \frac{\lambda_{s0}(\mathbf{s})}{\lambda_s(\mathbf{s})} - 1$$

which is zero if and only if $\lambda_s(\mathbf{s}) = \lambda_{s0}(\mathbf{s})$. As Equation (4.16) induces a closed form solution

$$\hat{\lambda}_s(\mathbf{s}) = \frac{\int_{\mathbf{s}'\in\mathcal{S},t\in[0,T],m\in\mathcal{M}} K_h(\mathbf{s}'-\mathbf{s})N(d\mathbf{s}'\times dt\times dm)}{T\int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S}} K(\mathbf{u})d\mathbf{u}},$$

we conclude $\hat{\lambda}_s(\mathbf{s}) \xrightarrow{P} \lambda_{s0}(\mathbf{s})$ if $h\to 0$ and $Th\to\infty$ as $T\to\infty$. For the consistency of $\hat{\theta}(\mathbf{s})$ when $\mathbf{s}$ is an interior point of $\mathcal{S}$, we consider

$$\begin{aligned}\lim_{h\to 0}\mu_{h,j}(\mathbf{s}) =& \int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S},m\in\mathcal{M}} K(\mathbf{u})\lambda_{s0}(\mathbf{s}+h\mathbf{u})\frac{\partial \log f_{\theta(\mathbf{s})}(m)}{\partial\theta_j(\mathbf{s})} f_{\theta_0(\mathbf{s}+h\mathbf{u})}(m)d\mathbf{u}dm \\ =& \lambda_{s0}(\mathbf{s})\int_{m\in\mathcal{M}} \frac{\partial \log f_{\theta(\mathbf{s})}(m)}{\partial\theta_j(\mathbf{s})} f_{\theta_0(\mathbf{s})}(m)d\mathbf{u}dm.\end{aligned}$$

According to the property of the *Kullback-Leibler information number* (e.g. Ferguson (1996), p. 112), the above is zero if and only if $\theta(\mathbf{s}) = \theta_0(\mathbf{s})$. Therefore, $\lim_{h\to 0}\mu_{h,j}(\mathbf{s}) = 0$ for $j = 1,\cdots,p$ if and only if $\theta(\mathbf{s}) = \theta_0(\mathbf{s})$. Let $\Omega_{\mathbf{s}} = \{\theta' : |\theta' - \theta_0(\mathbf{s})| \geq \gamma\}$ for a certain $\gamma > 0$. Since Θ is compact, $\Omega_{\mathbf{s}}$ is also compact. Let $\alpha = \min_{\theta'\in\Omega_{\mathbf{s}}}\{|E[T^{-1}\dot{\ell}_{h,j}(\theta', \lambda_{s0}(\mathbf{s}))]|\}$. Then, $\alpha > 0$ and

$$P(\liminf_{T\to\infty}\inf_{\theta'\in\Omega_{\mathbf{s}}}|T^{-1}\dot{\ell}_{h,j}(\theta', \lambda_{s0}(\mathbf{s}))| \geq \alpha) = 1.$$

Therefore, with probability one there exists a T' such that for all $T > T'$,

$$\inf_{\theta'\in\Omega_{\mathbf{s}}}|T^{-1}\dot{\ell}_{h,j}(\theta', \lambda_{s0}(\mathbf{s}))| \geq \alpha/2.$$

Then with probability one there is $|\hat{\theta}(\mathbf{s}) - \theta_0(\mathbf{s})| < \gamma$. Since γ is arbitrary, we have $\hat{\theta}(\mathbf{s}) \xrightarrow{P} \theta_0(\mathbf{s})$, and this is true for any interior $\mathbf{s}\in\mathcal{S}$. ♢

As long as the consistency is shown, we need to consider the asymptotic normality. To derive the asymptotic normality, we need two Taylor expansions. The first is

$$\lambda_{s0}(\mathbf{s}+h\mathbf{u}) = \lambda_{s0}(\mathbf{s}) + h\dot{\lambda}_{s0}^T(\mathbf{s})\mathbf{u} + \frac{h^2}{2}\mathbf{u}^T\ddot{\lambda}_{s0}(\mathbf{s})\mathbf{u} + o(h^2) \tag{4.21}$$

where $\dot{\lambda}_{s0}(\mathbf{s})$ is a two-dimensional gradient vector and $\ddot{\lambda}_{s0}(\mathbf{s})$ is a 2×2-dimensional Hessian matrix. The second is

$$\begin{aligned}
&\lambda_{s0}(\mathbf{s}+h\mathbf{u}) f_{\theta_0(\mathbf{s}+h\mathbf{u})}(m) \\
=&\lambda_{s0}(\mathbf{s}) f_{\theta_0(\mathbf{s})}(m) + h[\lambda_{s0}(\mathbf{s}) \dot{f}^T_{\theta_0(\mathbf{s})}(m) \dot{\theta}_0(\mathbf{s})^T \mathbf{u} + f_{\theta_0(\mathbf{s})}(m) \dot{\lambda}^T_{s0}(\mathbf{s}) \mathbf{u}] \\
&+ \frac{h^2}{2}[\lambda_{s0}(\mathbf{s}) \mathbf{u}^T \dot{f}_{\theta_0(\mathbf{s})}(m) \ddot{\theta}_0(\mathbf{s}) \mathbf{u} + \lambda_{s0}(\mathbf{s}) \mathbf{u}^T \dot{\theta}_0(\mathbf{s}) \ddot{f}_{\theta_0(\mathbf{s})}(m) \dot{\theta}^T_0(\mathbf{s}) \mathbf{u} \\
&\qquad + 2\mathbf{u}^T \dot{\lambda}_{s0}(\mathbf{s}) \dot{f}^T_{\theta_0(\mathbf{s})}(m) \dot{\theta}^T_0(\mathbf{s}) \mathbf{u}] + o(h^2),
\end{aligned} \tag{4.22}$$

where $\dot{f}_{\theta_0(\mathbf{s})}(m)$ is a p-dimensional gradient vector, $\ddot{f}_{\theta_0(\mathbf{s})}(m)$ is a $p \times p$-dimensional Hessian matrix, $\dot{\theta}_0(\mathbf{s})$ is a $p \times 2$-dimensional gradient matrix, and $\ddot{\theta}_0(\mathbf{s})$ is a $p \times 2 \times 2$-dimensional Hessian third-order tensor.

Theorem 3 *Assume Conditions (C1)–(C13) hold. Let* $\nu(\mathbf{s}) = (\nu_0(\mathbf{s}), \cdots, \nu_p(\mathbf{s}))$, *where*

$$\nu_0(\mathbf{s}) = \frac{1}{2} \int_{\mathbb{R}^2} K(\mathbf{u}) \mathbf{u}^T \ddot{\lambda}_s(\mathbf{s}) \mathbf{u} d\mathbf{u}$$

and

$$\begin{aligned}
\nu_j(\mathbf{s}) =& \frac{1}{2} \int_{\mathbf{u} \in \mathbb{R}^2, m \in \mathcal{M}} K(\mathbf{u})[\lambda_{s0}(\mathbf{s}) \mathbf{u}^T \dot{f}_{\theta_0(\mathbf{s})}(m) \ddot{\theta}_0(\mathbf{s}) \mathbf{u} + \lambda_{s0}(\mathbf{s}) \mathbf{u}^T \dot{\theta}_0(\mathbf{s}) \ddot{f}_{\theta_0(\mathbf{s})}(m) \dot{\theta}^T_0(\mathbf{s}) \mathbf{u} \\
&+ 2\mathbf{u}^T \dot{\lambda}_{s0}(\mathbf{s}) \dot{f}^T_{\theta_0(\mathbf{s})}(m) \dot{\theta}^T_0(\mathbf{s}) \mathbf{u}] dm d\mathbf{u}, j = 1, \cdots, p.
\end{aligned}$$

Let $\Sigma_0(\mathbf{s}) = (\sigma_{0,ij}(\mathbf{s}))_{(p+1)\times(p+1)}$ *for* $i, j = 0, \cdots, p$ *with* $\sigma_{0,00}(\mathbf{s}) = \xi / \lambda_{s0}(\mathbf{s})$, $\sigma_{0,0i}(\mathbf{s}) = 0$, *and* $\sigma_{0,ij}(\mathbf{s}) = \xi \lambda_{s0}(\mathbf{s}) I_{ij}(\theta_0(\mathbf{s}))$ *for* $i, j = 1, \cdots, p$. *If* $h = cT^{-1/6}$ *for a constant* c *as* $T \to \infty$, *then*

$$h\sqrt{T} \dot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s})) \xrightarrow{D} N(c^2 \nu(\mathbf{s}), \Sigma_0(\mathbf{s})/c^2). \tag{4.23}$$

Proof: Consider the bias and the variance of $T^{-1} \dot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s}))$ as $T \to \infty$ with $h \to 0$ and $Th \to \infty$. Let $\lambda_s(\mathbf{s}) = \lambda_{s0}(\mathbf{s})$ and $f_{\theta(\mathbf{s})}(\mathbf{s}) = f_{\theta_0(\mathbf{s})}(\mathbf{s})$ in Equations (4.18) and (4.19). The first-component of the bias is

$$\begin{aligned}
\nu_0(\mathbf{s}) =& \lim_{h \to 0} \frac{1}{h^2} \int_{\mathbf{s}+h\mathbf{u}} K(\mathbf{u})[\lambda_{s0}(\mathbf{s}+h\mathbf{u}) - \lambda_{s0}(\mathbf{s})] d\mathbf{s} \\
=& \lim_{h \to 0} \frac{1}{h^2} \int_{\mathbb{R}^2} K(\mathbf{u})[h \dot{\lambda}^T_{s0}(\mathbf{s}) \mathbf{u} + \frac{h^2}{2} \mathbf{u}^T \ddot{\lambda}_{s0}(\mathbf{s}) \mathbf{u}] d\mathbf{u} \\
=& \frac{1}{2} \int_{\mathbb{R}^2} K(\mathbf{u}) \mathbf{u}^T \ddot{\lambda}_{s0}(\mathbf{s}) \mathbf{u} d\mathbf{u}.
\end{aligned}$$

Since

$$\int_{\mathcal{M}} f_{\theta_0(\mathbf{s})}(m)\frac{\partial \log f_{\theta_0(\mathbf{s})}(m)}{\partial \theta_j(\mathbf{s})} f_{\theta_0(\mathbf{s})}(m)dm = 0, j = 1, \cdots, p,$$

the $(j+1)$-th component of the bias for $j \neq 0$ satisfies

$$\nu_j(\mathbf{s}) = \lim_{h\to 0}\frac{1}{h^2}\mu_{h,j}(\mathbf{s}).$$

The limit of $K_h(\mathbf{s}'-\mathbf{s}+h\mathbf{u})$ is not zero as $h \to 0$ if and only if $\mathbf{s}' = \mathbf{s}$, which induces the expression of the limit as $K(\mathbf{u}/h)/h^2$. To derive the variance of $T^{-1}\dot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s}))$, according to Equation (4.7) there is

$$\begin{aligned}
&\omega_0((\mathbf{s}', m'), (\mathbf{s}'', m''))\\
=&[\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m')][\lambda_{s0}(\mathbf{s}'')f_{\theta_0(\mathbf{s}'')}(m'')]\int_{\mathbb{R}}[g_0((\mathbf{s}', 0, m'), (\mathbf{s_2}, t, m_2)) - 1]dt\\
&+ [\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m')]\delta_{(\mathbf{s}',m')}(\mathbf{s}'', m'').
\end{aligned}$$

According to Condition (C13), only the second term is necessary to be considered in the variance expression of $T^{-1}\dot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s}))$ as $h \to 0$. Then,

$$\begin{aligned}
\lim_{h\to 0} h^2\sigma_{h,00} =& \lim_{h\to 0}\frac{h^2}{\lambda_{s0}^2(\mathbf{s})}\int_{\mathbf{s}'\in\mathcal{S},m'\in\mathcal{M}} K_h^2(\mathbf{s}'-\mathbf{s})\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m')dm'd\mathbf{s}\\
=& \lim_{h\to 0}\frac{1}{\lambda_{s0}^2(\mathbf{s})}\int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S}} K^2(\mathbf{u})\lambda_{s0}(\mathbf{s}+h\mathbf{u})d\mathbf{u}\\
=&\frac{1}{\lambda_{s0}(\mathbf{s})}\int_{\mathbb{R}^2} K^2(\mathbf{u})d\mathbf{u}\\
=&\sigma_{0,00}.
\end{aligned}$$

Using

$$\begin{aligned}
&\lim_{h\to 0} h^2\sigma_{h,0i}\\
&= \lim_{h\to 0}\frac{h^2}{\lambda_{s0}(\mathbf{s})}\int_{\mathbf{s}'\in\mathcal{S},m'\in\mathcal{M}} K_h^2(\mathbf{s}'-\mathbf{s})\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m')\frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,i}(\mathbf{s})(m')}dm'd\mathbf{s}'\\
&= \lim_{h\to 0}\frac{1}{\lambda_{s0}(\mathbf{s})}\int_{\mathbf{s}+h\mathbf{u}\in\mathcal{S},m'\in\mathcal{M}} K^2(\mathbf{u})\lambda_{s0}(\mathbf{s}+h\mathbf{u})f_{\theta_0(\mathbf{s}+h\mathbf{u})}(m')\frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,i}(\mathbf{s})(m')}dm'd\mathbf{u}\\
&= \int_{\mathbf{u}\in\mathbb{R}^2,m'\in\mathcal{M}} K^2(\mathbf{u})f_{\theta_0(\mathbf{s})}(m')\frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,i}(\mathbf{s})(m')}dm'd\mathbf{u}\\
&= 0,
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{h\to 0} h^2\sigma_{h,ij} \\
&= \lim_{h\to 0} h^2 \int_{\mathbf{s}'\in\mathcal{S},m'\in\mathcal{M}} K_h^2(\mathbf{s}'-\mathbf{s})\lambda_{s0}(\mathbf{s}')f_{\theta_0(\mathbf{s}')}(m')\frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,i}(\mathbf{s})(m')}\frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,j}(\mathbf{s})(m')}dm'd\mathbf{s}' \\
&= \lambda_{s0}(\mathbf{s})\int_{\mathbb{R}^2} K^2(\mathbf{u})[\int_{m'\in\mathcal{M}} \frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,i}(\mathbf{s})(m')}\frac{\partial \log f_{\theta_0(\mathbf{s})}(m')}{\partial\theta_{0,j}(\mathbf{s})(m')}dm']d\mathbf{u} \\
&= \sigma_{0,ij},
\end{aligned}$$

for $i,j = 1,\cdots,p$, the expression of $\nu_j(\mathbf{s})$ is derived. To make both limits of bias and variance exist, we should choose $h = cT^{-1/6}$. Then

$$T^{-\frac{1}{3}}\dot{\ell}_h(\theta_0(\mathbf{s}),\lambda_{s0}(\mathbf{s})) = \frac{h\sqrt{T}}{c}[\frac{1}{T}\dot{\ell}_h(\theta_0(\mathbf{s}),\lambda_{s0}(\mathbf{s})) - \mu_h(\mathbf{s})] + \frac{c^2}{h^2}\mu_h(\mathbf{s}).$$

Clearly, the first term of above is a stochastic term with mean zero and the limit of covariance equal to $\Sigma_0(\mathbf{s})/c^2$ as $T\to\infty$. The second term above is a deterministic term with the limit equal to $c^2\nu(\mathbf{s})$ as $T\to\infty$. Then, Equation (4.23) is drawn. $\diamondsuit$

Remark: The matrix $\Sigma_0(\mathbf{s})$ can be expressed as

$$\Sigma_0(\mathbf{s}) = \begin{pmatrix} \sigma_{0,00}(\mathbf{s}) & 0 \\ 0 & \xi I(\theta_0(\mathbf{s})) \end{pmatrix},$$

where $I(\theta_0(\mathbf{s}))$ is the Fisher information matrix.

Corollary 2 *Assume all the assumptions of Theorem 3 hold. If $h = cT^{-1/6}$ for a constant c as $T\to\infty$, then*

$$T^{\frac{1}{3}}\left[\begin{pmatrix}\hat{\lambda}_s(\mathbf{s})\\ \hat{\theta}(\mathbf{s})\end{pmatrix} - \begin{pmatrix}\lambda_{s0}(\mathbf{s})\\ \theta_0(\mathbf{s})\end{pmatrix}\right] \xrightarrow{D} N(c^2\Sigma_0^{-1}(\mathbf{s})\nu(\mathbf{s}), \Sigma_0^{-1}(\mathbf{s})/c^2). \tag{4.24}$$

Proof: Let $Z(\mathbf{s}) = (\hat{\lambda}_s(\mathbf{s}),\hat{\theta}(\mathbf{s}))^T$ and $z_0(\mathbf{s}) = (\lambda_{s0}(\mathbf{s}),\theta_0(\mathbf{s}))^T$. Consider the Taylor expansion

$$\begin{aligned}
&T^{-\frac{2}{3}}\dot{\ell}_h(\hat{\theta}(\mathbf{s}),\hat{\lambda}_s(\mathbf{s})) \\
=&T^{-\frac{2}{3}}\dot{\ell}_h(\theta_0(\mathbf{s}),\lambda_{s0}(\mathbf{s})) + [\frac{1}{T}\ddot{\ell}_h(\theta_0(\mathbf{s}),\lambda_{s0}(\mathbf{s}))][T^{\frac{1}{3}}(Z(\mathbf{s}) - z_0(\mathbf{s}))] + o_p(1).
\end{aligned}$$

Then,

$$T^{\frac{1}{3}}(Z(\mathbf{s}) - z_0(\mathbf{s})) = -[\frac{1}{T}\ddot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s}))]^{-1}[T^{-\frac{2}{3}}\dot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s}))] + o_p(1).$$

Note that

$$\frac{1}{T}\ddot{\ell}_h(\theta_0(\mathbf{s}), \lambda_{s0}(\mathbf{s})) \xrightarrow{P} \Sigma_0(\mathbf{s})$$

for all $\mathbf{s} \in \mathcal{S}$. Equation (4.24) is concluded. ♢

Remark: Corollary (2) recommends the optimal bandwidth in the KWCL approach is to choose $h = cT^{-1/6}$ for a certain c as $T \to \infty$, which may depend on the kernel function, the second-order partial derivatives of $\lambda_{s0}(\mathbf{s})$ and $f_{\theta_0(\mathbf{s})}(\mathbf{s})$, and the integrated covariance function $\omega_0((\mathbf{s}_1, m_1), (\mathbf{s}_2, m_2))$.

4.4 Bandwidth Selection

The most popular criterion to judge the accuracy of a bandwidth h in a kernel-based method is the *mean integrated square error* (MISE) (e.g., page 344 of van der Vaart, 1998). It can be defined as

$$MISE(\hat{f}) = \int E_f(\hat{f}(x) - f(x))^2 dx$$

$$= \int \mathrm{var}_f \hat{f}(x) dx + \int (E_f \hat{f}(x) - f(x))^2,$$

which is the sum of an integrated variance term and a bias term. Both terms need to be small for the MISE to be small.

Since the dimension of estimators in the KWCL approach is greater than one, we modify the classical MISE criterion and define

$$MISE(\hat{\lambda}_s(\mathbf{s}), \hat{\theta}(\mathbf{s})) = \int_{\mathcal{S}} E[(\hat{\lambda}_s(\mathbf{s}) - \lambda_{s0}(\mathbf{s}))^2 + \|\hat{\theta}(\mathbf{s}) - \theta_0(\mathbf{s})\|^2] d\mathbf{s}. \qquad (4.25)$$

According to Equation (4.24) in Corollary 2, if we choose $h = cT^{-1/6}$ as $T \to \infty$, then

$$\lim_{T\to\infty} T^{\frac{2}{3}} MISE(\hat{\lambda}_s(\mathbf{s}), \hat{\theta}(\mathbf{s})) = \int_{\mathcal{S}} \{c^4[\mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\nu(\mathbf{s})]^2 + \frac{1}{c^2}\mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\mathbf{1}\} d\mathbf{s},$$

where $\mathbf{1}$ is the $(p+1)$-dimensional vector with all of its components equal to one. As the right side of the above equation is maximized when

$$4c^3 \int_{\mathcal{S}} [\mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\nu(\mathbf{s})]^2 d\mathbf{s} - \frac{2}{c^3} \int_{\mathcal{S}} \mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\mathbf{1} d\mathbf{s} = 0, \tag{4.26}$$

we can solve for c:

$$c = \left\{ \frac{\int_{\mathcal{S}} \mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\mathbf{1} d\mathbf{s}}{2\int_{\mathcal{S}} [\mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\nu(\mathbf{s})]^2 d\mathbf{s}} \right\}^{\frac{1}{6}}. \tag{4.27}$$

The optimal bandwidth h_{opt} in the KWCL approach is

$$h_{opt} = cT^{-1/6} = \left\{ \frac{\int_{\mathcal{S}} \mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\mathbf{1} d\mathbf{s}}{2T\int_{\mathcal{S}} [\mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\nu(\mathbf{s})]^2 d\mathbf{s}} \right\}^{\frac{1}{6}},$$

which induces the optimal $MISE_{opt}$ is about

$$MISE_{opt} \approx \frac{3T^{-\frac{2}{3}}}{2^{\frac{2}{3}}} \left[\int_{\mathcal{S}} [\mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\nu(\mathbf{s})]^2 d\mathbf{s} \right]^{\frac{1}{3}} \left[\int_{\mathcal{S}} \mathbf{1}^T \Sigma_0^{-1}(\mathbf{s})\mathbf{1} d\mathbf{s} \right]^{\frac{2}{3}}.$$

The asymptotic distribution of the KWCLE of $\lambda_s(\mathbf{s})$ and $\theta(\mathbf{s})$ under the optimal bandwidth is obtained if (4.27) is used in Corollary 2.

We note that the h_{opt} depends on the length of the observed period T, the kernel function K, the unknown intensity function $\lambda_s(\mathbf{s}_0)$ for points, and the unknown function $\theta(\mathbf{s})$ in $f_{\theta(\mathbf{s})}$ for marks, which should be considered in practice.

4.5 Simulation Study

The performance of the KWCLEs of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ was examined by simulation studies. To evaluate the performance of proposed estimators, we generated the marked point process N from Model (4.2) on $\mathcal{S} \times [0, T] \times \mathcal{M}$, where $\mathcal{S} = [0, 1]^2$ represented the study region of points within the unit square, $\mathcal{M} = \mathbb{R}^+$ represented

the possible values of marks, and $[0, T]$ represented the period of observation. The length of the observation period was completely determined by T. According to our asymptotic studies, the KWCLE of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ should be close to the true values if T is large. Therefore, our study was focused on the behavior of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ as T increased. The marked point process N defined by Model (4.2) was partially stationary. Since its first-order intensity function $\lambda(\mathbf{s}, t, m)$ was independent of T, we only needed to focus on different choices of $\lambda_s(\mathbf{s})$ and $f_{\theta(\mathbf{s})}(m)$ in our simulation studies. As long as $\lambda_s(\mathbf{s})$ and $f_{\theta(\mathbf{s})}(m)$ were determined, the expected number of points within the study region was given as

$$\kappa = E[N([0,1]^2 \times [0,T] \times \mathbb{R}^+)] = T \int_{[0,1]^2} \lambda_s(\mathbf{s}) d\mathbf{s}, \tag{4.28}$$

which indicated that the expected number of points proportionally increased as T increased.

Note that only the first-order intensity function was described by Model (4.2). The choice of the second-order intensity function was flexible. In order to evaluate the impact of the second-order intensity function on the performance of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$, we considered two different types of marked point processes: the marked Poisson and the marked (Neyman-Scott) cluster processes. The definitions of these two processes can be easily generalized from the classical definitions of the Poisson point and the (Neyman-Scott) cluster point processes in many textbooks (e.g. p. 13 and p. 179 in Daley and Vere-Jones, 2003). As different types of second-order intensity functions were derived using these two processes, the comparison could reflect the influence of the second-order intensity functions on efficiency of our estimators.

In all of our simulation studies, we used

$$\lambda_s(\mathbf{s}) = 400\beta(s_x, a, a)\beta(s_y, a, a) \tag{4.29}$$

for both the marked Poisson and the marked cluster processes, where $\mathbf{s} = (s_x, s_y) \in [0,1]^2$ and $\beta(z, a, a)$ was the probability density function (PDF) of $Beta(a, a)$ distribution. According to this setting, the expected number of points of a unit time

interval was 400 and the intensity function of N_s was $T\lambda_s(\mathbf{s})$. The parameter T was chosen to be 2.5, 5, 12.5 and 25 such that we had the expected number of points κ to be 1000, 2000, 5000 and 10000 respectively. We also used $a = 1$ and $a = 2$ for uniform and non-uniform $\lambda_s(\mathbf{s})$, respectively.

After the first-order intensity function was chosen, we generated both the marked Poisson and the marked cluster processes. For the marked Poisson process, we independently generated points from $[0,1]^2 \times [0,T]$ with the intensity function $T\lambda_s(\mathbf{s})$, where $\lambda_s(\mathbf{s})$ was given by (4.29). Example of two 3-dimensional Poisson processes (no cluster) are shown in the upper panel of Figure 4.1, where $a = 1$ represents uniform $\lambda_s(\mathbf{s})$ and $a = 2$ represents non-uniform $\lambda_s(\mathbf{s})$. For illustration we set the expected number of points κ to be 250 here. As long as points were derived, we independently generated marks from the exponential distribution with the PDF given by $f_{\theta(\mathbf{s})}(m) = \theta(\mathbf{s})e^{-\theta(\mathbf{s})m}$, where

$$\theta(\mathbf{s}) = e^{\frac{\|\mathbf{s}-\mathbf{s}_0\|^2}{3}} \tag{4.30}$$

with the center $\mathbf{s}_0 = (0.5, 0.5)$. According to (4.30), the localized parameter $\theta(\mathbf{s})$ attained its minimum value 1 at $\mathbf{s}_0$ or maximum value 5.294 at $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$, respectively. Smaller values of $\theta(\mathbf{s})$ corresponded to higher probability of getting large marks. From this setting, we expected to see more large marks if points were close to $\mathbf{s}_0$; otherwise for points close to the boundary, more small marks were expected. The evaluation of the exponential distribution for marks was important since it was resulted from the well-known power-law (Turcotte *et. al*, 2006) and Gutenberg-Richter law (Gutenberg and Richter, 1954) for frequency-size relationships in forest wildfires and earthquakes, respectively.

For the marked cluster process, we first independently generated the parent points from a Poisson process on $[0,1]^2 \times [0,T]$ with the intensity function equal to $T\lambda_s(\mathbf{s})/k$, where $\lambda_s(\mathbf{s})$ was given by (4.29). After the parent points were derived, we then generated the offspring points using the definition of the cluster process. We generated a Poisson number with mean b of offspring points for each parent point, where the

position of offspring points relative to its parent point was generated from an independent bivariate normal distribution with standard deviation σ. In our simulations, we fixed $b = 4$ and chose $\sigma = 0.02$ or $\sigma = 0.04$. Therefore, the expected number of parent points was κ/b, the expected number of offspring points was κ, and the intensity function of the offspring points was exactly equal to $T\lambda_s(\mathbf{s})$. Example of two 3-dimensional Poisson cluster processes are given in the lower panel of Figure 4.1, where $\sigma = 0.02$ and 0.04. For illustration we set the mean number of offspring b to be 2 and the expected number of points $\kappa = 500$. The cluster structure can be observed more easily when we set smaller standard deviation σ to generate the offspring points. As long as the points in the marked cluster process were derived, we used the same way as we did for the marked Poisson process to generate the marks.

After points and marks in the marked Poisson or marked cluster process were derived, we computed the KWCLE $\hat{\theta}(\mathbf{s})$ of $\theta(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ of $\lambda_s(\mathbf{s})$ by Equation (4.15). To apply the equation, we chose $\underline{m} = 0$ and used the isotropic Gaussian kernel as

$$K(\mathbf{s}) = \phi(s_x)\phi(s_y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\|\mathbf{s}\|^2}{2}},$$

where ϕ represented the PDF of $N(0,1)$. To apply Equation (4.15), we needed to figure out a method to compute the denominator of $\hat{\lambda}_s(\mathbf{s})$. For the Gaussian kernel, the answer was clear as we had

$$\begin{aligned}\int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) d\mathbf{s}' &= \int_{[0,1]^2} \frac{1}{h^2} \phi\left(\frac{\mathbf{s}'_x - \mathbf{s}_x}{h}\right) \phi\left(\frac{\mathbf{s}'_y - \mathbf{s}_y}{h}\right) d\mathbf{s}_x d\mathbf{s}_y \\ &= \left[\Phi\left(\frac{1-\mathbf{s}_x}{h}\right) - \Phi\left(\frac{0-\mathbf{s}_x}{h}\right)\right]\left[\Phi\left(\frac{1-\mathbf{s}_y}{h}\right) - \Phi\left(\frac{0-\mathbf{s}_y}{h}\right)\right],\end{aligned}$$

where Φ represented the CDF of $N(0,1)$.

For comparison, we computed the mean integrated square error (MISE) of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ in each replication of our simulations. To derive the MISEs, we needed to obtain the mean square error (MSE) of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ at each selected $\mathbf{s} \in [0,1]^2$. The MISE of $\hat{\theta}(\mathbf{s})$ was defined as

$$MISE(\hat{\theta}(\mathbf{s})) = E \int_{[0,1]^2} \left(\hat{\theta}(\mathbf{s}) - \theta(\mathbf{s})\right)^2 d\mathbf{s} = \int_{[0,1]^2} \mathrm{MSE}(\hat{\theta}(\mathbf{s})) d\mathbf{s}.$$

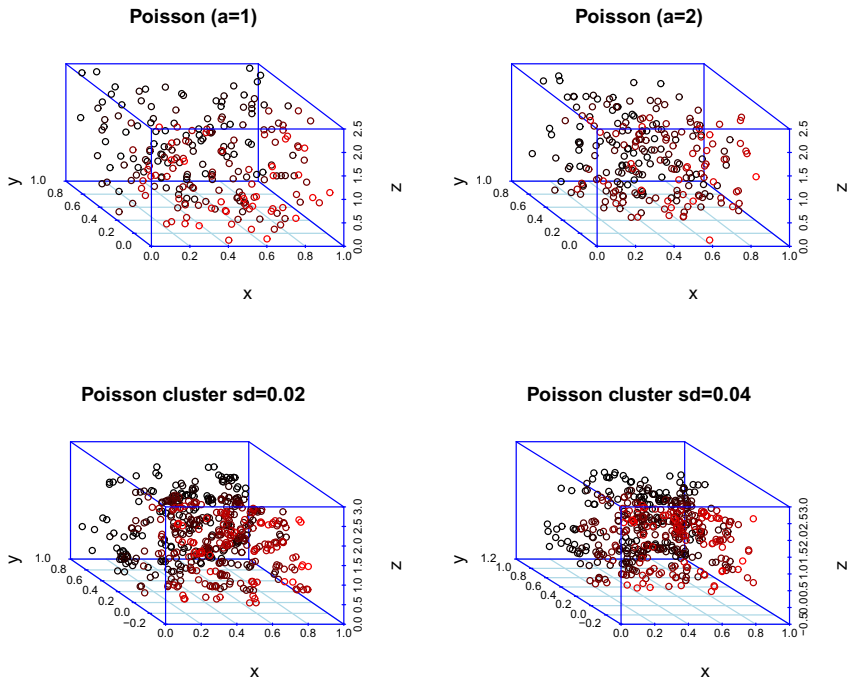

Figure 4.1. Realizations of 3D Poisson processes with uniform ($a = 1$) and non-uniform intensity ($a = 2$) in the upper panel and Poisson cluster processes with standard deviation $\sigma = 0.02$ and 0.04 in the lower panel.

The MISE of $\hat{\lambda}_s(\mathbf{s})$, denoted by $MISE(\hat{\lambda}_s(\mathbf{s}))$, was defined similarly.

To evaluate the integral contained in the formulae of MISE($\hat{\theta}(\mathbf{s})$) and MISE($\hat{\lambda}_s(\mathbf{s})$), we evenly selected $L = 50 \times 50$ points $\{\mathbf{s}_{0i} : i = 1, \cdots, L\}$ in $[0,1]^2$ and computed

$$\widehat{MISE}(\hat{\theta}(\mathbf{s})) = \frac{1}{L}\sum_{i=1}^{L}[\hat{\theta}(\mathbf{s}_{0i}) - \theta(\mathbf{s}_{0i})]^2$$

to estimate the MISE of $\hat{\theta}(\mathbf{s})$. A similar method was used to estimate the MISE of $\hat{\lambda}_s(\mathbf{s})$. In our simulations, we chose $\mathbf{s}_{0i}$ as points on a lattice defined by $\{\mathbf{s} = (s_x, s_y) : s_x = 0.01 + 0.02(j-1), s_y = 0.01 + 0.02(k-1), j, k = 1, \cdots, 50\}$.

We simulated 1000 realizations from the marked Poisson and marked cluster processes for each set of selected a and κ. As long as points and marks were generated, we computed the values of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ using different bandwidth h. For each h, we computed the MISE values of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ and the optimal h was determined by the one with the minimum MISE value. For the selected cases displayed in Table 4.1 and Table 4.2, we found that the optimal h for $\kappa = 1000, 2000, 5000$ and 10000 was almost equal to $0.075, 0.060, 0.058$ and 0.046 respectively. To compare, we also displayed the results from different choices of h in the tables. It showed that the MISE values were significantly enlarged if h moved away from the optimal values. The only exceptional case was the MISE of $\hat{\lambda}_s(\mathbf{s})$ when $a = 1$. The interpretation of this case was easy because the underlying point process $N_{\mathbf{s}}$ was homogeneous when $a = 1$. In this case, $\lambda_s(\mathbf{s})$ was constant in $[0,1]^2 \times [0,T]$, which indicated that the minimum MISE value of $\hat{\lambda}_s(\mathbf{s})$ was attained if h was maximized (i.e. equal to infinite). From the simulations, we found that the optimal h for $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$ might be different. This would make the determination of the bandwidth more complicated in practice. Since estimation of $\hat{\theta}(\mathbf{s})$ is more important than estimation of $\hat{\lambda}_s(\mathbf{s})$, we should focus on the best strategy of the bandwidth selection in $\hat{\theta}(\mathbf{s})$ in applications.

4.6 Case Study

We apply our approach to natural hazard data. Natural hazard data often include the locations and magnitudes of events, which can be considered as points and marks respectively in a marked point process (Peduzzi and Herold, 2004). We focus our approach to applications in forest wildfire and earthquake studies. Specifically, the *Alberta Forest wildfire* data and the *Japan earthquake* data will be considered for illustrative purposes.

4.6.1 Alberta Forest Wildfire Data

The *Alberta Forest wildfire* data consisted of historical forest wildfire occurrences in Alberta, Canada, from 1931 to 2012. The Canadian Alberta Forest Service initiated the modern era of wildfire record keeping in 1931. Over the years, this fire information has been recorded, stored and made available in different formats. Beginning in 1996, paper copies of wildfire historical information were no longer retained. The wildfire historical data were entered at the field level on the Fire Information Resource Evaluation System (FIRES), which are available at *http://www.srd.alberta.ca/Wildfire*.

We collected the historical forest wildfire data from 1996 to 2010 from the website. The dataset contained forest wildfire activities with area burned greater than or equal to 0.01 hectares. There were 1954 fires recorded in 2006 which was the highest annual frequency among 15 years, and 1840 fires in 2010 was the second highest. However, the total burned area in 1998 (7269.67km^2) was the largest among those 1692 fires. Overall, the largest wildfire occurred at 111.77 longitude west and 55.47 latitude north with burned area 2388.67km^2 in 2002.

We treated the spatiotemporal locations of wildfire occurrences as points and area burned as marks in our approach. We assumed that the area burned followed an exponential distribution with parameter $\theta(\mathbf{s})$ described by Model (4.14) with $\underline{m} = 0.01$ (hectares), which only depended on the spatial locations of events. We used Equation (4.15) to compute the KWCLE of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$, where the spherical distance and the

Gaussian kernel were used. We investigated multiple choices of h in the computation of the KWCLE. We found that $h = 100$km was almost the optimal in our method. Therefore, we chose $h = 100$km in the computation of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$.

Since the study region was irregular and not convex, we should take the edge effect into account (Berman and Diggle, 1989). The edge effect was interpreted by the integral term on denominator of $\hat{\lambda}_s(\mathbf{s})$ in Equation (4.15), which must be estimated. To compute the integral, we randomly sampled 10^6 points inside and outside of the study area as base points. For each base point $\mathbf{s}'$, we computed the values of $K_h(\mathbf{s}'-\mathbf{s})$ for every $\mathbf{s}' \in \mathcal{S}$ or $\mathbf{s}' \notin \mathcal{S}$, respectively. Then, we had

$$\int_{\mathcal{S}} K_h(\mathbf{s}' - \mathbf{s}) d\mathbf{s}' \approx \frac{\sum_{\mathbf{s}' \in \mathcal{S}} K_h(\mathbf{s}' - \mathbf{s})}{\sum_{\mathbf{s}'} K_h(\mathbf{s}' - \mathbf{s})}.$$

This equation was used at every $\mathbf{s} \in \mathcal{S}$ for the computation of $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}_s(\mathbf{s})$.

The estimated surfaces of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ from the Gaussian kernel are displayed in the upper panel of Figure 4.2. We found the estimated values of $\theta(\mathbf{s})$ were smaller in the North, indicating that large fires were more likely to occur in the North than in the South. The intensity of fires was higher in the South because of the high frequencies of small fires there. This was generally consistent with earlier findings that the forests in the North were with higher risks than in the South using the power-law (Jiang *et. al*, 2009) and local odds ratio (Zhang and Zhuang, 2014) analyses for fire frequency and area burned of Canadian forest ecosystems.

4.6.2 Japan Earthquake Data

We can find many sources of earthquake data available online. For instance, the website of the North California Earthquake Data Center (NCEDC) at *http://www.ncedc.org/anss/catalog-search.html*, or the United States National Geophysical (USGS) data center at *http://earthquake.usgs.gov/research/data/*, and there are still many other websites with information about earthquake activities. They recorded the date, time, longitude, latitude, and magnitude of earthquakes starting from thousands of years ago until now.

We collected the historical earthquake data from the NCEDC website. We focused on earthquakes in Japan and the ocean area nearby because Japan was one of the countries in the world with highest risk of earthquakes. Data of earthquakes of magnitude 4.0 and larger were chosen from the region between latitude 30 to latitude 45 north and longitude 130 to 150 east for the time span 1961 to 2013. This study region was the area that most earthquakes occurred.

Total there were $29,032$ earthquakes in our dataset. Among those, 405 of them had magnitude between six and seven, 29 of them had magnitude between seven and eight, and 4 of them had magnitude higher than eight. There are a few large earthquakes which resulted in severe damage. The 1993 Southwest-off Hokkaido earthquake occurred on July 12, 1993 had a magnitude of 7.6 and caused a total fo 230 fatalities. The Great Hanshin earthquake (or Kobe earthquake) occurred on January 16, 1995 had a magnitude of 6.8 which made approximately 6437 people lose their lives. The largest earthquake (Tohoku earthquake) took place in the ocean area near the north-eastern Japan on March 11, 2011 where the epicentre was at 38.30 latitude north and 142.37 longitude east with a magnitude of 9.1. Over fifteen thousand deaths as well as much more injuries and missing people were reported by Japanese National Police Agency. In addition, the earthquake triggered powerful tsunami waves and caused a nuclear accident of the Fukushima Daiichi Nuclear Power Plant which affected hundreds of thousands of residents. The damage caused by the earthquake and tsunami were enormous.

We treated the spatiotemporal locations of earthquakes as points and magnitude as marks. As we did for the *Alberta Forest wildfire* data, we used the spherical distance and Gaussian kernel to estimate $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$. We used the similar method to determine the bandwidth h and found that $h = 100$km was almost the optimal one. The estimated surfaces of $\theta(\mathbf{s})$ and $\lambda_s(\mathbf{s})$ are displayed in the lower panel of Figure 4.2. We identified some areas that were more dangerous than the rest. The region between latitude 42 to 45 north and longitude 145 to 150 east had small $\hat{\theta}(\mathbf{s})$ and large $\lambda_s(\mathbf{s})$ indicating both the proportion of large earthquakes and the occurrence rates

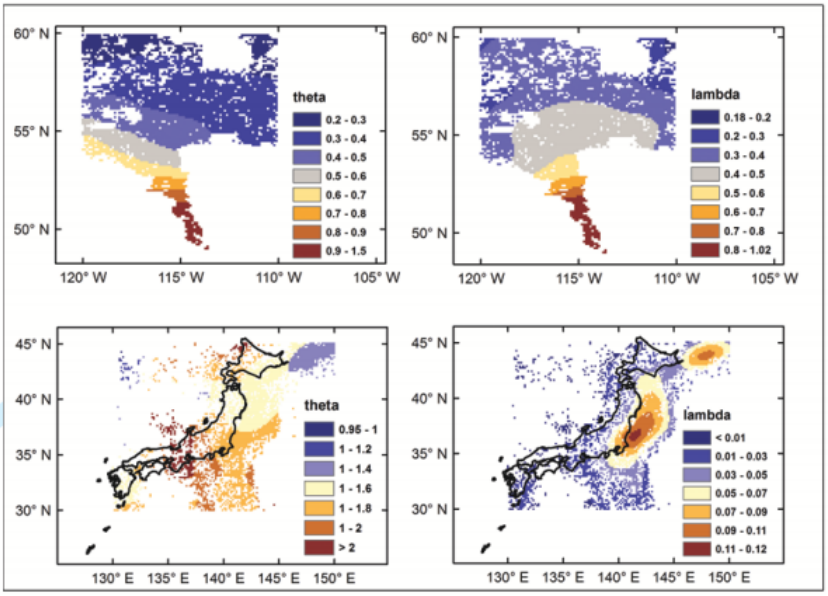

Figure 4.2. Contour plots of $\hat{\theta}(\mathbf{s})$ (left) and $\hat{\lambda}(\mathbf{s})$ (right) in the Alberta wildfire (the upper panel) and Japan earthquake data (the lower panel).

were high. Thus, it was considered as a risky area. The Kuril Islands earthquake occurred in 1994 with magnitude 8.1 was covered in this region. Another risky area was between latitude 35 to 39 north and longitude 140 to 144 east with moderate $\hat{\theta}(\mathbf{s})$ but almost the largest $\lambda_s(\mathbf{s})$. The largest earthquake in Japan occurred in 2011 with magnitude 9.1 and the Iwate-Miyagi Nairiku earthquake in 2008 with magnitude 6.9 were both covered in this region. The third area we noticed was between latitude 37 to 40 north and longitude 137 to 140 east with a few small $\hat{\theta}(\mathbf{s})$. Three earthquakes in 1983, 2004 and 2007 with magnitudes 7.7, 6.6, and 6.6 were covered in this region.

In summary, there is a notable interest in proper analysis of historical wildfire and earthquake data. It is important to study their patterns and pay more attention to the area where serious natural hazards are more likely to occur, and try to lower the degree and extent of damage caused by these natural hazards beforehand.

4.7 Discussion

In this chapter, we have proposed a partial stationarity approach to analyze SMPP data. The approach is developed based on the concept of partially stationary models for SMPP data. Specifically, if an SMPP is partially stationary, then its intensity functions of the SMPP does not vary in time, which indicates that the joint distribution of marks and points can only depend on the spatial locations. The model we consider in this chapter is a local parametric model, which is composed of a local parametric component for marks and a nonparametric component for points. To make conditions of partial stationarity hold, we require that the local parametric and the nonparametric components depend on the space but not time. The KWCL approach, which can be understood as an extension of the local likelihood approach, is proposed to estimate these two components. The KWCL approach may reduce to the KWL approach if the SMPP is Poisson.

According to the nature of SMPPs, the intensity function of points reflects the frequency and the distribution of marks reflects the magnitude of occurrences. Therefore, both are important in applications. For instance, if the intensity function in a particular region is high, then it is expected to see more event occurrences in the region. However, the region may not be risky if magnitudes of marks are low. This indicates that the analysis of the connection between points and marks are more important than the separate analysis for points or for marks only. Therefore, it is recommended that one should jointly look at the performance of the estimated surfaces for marks (i.e., $\hat{\theta}(\mathbf{s})$) and points (i.e., $\hat{\lambda}(\mathbf{s})$) in the study region.

We expect that the proposed approach will find wide applications in natural hazard studies as many of natural events satisfy the partial stationarity conditions proposed in this article. A quick answer can be drawn based on the two examples that we have discussed in Section 4.6. In a usual forest wildfire study, the main interest is its impact on global climate change and carbon cycle in the atmosphere (Girardin, 2007). Therefore, it is more important to study the overall effect of forest wildfires

in a study region. However, the main issue in the earthquake study is the impact of extremely large earthquakes occurred in a study region. The study of extreme large marks are more important than the study of frequencies of marks. If Model 4.14 is used, our approach suggests that risk studies should pay more attention to the area that has small values of $\theta(\mathbf{s})$.

There are a few possible extensions to our approach. First, the partially stationary model proposed in this article mostly focuses on the first-order intensity function and only uses simple properties of the second-order intensity function. Therefore, it is hard to interpret higher-order (i.e. three-order or higher) dependence of the event occurrences in an SMPP. Second, although we have provided a method to estimate the distribution of marks and the intensity function of points, we have not considered any explanatory variables (e.g., in forest wildfire studies, the effects of drought condition and fuel loadings). It will be important to involve explanatory variables in the analysis. Third, we have only considered the kernel weighted likelihood approach in estimation of our model. It is possible to have other nonparametric methods, such as the smoothing spline methods (Gu, 2012), in estimation of our model. To extend our proposed approach by considering these issues will be important to wide applications and deserves further investigations.

Table 4.1

Simulated values of root $\widehat{\text{MISE}}$ for $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}(\mathbf{s})$ using Gaussian kernel for $a = 1$ and $\kappa = 1000, 2000, 5000, 10000$ in the marked Poisson process and marked Poisson cluster processes ($\sigma = 0.02, 0.04$), where the bandwidth was chosen as $h = w \times \kappa^{-1/3}$ with $\omega = 0.5, 0.75, 1.0, 1.25, 1.5$.

			Marked PP		Marked Poisson Cluster Process			
					$\sigma = 0.02$		$\sigma = 0.04$	
a	$E(N)$	w	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\lambda}$
1	10000	1.50	0.1136	0.0439	0.1128	0.0968	0.1087	0.0922
		1.25	0.0980	0.0525	0.0975	0.1121	0.0935	0.1079
		1.00	0.0912	0.0649	0.0913	0.1354	0.0888	0.1237
		0.75	0.1010	0.0849	0.1022	0.1701	0.1026	0.1490
		0.50	0.1428	0.1256	0.1472	0.2285	0.1508	0.1905
	5000	1.50	0.1728	0.0730	0.1462	0.1116	0.1427	0.1072
		1.25	0.1254	0.0602	0.1252	0.1303	0.1215	0.1231
		1.00	0.1123	0.0737	0.1130	0.1579	0.1108	0.1444
		0.75	0.1182	0.0966	0.1202	0.2006	0.1191	0.1747
		0.50	0.1642	0.1423	0.1700	0.2735	0.1731	0.2244
	2000	1.50	0.2018	0.0624	0.2037	0.1372	0.1993	0.1327
		1.25	0.1735	0.0727	0.1738	0.1588	0.1709	0.1522
		1.00	0.1496	0.0884	0.1518	0.1919	0.1489	0.1790
		0.75	0.1469	0.1153	0.1521	0.2467	0.1498	0.2216
		0.50	0.1968	0.1683	0.2080	0.3433	0.2099	0.2902
	1000	1.50	0.2498	0.0712	0.2516	0.1563	0.2494	0.1545
		1.25	0.2150	0.0839	0.2190	0.1832	0.2153	0.1784
		1.00	0.1833	0.1017	0.1897	0.2230	0.1876	0.2134
		0.75	0.1761	0.1311	0.1834	0.2870	0.1795	0.2627
		0.50	0.2278	0.1899	0.2440	0.3998	0.2445	0.3519

Table 4.2

Simulated values of root $\widehat{\text{MISE}}$ for $\hat{\theta}(\mathbf{s})$ and $\hat{\lambda}(\mathbf{s})$ using Gaussian kernel for $a = 2$ and $\kappa = 1000, 2000, 5000, 10000$ in the marked Poisson process and marked Poisson cluster processes ($\sigma = 0.02, 0.04$), where the bandwidth was chosen as $h = w \times \kappa^{-1/3}$ with $\omega = 0.5, 0.75, 1.0, 1.25, 1.5$.

			Marked PP		Marked Poisson Cluster Process			
					$\sigma = 0.02$		$\sigma = 0.04$	
a	$E(N)$	w	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\theta}$	$\hat{\lambda}$
2	10000	1.50	0.1964	0.1270	0.1949	0.1544	0.1846	0.1625
		1.25	0.1689	0.1059	0.1682	0.1449	0.1573	0.1494
		1.00	0.1569	0.0915	0.1577	0.1493	0.1469	0.1442
		0.75	0.1989	0.0932	0.1928	0.1711	0.1884	0.1516
		0.50	0.4348	0.1244	0.4156	0.2237	0.4276	0.1803
	5000	1.50	0.2422	0.1701	0.2429	0.1974	0.2317	0.2059
		1.25	0.2088	0.1396	0.2077	0.1808	0.1976	0.1866
		1.00	0.1823	0.1172	0.1852	0.1798	0.1739	0.1767
		0.75	0.1995	0.1117	0.2045	0.2036	0.2016	0.1856
		0.50	0.4199	0.1416	0.4126	0.2677	0.4180	0.2199
	2000	1.50	0.3089	0.2458	0.3098	0.2746	0.3035	0.2803
		1.25	0.2706	0.2030	0.2733	0.2442	0.2634	0.2512
		1.00	0.2299	0.1646	0.2346	0.2315	0.2254	0.2319
		0.75	0.2202	0.1453	0.2296	0.2535	0.2220	0.2407
		0.50	0.4149	0.1694	0.3701	0.3353	0.4107	0.2892
	1000	1.50	0.3601	0.3193	0.3612	0.3499	0.3571	0.3526
		1.25	0.3217	0.2652	0.3245	0.3107	0.3164	0.3128
		1.00	0.2771	0.2141	0.2832	0.2882	0.2740	0.2902
		0.75	0.2472	0.1819	0.2607	0.3008	0.2532	0.2897
		0.50	0.3862	0.1972	0.3720	0.3916	0.4048	0.3487

5. SUMMARY

In the previous chapters, we have introduced different approaches to model dependence between points and marks. In this chapter we give a discussion of current results and possible future work.

5.1 Discussion

Our first approach is based on a newly derived covariance function of additive model for spatial mark point processes. When the points and marks are dependent, we expect the newly derived covariance function will help us to obtain better prediction results when kriging prediction is chosen instead of point process method. Exponential covariance for $C_Z(h)$ is used in this approach. It is the first component of the newly derived covariance. We can also consider other commonly used covariance functions like Matérn covariance. We use weighted least squares estimator to estimate the parameters in the covariance function, but other more sophisticated estimation methods as mentioned in Section 2.2 can also be considered.

In the second approach we consider modeling of univariate and bivariate intensity-dependent marked log Gaussian Cox processes. In this dissertation we only use normal for the mark distribution, but other choices like exponential or gamma distribution are also possible. Due to the reason that kernel methods are sensitive to the choice of bandwidth, we update both the intensity and other mark-related parameters in our MCMC algorithm to avoid the kernel estimates of intensity function. Nevertheless, the estimates of intensity still affects parameter b and makes the convergence slower.

The computation time may be reduced if we use common LGCP modeling methodology which considers marks as covariates to estimate intensity $\Lambda(x)$. However, for forestry data it makes more sense to let the variables like height and diameter to

depend on the intensity instead of the other way around, since the tree growth characteristics should be affected by how crowed the neighbor trees are as they compete for nutrients and light.

For both univariate and bivariate marked point processes, it is more flexible to add the dependence structure between marks than the conditional independence assumption. Especially when we deal with highly correlated bivariate marks, better prediction results are obtained if our model consist of dependence modeling between them. But this way we need to estimate more parameters from bivariate cross-covariance function than the univariate case, and have to take care of the constraints between parameters to make sure the covariance matrix is positive definite.

In the third approach, we consider a partial stationarity approach to model spatiotemporal marked point processes (SMPPs). If an SMPP is partially stationary, then its intensity functions are constant from time to time, indicating that the joint distribution of points and marks only depend on the spatial locations. The model we consider here is a local parametric model, which consists of a nonparametric component for points and a local parametric component for marks. The kernel weighted composite likelihood (KWCL) approach is proposed to estimate these two components. It can be thought of as an extension of the local likelihood approach. If the SMPP follows Poisson distribution, then the KWCL approach is reduced to the kernel weighted likelihood approach.

The intensity function of points reflects the frequency and the distribution of marks reflects the magnitude of events in an SMPP. These are both important in applications. High intensity function in a particular region indicates that it is likely to observe more events in the region, though we may not need to worry about it if magnitudes of marks are low. Areas with high intensity as well as high magnitudes of marks are the risky regions we need to keep an eye on. If Model 4.14 is used, our approach suggests that the area with small values of $\theta(\mathbf{s})$ are more likely for higher magnitudes of marks to occur. We should jointly look at the estimated surfaces for marks (i.e., $\hat{\theta}(\mathbf{s})$) and points (i.e., $\hat{\lambda}(\mathbf{s})$) in the study region to identify the risky

places with both high intensity and high magnitudes of marks. The analysis of the connection between points and marks are more important than the analysis for points or marks separately.

We expect the proposed approach to be useful in the study of natural hazards. Severe wildfires and earthquakes result in an increase in the cost of wildfire and earthquake protection. Usually the impact of extremely large marks are essential in the study of natural hazards. It is important to model the patterns of large marks and attempt to reduce the destruction caused by these natural hazards.

5.2 Future Work

We derive a covariance function for univariate additive marked point process, it is possible to consider the bivariate case of additive model. Furthermore, we can use different characteristics related to dependence between points and marks in the additive model. In this dissertation we focus on the number of points within a certain distance in the additive model. Nearest neighbor distance may be considered in the additive model as well.

One potential application of the mark-point dependent models is in the study of extreme events, where the points represent the locations of times of the occurrence of extreme events and the mark represents the magnitude. We also want to know about the difference between using point process models and geostatistical methodology. From the nature of point process data, we should not use geostatistical methodology since the sampling locations are random. We will not be able to describe the dependence between points and marks or between intensity and marks from geostatistical analysis. However, we can still apply geostatistical methods to model and predict the marks without considering the point process structure. It is still unclear how much difference the two approaches will generate, and this will be interesting to study both theoretically and in applications.

For the partially stationary model we propose for spatiotemporal marked point processes, we mainly focus on first-order intensity for now. It will be interesting to study and interpret second-order or higher order dependence of the events in an SMPP. Another possible extension is to include explanatory variable in our model. For instance, the drought condition in a forest should be important to consider in wildfire studies. Other than the kernel weighted likelihood approach in our study, different nonparametric methods such as smoothing spline methods may be applied.

Currently we deal with quantitative marks in spatial or spatiotemporal marked point processes, it is also of interest to take account of categorical marks like the species of trees for forestry data, or types of disease in epidemiological data. It is worth investigating how the different marks can be dependent on each others.

LIST OF REFERENCES

LIST OF REFERENCES

D.K. Agarwal and A.E. Gelfand. Slice sampling for simulation based fitting of spatial data models. *Statistics and Computing*, 15:61-69, 2005.

J. Aitchison and J.A.C. Brown. *The Lognormal Distribution*. Cambridge University Press, UK, 1957.

A. Baddeley and R. Turner. `spatstat`: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1-42, 2005.

M. Berman and P.J. Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society Series B*, 51(1):81-92, 1989.

J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24(3):179-195, 1975.

P. Billingsley. *Probability and Measure*. Wiley, New York, 1995.

A. Brix. Generalized gamma measures and shot-noise Cox processes, *Advances in Applied Probability*, 31:929-953, 1999.

A. Brix and P.J. Diggle. Spatio-temporal prediction for log-Gaussian Cox processes, *Journal of the Royal Statistical Society Series B*, 63:823-841, 2001.

A. Brix and J. Møller. Space-time multitype log Gaussian Cox processes with a view to modelling weed data, *Scandinavian Journal of Statistics*, 28:471-488, 2001.

P. Coles and B. Jones. A lognormal model for the cosmological mass distribution, *Monthly Notices of the Royal Astronomical Society*, 248:1-13, 1991.

D.R. Cox. Some statistical models related with series of events, *Journal of the Royal Statistical Society Seris B*, 17:129-164, 1955.

D.R. Cox. Partial likelihood. *Biometrika*, 62:269-276, 1975.

D.R. Cox and V. Isham. *Point Processes*, Chapman & Hall, London, 1980.

N. Cressie. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, 17:563-586, 1985.

N. Cressie. *Statistics for Spatial Data*, 2nd Edition. Wiley, New York, 1993.

F.C. Curriero and S. Lele. A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological and Environmental Statistics*, 4(1):9-28, 1999.

D.J. Daley and D. Vere-Jones. *An Introduction to Theory of Point Processes: Volume I: Elementary Theory and Methods*, 2nd Edition. Springer, New York, 2003.

P.J. Diggle. A kernel method for smoothing point process data. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 34(2):138-147, 1985.

P.J. Diggle. *Statistical Analysis of Spatial Point Patterns*, 2nd Edition. Arnold, London, 2003.

P.J. Diggle, R. Menezes, and T.-L. Su. Geostatistical analysis under preferential sampling. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 59(2):191-232, 2010.

P. Elliott, J. Wakefield, N. Best, and D. Briggs (eds). *Spatial Epidemiology: Methods and Applications*. Oxford University Press, Oxford, UK, 2001.

T.S. Ferguson. *A course in large sample theory*, Chapman & Hall/CRC, New York, 1996.

P. Fiorucci, F. Gaetani, and R. Minciardi. Regional partitioning for wildfire regime characterization. *Journal of Geophysical Research*, 113, F02013, 2008.

J.M. Flegal and G.L. Jones. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2):1034-1070, 2010.

M.P. Girardin. Interannual to decadal changes in area burned in Canada from 1781 to 1982 and the relationship to North Hemisphere land temperatures. *Global Ecology and biogeography*, 16(5):557-566, 2007.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105:1167-1177, 2010.

C. Gu. *Smoothing Spline ANOVA Models*, 2nd Edition. Springer, New York, 2012.

Y. Guan. Tests for independence between marks and points of a marked Point Process. *Biometrics*, 62:126-134, 2006.

Y. Guan and D.R. Afshartous. Tests for independence between marks and points of marked Point Processes: a subsampling approach. *Environmental and Ecological Statistics*, 14(2):101-111, 2007.

Y. Guan and Y. Shen. A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, 97(4):867-880, 2010.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, 93(4):989-995, 2006.

B. Gutenberg and C.F. Richter. *Seismicity of the Earth and Associated Phenomena*, 2nd Edition. Princeton University Press, Princeton, New Jersey, 1954.

L.P. Ho and D. Stoyan. Modeling marked point patterns by intensity-marked Cox processes. *Statistics and Probability Letters*, 78(10):1194-1199, 2008.

L. Holden, S. Sannan, and H. Bungum. A stochastic marked point process model for earthquakes. *Natural Hazards and Earth System Sciences*, 3:95-101, 2003.

I.A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability and its applications*, 7(4):349-382, 1962.

V. Isham. Marked point processes and their correlations. *Spatial Processes and Spatial Time Series Analysis* (ed. E Droesbeke). Publications des Facultés universitaires, Saint-Louis, Brussels, 1985.

G. Ivanoff. Central limit theorems for point processes. *Stochastic Processes and Their Applications*, 12:171-186, 1982.

H. Jeffreys. *Theory of probability*, Oxford University Press, Oxford, UK, 1961.

Y. Jiang, Q. Zhuang, M.D. Flannigan, and J.M. Little. Characterization of wildfire regimes in Canadian boreal terrestrial ecosystems. *International Journal of Wildland Fire*, 18(8):992-1002, 2009.

A.F. Karr. *Point Processes and Their Statistical Inference*, 2nd Edition. Marcel Dekker, Inc. New York, 1991.

R.E. Kass, and A.E. Raftery. Bayes Factor. *Journal of the American Statistical Association*, 90:773-795, 1995.

S. Lele and M. Taper. A composite likelihood approach to (co)variance components estimation. *Journal of Statistical Planning and Inference*, 103:117-135, 2002.

B. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:220-239, 1988.

D. Liu, M. Kelly, P. Gong, and Q. Guo. Characterizing spatialtemporal tree mortality patterns associated with a new forest disease. *Forest Ecology and Management*, 253:220-231, 2007.

C.R. Loader. Local likelihood density estimation. *Annals of Statistics*, 24(4):1602-1618, 1996.

B.D. Malamud, J.D.A. Millington, and G.L.W. Perry. Characterizing wildfire regimes in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13):4694-4699, 2005.

B.D. Malamud and D.L. Turcotte. An inverse cascade explanation for the power-law frequency-area statistics of earthquakes, landslides and wildfires. *Geological Society*, 261:1-9, 2006.

A. Malinowski, M. Schlather, and Z. Zhang. Intrinsically weighted means of marked point processes. Arxiv 1210.1335.

K.V. Mardia and R.J. Marshall. Maximum likelihood estimation of models for residual covariance in spatial statistics. *Biometrika*, 71(1):135-146, 1984.

S. Mase. The threshold method for estimating total rainfall. *Annals of the Institute of Statistical Mathematics*, 48:201-213, 1996.

B. Matérn. *Spatial Variation*, 2nd Edition (Lecture Notes in Statistics 36). Springer, Berlin, 1986.

M.A. McCarthy and A.M. Gill. Fire modeling and biodiversity. In 'Frontiers in Ecology: Building the Links 'Proceedings of the 1997 Conference of the Ecological Society of Australia', 1-3 October 1997, Albury, Australia. (Eds N Klomp, I Lunt) 79-88, 1997. (Elsevier: Oxford, UK).

G. Molenberghs and G. Verbeke. *Models for Discrete Longitudinal Data*. Springer, New York, 2005.

J. Møller, A.R. Syversveen, and R.P. Waagepetersen. Log Gaussian Cox processes, *Scandinavian Journal of Statistics*, 25:451-482, 1998.

J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall/CRC, Boca Raton, 2004.

M. Myllymäki and A. Penttinen. Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. *Statistica Neerlandica*, 63(4):450-473, 2009.

M. Niklasson and A. Granstrom. Numbers and sizes of fires: long-term spatially explicit fire history in a Swedish boreal landscape. *Ecology*, 81(6):1484-1499, 2000.

Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9-27, 1988.

Y. Ogata and K. Katsura. Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogs. *Geophysical Journal International*, 113:727-738, 1993.

Y. Ogata. Space-time point process models for earthquake occurrences. *Annals of the Institution Statistical Mathematics*, 50:379-402, 1998.

J. Ohser and D. Stoyan. On the second-order and orientation analysis of planar stationary point processes. *Biometrical Journal*, 23(6):523-533, 1981.

M. Parzen, S. Lipsitz, G. Fitzmaurice, J. Ibrahim, A. Troxel and G. Molenberghs. Pseudo-likelihood methods for the analysis of longitudinal binary data subject to nonignorable non-monotone missingness. *Journal of Data Science*, 5:1-21, 2007.

P. Peduzzi and H.D. Herold. Mapping disastrous natural hazards using global datasets. *Natural Hazards*, 35:265-289, 2004.

A. Penttinen, D. Stoyan, and H.M. Henttonen. Marked point processes in forest statistics. *Forest Science*, 38(1):806-824, 1992.

C. Ricotta, G. Avena, and M. Marchetti, M. The flaming sand pile: self organized criticality and wildfires. *Ecological Modelling*, 119:73-77, 1999.

C. Ricotta, M. Arianoutsou, R. Diaz-Delgado, B. Duguy, F. Lloret, E. Maroudi, S. Mazzoleni, J.M. Moreno, S. Rambal, R. Vallejo, A. Vázquez. Self-organized criticality of wildfires ecologically revisited. *Ecological Modelling*, 141:307-311, 2001.

B.D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2):255-266, 1976.

M. Rosenblatt. A central llimt theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42:43-47, 1956.

M. Schlather, P.J. Ribeiro, and P.J. Diggle. Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society B*, 99:250-261, 2004.

F.P. Schoenberg. Testing separability in spatial-temporal marked point processes. *Biometrics*, 60:471-481, 2004

F.P. Schoenberg. Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference*, 128:79-93, 2005.

W.G. Song, W.C. Fan, B.H. Wang, and J.J. Zhou. Self-organized criticality of forest fire in China. *Ecological Modelling*, 145:61-68, 2001.

M. Stein, Z. Chi and L. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society B*, 66:275-296, 2004.

D. Stoyan, W. S. Kendall, J. Mecke, and L. Ruschendorf. *Stochastic geometry and its applications*, Volume 2. Wiley Chichester, 1995.

D. Stoyan and A. Penttinen. Recent Applications of Point Process Methods in Forestry Statistics. *Statistical Science*, 15(1): 61-78, 2000.

H. Takahata. Nonparametric density estimations for a class of marked point processes. *New York Journal of Mathematics*, 41:127-152, 1994.

R. Tibshirani and T. Hastie. Local likelihood estimation. *Journal of the American Statistical Association*, 82:559-567, 1987.

D.L. Turcotte, B.D. Malamud, F. Guzzetti, and F. Reichenbach. A general landslide distribution applied to a small inventory in Todi, Italy. In: G. Cello, B.D. Malamud (eds). *Fractal Analysis for Natural Hazards.* Geological Society, London, Special Publications, 261:105-111, 2006.

A.W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK, 1998.

C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21:5-42, 2011.

D. Vere-Jones. Statistical methods for the description and display of earthquake catalogs. In *Statistics in Environmental and Earth Sciences*, A.Walden and P. Guttorp (eds), 220-236. Edward Arnold, London, 1995.

R.P. Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63:252-258, 2007.

M. Wang and J.M. Williamson. Generalization of the Mantel-Haenszel estimating function for sparse clustered binary data. *Biometrics*, 61:973-981, 2005.

Y. Yi, L. Zeng and R.J. Cook. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *Canadian Journal of Statistics*, 39(1):34-51, 2011.

H. Zhang. Inconsistent estimation and asymptotically equal interpolations in Model-Based Geostatistics. *Journal of the American Statistical Association*, 99:250-261, 2004.

J.H. Zhang, M.J. Wooster, O. Tutubalina, and G.L.W. Perry. Monthly burned area and forest fire carbon emission estimates for the Russian Federation from SPOT VGT. *Remote Sensing of Environment*, 87:1-15, 2003.

T. Zhang and Q. Zhuang. On the local odds ratio between points and marks in marked point processes. *Spatial Statistics*, 9:20-37, 2014.

VITA

VITA

Yen-Ning Huang was born in Troy, Michigan. She received her B.S. degree from Department of Mathematics at National Tsing Hua University in June 2007. She started her M.S. degree in Statistics at Purdue University in August 2007 and graduated in May 2009. After receiving the degree, she continued her Ph.D. study in Statistics at Purdue University.

During her stay at Purdue University, Yen-Ning worked with Professor Hao Zhang and Professor Tonglin Zhang in the field of spatial statistics focusing on developing flexible models for both spatial and spatiotemporal marked point processes.