NONPARAMETRIC METHODS FOR POINT PROCESSES AND GEOSTATISTICAL DATA

A Dissertation

by

ELIZABETH YOUNG KOLODZIEJ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Statistics

NONPARAMETRIC METHODS FOR POINT PROCESSES AND GEOSTATISTICAL DATA

A Dissertation

by

ELIZABETH YOUNG KOLODZIEJ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Michael Sherman
Committee Members,	Jeffrey Hart
	Mikyoung Jun
	Carl Laird
Head of Department,	Simon Sheather

August 2010

Major Subject: Statistics

ABSTRACT

Nonparametric Methods for Point Processes and Geostatistical Data. (August 2010)

Elizabeth Young Kolodziej, B.S., The University of Georgia;

M.S., The University of Georgia

Chair of Advisory Committee: Dr. Michael Sherman

In this dissertation, we explore the properties of correlation structure for spatio-temporal point processes and a quantitative spatial process. Spatio-temporal point processes are often assumed to be separable; we propose a formal approach for testing whether a particular data set is indeed separable. Because of the resampling methodology, the approach requires minimal conditions on the underlying spatio-temporal process to perform the hypothesis test, and thus is appropriate for a wide class of models.

Africanized Honey Bees (AHBs, *Apis mellifera scutellata*) abscond more frequently and defend more quickly than colonies of European origin. That they also utilize smaller cavities for building colonies expands their range of suitable hive locations to common objects in urban environments. The aim of the AHB study is to create a model of this quantitative spatial process to predict where AHBs were more likely to build a colony, and to explore what variables might be related to the occurrences of colonies. We constructed two generalized linear models to predict the habitation of water meter boxes, based on surrounding landscape classifications, whether there were colonies in surrounding areas, and other variables. The presence of colonies in the area was a strong predictor of whether AHBs occupied a water meter box, suggesting that AHBs tend to form aggregations, and that the removal of a colony from a water meter box may make other nearby boxes less attractive to the bees.

To Scott

ACKNOWLEDGMENTS

I would like to express my deepest thanks to my advisor, Dr. Michael Sherman, for his continuing guidance and support through my studies. I am so fortunate to have experienced graduate school with Dr. Sherman as my chair; he has made the journey an adventure. Without his guidance and persistent help, this dissertation would have been impossible.

Many thanks to my committee members, Dr. Jeffrey Hart, Dr. Mikyoung Jun, and Dr. Carl Laird, for their willingness to serve on my committee and for their helpful corrections and guidance.

I would like to thank the Knowledge and Engineering Lab at Texas A&M University for their permission to use their data set. They have also provided valuable assistance and guidance with respect to the ecological and entomological applications of the data.

My mother has been my guide and support, my counselor and source of encouragement throughout the graduate school experience. I cannot thank her enough for this. I also thank my father for his love and support through the years, believing in me through thick and thin.

My husband Scott has been my source of joy these happy months together. He makes me smile even during the darkest times. I am so lucky to have him as my soul mate through this journey of graduate school and beyond.

To God be the glory!

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

Spatial statistics concerns data collected on a two-dimensional plane or on a higher dimensional space. While some models mathematically make use of multiple dimensions, spatial statistics is a set of methodologies designed specifically for answering questions about where the data is located, and how much of the data is in what locations. Spatial statistics is also concerned with measuring particular attributes, estimating parameters summarizing those attributes spatially, and interpolating values for them. We see, then, that there are two main branches of spatial statistics: point processes, which focus on the locations of the data, and quantitative spatial processes, which measure attributes variables in varying locations.

One major difference between spatially generated data and some other types of data is the often present correlation structure. Many models are built around the assumption that data has been sampled randomly from a population, making the observations independent of one another. Time series data, on the other hand, is correlated, but time has a natural ordering to it, making the orderless quality of spatial correlation unique. Spatial data also is more often unequally spaced than time series data, prompting the common assumption of some continuous function for correlation at distance d. These unique challenges prompt the necessity of using spatial statistics. In this dissertation, we explore the properties of correlation structure for point processes and a quantitative spatial process.

The journal model is *IEEE Transactions on Automatic Control.*

A. Spatio-temporal Processes

Many processes are measured in both space and time. Many times the data is thus analyzed in one of several ways: separate spatial analyses for each time point or a single spatial analysis averaging values over time; separate temporal analyses for each spatial location or a single temporal analysis averaging values over space; or spatio-temporal analysis allowing for both the spatial and temporal structure. The last option is of course preferable, as the first two involve the loss of information, and Chapter II introduces one new method of many being developed [e.g., 1] to make it more feasible.

Spatio-temporal methodologies are distinctly different from, although usually mathematically identical to, methods developed for $k+1$ dimensional space. Time is not modeled as the next spatial dimension because time is ordered and space is not. While it is appropriate to use all surrounding spatial data to interpolate temperatures at a given location, for example, users of the statistical methodologies would not be interested in using tomorrow's temperature to predict today's. Another reason the methods for spatio-temporal data may be different is that units in space and time are not comparable. For example, if one is interested in developing a kernel estimator, as in Chapter II, separate bandwidths are used for space and time.

B. Spatial Point Processes

The origin of spatial point processes is in counting illustrations, modeling the number of events within a length of pipe, region of forest, or interval of time. Spatial point processes primarily measure location, as opposed to quantitative spatial processes. A spatial point process is completely spatially random if the number of events in any subregion is Poisson distributed, thus popularizing the Poisson process [2]. Examples of Poisson processes include locations of lightning strikes, woodpeckers, or pine trees, and point processes have found applications in many fields, including ecology, forestry, and epidemiology.

First- and second-order intensities in space are defined by the following functions. The first-order intensity function denotes the average number of points at a given location, while the second-order intensity function denotes the correlation between the number of points at two locations. Let us consider an infinitesimal region D in $\mathbb{R}^d$, centered at $(\mathbf{s})$: $D = d\mathbf{s}$. Then a process N is first-order stationary if

$$\lambda(\mathbf{s}) = \lim_{|D| \to 0} \frac{E[N(D)]}{|D|} = \nu,$$

where $N(D)$ measures the number of events in the Borel set D, and $|d\mathbf{s}|$ denotes the area of $d\mathbf{s}$. Similarly, the process is second-order stationary if

$$\lambda_2(\mathbf{s}_1, \mathbf{s}_2) = \lim_{|D_1|, |D_2| \to 0} \frac{E[N(D_1)N(D_2)]}{|D_1||D_2|} = \Psi(\mathbf{s}_1 - \mathbf{s}_2),\ \forall\, \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d,$$

for some function $\Psi(\cdot)$, where D_1 and D_2 are infinitesimal regions centered at $(\mathbf{s}_1)$ and $(\mathbf{s}_2)$. In Chapter II, we investigate testing whether a spatio-temporal point process is separable, meaning its second-order intensity can be separated into spatial and temporal components. Because of the resampling methodology, the approach requires minimal conditions on the underlying spatio-temporal process to perform the hypothesis test.

C. Quantitative Spatial Processes

Quantitative spatial processes primarily measure some attribute such as concentration, temperature, or ore reserves. Measurements of a quantitative spatial process may or may not be made in random locations. In this dissertation, we focus on a data set collected by the city of Tucson, Arizona. The city checks water meters monthly for billing purposes. Each time a colony of Africanized honey bees was found in a water meter, the city recorded the address and date of the finding, creating a rich data set to analyze. We did not consider the locations of the colonies to be a point process because the locations of the water meters were not random; rather, they were constrained to the requirement of one per lot. Lot sizes varied widely in size, so apparent clustering of the colonies in water meters could be due to the fact that the water meters themselves were clustered in some areas. Therefore, we treated the data as quantitative, measuring at each location whether or not a colony had been formed during the 12 year study period, and measuring the number of times colonies had been formed. For the former, we used a logistic regression model, and for the latter, a Poisson model.

D. Overview

Chapter II develops a test for spatio-temporal separability for point processes. Chapter III applies two generalized linear models to a set of locations of Africanized honey bees in the Tucson metropolitan area, using a resampling method to account for correlation between observations. Lemmas and proofs of the theorems in Chapter II are described in the first appendix. Tables and figures for the Africanized honey bee application in Chapter III are in the second appendix.

CHAPTER II

SEPARABILITY FOR SPATIO-TEMPORAL POINT PROCESSES

A. Introduction

Spatial point processes began with counting illustrations, modeling the number of events within a length of pipe, region of forest, or interval of time. The student's first exposure to these types of problems almost certainly begins with the Poisson process [2]. Spatial analysis has expanded to include geostatistical data, lattice data, point processes, and spatio-temporal processes, and applications in the spatial field now include weather patterns, disease spread, crime waves, locations of promising oil wells, and mineral concentrations in soil. For more illustrations, see, e.g., Schabenberger and Gotway [3]. To determine whether a spatio-temporal point process is separable, we will use the second-order intensity function. The first-order intensity function gives information about the number of events that occur per region; the second-order intensity function gives information about the probability that there is one observed point in each of two infinitesimal regions. That is, second-order intensity functions measure covariance. If the covariance between space and time is zero, we call it separable, and it is appropriate to model the space and time correlations separately. For example, if we were modeling the locations of bee colonies using a point process, we would want to know whether we should model their spatial behavior differently at distinct time points; if not, it is permissible to create a single spatial model using all of the data, rather than creating separate models at different times.

To test the separability of time and space, we will need to construct a test statistic. We propose a test statistic that has a limiting Gaussian distribution based on a version of the Central Limit Theorem. We begin with some definitions, and then propose a test statistic.

B. Definitions

The first order intensity $\lambda(\mathbf{s},t)$ of a spatial point process gives the average number of events per unit volume at a spatial location $\mathbf{s}$ and time t [3]. Let us consider an infinitesimal cylinder in $\mathbb{R}^{d+1}$, centered at $(\mathbf{s},t)$: $A = d\mathbf{s} \times dt$; we will also write in general $|A| = |D| \times T$, $|D|$ being the size of the spatial domain of interest and T the time domain of interest. Then the first order intensity function is defined as:

$$\lambda(\mathbf{s},t) = \lim_{|A|\to 0} \frac{E[N(A)]}{|A|},$$

where $N(A)$ is measuring the number of events in the Borel set A. Similarly, the second order intensity is defined as:

$$\lambda_2(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) = \lim_{|A_1|,|A_2|\to 0} \frac{E[N(A_1)N(A_2)]}{|A_1||A_2|},$$

where A_1 and A_2 are infinitesimal cylinders centered at $(\mathbf{s}_1,t_1)$ and $(\mathbf{s}_2,t_2)$. We now consider processes that are both second-order stationary and isotropic. A process is said to be *second-order stationary* if $\lambda(\mathbf{s},t)$ is constant for all $(\mathbf{s},t)$, and $\lambda_2(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) \equiv \Psi(\mathbf{s}_1-\mathbf{s}_2,t_1-t_2)$ for some function $\Psi(\cdot)$. Assuming second-order stationarity, only the distance between two points and direction will be important in the second order intensity, so we can define $\mathbf{s} = \mathbf{s}_1 - \mathbf{s}_2$, and $t = t_1 - t_2$. A process is said to be *isotropic* if whenever $||(\mathbf{s},t)|| = ||(\mathbf{s},t)'||$, $\Psi(\mathbf{s},t) = \Psi(\mathbf{s},t)'$. If a process is isotropic, only distance between two points is important. Under second-order stationarity and isotropy, we can write

$$\Psi(\mathbf{s},t) = \lim_{|A_1|,|A_2|\to 0} \frac{E[N(A_1)N(A_2)]}{|A_1||A_2|},$$

where A_1 and A_2 are any two infinitesimal cylinders separated by the vector $(\mathbf{s}, t)$.

Next we define a kernel estimator of the second order intensity function, and then we can create a test statistic based on that estimator. Our kernel estimator is:

$$\hat{\Psi}(\mathbf{s}, t) = \int_{\mathbf{x}_1 \in D} \int_{\mathbf{x}_2 \in D} \int_{t_1 \in T} \int_{t_2 \in T} \frac{K[(\mathbf{s} - \mathbf{x}_1 + \mathbf{x}_2)/h_1, (t - q_1 + q_2)/h_2]}{|A \cap B| \times h_1^2 \times h_2} \, \mathrm{d}N^{(2)},$$

where we define the following: 1.) $B = \{\mathbf{y}, k : \mathbf{y} = \mathbf{z} - \mathbf{x}_1 + \mathbf{x}_2,\ \mathbf{z} \in D,\ k = l - q_1 + q_2,\ l \in T\}$, 2.) $\mathrm{d}N^{(2)} = N^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)) = N(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1) N(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2) I(\mathbf{x}_1 \neq \mathbf{x}_2) I(q_1 \neq q_2)$, where I is the indicator function, 3.) $K(\mathbf{x}, l)$ is a kernel density, and 4.) h_1 and h_2 are spatial and temporal bandwidths, respectively. The asymptotic mean and variance of our kernel estimator are found in Theorem 1, and asymptotic normality is shown in Theorem 2 in the next section.

The covariance of a spatio-temporal point process is said to be *separable* if it can be decomposed into spatial and temporal components. An example of a separable second-order intensity is given below.

$$\Psi(\mathbf{s}, t) = C\, \Psi(\mathbf{s}, 0)\, \Psi(\mathbf{0}, t),$$

where C is some constant. A non-separable covariance function is interpreted in a similar manner to interactions in linear models: we may have different spatial covariances for each time instance and different temporal covariances for each spatial location. Using a separable covariance function simplifies interpretation of the model. A hypothesis test for testing separability is formulated in section E.

C. Asymptotic Features of the Sample Second-order Spatio-temporal Intensity Function

First we discuss the asymptotic bias and covariance of our kernel estimator, and then we show that it is asymptotically normally distributed. To measure covariance, we use cumulants. Cumulants are comparable to a measurement of the spatial and temporal dependence of the process. Define the kth-order cumulant function as

$$\begin{aligned} & C_N^{(k)}(\mathbf{x}_2 - \mathbf{x}_1, \ldots, \mathbf{x}_k - \mathbf{x}_1, q_2 - q_1, \ldots, q_k - q_1) \\ & \equiv \lim_{|d\mathbf{x}_1|, \cdots, |d\mathbf{x}_k|, |dq_1|, \cdots, |dq_k| \to 0} \left[\frac{\mathrm{Cum}[N(d\mathbf{x}_1, dq_1), \ldots, N(d\mathbf{x}_k, dq_k)]}{|d\mathbf{x}_1| \times \cdots \times |d\mathbf{x}_k| \times |dq_1| \times \cdots \times |dq_k|} \right] \end{aligned}$$

where $\mathrm{Cum}(Y_1, \ldots, Y_k)$ is given by the coefficient of $(it_1, \cdots, i^k t_k)$ in the Taylor series expansion of $\log\{E[\exp(i \sum_{j=1}^k Y_j t_j)]\}$ about the origin [see 4]. For example, if N is Poisson, then all $C_N^{(k)}(\mathbf{x}_2 - \mathbf{x}_1, \ldots, \mathbf{x}_k - \mathbf{x}_1, q_2 - q_1, \ldots, q_k - q_1)$ will be equal to zero if any of $\mathbf{x}_j - \mathbf{x}_1 \neq \mathbf{0}$ and $q_j - q_1 \neq 0, j = 2, \ldots, k$. We consisder a sequence of random fields $A_n = D_n \times T_n$, and let $\hat{\Psi}_n(\mathbf{s}, t)$ be the estimator of $\Psi(\mathbf{s}, t)$ over A_n. We investigate the large sample properties of this estimator, beginning with the following theorem.

Theorem 1: Assume that:

1. $C_N^{(2)}(\cdot, \cdot)$ and $C_N^{(3)}(\cdot, \cdot, \cdot, \cdot)$ are bounded and $C_N^{(2)}(\cdot, \cdot)$ is continuous and integrable.
2. $\int_{\mathbb{R}^2} \int_T |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_2, r_1, r_2)| \mathrm{d}\mathbf{u}_1 \, \mathrm{d}r_1 < \infty$ for finite $\mathbf{u}_2, r_2$,
   $\int_{\mathbb{R}^2} \int_T |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_1 + \mathbf{u}_2, r_1, r_1 + r_2)| \mathrm{d}\mathbf{u}_1 \, \mathrm{d}r_1 < \infty$ for finite $\mathbf{u}_2, r_2$, and
   $\int_{\mathbb{R}^2} \int_T |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_3, r_1, r_2, r_2 + r_3)| \mathrm{d}\mathbf{u}_2 \mathrm{d}r_2 < \infty$ for finite $\mathbf{u}_1, \mathbf{u}_3, r_1, r_3$.
3. $|A_n| = O(n^2), |\partial A_n| = O(n)$, where ∂A_n denotes the boundary of A_n, and $|\partial A|$ denotes the length of ∂A. This accounts for the shape of the random field from which we sample data.

4. The bandwidths, h_{1n} and h_{2n}, have the condition that $h_{in} = O(n^{-\beta})$ for some $\beta \in (0, 1)$, for $i = 1, 2$.

5. The kernel function $K(\cdot, \cdot)$ is a bounded, nonnegative, isotropic density function which takes positive values only on a finite support, C.

Let N be a time-space stationary point process observed on domain A_n. Then:

$$E[\hat{\Psi}(\mathbf{s}, t)] = \int_C K(\mathbf{s}, t)\, \Psi(\mathbf{s} - h_{1n}\mathbf{x}, t - h_{2n}q)\, \mathrm{d}q\, \mathrm{d}\mathbf{x} \to \Psi(\mathbf{s}, t)$$

and:

$$\lim_{n\to\infty} |A_n| \times h_{1n}^2 \times h_{2n} \times \mathrm{Cov}\{\hat{\Psi}_n(\mathbf{s}_i, t_i), \hat{\Psi}_n(\mathbf{s}_j, t_j)\} \to \sigma_{ij},$$

where:

$$\sigma_{ij} = \begin{cases} \int K^2(\mathbf{x}, q)\, \mathrm{d}\mathbf{x}\, \mathrm{d}q \times \Psi(\mathbf{s}_i, t_i) & , \quad \mathbf{s}_i = \pm\mathbf{s}_j, t_i = \pm t_j \\ 0 & , \quad o.w. \end{cases}$$

Proof: See Appendix A.

To evaluate separability, we will measure the second-order intensity function at a set of user-chosen lags $\mathbf{\Lambda}$. Define $\mathbf{G}_n \equiv \{\Psi_n(\mathbf{s}, t) : (\mathbf{s}, t) \in \Lambda\}$ to be the vector of second-order intensity functions at lags in Λ. Then $\hat{\mathbf{G}}_n \equiv \{\hat{\Psi}_n(\mathbf{s}, t) : (\mathbf{s}, t) \in \Lambda\}$ are the estimators of $\mathbf{G}$. To show that $\hat{\mathbf{G}}_n$ is asymptotically normal, we must quantify the strength of the dependence between locations on the spatio-temporal field using the following mixing coefficient [5].

$$\alpha_N(p; k; j) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_N(E_1),\ A_2 \in \mathcal{F}_N(E_2),$$
$$E_2 = E_1 + (\mathbf{s}, t),\ |E_1| = |E_2| \le p,\ d(E_1, E_2) \ge k,\ t(E_1, E_2) \ge j\}$$

where the supremum is taken over all compact, convex subsets $E_1 \subset \Re^2$, and over all E_2 such that the distance $d(E_1, E_2) \ge k$ and $t(E_1, E_2) \ge j$. Here we define the following: $\mathcal{F}_N(E)$ denotes the $\sigma-$algebra generated by the events $\{(\mathbf{x}, q) : (\mathbf{x}, q) \in E\}$, $d(\cdot, \cdot)$ denotes the maximal Euclidean distance between disjoint sets of points, and $t(\cdot, \cdot)$ denotes the maximal distance in time between disjoint sets of points. If, for example, N is Poisson, $\alpha_N(p; k; j) = 0$ for all $k > 0$ or $j > 0$. Our mixing condition is thus

$$\sup_p \frac{\alpha_N(p;k;j)}{p} = O(k^{-\epsilon} j^{-\delta}) \,\text{for some}\, \epsilon > 2, \delta > 1 \tag{2.1}$$

The mixing condition says that as disjoint groups of points are separated by larger distances in space or time, dependence decreases at some rate depending on the volume p. That is, we require $\alpha_n(p; k; j)$ to approach 0 for large k or j at some rate, depending on p. Put another way, at a fixed distance in space k, dependence may increase as the volume increases at a rate controlled by p.

In addition to the mixing condition, we also require the following mild moment condition.

$$\sup_n E\left\{\left|\sqrt{|A_n|} \times h_{1n} \times \sqrt{h_{2n}} \times \left[\hat{\Psi}_n(\mathbf{s}, t) - \Psi(\mathbf{s}, t)\right]\right|^{2+\delta}\right\} \leq C_\delta, \tag{2.2}$$

for some $\delta > 0$ and $C_\delta < \infty$.

Theorem 2: In addition to the conditions in Theorem 1, assume that our mixing condition (2.1) holds and our moment condition (2.2) holds. Denote the size of the spatio-temporal random field as $r_n^2 \times n$, so that the size of D_n is r_n, and $A_n = D_n \times T_n$, as before. The temporal domain, then, expands at rate n, while the spatial domain expands at a rate which is some function of n. We assume this function to be a monotone, increasing, unbounded function in n such that $\lim_{n\to\infty} r_n = \infty$. Let N be a stationary spatio-temporal point process observed on domain A_n. Then $\sqrt{|A_n|} \times h_{1n} \times \sqrt{h_{2n}} \times \{\hat{\mathbf{G}}_n - E(\hat{\mathbf{G}}_n)\}$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix Σ, where the elements are given in Theorem 1.

Proof: See Appendix A.

D. Covariance Estimation

To form our test statistic, we need to know the covariance of $\hat{\mathbf{G}}_n$, denoted $\mathbf{\Sigma}$. As $\mathbf{\Sigma}$ is generally unknown, we find an estimator based on the data. While the off-diagonal elements in the covariance matrix are shown to be asymptotically zero in Theorem 1, they may be non negligible in finite samples, so a plug-in method may be overly simplistic. We therefore apply a subsampling technique to estimate covariance.

Resampling methods began with U-statistics [6], the Jackknife [7, 8], and the bootstrap [9]. Since then, resampling methods have been extended to many different types of parameter estimation situations, including that of data correlated in space and time. Carlstein introduced the idea of using subseries, or "windows" to compute asymptotic standard errors [10]. These windows are contiguous regions, sections of time or subshapes of space, that represent smaller portions of the original larger process. Künsch [11] suggested the use of overlapping windows, and Hall and Jing [12] introduced the idea of an overlapping, non-independent window for dependent (temporal or spatial) data. Kaiser *et al.* explored inference on the spatial cumulative distribution function using subsampling [13]. Sherman extended the subseries method to linear models, and in 1996 showed for lattice data, the optimal subshape size is proportional to $n^{1/2}$ [14, 15]. Nordman and Lahiri further explored the issue, showing that the optimal block size depends on the shape of the sampling region and characteristics of the random field [16]. Guan *et al.* considers subsampling to estimate covariance for spatial point processes; we extend the idea here to spatio-temporal point processes [16].

Let $D_{m(r_n)} \times T_{l(n)}$ be a subshape congruent to $D_n \times T_n$ in shape and orientation, rescaled such that $m(r_n) = c_1\, r_n^{\alpha}$ is the length of one side of $|D_n|$, $l(n) = c_2\, n^{\beta} = |T_n|$, $\alpha \in (0,1)$, $\beta \in (0,1)$, and c_1 and c_2 positive constants. We need to allow $D_{m(r_n)} \times T_{l(n)}$ to become large for our asymptotic considerations, but we also need many subshapes. Thus we assume that $c_1 \to 0$, $c_2 \to 0$, $c_1\, r_n^{\alpha} \to \infty$, and $c_2\, n^{\beta} \to \infty$ as $n \to \infty$. For example, we could use $c_1 = r_n^{-2\alpha}$ and $c_2 = n^{-2\beta}$. Then a displaced copy of $D_{m(r_n)} \times T_{l(n)}$ is $D_{m(r_n)} \times T_{l(n)} + (\mathbf{x}, q) \equiv \{(\mathbf{s}, t) + (\mathbf{x}, q) : (\mathbf{s}, t) \in D_{m(r_n)} \times T_{l(n)}\}$, where $(\mathbf{x}, q) \in D_n^{1-c} \times T_n^{1-c}$ and $D_n^{1-c} \times T_n^{1-c} \equiv \{(\mathbf{x}, q) \in D_n \times T_n : D_{m(r_n)} \times T_{l(n)} + (\mathbf{x}, q) \subset D_n \times T_n\}$. Then we define $\hat{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q)$ as the sample second-order intensity function estimated at lags $\mathbf{\Lambda}$ on the displaced subshape, and we define $h_{1,m(r_n)}$ and $h_{2,l(n)}$ as the spatial and temporal bandwidths, respectively, used to obtain $\hat{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q)$. Then our subsampling estimator denoted by $\hat{\mathbf{\Sigma}}_n$ is as follows:

$$\begin{aligned}
\frac{1}{|D_n^{1-c}| \times |T_n^{1-c}|} \quad & \times \int_{T_n^{1-c}} \int_{D_n^{1-c}} |D_{m(r_n)}|\, |T_{l(n)}|\, h_{1,m(r_n)}^2\, h_{2,l(n)} \\
& \times \left(\hat{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q) - \bar{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q) \right) \\
& \times \left(\hat{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q) - \bar{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q) \right)' \, \mathrm{d}\mathbf{x}\, \mathrm{d}q,
\end{aligned}$$

where $\bar{\mathbf{G}}_{m(r_n),l(n)} \equiv 1/\left\{|D_n^{1-c}| \times |T_n^{1-c}|\right\} \int_{T_n^{1-c}} \int_{D_n^{1-c}} \hat{\mathbf{G}}_{m(r_n),l(n)}(\mathbf{x}, q)\, \mathrm{d}\mathbf{x}\, \mathrm{d}q$ In practice, this integral must be approximated by a finite sum. Then we find that every element of the subsampling estimator is L_2 consistent for the appropriate element of $\mathbf{\Sigma}$.

Theorem 3: Assume that the conditions for Theorems 1 and 2 hold, along with condition 2.2 for $\delta > 2$. Then $\hat{\mathbf{\Sigma}}_n$ is an L_2 consistent estimator for $\mathbf{\Sigma}$.

Proof: See Appendix A.

E. Assessment of Separability

Analogous to Li *et al.*, we can create our test statistic [17]. We can write our null hypothesis that the covariance function is separable as

$$H_0 : \mathbf{Af}(\mathbf{G}) = \mathbf{0},$$

where $\mathbf{A}$ is a contrast matrix of row rank q and $\mathbf{f} = (f_1, \ldots, f_r)$ are real-valued functions differentiable at $\mathbf{G}$. Then we choose lags:

$$\Lambda = \{(\mathbf{s}_i, t_j), (\mathbf{s}_i, 0), (\mathbf{0}, t_j)\}, i = 1, \ldots, k, j = 1, \ldots, l.$$

One function, for example, might be:

$$\mathbf{f}(\mathbf{G}) = \left(\frac{\Psi(\mathbf{s}_1, t_1)}{\Psi(\mathbf{s}_1, 0)\Psi(\mathbf{0}, t_1)}, \frac{\Psi(\mathbf{s}_2, t_2)}{\Psi(\mathbf{s}_2, 0)\Psi(\mathbf{0}, t_2)} \right)^T.$$

Then where $A = [1 - 1]$, we can see that $\mathbf{Af}(\mathbf{G}) = \frac{\Psi(\mathbf{s}_1, t_1)}{\Psi(\mathbf{s}_1, 0)\Psi(\mathbf{0}, t_1)} - \frac{\Psi(\mathbf{s}_2, t_2)}{\Psi(\mathbf{s}_2, 0)\Psi(\mathbf{0}, t_2)}$, which is equal to zero under our null hypothesis.

Then by the multivariate delta theorem [18], we can see that:

$$\sqrt{|A_n|} \times h_{1n} \times \sqrt{h_{2n}} \times \left\{ \mathbf{f}(\hat{\mathbf{G}}_n) - \mathbf{f}\left(E\left[\hat{\mathbf{G}}_n \right] \right) \right\} \xrightarrow{d} N_r(\mathbf{0}, \mathbf{B}^T \mathbf{\Sigma} \mathbf{B}),$$

where $\mathbf{B}_{ij} = \partial f_j / \partial \mathbf{G}_i, i = 1, \ldots, m, j = 1, \ldots, r$. Thus because it is L_2-consistent, we simply use our subsampling estimate of Σ, defined in the previous section, to obtain our test statistic:

$$TS = |A_n| \times h_{1n}^2 \times h_{2n} \times \left\{ \mathbf{Af}(\hat{\mathbf{G}}_\mathbf{n}) \right\}^T \left(\mathbf{AB^T \hat{\Sigma}_n BA^T} \right)^{-1} \left\{ \mathbf{Af}(\hat{\mathbf{G}}_\mathbf{n}) \right\}.$$

Finally, we know that $TS \xrightarrow{d} \chi_q^2$ as $n \to \infty$ by the multivariate Slutsky's theorem [19]. An approximate α-level hypothesis test for separability rejects the null hypothesis if the test statistic is greater than the upper α percentage point of a χ^2 distribution with q degrees of freedom.

CHAPTER III

SPATIAL DISTRIBUTION OF AFRICANIZED HONEY BEES IN TUCSON, ARIZONA

A. Introduction

Africanized honey bees (*Apis mellifera scutellata*)have received media attention due to their highly defensive characteristics, including a much shorter reaction time (8-10 s compared to the European 55-62 s), and a larger intensity of nest defense [20]. Therefore the bees' arrival in Tucson, Arizona was cause for concern about residents' safety. The purpose of this study is twofold: First, it investigates explanatory variables such as the influences of the age of buildings, whether a residence with a pool was located within half a mile, acreage, distance from each land class, whether a location was zoned as residential or commercial, and the proportion of the locations within a half-mile radius that had ever been occupied by a colony on the locations of Africanized honey bee colonies. The resulting models can also be used to predict which locations are most likely to be inhabited by an Africanized honey bee colony, thereby giving any employees who check the water meters some warning about the possible presence of a colony.

African honey bees had been known to produce more honey in tropical climates than European honey bees; thus scientists attempted in 1956 to produce a honey bee better suited to the tropical climates by importing African queens and breeding them. Unfortunately, in 1957, 26 colonies established from the imported queens swarmed in Brazil and established feral populations. Beekeepers in southern Brazil also acquired African queens, whose progeny also contributed to the feral population. Queens from beekeepers' managed European colonies have mated with the feral African drones, creating Africanized honey bees, hybrids of the African honey bee and European honey bees like *A. m. ligustica* and *A. m. iberiensis*. The tremendous success of these feral populations has resulted in a rapid spread of the population of these Africanized Honey Bees up through Central America and Mexico to the United States in 1990, and to Arizona in 1993. Hunter *et al.*; Rubink *et al.*; Guzman-Novoa and Page, Jr; and Loper discuss this migration of the Africanized honey bee [21, 22, 23, 24].

Temperature, degree of insolation, humidity, and rainfall are correlated with requests for colony removal [25]. Mistro *et al.* suggest that factors which may influence the bees' absconding include fires; heavy rain; predators; excessive heat, cold, or humidity; scarcity of resources; and human manipulation [26]. Similarly, this study explores whether land class types and other variables are related to colony building in water meter boxes in Tucson.

B. Data Collection

While European honey bees are more selective when choosing nest sites, Africanized honey bees do not need to store as much honey, and so build nests in smaller locations, giving them a wider range of choices in location. Baum *et al.* and Winston discuss that colonies may therefore be found in sewer manholes, flower pots, garbage cans, and water meter boxes [27, 28]. The Tucson Water Department began recording the removal of the Africanized honey bee colonies from water meter boxes in April of 1996 while they were checking the meters monthly for billing purposes, and the most recent record obtained of their removal is from May 2008. The data set Tucson Water Department provided contains the date a colony was found in a water meter and the address of the location. This study is part of ongoing research using the same data set at the Knowledge Engineering Laboratory in the Department of Entomology at

Texas A&M University. This data set is the main data set of interest; other data collected were obtained for the purpose of explaining why the bees were choosing certain water meter boxes over others.

An important part of learning why the bees choose to build colonies where they do is to find out where the bees are not building colonies. To that end, a large set of records was acquired from the Pima County Tax Assessor's Office, denoting the year the building was built, the square footage of the house, whether or not the house had a pool, and a few other details not of interest or too incomplete to be useful to this study. The records were divided into commercial and residential buildings, the source of the corresponding indicator variable. Individual buildings were identified by their parcel number, and so could be matched in a commercial GIS (Geographic Information System) product to the correct locations by using a data set containing the land parceling divisions as polygons. This parcel data came from the Pima County Department of Transportation, Geographic Information Services Division, and contained the acreage of each parcel of land. The water meter data set was also mapped to parcels of land, this time by the address given by the water meter company. Not every parcel of land in Tucson had a matching building in the Tax Assessor data; the main reason for that is that some parcels of land were not yet developed and so had no buildings. If a parcel of land has no building, it also has no water meter, and so those parcels were removed from our data.

It was also of interest to determine whether vegetative land classes influenced the bees' colony home choices. Shaw used aerial photographs to classify areas of Tucson into categories such as residential land, natural land, and watercourses [29]. Parameter estimates must be well-defined on each subspace (a small section of Tucson, as described in Section F), so every variable must be found at more than one level in every subspace. Therefore rather than treating the land class variables as categorical, it was decided to use the distance from each land class as a continuous variable.

C. Data Cleaning

The study was focused on the metropolitan area close to the middle of town in Tucson, rather than using the data from the entirety of Pima County, thus eliminating 581 of the colony occupation records. Data from the water meter boxes, data from the Tax Assessor's Office, the parcel polygons, and the vegetation data were combined using a commercial GIS product. Not every building was in the Tax Assessor data. Among the commercial properties, only those properties that had been improved were posted online. Because over 90% of the buildings in Tucson were in the Tax Assessor data set, any bias incurred due to any pattern of missing data, if it exists, should be minimal. Therefore we assume the data is missing completely at random, an assumption that is at least approximately appropriate.

The data from the water meter boxes presented more of a challenge; about 10% of the addresses did not match the land parcels. Sometimes this was due to a difference in spelling of the address or the addition of an apartment number in one data set that wasn't in the other; those differences were easily remedied. At other times, the parcel data was missing an address. Sometimes parcels in the vicinity had addresses and one could note the pattern of street numbering in order to match the address to the parcel, while at other times Google Maps™ mapping service and Google Earth™ mapping service were employed to find addresses.

R was used to find the proportion of water meters within a half mile radius that had been occupied by a colony at any point during the 12-year study period. Land classes were stored as polygons in Arc GIS® software, and distances from each land class were calculated by finding the Euclidean distance between the centroid of the land parcel on which the water meter was located and the nearest edge of the land class polygon. If the centroid was inside a land class, distance from that class was calculated to be zero.

D. Data Analysis

The goal of this study is to create a model predicting the locations of Africanized honey bees using explanatory variables such as the age of the building corresponding to the water meter, whether a residence with a pool was located within half a mile, acreage, distance from each land class, and the proportion of the locations within a half-mile radius that had ever been occupied by a colony. An important point to be made is that each month, the water meter company not only records the location of the colony infestation; they also remove the colony from the water meter box in order to read the meter. Therefore the model is not predicting the growth of the Africanized honey bee population, but rather it predicts the recovery or rebuilding of the population after the colonies found in the boxes were eliminated. Of the 275,877 locations, only 5,640 had colonies in their water meter boxes during the 12 year study. Thus the prediction for most buildings is that a colony is quite unlikely to be found there; however, it is of interest to learn why some houses are predicted to be more likely to have colonies in the water meter boxes than others.

The usual logistic and Poisson regression models assume independence of the observations. An important aspect of this data set is that the presence of colonies is spatially correlated; therefore methods developed by Heagerty and Lumley are used to subsample from the data in order to obtain variance parameter estimators that do not assume independence [30]. The approach involves estimating functions, and does not require that parameter estimates be computed on each subseries. Therefore if a subspace has no records of colonies, this method does not have the usual problem in logistic regression of leading to estimates of $\pm\infty$.

In order to use the logistic and Poisson models described in section E, certain assumptions must be made. First, it is assumed that the data from the water meters is complete: that is, no individual removed a colony from a water meter box before the water meter company found it when they made their rounds for the month, nor did the person in charge of checking the meter fail to record it. If there is a socioeconomic pattern to the failures to record colony removal, some social bias is incurred. Social bias is a smaller problem for this data set, however, than for others that have been collected from pest removal companies e.g., Baum *et al.* [27]: every building has a water meter box that is checked by the city, while not every family has the financial resources to pay for pest removal. The assumption is that every month there is no record of the bees' presence, they are absent. Second, it is assumed that when records are at most 30 days apart, those are multiple records concerning the same colony that has not yet been removed or that was found on the first date but cleared out on the second. If, however, the records are more than 30 days apart, it is assumed that there were two separate colonies: one swarm made their colony in the water meter box, then they were cleared out, and another colony came in the very next month. Thirty days was chosen in part because the Africanized honey bees require that amount of time to build a new colony [31] and in part because the water meter company comes to check the water meters once a month.

Third, while records of the years the buildings were built have been obtained, records for when the water meters were turned on or off were not. If the water meter company were to turn off a meter for non-payment or a vacation, and then not check the water meter for three months, they would not be checking for bees in that water meter for that time. Similarly, the newer houses are recorded in the data set as not having bees from the year they were built onward. However, it may take some houses in a new neighborhood longer to sell than others, and so there might be a water meter that is again not being checked every month, as the water has not been turned on yet. Because data on the water meter status changes has not been obtained, this model might again incur some patterns of social bias by assuming there were no bees in the water meters during the time the company wasn't checking the meters. As stated earlier, any social bias incurred is smaller than for data collected from pest removal companies. Finally, it is assumed that there is a linear relationship between the explanatory variables and the log odds of a colony occupying a water meter for the logistic regression model, and that there is a linear relationship between the log of the mean number of times colonies have been built in a water meter and the explanatory variables for the Poisson regression model. There is not an assumption of the independence of the observations; instead Heagerty and Lumley's [30] resampling method is used to find variance estimates for the parameters.

E. Parameter Estimation

Two models are implemented for the data: the logistic model, which models the odds of a colony occupying a particular location, and the Poisson model, which models the number of times a colony has been occupied. The data is quite sparse, yet some locations have been inhabited by colonies over ten times, meaning the Poisson model requires the additional dispersion parameter. Also, the logistic model does not require the assumption that records more than 30 days apart are necessarily separate colonies. Both models are used in part to ascertain whether results are similar.

1. Parameter Estimation: Logistic Model

The logistic regression model predicts the presence or absence of bees in water meters over the course of the entire 12 year study. In this case, the response variable y_s is a Bernoulli random variable denoting the presence or absence of a colony at any time during the study period at location s. The log odds of the presence of a colony in a particular location are modeled:

$$\log\left(\frac{\pi_s}{1-\pi_s}\right) = \sum_{j=1}^{k} \beta_j x_{s,j},$$

where π_s is the probability of observing a colony during the study period at location s. The first explanatory variable, $\mathbf{x}_1$, is simply the 1-vector, yielding an intercept for the model, β_1. Other potential explanatory variables, $\mathbf{x}_2, \ldots, \mathbf{x}_k$, (suppose that the parameter vector is of dimension k) include the age of the building corresponding to the water meter, whether a residence with a pool was located within half a mile, acreage, distance from each land class, whether a building was residential, and the proportion of the locations within a half-mile radius that had ever been occupied by a colony. The β's are the corresponding parameters. If $\mathbf{x}_2$ denotes acreage, for example, a positive estimate for β_2 would indicate that buildings with larger lot sizes are predicted to have a higher probability of occupancy.

To find parameter estimates for the logistic model, the default method of Iterative Reweighted Least Squares [32] for generalized linear models can be implemented in R. With such a large sample size, concerns about power to detect significance are not large. The main question, rather, is whether the statistically significant parameters are meaningful in the context of the problem. We must find variance estimates for the parameters before testing their significance; variance estimation is described in more detail in section F.

2. Parameter Estimation: Poisson Model

To account for different numbers of occupancies in different water meters, the over-dispersed Poisson regression model is implemented: $\log(\mu_s) = \sum_{j=1}^{k} \beta_j x_{sj}$. The response variable, y_s, denotes the number of times a colony had been built at location s. The same explanatory variables $\mathbf{x}_1, \ldots, \mathbf{x}_k$ as in the Poisson model are used. For the Poisson model, a positive estimate for the parameter β_2 corresponding to acreage would indicate that buildings with larger lot sizes are expected to be occupied more often during the 12 year study period. Because of the natural high zero count of the data (that is, there were many locations that had never had bees at all), an overdispersion parameter that accounts for the mean and variance not being equal in the Poisson model is included.

The Iterative Reweighted Least Squares method is implemented in the Poisson model as in the logistic model to estimate the regression parameters. After estimating them, the final parameter to be estimated is the over-dispersion parameter; as in McCullagh and Nelder [33], it is estimated as:

$$\hat{\phi} = \frac{1}{n-p} \sum_{s=1}^{n} \frac{(Y_s - \hat{\mu}_s)^2}{\hat{\mu}_s},$$

where p is the length of β, $\hat{\mu}_s$ is defined as $\exp\left\{\sum_j \hat{\beta}_j x_{js}\right\}$, and x_{js} is the value of covariate j corresponding to location s. The overdispersion parameter is essentially a ratio of the variance estimate to the estimate of the mean. When this parameter is equal to 1, we say there is no over-dispersion. For our data set, the parameter is estimated to be 1.52. To test for significance of the β parameters, variance estimates of the parameter estimates are obtained in section F.

F. Variance Estimation

Estimating the variances of the parameter estimates in order to find significance levels is more complex when the observations are correlated in space or time. Schabenberger and Gotway discuss the challenges in fitting generalized linear models in the case of spatial models for non-Gaussian data [3]. In the Gaussian case, the joint distribution is multivariate Gaussian, but for other distributions, the joint distributions may be unknown. Thus methods like maximum likelihood or the method of moments may be difficult or impossible, so a non-parametric approach to covariance estimation is sought.

Resampling methods began with U-statistics [6], the Jackknife [7, 8], and the bootstrap [9]. Since then, resampling methods have been extended to many different types of parameter estimation situations, including that of data correlated in space and time. Carlstein introduced the idea of using subseries, or "windows" to compute asymptotic standard errors [10]. These windows are contiguous regions, sections of time or subshapes of space, that represent smaller portions of the original larger process. Künsch [11] suggested the use of overlapping windows, and Hall and Jing [12] introduced the idea of an overlapping, non-independent window for dependent (temporal or spatial) data. Kaiser *et al.* explored inference on the spatial cumulative distribution function using subsampling [13]. Sherman extended the subseries method to linear models, and in 1996 showed for spatial data, the optimal subshape size is proportional to $n^{1/2}$ [14, 15]. Nordman and Lahiri further explored the issue, showing that the optimal block size depends on the shape of the sampling region and characteristics of the random field [16].

Extending the idea of subseries to the generalized linear model introduces new challenges. Our logistic regression model is most challenged by the fact that it is possible to have well-defined parameter estimates on the entire data set, while on certain windows, the binary response variable Y_i only takes on the value 0, necessitating a different approach. Heagerty and Lumley have introduced a method that utilizes estimating functions, functions which have as their root the parameter [30]. The estimating functions are suggested by Heyde to unify the theory behind least squares and maximum likelihood under the larger umbrella of quasi-likelihood [34]. Heagerty and Lumley's method is implemented, as shown here.

Accordingly, suppose the function $\mathbf{U}_s(Y_s, \mathbf{x}_s, \beta) \in \mathbb{R}^p$ is a general vector-valued estimating function of the data, where Y_s denotes the response variables, covariates are denoted $\mathbf{x}_s$ for $s \in \mathcal{D}_n$, and parameter $\beta \in \mathbb{R}^p$. $\mathcal{D}_n$ is a subset of $\mathbb{Z}^d$. It is required that the estimating functions be unbiased; that is, $E_\beta[\mathbf{U}_s(Y_s, \mathbf{x}_s, \beta)] = \mathbf{0}$ for all β. Denoting the cardinality of $\mathcal{D}_n$ as $|\mathcal{D}_n|$, our estimating function is

$$\bar{\mathbf{U}}_n(\beta) = \frac{1}{|\mathcal{D}_n|} \sum_{s \in \mathcal{D}_n} \mathbf{U}_s(Y_s, \mathbf{x}_s, \beta).$$

When evaluation at Y_s and $\mathbf{x}_s$ is implicit, the notation $\bar{\mathbf{U}}_n(\beta) = |\mathcal{D}_n|^{-1} \sum_s \mathbf{U}_s(\beta)$ is used, and when evaluation at β is implicit, the notation $\mathbf{U}_n(\beta) = \mathbf{U}_n$ is used.

By a theorem in Heagerty and Lumley's (2000) paper, letting $\mathbf{\Sigma}_n = \text{Cov}(\bar{\mathbf{U}}_n)$, convergence in distribution is obtained $\mathbf{\Sigma}_n^{-1/2}\bar{\mathbf{U}}_n \rightarrow \mathbf{N}(\mathbf{0}, \mathbf{I}_{p\times p})$, $\mathbf{I}$ being the identity matrix. So where $\mathbf{H}_n$ is denoted $\mathbf{H}_n = \mathbf{H}_n(\beta) = (\partial/\partial\beta)\bar{\mathbf{U}}_n$, convergence in distribution of the parameter estimates is obtained to find p-values for testing hypotheses about the parameters β: $\mathbf{\Sigma}_n^{-1/2}\mathbf{H}_n(\hat{\beta}_n - \beta) \rightarrow \mathbf{N}(\mathbf{0}, \mathbf{I}_{p\times p})$. As $\mathbf{H}_n$ is directly estimable, all that is necessary is to find an estimator for $\mathbf{\Sigma}_n$, and then an estimator of the asymptotic covariance of the parameter estimates will be obtained: $V(\hat{\beta}_n) \approx \hat{\mathcal{I}}_n = \mathbf{H}_n^{-1}(\hat{\beta})\hat{\mathbf{\Sigma}}_n\mathbf{H}_n^{-1}(\hat{\beta})$. The matrix $\mathbf{H_n}$ is directly estimable because it involves only the original data and parameter estimates of the original model, which

Heagerty and Lumley argue are easily calculated from readily available software.

Here Heagerty and Lumley [30] use a window subsampling estimator, as Garcia-Soidan and Hall [35] and Sherman [15] have used in other situations. Notation used is similar to that of Sherman's paper: $S \subset (-1,1]^d \subset \mathbb{R}^d$ "represents the interior of a simple closed curve with finite boundary and positive volume that is used as a template for $\mathcal{D}_n$." For any positive $m \in \mathbb{R}$, let S_m be the expansion of S over $(-\lceil (m/2)^{1/d} \rceil, \lceil (m/2)^{1/d} \rceil]^d$; then assume $\mathcal{D}_n = \{i : i \in S_n \cap \mathbb{Z}^d\}$. The contraction $\mathcal{D}^i_{l_n}$, where $l_n = \lceil \gamma n^{\delta/d} \rceil^d$ for some scalar $\gamma > 0$ and $0 < \delta < 1$, is defined as the translated subshape centered at $i \in \mathbb{Z}^d$ containing $\{j : j \in i + S_{l_n} \cap \mathbb{Z}^d\}$. The domain $\mathcal{D}_n$ contains the subregions $\mathcal{D}^i_{l_n}$; let m_n denote the number of subregions contained entirely in $\mathcal{D}_n$. Hall *et al.* showed that the use of nonoverlapping subregions often results in a less efficient estimator, although it is permissible to use nonoverlapping subregions [36]. Thus overlapping subregions are used.

Finally, the window subsampling empirical variance estimator is defined as:

$$\hat{\mathbf{\Sigma}}^{(0)}_\infty = \frac{1}{m_n} \sum_i^{m_n} \mathbf{h}_{\hat{\beta}_n}(\mathcal{D}^i_{l_n}),$$

where:

$$\begin{aligned} \mathbf{h}_{\hat{\beta}_n}(\mathcal{D}^i_{l_n}) &= |\mathcal{D}^i_{l_n}|(\bar{\mathbf{U}}^i_{l_n})(\bar{\mathbf{U}}^i_{l_n})^T \\ &= \frac{1}{|\mathcal{D}^i_{l_n}|} \sum_{j,k \in \mathcal{D}^i_{ln}} \mathbf{U}_j(\hat{\beta}_n)\mathbf{U}_k(\hat{\beta}_n)^T \end{aligned}$$

and

$$\bar{\mathbf{U}}^i_{l_n} = \sum_{j \in \mathcal{D}^i_{l_n}} \mathbf{U}_j(\hat{\beta}_n)/|\mathcal{D}^i_{l_n}|.$$

Then to estimate the asymptotic covariance of $\hat{\beta}_n$, Heagerty and Lumley (2000) use $\hat{\mathcal{I}}_n = \mathbf{H}^{-1}_n(\hat{\beta})\hat{\mathbf{\Sigma}}^{(0)}_n\mathbf{H}^{-1}_n(\hat{\beta})$, where $\hat{\mathbf{\Sigma}}^{(0)}_n = \hat{\mathbf{\Sigma}}^{(0)}_\infty/|\mathcal{D}_n|$.

1. Variance Estimation: Logistic Model

To find the estimating equations for the logistic model, the Maximum Likelihood method can be implemented. Defining $\ell(\beta)$ to be the log likelihood function for the logistic model, the first derivative is found:

$$\ell(\beta) = \sum_i (\sum_j \beta_j x_{s,j}) y_s - \sum_s \log \left[1 + \exp \left\{ \sum_j \beta_j x_{s,j} \right\} \right]$$
$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_s y_s x_{s,j} - \sum_s x_{s,j} \frac{\exp\{\sum_l \beta_l x_{s,l}\}}{1 + exp\{\sum_l \beta_l x_{s,l}\}}.$$

For simplicity, the mean is defined:

$$\pi_s = \frac{\exp\{\sum_l \beta_l x_{s,l}\}}{1 + exp\{\sum_l \beta_l x_{s,l}\}},$$

and $\pi_s = \mu_s$. Then the estimating equation vector is:

$$U_s = \begin{bmatrix} x_{s,1}(y_s - \mu_s) \\ \vdots \\ x_{s,k}(y_s - \mu_s) \end{bmatrix}$$

It is necessary to insert parameter estimates into the U function to be able to calculate covariance estimates, so parameter estimates based on the entire data are used. Thus the estimate for the mean is $\hat{\mu}_s = \exp\{\sum_k \hat{\beta}_k x_{s,k}\}/(1 + \exp\{\sum_k \hat{\beta}_k x_{s,k}\})$.

Next the $\mathbf{H}_n(\beta)$ matrix is found; differentiating, the (a, b) element is denoted

$$\frac{1}{n} \sum_s x_{s,a}\, x_{s,b} \left(\frac{\exp\left\{\sum_l \hat{\beta}_l x_{s,l}\right\}}{\left(1 + \exp\left\{\sum_l \hat{\beta}_l x_{s,l}\right\}\right)^2} \right),$$

where parameter estimates $\hat{\beta}$ from the original data set are used (rather than, say, averaging the window subsamples) as described in the parameter estimation section to get the $\mathbf{H}_n(\hat{\beta})$ matrix. Then after finding the $\mathbf{H}_n(\hat{\beta})$ matrix, the next step is to find an estimate $\hat{\mathbf{\Sigma}}_n$ of the covariance matrix $\mathbf{\Sigma}_n$. U-functions are thus calculated on each of the subsamples to get the window subsampling estimator of the covariance matrix. After finding $\hat{\Sigma}_n$, since $\mathbf{\Sigma}_n^{-1/2}\mathbf{H}_n(\hat{\beta}_n - \beta) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{I}_{p\times p})$, finally the variance matrix is calculated $V(\hat{\beta}_n) \approx \hat{\mathcal{I}}_n = \hat{\mathbf{H}}_n^{-1}\hat{\mathbf{\Sigma}}_n\hat{\mathbf{H}}_n^{-1}$. P-values are then calculated from $2P(T > |t_{obs}|)$, where $t_{obs} = \hat{\beta}/se(\hat{\beta})$, and $se(\hat{\beta}) = \sqrt{\text{Var}(\hat{\beta})}$. The test statistic t_{obs} has a large sample standard normal distribution by the Multivariate Slutsky's Theorem [19] since $\hat{\mathbf{\Sigma}}_n$ is L_2-convergent.

2. Variance Estimation: Poisson Model

To estimate the variances of the parameters, first the estimating equations are found, then calculate the $\mathbf{H}_n$ matrix (this time also denoted $\mathbf{H_n}(\beta, \phi)$ because of the extra parameter being estimated), and finally the covariance matrix can be calculated as for the logistic model. McCullagh, P. and Nelder state that the estimating equations for the Poisson model are [33]:

$$\mathbf{U}(\beta) = \mathbf{D}^T\mathbf{V}^{-1}(Y - \mu)\sigma^2,$$

where the diagonal elements in the covariance matrix $V_{ss} = Var(Y_s) = \phi\mu_s$, ϕ is the over-dispersion parameter, $D_{sr} = \delta\mu_s/\delta\beta_r$, and μ_s is the expected number of times bee colonies are built at location s. Recall that the mean is $\mu_s = \exp\{\sum_{j=1}^k x_{sj}\beta_j\}$. Then if $V(\mu) = \text{diag}(\mu_s)$ and $\sigma^2 = \phi$, note that:

$$\begin{aligned} D_{sr} &= \frac{\delta\mu_s}{\delta\beta_r} \\ &= \frac{\delta}{\delta\beta_r}\exp\left\{\sum_{j=1}^k x_{sj}\beta_j\right\} \\ &= \exp\left\{\sum_{j=1}^k x_{sj}\beta_j\right\} x_{sr} \\ &= \mu_s x_{sr} \end{aligned}$$

and thus because the μ_s's cancel, the s^{th} entry in the $\mathbf{U}$ vector is:

$$U_s(\beta) = \sum_{k=1}^{n} x_{ks}(y_k - \mu_k)/\phi.$$

This matches the estimating equations given by Fleiss *et al.* [37]. The overdispersion parameter is estimated as described in the parameter estimation section. Then $\mathbf{H}_n$, the matrix of partials, is found by differentiating: the (a, b) entry in the matrix is defined by:

$$\mathbf{H}_{n(a,b)}(\beta, \phi) = \frac{1}{n} \sum_{m=1}^{n} \frac{x_{ma}\, x_{mb}}{\phi} \exp\left\{\sum_j x_{mj}\, \beta_j\right\}.$$

In both of the above equations, again the estimating equations depend on β and ϕ. Thus $\mathbf{H}_n(\beta, \phi)$ is approximated by $\mathbf{H}_n(\hat{\beta}, \hat{\phi})$, using the parameter estimates from the original data set, as for the logistic model.

Finally, having obtained the $\mathbf{H}_n(\hat{\beta}, \hat{\phi})$ matrix, all that is left is to find the estimate of $\mathbf{\Sigma}_n$. Then Heagerty and Lumley's subsampling method is implemented as for the logistic model, computing the U-functions on each of the subsamples. Variances of the parameter estimates and test statistics are also computed as for the logistic model to obtain p-values.

G. Results

Table I in Appendix B displays the mean and standard deviation for all of the explanatory variables in the model. While the average percentage of neighbors with bees seems quite small, it is one of the most influential predictor variables. Few locations had been infested by an AHB colony, but the percentage of neighbors that had been infested was an important predictor in the model. The mean of indicator variables is a proportion: for example, for the pool indicator variable, 0.742 indicates that there was a pool within half a mile of 74.2% of the locations in Tucson. The variable distance from “water” is in quotation marks because most of the land parcels classified as the water land class do not actually hold water for much of the year; some are washes which fill with water only during the monsoon season [29]. On average, bees had access to almost all the land classes in under a mile; agricultural land was on average more than 2 miles from a given location. The average construction year for the buildings was 1979, and the oldest house in the data set was built in 1875.

The results from the logistic model are shown in Table II in Appendix B, and the results from the Poisson model are shown in Table III. The p-values for the two models are similar; in both models, the proportion of neighbors with colonies, construction year, acreage, distance from vacant land, and the indicator of whether there is a pool in the neighborhood are significant. Distance from natural land is also significant for only the logistic model at the 0.05 level.

With such a large sample size, power was less of a concern than finding statistically significant but meaningless results. The column “Change” is used to discern whether significant effects were meaningful in context of the data. That column represents an increase in the explanatory variable by the amount specified, an amount within the range of available data. The column “Factor” shows what the odds of occupancy, in the case of the logistic model, or the average number of occupations, for the Poisson model, should be multiplied by when the explanatory variable increases by the amount indicated by “Change.” All significant variables seemed to have a meaningful effect on the response. That is, the odds or expected number of occupancies for all significant variables except the residential indicator should be multiplied by a factor at least 1.3 or at most 0.7 when the explanatory variable is increased by an amount within the range of observed values for that variable. When a location is changed from commercial to residential, its odds of occupancy should be multiplied by 1.17, and the expected number of occupancies should be multiplied by 1.13, factors closer to 1 and thus somewhat less important.

In order to make appropriate conclusions, careful interpretations must be made. Because observational rather than experimental data was collected, the possibility of lurking variables means it cannot be concluded that the presence of colonies in the neighborhood, or any other explanatory variable, *causes* an increase or decrease in the predicted odds of a colony being present or number of expected colonies built. All the significant variables are also associated with distance from the center of town, and indeed both models predict that locations in the center of town to have more colony occupations. There may be another variable not measured that is associated with distance from the center of town that causes the explanatory variables to be significant. As aforementioned, bias is a possibility, though perhaps less of a problem with the logistic model, which does not account for repeated visits. The logistic model may still have some social bias if there is a failure to record a colony removal, however, it does not have mistakes from repeated records of visits that did not actually represent new colonies.

Some variables were likely surrogates for other, more influential, variables. For example, construction year may be a surrogate for whether a building had available openings providing opportunities for nest sites. Lot size may indicate the availability and variety of vegetation accessible to the colonies. The indicator variable for whether a pool is in the neighborhood may also indicate vegetation availability; or it may indicate that water is more readily available, perhaps from water hoses or pumps used to fill the pools. Whether a location is residential may be indicative of a difference in vegetation. The fact that the percent of neighbors with colonies was significant may indicate that the bees tend to form aggregations, perhaps because of increased mating efficiency or increased colony defense [38, 39]. It may also be true that a nearby established colony in another type of cavity regularly produced swarms with colonized water meter boxes. This explanation would indicate the importance of removing established colonies to control the population of Africanized honey bees.

CHAPTER IV

CONCLUSIONS

A. Summary

A method for testing spatiotemporal separability for point processes has been presented in this dissertation. The testing approach can be used in many settings because it requires only mild moment and weak dependence assumptions about the underlying process. To develop our test statistic, we established the asymptotic distribution of the second-order intensity estimators along with an L_2 consistent subsampling estimator for the covariance of our estimators, allowing us to show the asymptotic distribution of our test statistic to be χ^2 assuming separability.

Our Africanized honey bee study has identified several variables significantly associated with the occupation of colonies in a water meter box. The Poisson and logistic regression models that have been developed can be used to identify water meter boxes that are more likely to contain colonies, thus potentially protecting workers who read the meters. Identification of higher risk water meters can also lead to discovery of established colonies in nearby areas. Removal of these established colonies will prevent their forming new swarms.

B. Future Research

Because our test statistic has a χ^2 distribution only asymptotically, we should run simulations to find out what sample sizes are necessary under given conditions. We will want to know how much power we have to detect particular deviations from the assumption of separability. In addition, subspace size, bandwidth for the kernel estimator, and lags at which to calculate the second-order intensity must be chosen by the user. Some simulations under different covariance structures would help the user decide which choices are optimal for a particular data set.

Several aspects of the Africanized honey bee study warrant further investigation. Exploration of the land class variables may identify particular types of vegetation hospitable to Africanized honey bees. Measuring aspects of the buildings in Tucson may help explain why older homes are more likely to be occupied by colonies. The subspaces used to form the sample covariances of the parameter estimates have been user chosen; simulation may help decide the optimal subspace size for this data set.

The temporal aspect of the data has also been ignored. Further analysis develops a spatiotemporal model for the data modeling the log odds of the presence of a colony in a particular location on a particular month:

$$\log\left(\frac{\pi_{(\mathbf{s},t)}}{1-\pi_{(\mathbf{s},t)}}\right) = \sum_{j=1}^{k} \beta_j x_{(\mathbf{s},t),j},$$

where $\pi_{(\mathbf{s},t)}$ is the probability of observing a colony during the study period at location s and time t and the explanatory variables may include those mentioned for the spatial model, along with temporal variables such as season, temperature, rainfall, and pollen levels. Variance estimates may be found in a way similar to our spatial logistic model. The further sparsity of the data using the temporal aspect (8,211 colonies removed from 275,877 parcels in 145 months means only 0.02% of the locations were occupied during any given month) reemphasizes the need for a method like Heagerty and Lumley's [30] method which does not require all the subspaces to have occupied water meters.

REFERENCES

[1] N. Cressie and H.-C. Huang, "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, vol. 94, pp. 1330–1340, 1999.

[2] S. Poisson, *Recherches sur la Probabilité des Jugement en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités.* Paris: Bachelier, 1837.

[3] O. Schabenberger and C. Gotway, *Statistical Methods for Spatial Data Analysis.* Boca Raton, FL: Chapman & Hall/CRC, 2005.

[4] D. R. Brillinger, *Time Series: Data Analysis and Theory.* New York: Rinehart & Winston, 1975.

[5] M. Rosenblatt, "A central limit theorem and a strong mixing condition," *Biometrics*, vol. 37, pp. 531–539, 1956.

[6] W. Hoeffding, "A class of statistics with asymptotically normal distribution," *The Annals of Mathematical Statistics*, vol. 19, pp. 293–325, 1948.

[7] M. H. Quenouille, "Notes on bias in estimation," *Biometrika*, vol. 43, pp. 353–360, 1956.

[8] J. Tukey, "Bias and confidence in not-quite large samples," *The Annals of Mathematical Statistics*, vol. 29, p. 614, 1958.

[9] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, vol. 7, pp. 1–26, 1979.

[10] E. Carlstein, “The use of subseries values for estimating the variance of a general statistic from a stationary sequence,” *The Annals of Statistics*, vol. 14, pp. 1171–1179, 1986.

[11] H. Künsch, “The jackknife and bootstrap for general stationary observations,” *The Annals of Statistics*, vol. 17, pp. 1217–1241, 1989.

[12] P. Hall and B. Jing, “On sample reuse methods for dependent data,” *Journal of the Royal Statistical Society*, vol. 58, pp. 727–737, 1996.

[13] M. Kaiser, N. Hsu, N. Cressie, and S. Lahiri, “Inference for spatial processes using subsampling: A simulation study.” *Environmetrics*, vol. 8, pp. 485–502, 1997.

[14] M. Sherman, “Subseries methods in regression,” *Journal of the American Statistical Association*, vol. 92, pp. 1041–1048, 1997.

[15] ——, “Variance estimation for statistics computed from spatial lattice data,” *Journal of the Royal Statistical Society*, vol. 58, pp. 509–523, 1996.

[16] D. Nordman and S. Lahiri, “On optimal spatial subsample size for variance estimation,” *The Annals of Statistics*, vol. 32, pp. 1981–2027, 2004.

[17] B. Li, M. Genton, and M. Sherman, “A nonparametric assessment of properties of space-time covariance functions,” *Journal of the American Statistical Association*, vol. 102, pp. 736–744, 2007.

[18] R. Serfling, *Approximation Theorems of Mathematical Statistics.* Hoboken, NJ: John Wiley and Sons, Inc., 1980.

[19] T. Ferguson, *A Course in Large Sample Theory.* Boca Raton, FL: Chapman & Hall, 1996.

[20] E. Guzman-Novoa and R. Page, Jr, "Backcrossing Africanized honey bee queens to European drones reduces colony defensive behavior," *Annals of the Entomological Society of America*, vol. 86, pp. 352–355, 1993.

[21] L. Hunter, J. Jackman, and E. Sugden, "Detection records of Africanized honey bees in Texas during 1990, 1991,and 1992," *Southwestern Entomologist*, vol. 18, pp. 79–89, 1993.

[22] W. Rubink, P. Leuvano-Martinez, E. Sugden, W. Wilson, and A. Collins, "Subtropical Apis mellifera (Hymenoptera: Apidae) swarming dynamics and Africanization rates in northeastern Mexico and southern Texas," *Annals of the Entomological Society of America*, vol. 89, pp. 243–251, 1996.

[23] E. Guzman-Novoa and R. Page, Jr, "The impact of Africanized bees on Mexican beekeeping," *American Bee Journal*, vol. 58, pp. 727–737, 1994.

[24] G. Loper, "Genetic evidence of the Africanization of feral colonies in S. Arizona between 1993 and 1995," *American Bee Journal*, vol. 146, pp. 521–524, 1997.

[25] M. de Mello, E. da Silva, and D. Natal, "Abelhas Africanizadas em área metropolitana do Brasil: Abrigos e influências," *Revista Saúde Pública*, vol. 37, pp. 237–241, 2003.

[26] D. Mistro, L. Rodrigues, and W. Ferreira, Jr, "The Africanized honey bee dispersal: A mathematical zoom," *Bulletin of Mathematical Biology*, vol. 67, pp. 281–312, 2005.

[27] K. Baum, M. Tchakerian, S. Thoenes, and R. Coulson, "Africanized honey bees in urban environments: A spatio-temporal analysis," *Landscape and Urban Planning*, vol. 85, pp. 123–132, 2007.

[28] M. L. Winston, "The biology and management of Africanized honey bees," *Annual Review Entomology*, vol. 37, pp. 173–193, 1992.

[29] W. K. Shaw, L. Harris, and M. Livingston, "Vegetative characteristics of urban land covers in metropolitan Tucson," *Urban Ecosystems*, vol. 2, pp. 65–73, 1998.

[30] P. Heagerty and T. Lumley, "Window subsampling of estimating functions with application to regression models," *Journal of the American Statistical Association*, vol. 95, pp. 197–211, 2000.

[31] L. Gerig and A. Imdorf, "Development of the bee population in two swarming colonies and their swarms," *Schweizerische Bienen-Zeitung*, vol. 107, pp. 309–313, 1984.

[32] J. Scales and A. Gersztenkorn, "Robust methods in inverse theory," *Inverse Problems*, vol. 4, pp. 1071–1091, 1988.

[33] P. McCullagh and J. Nelder, *Generalized Linear Models*, 2nd ed. New York: Chapman & Hall/CRC, 1989.

[34] C. Heyde, *Quasilikelihood and Its Applications: A General Approach to Optimal Parameter Estimation*. New York: Springer-Verlag, 1997.

[35] P. Garcia-Soidan and P. Hall, "On sample reuse methods for spatial data," *Biometrics*, vol. 53, pp. 273–281, 1997.

[36] P. Hall, J. Horowitz, and B. Jing, "On blocking rules for the bootstrap with dependent data," *Biometrika*, vol. 82, pp. 561–574, 1995.

[37] J. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*. Indianapolis, IN: Wiley, 2003.

[38] T. Seeley, R. Seeley, and P. Akratanakul, “Colony defense strategies of the honeybees in Thailand,” *Ecological Monographs*, vol. 52, pp. 43–63, 1982.

[39] B. Oldroyd, A. Smolenski, S. Lawler, A. Estoup, and R. Crozier, “Colony aggregations in Apis mellifera L.” *Apidologie*, vol. 26, pp. 119–130, 1995.

[40] Y. Guan, M. Sherman, and J. Calvin, “A nonparametric test for spatial isotropy using subsampling,” *Journal of the American Statistical Association*, vol. 99, pp. 810–821, 2004.

[41] D. Politis and M. Sherman, “Moment estimation for statistics from marked point processes,” *Journal of the Royal Statistical Society Series B*, vol. 63, pp. 261–275, 2001.

APPENDIX A

LEMMAS AND PROOFS OF THEOREMS IN CHAPTER II

Lemmas

Lemma A.1: Assume the intensity functions of the point process exist up to order four. Then for $\mathbf{x}_1 \neq \mathbf{x}_2$, $\mathbf{y}_1 \neq \mathbf{y}_2$, $q_1 \neq q_2$, $p_1 \neq p_2$,

$$
\begin{aligned}
& E\{N^{(2)}[(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)] \times N^{(2)}[(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)]\} \\
= \; & \lambda_4(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1,\, p_2 - q_1) \\
& \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \mathrm{d}p_2 \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_2\, \epsilon_{\mathbf{x}_1,q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \epsilon_{\mathbf{x}_1,q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_2\, \epsilon_{\mathbf{x}_2,q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \epsilon_{\mathbf{x}_2,q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
+ \; & \lambda(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \epsilon_{\mathbf{x}_1,q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)\, \epsilon_{\mathbf{x}_2,q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
+ \; & \lambda(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \epsilon_{\mathbf{x}_1,q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)\, \epsilon_{\mathbf{x}_2,q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1),
\end{aligned}
$$

where $\epsilon_{\mathbf{x},q}(\cdot,\cdot)$ is a point measure and λ_k denotes the kth order intensity function, $k = 3, 4$.

Proof of Lemma A.1: Let $I_B(\mathbf{s}, t) = \begin{cases} 1 & , \quad (\mathbf{s}, t) \in B \\ 0 & , \quad \textit{otherwise} \end{cases}$. Define the kth factorial moment measure as

$$\alpha^{(k)}(B_1, B_2, \ldots, B_k) =$$
$$E\left[\sum_{(\mathbf{s}_1,t_1)\neq(\mathbf{s}_2,t_2)\neq\ldots\neq(\mathbf{s}_k,t_k)\in N}\cdots\sum I_{B_1}(\mathbf{s}_1,t_1)\cdot I_{B_2}(\mathbf{s}_2,t_2)\cdots I_{B_k}(\mathbf{s}_k,t_k)\right].$$

Because

$$\begin{aligned} & E[N(d\mathbf{x}_1, dq_1), N(d\mathbf{x}_2, dq_2), N(d\mathbf{y}_1, dp_1), N(d\mathbf{y}_2, dp_2)] \\ = \quad & E\Bigg[\sum\sum\sum\sum_{(\mathbf{s}_1,t_1),(\mathbf{s}_2,t_2),(\mathbf{s}_3,t_3),(\mathbf{s}_4,t_4)\in N} I_{d\mathbf{x}_1 dq_1}(\mathbf{s}_1,t_1) \\ & I_{d\mathbf{x}_2 dq_2}(\mathbf{s}_2,t_2) I_{d\mathbf{y}_1 dp_1}(\mathbf{s}_3,t_3) I_{d\mathbf{y}_2 dp_2}(\mathbf{s}_4,t_4)\Bigg] \end{aligned}$$

and

$$\begin{aligned} & \{(\mathbf{s}_1,t_1),(\mathbf{s}_2,t_2),(\mathbf{s}_3,t_3),(\mathbf{s}_4,t_4)\in N\} \\ & = \{(\mathbf{s}_1,t_1)\neq(\mathbf{s}_2,t_2)\neq(\mathbf{s}_3,t_3)\neq(\mathbf{s}_4,t_4)\}\cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_2,t_2)\neq(\mathbf{s}_3,t_3)\neq(\mathbf{s}_4,t_4)\} \\ & \cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_3,t_3)\neq(\mathbf{s}_2,t_2)\neq(\mathbf{s}_4,t_4)\}\cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_2,t_2)\neq(\mathbf{s}_3,t_3)\} \\ & \cup\{(\mathbf{s}_2,t_2)=(\mathbf{s}_3,t_3)\neq(\mathbf{s}_1,t_1)\neq(\mathbf{s}_4,t_4)\}\cup\{(\mathbf{s}_2,t_2)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_1,t_1)\neq(\mathbf{s}_3,t_3)\} \\ & \cup\{(\mathbf{s}_3,t_3)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_1,t_1)\neq(\mathbf{s}_2,t_2)\}\cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_2,t_2)=(\mathbf{s}_3,t_3)\neq(\mathbf{s}_4,t_4)\} \\ & \cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_2,t_2)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_3,t_3)\}\cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_3,t_3)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_2,t_2)\} \\ & \cup\{(\mathbf{s}_2,t_2)=(\mathbf{s}_3,t_3)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_1,t_1)\}\cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_2,t_2)\neq(\mathbf{s}_3,t_3)=(\mathbf{s}_4,t_4)\} \\ & \cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_3,t_3)\neq(\mathbf{s}_2 t_2)=(\mathbf{s}_4,t_4)\}\cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_4,t_4)\neq(\mathbf{s}_2 t_2)=(\mathbf{s}_3,t_3)\} \\ & \cup\{(\mathbf{s}_1,t_1)=(\mathbf{s}_2,t_2)=(\mathbf{s}_3 t_3)=(\mathbf{s}_4,t_4)\}, \end{aligned}$$

so we observe that $E\left[N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1)\times N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2)\times N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1)\times N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\right]$ can be written as fifteen terms. By definition, the first term

$$E\left[\sum\sum\sum\sum_{(\mathbf{s}_1,t_1)\neq(\mathbf{s}_2,t_2)\neq(\mathbf{s}_3,t_3)\neq(\mathbf{s}_4,t_4)} I_{\mathrm{d}\mathbf{x}_1\mathrm{d}q_1}(\mathbf{s}_1,t_1)I_{\mathrm{d}\mathbf{x}_2\mathrm{d}q_2}(\mathbf{s}_2,t_2)I_{\mathrm{d}\mathbf{y}_1\mathrm{d}p_1}(\mathbf{s}_3,t_3)I_{\mathrm{d}\mathbf{y}_2\mathrm{d}p_2}(\mathbf{s}_4,t_4)\right]$$
$$= \alpha^{(4)}((\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),(\mathrm{d}\mathbf{y_2},\mathrm{d}p_2)).$$

Consider the second term:

$$\begin{aligned} & E\Bigg[\sum\sum\sum\sum_{(\mathbf{s}_1,t_1)=(\mathbf{s}_2,t_2)\neq(\mathbf{s}_3,t_3)\neq(\mathbf{s}_4,t_4)} I_{\mathrm{d}\mathbf{x}_1\mathrm{d}q_1}(\mathbf{s}_1,t_1) \\ & I_{\mathrm{d}\mathbf{x}_2\mathrm{d}q_2}(\mathbf{s}_2,t_2)I_{\mathrm{d}\mathbf{y}_1\mathrm{d}p_1}(\mathbf{s}_3,t_3)I_{\mathrm{d}\mathbf{y}_2\mathrm{d}p_2}(\mathbf{s}_4,t_4)\Bigg] \\ = \; & E\left[\sum\sum\sum_{(\mathbf{s}_1,t_1)\neq(\mathbf{s}_3,t_3)\neq(\mathbf{s}_4,t_4)} I_{(\mathrm{d}\mathbf{x}_1\mathrm{d}q_1)\cap(\mathrm{d}\mathbf{x}_2\mathrm{d}q_2)}(\mathbf{s}_1,t_1)I_{\mathrm{d}\mathbf{y}_1\mathrm{d}p_1}(\mathbf{s}_3,t_3)I_{\mathrm{d}\mathbf{y}_2\mathrm{d}p_2}(\mathbf{s}_4,t_4)\right] \\ = \; & \alpha^{(3)}((\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\epsilon_{\mathbf{x}_1}(\mathrm{d}\mathbf{x}_2) \end{aligned}$$

We find the second equality from the fact that $(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1)\cap(\mathrm{d}\mathbf{x_2},\mathrm{d}q_2)$ is empty unless $(\mathbf{x}_1,q_1)=(\mathbf{x}_2,q_2)$ (i.e. two infinitesimally small discs centered at $(\mathbf{x}_1,q_1)$ and $(\mathbf{x}_2,q_2)$ are disjoint if $(\mathbf{x}_1,q_1)\neq(\mathbf{x}_2,q_2)$) and the definition of $\alpha^{(3)}(\cdot,\cdot)$.

Continuing in this way, we obtain

$$
\begin{aligned}
&E[N(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), N(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), N(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), N(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)] \\
&= \alpha^{(4)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2))\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{y}_1 p_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2))\epsilon_{\mathbf{x}_1 q_1}((\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{x}_1 q_1}((\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2))\epsilon_{\mathbf{x}_1 q_1}((\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2))\epsilon_{\mathbf{x}_2 q_2}((\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)\epsilon_{\mathbf{y}_1 p_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)) \\
&+ \lambda\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}q_1\, \epsilon_{\mathbf{x}_1 q_1}((\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)),
\end{aligned}
$$

where λ is the first-order intensity of the process. Further imposing the condition that $\mathbf{x}_1 \neq \mathbf{x}_2$ and $\mathbf{y}_1 \neq \mathbf{y}_2$, we obtain

$$
\begin{aligned}
&E\left[N^{(2)}(\mathrm{d}\mathbf{x}_1, \mathrm{d}\mathbf{x}_2) \times N^{(2)}(\mathrm{d}\mathbf{y}_1, \mathrm{d}\mathbf{y}_2)\right] \\
&= \alpha^{(4)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2))\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
&+ \alpha^{(3)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2), (\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)) \\
&+ \alpha^{(2)}((\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2))\epsilon_{\mathbf{x}_1 q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)\epsilon_{\mathbf{x}_2 q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1))
\end{aligned}
$$

By the definition of $\alpha^{(k)}$, then, we see that the above term is equal to:

$$
\begin{aligned}
&\lambda_4(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1,\, p_2 - q_1) \\
&\mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \mathrm{d}p_2 \\
&+ \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_2\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
&+ \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_2\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
&+ \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \lambda(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
&+ \lambda(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1),
\end{aligned}
$$

as desired. ■

Lemma A.2: Assume the intensity functions of the point process exist up to order four. Then $\lambda(\mathbf{x}_2 - \mathbf{x_1}, q_2 - q_1) = C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1, q_2 - q_1) + \lambda^2$ and

$$
\begin{aligned}
&\lambda_4(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1,\, p_2 - q_1) \\
&= C_N^{(4)}(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1,\, p_2 - q_1) \\
&+ \lambda C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1) \\
&+ \lambda C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1) \\
&+ \lambda C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, p_1 - q_1,\, p_2 - q_1) \\
&+ \lambda C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_2,\, \mathbf{y}_2 - \mathbf{x}_2,\, p_1 - q_2,\, p_2 - q_2) \\
&+ C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1) C_N^{(2)}(\mathbf{y}_2 - \mathbf{y}_1,\, p_2 - p_1) \\
&+ C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1,\, p_1 - q_1) C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2,\, p_2 - q_2) \\
&+ C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1,\, p_2 - q_1) C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2,\, p_1 - q_2) \\
&+ \lambda^2 C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1) + \lambda^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1,\, p_1 - p_2) \\
&+ \lambda^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1,\, p_2 - q_1) + \lambda^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2,\, p_1 - q_2) \\
&+ \lambda^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2,\, p_2 - q_2) + \lambda^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{y}_1,\, p_2 - p_1) + \lambda^2.
\end{aligned}
$$

Proof of Lemma A.2: We repeatedly use the relationship between moments and cumulants (e.g. McCullagh 1987).

$$
\begin{aligned}
&E[N(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1) N(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)] \\
&= \mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), N(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)) + \mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1)) \mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)) \\
&= C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1, q_2 - q_1) \mathrm{d}\mathbf{x}_1 \mathrm{d}\mathbf{x}_2 \mathrm{d}q_1 \mathrm{d}q_2 + \lambda^2 \mathrm{d}\mathbf{x}_1 \mathrm{d}\mathbf{x}_2 \mathrm{d}q_1 \mathrm{d}q_2
\end{aligned}
$$

$$
\begin{aligned}
&E[N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1)N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2)N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1)N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)]\\
=&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2))\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1))\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1),N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2)\\
+&\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_1,\mathrm{d}q_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{x}_2,\mathrm{d}q_2))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_1,\mathrm{d}p_1))\mathrm{Cum}(N(\mathrm{d}\mathbf{y}_2,\mathrm{d}p_2).
\end{aligned}
$$

The lemma is then proved by using the definition of cumulant functions. ■

Proof of Theorems

Proof of Theorem 1: Let $K_n(\mathbf{x}, q) \equiv h_{1n}^{-2} h_{2n}^{-1} K(\mathbf{x}/h_{1n}, q/h_{2n})$. For large n such that

$C \in A_n - A_n$,

$$\begin{aligned} E[\hat{\lambda}_{2n}(\mathbf{s},t)] &= \int_{\mathbf{x}_1 \in D_n} \int_{\mathbf{x}_2 \in D_n} \int_{q_1 \in T_n} \int_{q_2 \in T_n} \frac{K_n[(\mathbf{s} - \mathbf{x}_1 + \mathbf{x}_2), (t - q_1 + q_2)]}{|A_n|} \\ &\quad \times \lambda_2(\mathbf{x}_2 - \mathbf{x}_1) \, \mathrm{d}q_2 \, \mathrm{d}q_1 \, \mathrm{d}\mathbf{x}_2 \, \mathrm{d}\mathbf{x}_1 \\ &= \int_{D_n - D_n} \int_{T_n - T_n} K_n((\mathbf{s} + \mathbf{v}), (t + u)) \lambda_2(\mathbf{v}, u) \, \mathrm{d}u \, \mathrm{d}\mathbf{v} \\ &= \int_C K(\mathbf{z}, y) \, \lambda_2(\mathbf{s} - h_{1n}\mathbf{z}, t - h_{2n}y) \, \mathrm{d}y \, \mathrm{d}\mathbf{z} \\ &\to \lambda_2(\mathbf{s}, t), \end{aligned}$$

where the last limit is found by Lebesgue's Dominated Convergence theorem.

Now to derive the variance, consider two lags, $(\mathbf{s}, t)$ and $(\mathbf{s}', t')$, where $(\mathbf{s}, t)$ and $(\mathbf{s}', t')$ are elements of Λ, and Λ is a user-chosen lag set of interest. Then $\mathrm{Cov}(\hat{\lambda}_n(\mathbf{s}, t), \hat{\lambda}_n(\mathbf{s}', t'))$

can be written as:

$$
\begin{aligned}
& E[\hat{\lambda}_{2n}(\mathbf{s}, t) \times \hat{\lambda}_{2n}(\mathbf{s}', t')] - E[\hat{\lambda}_{2n}(\mathbf{s}, t)]\, E[\hat{\lambda}_{2n}(\mathbf{s}', t')] \\
= \; & \frac{1}{|A_n| \times |A_n'|} \underset{D_n}{\iiiint} \underset{T_n}{\iiiint} K_n[(\mathbf{s} - \mathbf{x}_1 + \mathbf{x}_2), (t - q_1 + q_2)] \\
& \times K_n[(\mathbf{s}' - \mathbf{y}_1 + \mathbf{y}_2), (t' - p_1 + p_2)] \\
& \times E\{N^{(2)}[(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)]\, N^{(2)}[(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)]\} \\
& - \frac{1}{|A_n| \times |A_n'|} \underset{D_n}{\iiiint} \underset{T_n}{\iiiint} K_n[(\mathbf{s} - \mathbf{x}_1 + \mathbf{x}_2), (t - q_1 + q_2)] \\
& \times K_n[(\mathbf{s}' - \mathbf{y}_1 + \mathbf{y}_2), (t' - p_1 + p_2)] \\
& \times E\{N^{(2)}[(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)]\}\, E\{N^{(2)}[(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)]\} \\
= \; & \frac{1}{|A_n| \times |A_n'|} \underset{D_n}{\iiiint} \underset{T_n}{\iiiint} K_n[(\mathbf{s} - \mathbf{x}_1 + \mathbf{x}_2), (t - q_1 + q_2)] \\
& \times K_n[(\mathbf{s}' - \mathbf{y}_1 + \mathbf{y}_2), (t' - p_1 + p_2)] \\
& \times \big(E\{N^{(2)}[(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)]\, N^{(2)}[(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)]\} \\
& - E\{N^{(2)}[(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)]\}\, E\{N^{(2)}[(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)]\}\big)
\end{aligned}
$$

Then from the results of Lemma 1, we find that:

$$
\begin{aligned}
& E\{N^{(2)}[(\mathrm{d}\mathbf{x}_1, \mathrm{d}q_1), (\mathrm{d}\mathbf{x}_2, \mathrm{d}q_2)] \times N^{(2)}[(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1), (\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)]\} \\
= \; & [\lambda_4(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1,\, p_2 - q_1) \\
- \; & \lambda(\mathbf{x}_2 - \mathbf{x}_1, q_2 - q_1)\lambda(\mathbf{y}_2 - \mathbf{y}_1, p_2 - p_1)]\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \mathrm{d}p_2 \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_2\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_2 - \mathbf{x}_1,\, q_2 - q_1,\, p_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_2\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1) \\
+ \; & \lambda_3(\mathbf{x}_2 - \mathbf{x}_1,\, \mathbf{y}_1 - \mathbf{x}_1,\, q_2 - q_1,\, p_1 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}\mathbf{y}_1\, \mathrm{d}q_1\, \mathrm{d}q_2\, \mathrm{d}p_1\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
+ \; & \lambda(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1)\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2) \\
+ \; & \lambda(\mathbf{x}_2 - \mathbf{x}_1,\, q_2 - q_1)\, \mathrm{d}\mathbf{x}_1\, \mathrm{d}\mathbf{x}_2\, \mathrm{d}q_1\, \mathrm{d}q_2\, \epsilon_{\mathbf{x}_1, q_1}(\mathrm{d}\mathbf{y}_2, \mathrm{d}p_2)\, \epsilon_{\mathbf{x}_2, q_2}(\mathrm{d}\mathbf{y}_1, \mathrm{d}p_1). \qquad \text{(A.1)}
\end{aligned}
$$

So we see the covariance can be written as seven terms; we label them as terms 1-7, respectively. First we focus on the first term. From the results of Lemma 2, we have:

$$
\begin{aligned}
& \lambda_4(\mathbf{x}_2-\mathbf{x}_1,\, \mathbf{y}_1-\mathbf{x}_1,\, \mathbf{y}_2-\mathbf{x}_1,\, q_2-q_1,\, p_1-q_1,\, p_2-q_1) \\
- \; & \lambda(\mathbf{x}_2-\mathbf{x}_1, q_2-q_1)\lambda(\mathbf{y}_2-\mathbf{y}_1, p_2-p_1) \\
= \; & C_N^{(4)}(\mathbf{x}_2-\mathbf{x}_1, \mathbf{y}_1-\mathbf{x}_1, \mathbf{y}_2-\mathbf{x}_1, q_2-q_1, p_1-q_1, p_2-q_1) \\
+ \; & \lambda C_N^{(3)}(\mathbf{x}_2-\mathbf{x}_1, \mathbf{y}_2-\mathbf{x}_1, q_2-q_1, p_1-q_1) \\
+ \; & \lambda C_N^{(3)}(\mathbf{x}_2-\mathbf{x}_1, \mathbf{y}_2-\mathbf{x}_1, q_2-q_1, p_2-q_1) \\
+ \; & \lambda C_N^{(3)}(\mathbf{y}_1-\mathbf{x}_1, \mathbf{y}_2-\mathbf{x}_1, p_1-q_1, p_2-q_1) \\
+ \; & \lambda C_N^{(3)}(\mathbf{y}_1-\mathbf{x}_2, \mathbf{y}_2-\mathbf{x}_2, p_1-q_2, p_2-q_2) \\
+ \; & C_N^{(2)}(\mathbf{y}_1-\mathbf{x}_1, p_1-q_1)\, C_N^{(2)}(\mathbf{y}_2-\mathbf{x}_2, p_2-q_2) \\
+ \; & C_N^{(2)}(\mathbf{y}_2-\mathbf{x}_1, p_2-q_1)\, C_N^{(2)}(\mathbf{y}_1-\mathbf{x}_2, p_1-q_2) \\
+ \; & \lambda^2 C_N^{(2)}(\mathbf{y}_1-\mathbf{x}_1, p_1-q_1) + \lambda^2 C_N^{(2)}(\mathbf{y}_2-\mathbf{x}_1, p_2-q_1) \\
+ \; & \lambda^2 C_N^{(2)}(\mathbf{y}_1-\mathbf{x}_2, p_1-q_2) + \lambda^2 C_N^{(2)}(\mathbf{y}_2-\mathbf{x}_2, p_2-q_2)
\end{aligned}
$$

Above we have eleven terms; we denote them as (1.1) - (1.11). We need to show that all eleven terms are of order $\frac{1}{|A_n|}$. From here on, because we are interested only in the speed of convergence, we will assume that $\lambda = 1$. First consider (1.1):

$$\iiiint\limits_{D_n} \iiiint\limits_{T_n} \frac{K_n[(\mathbf{s}-\mathbf{x}_1+\mathbf{x}_2),(t-q_1+q_2)]}{|D_n \times T_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2) \times (T_n - q_1 + q_2)|}$$

$$\times \quad \frac{K_n[(\mathbf{s}'-\mathbf{y}_1+\mathbf{y}_2),(t'-p_1+p_2)]}{|D_n \times T_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2) \times (T_n - p_1 + p_2)|}$$

$$\times \quad C_N^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, q_2 - q_1, p_1 - q_1, p_2 - q_1)$$

$$\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}p_1\,\mathrm{d}p_2\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2\,\mathrm{d}\mathbf{y}_1\,\mathrm{d}\mathbf{y}_2$$

$$= \quad \iiint\limits_{D_n - D_n} \iiint\limits_{T_n - T_n} |D_n \times T_n \cap (D_n - \mathbf{u}_1) \times (T_n - v_1) \cap (D_n - \mathbf{u}_2) \times (T_n - v_2)$$

$$\cap (D_n - \mathbf{u}_3) \times (T_n - v_3)| / |D_n \times T_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2) \times (T_n - q_1 + q_2)|$$

$$\times \quad |D_n \times T_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2) \times (T_n - p_1 + p_2)|$$

$$\times \quad K_n(\mathbf{s} + \mathbf{u}_1, t + v_1) \times K_n(\mathbf{s}' + \mathbf{u}_3 - \mathbf{u}_2, t' + v_3 - v_2)$$

$$\times \quad C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, v_1, v_2, v_3)$$

$$\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_2\,\mathrm{d}\mathbf{u}_3$$

$$\leq \quad \iiint\limits_{D_n - D_n} \iiint\limits_{T_n - T_n} \frac{K_n(\mathbf{s} + \mathbf{u}_1, t + v_1) \times K_n(\mathbf{s}' + \mathbf{u}_3 - \mathbf{u}_2, t' + v_3 - v_2)}{|D_n \times T_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2) \times (T_n + v_3 - v_2)|}$$

$$\times \quad |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, v_1, v_2, v_3)|\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_2\,\mathrm{d}\mathbf{u}_3$$

$$\leq \quad \iiint\limits_{\Re^2} \iiint\limits_{\Re} \frac{K_n(\mathbf{s} + \mathbf{u}_1, t + v_1) \times K_n(\mathbf{s}' + \mathbf{u}_4, t' + v_4)}{|D_n \times T_n \cap (D_n + \mathbf{u}_4) \times (T_n + v_4)|}$$

$$\times \quad |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_4, v_1, v_2, v_2 + v_4)|\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_4\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_2\,\mathrm{d}\mathbf{u}_4$$

$$\leq \quad C_1 \times \iint\limits_{\Re^2} \iint\limits_{\Re} \frac{K_n(\mathbf{s} + \mathbf{u}_1, t + v_1) \times K_n(\mathbf{s}' + \mathbf{u}_4, t' + v_4)}{|D_n \times T_n \cap (D_n + \mathbf{u}_4) \times (T_n + v_4)|}\,\mathrm{d}v_1\,\mathrm{d}v_4\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_4$$

$$= \quad \left(\frac{1}{|D_n \times T_n|}\right).$$

Next we look at (1.2):

$$
\begin{aligned}
& \iiiint\limits_{D_n} \iiiint\limits_{T_n} \frac{K_n[(\mathbf{s}-\mathbf{x}_1+\mathbf{x}_2),(t-q_1+q_2)]}{|D_n\times T_n\cap(D_n-\mathbf{x}_1+\mathbf{x}_2)\times(T_n-q_1+q_2)|} \\
& \times\frac{K_n[(\mathbf{s}'-\mathbf{y}_1+\mathbf{y}_2),(t'-p_1+p_2)]}{|D_n\times T_n\cap(D_n-\mathbf{y}_1+\mathbf{y}_2)\times(T_n-p_1+p_2)|} \\
& \times C_N^{(3)}(\mathbf{x}_2-\mathbf{x}_1,\mathbf{y}_1-\mathbf{x}_1,q_2-q_1,p_1-q_1)\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}p_1\,\mathrm{d}p_2\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2\,\mathrm{d}\mathbf{y}_1\,\mathrm{d}\mathbf{y}_2 \\
= \quad & \iiint\limits_{D_n-D_n}\iiint\limits_{T_n-T_n}|D_n\times T_n\cap(D_n-\mathbf{u}_1)\times(T_n-v_1)\cap(D_n-\mathbf{u}_2)\times(T_n-v_2) \\
& \cap(D_n-\mathbf{u}_3)\times(T_n-v_3)| \\
& /|D_n\times T_n\cap(D_n-\mathbf{x}_1+\mathbf{x}_2)\times(T_n-q_1+q_2)| \\
& \times|D_n\times T_n\cap(D_n-\mathbf{y}_1+\mathbf{y}_2)\times(T_n-p_1+p_2)| \\
& \times K_n(\mathbf{s}+\mathbf{u}_1,t+v_1)\times K_n(\mathbf{s}'+\mathbf{u}_3-\mathbf{u}_2,t'+v_3-v_2) \\
& \times C_N^{(3)}(\mathbf{u}_1,\mathbf{u}_2,v_1,v_2)\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_2\,\mathrm{d}\mathbf{u}_3 \\
\leq \quad & \iiint\limits_{D_n-D_n}\iiint\limits_{T_n-T_n}\frac{K_n(\mathbf{s}+\mathbf{u}_1,t+v_1)\times K_n(\mathbf{s}'+\mathbf{u}_3-\mathbf{u}_2,t'+v_3-v_2)}{|D_n\times T_n\cap(D_n+\mathbf{u}_3-\mathbf{u}_2)\times(T_n+v_3-v_2)|} \\
& \times|C_N^{(3)}(\mathbf{u}_1,\mathbf{u}_2,v_1,v_2)|\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_2\,\mathrm{d}\mathbf{u}_3 \\
\leq \quad & C_2\times\iint\limits_{\Re^2}\iint\limits_{\Re}\frac{K_n(\mathbf{s}+\mathbf{u}_1,t+v_1)\times K_n(\mathbf{s}'+\mathbf{u}_4,t'+v_4)}{|D_n\times T_n\cap(D_n+\mathbf{u}_4)\times(T_n+v_4)|}\,\mathrm{d}v_1\,\mathrm{d}v_4\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_4 \\
= \quad & O\left(\frac{1}{|D_n\times T_n|}\right).
\end{aligned}
$$

Similarly we can prove that terms (1.3) - (1.5) are all of order $\frac{1}{|D_n|}$. Now let's consider

(1.6).

$$\begin{aligned}
&\iiiint\!\!\int_{D_n} \iiiint\!\!\int_{T_n} \frac{K_n[(\mathbf{s}-\mathbf{x}_1+\mathbf{x}_2),(t-q_1+q_2)]}{|D_n \times T_n \cap (D_n-\mathbf{x}_1+\mathbf{x}_2)\times(T_n-q_1+q_2)|} \\
&\times \frac{K_n[(\mathbf{s}'-\mathbf{y}_1+\mathbf{y}_2),(t'-p_1+p_2)]}{|D_n \times T_n \cap (D_n-\mathbf{y}_1+\mathbf{y}_2)\times(T_n-p_1+p_2)|} \\
&\times C_N^{(2)}(\mathbf{x}_2-\mathbf{x}_1,q_2-q_1)\times C_N^{(2)}(\mathbf{y}_1-\mathbf{x}_2,p_1-q_2)\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}p_1\,\mathrm{d}p_2\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2\,\mathrm{d}\mathbf{y}_1\,\mathrm{d}\mathbf{y}_2 \\
\leq\;& \iiint_{D_n-D_n}\iiint_{T_n-T_n} |D_n\times T_n \cap (D_n-\mathbf{u}_1)\times(T_n-v_1)\cap(D_n-\mathbf{u}_2)\times(T_n-v_2) \\
&\cap(D_n-\mathbf{u}_3)\times(T_n-v_3)| \\
&/|D_n\times T_n\cap(D_n-\mathbf{x}_1+\mathbf{x}_2)\times(T_n-q_1+q_2)| \\
&\times|D_n\times T_n\cap(D_n-\mathbf{y}_1+\mathbf{y}_2)\times(T_n-p_1+p_2)| \\
&\times K_n(\mathbf{s}+\mathbf{u}_1,t+v_1)\times K_n(\mathbf{s}'+\mathbf{u}_3-\mathbf{u}_2,t'+v_3-v_2) \\
&\times|C_N^{(2)}(\mathbf{u}_2,v_2)|\,\mathrm{d}v_1\,\mathrm{d}v_2\,\mathrm{d}v_3\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_2\,\mathrm{d}\mathbf{u}_3 \\
\leq\;& C_3\times\iint_{\Re^2}\iint_{\Re}\frac{K_n(\mathbf{s}+\mathbf{u}_1,t+v_1)\times K_n(\mathbf{s}'+\mathbf{u}_4,t'+v_4)}{|D_n\times T_n\cap(D_n+\mathbf{u}_4)\times(T_n+v_4)|}\,\mathrm{d}v_1\,\mathrm{d}v_4\,\mathrm{d}\mathbf{u}_1\,\mathrm{d}\mathbf{u}_4 \\
=\;& O\left(\frac{1}{|D_n\times T_n|}\right).
\end{aligned}$$

Similarly, terms (1.6) - (1.11) are all of order $\frac{1}{|D_n|}$. Thus we conclude that all eleven terms of the first term in (A.1) are all of order $\frac{1}{|D_n|}$. Now we proceed to the other six terms in (A.1). Terms 2-5 can be shown all of order $\frac{1}{|D_n|}$ due to the fact that λ_3 is finite. Now we consider the sixth term.

$$\iint_{D_n}\iint_{T_n} \frac{K_n(\mathbf{s}-\mathbf{x}_1+\mathbf{x}_2, t-q_1+q_2)\times K_n(\mathbf{s}'-\mathbf{x}_1+\mathbf{x}_2, t'-q_1+q_2)}{|(D_n\times T_n\cap[(D_n-\mathbf{x}_1+\mathbf{x}_2)\times(T_n-q_1+q_2)]|^2}$$
$$\times\lambda_2(\mathbf{x}_2-\mathbf{x}_1, q_2-q_1)\,\mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2$$
$$= \int_{D_n-D_n}\int_{T_n-T_n} \frac{K_n(\mathbf{s}+\mathbf{u}, t+v)\times K_n(\mathbf{s}'+\mathbf{u}, t'+v)}{|(D_n\times T_n)\cap[(D_n-\mathbf{u})\times(T_n-v)]|}\times\lambda_2(\mathbf{u},v)\,\mathrm{d}v\,\mathrm{d}\mathbf{u}$$
$$= \int_{D_n-D_n}\int_{T_n-T_n} \frac{K(\mathbf{w},z)\times K(\mathbf{w}+(\mathbf{s}'-\mathbf{s})/h_{1n}, z+(t'-t)/h_{2n})}{|(D_n\times T_n)\cap[(D_n+\mathbf{s}-h_{1n}\mathbf{w})\times(T_n+t-h_{2n}z)]|\times h_{1n}^2\times h_{2n}}$$
$$\times\lambda_2(h_{1n}\mathbf{w}-\mathbf{s}, h_{2n}z-t)\,\mathrm{d}z\,\mathrm{d}\mathbf{w}.$$

Thus $\lim_{n\to\infty}|D_n|\times|T_n|\times h_{1n}^2\times h_{2n}\times$ (A.1.6) $=\iint_C K^2(\mathbf{w},z)\,\mathrm{d}z\,\mathrm{d}\mathbf{w}\times\lambda_2(\mathbf{s},t)\times I(\mathbf{s}=\mathbf{s}', t=t')$, where (A.1.6) denotes the sixth term of Equation (A.1). Similarly we can show $\lim_{n\to\infty}|D_n|\times|T_n|\times h_{1n}^2\times h_{2n}\times(7)=\iint_C K^2(\mathbf{w},z)\,\mathrm{d}z\,\mathrm{d}\mathbf{w}\times\lambda_2(\mathbf{s},t)\times I(\mathbf{s}=\mathbf{s}', t=-t')$. Hence we find that

$$\lim_{n\to\infty}|A_n|\times h_{1n}^2\times h_{2n}\times\mathrm{Cov}[\hat{\lambda}_{2n}(\mathbf{s}_i,t_i),\hat{\lambda}_{2n}(\mathbf{s}_j,t_j)]$$
$$= \begin{cases} \int K^2(\mathbf{x},q)\,\mathrm{d}\mathbf{x}\,\mathrm{d}q\times\lambda_2(\mathbf{s}_i,t_i) & ,\quad \mathbf{s}_i=\pm\mathbf{s}_j, t_i=\pm t_j \\ 0 & ,\quad o.w. \end{cases}$$

■

Proof of Theorem 2: Our statistic of interest is $S_n\equiv\sqrt{|A_n|}\times h_{1n}\times\sqrt{h_{2n}}\times\{\hat{\lambda}_{2n}(\mathbf{s},t)-E[\hat{\lambda}_{2n}(\mathbf{s},t)]\}$; we shall prove that $S_n\xrightarrow{D}N(0,\sigma^2)$, where $\sigma^2\equiv\int_C K^2(\mathbf{u},v)\,\mathrm{d}\mathbf{u}\,\mathrm{d}v\times\lambda_2(\mathbf{s},t)$. We use a blocking technique (e.g. Ibragimov & Linnik, 1971) to do this.

Take $d=2$, for the spatial domain. Divide the original domain $A_n=D_n\times T_n$ into nonoverlapping partitions of size $m(r_n)\times m(r_n)\times l(n)$, where $m(r_n)=r_n^\alpha$ is the length of one side of D_n and $l(n)=n^\beta=|T_n|$. Call the partitioned square cuboids $A^i_{m(r_n),l(n)}, i=1,\ldots,k_n$. Within each partition further obtain subcuboids $A^i_{m(r_n)',l(n)'}$, where $m(r_n)' = r_n^\alpha - r_n^\eta$ and $l(n)' = n^\beta - n^\theta$, for some $4/(2+\epsilon) < \eta < \alpha < 1$ and $2/(1+\delta) < \theta < \beta < 1$. The subcubes $A^i_{m(r_n)',\, l(n)'}$ should have the same centers as the original $A^i_{m(r_n),\, l(n)}$. Thus for $i \neq j$, $d(A^i_{m(r_n)',\, l(n)'}, A^j_{m(r_n)',\, l(n)'}) \geq \min(r_n^\eta, n^\theta)$.

Now we'll use the following statistics in addition to our S_n from above:

$$s_n \equiv \textstyle\sum_{i=1}^{k_n} s_n^i/\sqrt{k_n}$$
$$s_n' \equiv \textstyle\sum_{i=1}^{k_n} (s_n^i)'/\sqrt{k_n}$$

where:

$$\begin{aligned} s_n^i &\equiv m(r_n) \times \sqrt{l(n)} \times h_{1n} \times \sqrt{h_{2n}} \\ &\quad \times \{\hat{\lambda}^i_{2,\, m(r_n),\, l(n)}(\mathbf{s}, t) - E[\hat{\lambda}^i_{2,\, m(r_n),\, l(n)}(\mathbf{s}, t)]\} \\ (s_n^i)' &\equiv m(r_n)' \times \sqrt{l(n)'} \times h_{1n} \times \sqrt{h_{2n}} \\ &\quad \times \{\hat{\lambda}^i_{2,\, m(r_n)',\, l(n)'}(\mathbf{s}, t) - E[\hat{\lambda}^i_{2,\, m(r_n)',\, l(n)'}(\mathbf{s}, t)]\}. \end{aligned}$$

Note that the $(s_n^i)'$ have the same marginal distributions as the s_n^i, but the $(s_n^i)'$ are independent of one another. Letting $\phi_n(x)$ and $\phi_n'(x)$ be the characteristic functions of s_n and s_n', respectively, the proof continues with the following steps:

1. $S_n - s_n \xrightarrow{p} 0$

2. $\phi_n'(x) - \phi_n(x) \to 0$

3. $s_n' \xrightarrow{D} N(0, \sigma^2)$.

Proof of 1: We need only to show that $\mathrm{Var}(S_n - s_n) \to 0$. We do this by noting that $\mathrm{Var}(S_n - s_n) = \mathrm{Var}(S_n) + \mathrm{Var}(s_n) - 2 \times \mathrm{Cov}(S_n, s_n)$. First, observe $\mathrm{Var}(S_n) \to \sigma^2$ as $n \to \infty$.

Second, let $D^{m(r_n)',\, l(n)'}$ denote the union of all the $D^i_{m(r_n)',\, l(n)'}$; let $T^{m(r_n)',\, l(n)'}$ denote the union of all the $T^i_{m(r_n)',\, l(n)'}$; and let $A^{m(r_n)',\, l(n)'} = D^{m(r_n)',\, l(n)'} \times T^{m(r_n)',\, l(n)'}$.

Observe that:

$$
\begin{aligned}
s_n &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} m(r_n) \times \sqrt{l(n)} \times h_{1n} \times \sqrt{h_{2n}} \\
&\quad \times \{\hat{\lambda}^i_{2,m(r_n),l(n)}(\mathbf{s},t) - E[\hat{\lambda}^i_{2,m(r_n),l(n)}(\mathbf{s},t)]\} \\
&= \sqrt{\frac{m(r_n) \times m(r_n) \times l(n)}{k_n}} \sum_{i=1}^{k_n} h_{1n} \times \sqrt{h_{2n}} \\
&\quad \times \{\hat{\lambda}^i_{2,m(r_n),l(n)}(\mathbf{s},t) - E[\hat{\lambda}^i_{2,m(r_n),l(n)}(\mathbf{s},t)]\} \\
&= \sqrt{|A^{m(r_n)',\,l(n)'}|} \times h_{1n} \times \sqrt{h_{2n}} \\
&\quad \times \{\hat{\lambda}_{A^{m(r_n)',\,l(n)'}}(\mathbf{s},t) - E[\hat{\lambda}_{A^{m(r_n)',\,l(n)'}}(\mathbf{s},t)]\}
\end{aligned}
$$

Then if we see that that $A^{m(r_n)',l(n)'}$ is the union of a set of disjoint cuboids whose sizes tend to infinity, we notice that it satisfies the third condition of Theorem 1. Assuming again the first two conditions of Theorem 1, then, we see that $\text{Var}(s_n) \to \sigma^2$ by Theorem 1.

Finally, notice that $D^{m(r_n)',\,l(n)} \subset D_n$ and $T^{m(r_n)',\,l(n)} \subset T_n$, and that $|D^{m(r_n)',\,l(n)'}| / |D_n| \to 1$ and $|T^{m(r_n)',\,l(n)'}|/T_n| \to 1$ by Lemma 3; therefore from the proof of Theorem 1, we conclude that $\text{Cov}(S_n, s_n) \to \sigma^2$. Hence $\text{Var}(S_n - s_n) \to 0$, implying that $S_n - s_n \xrightarrow{p} 0$ as $n \to \infty$.

Proof of 2: By an extension of Ibragimov and Linnik's (1971) telescoping argument and because $O(k_n) = O(\frac{r_n^2 \times n}{r_n^{2\alpha} n^{\beta}}) = O(r_n^{2-2\alpha} \times n^{1-\beta})$ and since $\alpha(r_n^2 \times n; r_n^{\eta}; n^{\theta}) \leq r_n^2 n O(r_n^{-\eta\epsilon} n^{-\theta\delta})$, we see

$$
|\phi_n'(x) - \phi_n(x)| \leq 16 k_n O(r_n^{-\epsilon\eta} n^{-\delta\theta}) = O(r_n^{4-2\alpha-\epsilon\eta} n^{2-\beta-\delta\theta}).
$$

Because by assumption $4/(2+\epsilon) < \eta < \alpha < 1$ and $2/(1+\delta) < \theta < \beta < 1$, we know $4 - 2\alpha - \epsilon\eta < 0$ and $2 - \beta - \theta\delta < 0$, so as both exponents are negative, we conclude

$|\phi_n'(x) - \phi_n(x)| \to 0$.

Proof of 3: Recall that the $(s_n^i)'$ are independent and have the same distribution as the s_n^i. Because of this, $\mathrm{Var}[\sum_{i=1}^{k_n}(s_n^i)'] = k_n \mathrm{Var}[(s_n^i)']$. Also, if we define $\sigma_n^2 = \mathrm{Var}[(s_n^i)']$, we note $\sigma_n^2 \to \sigma^2$ from the proof of 1, and from the fact that the $(s_n^i)'$ have the same distribution as the s_n^i. Then because of assumption 2, we have:

$$\begin{aligned}
& \lim_{n\to\infty} \sum_{i=1}^{k_n} \frac{E[|(s_n^i)'|^{2+\delta}]}{\sqrt{\{\mathrm{Var}[\sum_{i=1}^{k_n}(s_n^i)']\}^{2+\delta}}} \\
\leq & \lim_{n\to\infty} \sum_{i=1} k_n \frac{C_\delta}{(k_n\sigma_n^2)^{(2+\delta)/2}} \\
= & \lim_{n\to\infty} C_\gamma \frac{k_n}{(k_n\sigma_n^2)^{(2+\delta)/2}} \\
= & \ 0.
\end{aligned}$$

Therefore, by Lyapounov's theorem, we conclude that

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} (s_n^i)' \xrightarrow{d} N(0, \sigma^2).$$

Finally, notice that the Cramér-Wold device proves the joint normality. ■

Proof of Theorem 3: First we consider the univariate case; then $\mathbf{G}$ and $\hat{\mathbf{G}}_n$ are $\Psi(\mathbf{s}, t)$ and $\hat{\Psi}(\mathbf{s}, t)$, respectively. Then the subsampling estimator becomes

$$\begin{aligned}
\hat{\sigma}^2 \ = \ & \frac{1}{|D_n^{1-c}| \times |T_n^{1-c}|} \int_{T_n^{1-c}} \int_{D_n^{1-c}} |D_{m(r_n)}| \times |T_{l(n)}| \, h_{1,m(r_n)}^2 \, h_{2,l(n)} \\
& \left(\hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}, q) - \bar{\Psi}_{m(r_n),l(n)}\right)^2 \mathrm{d}\mathbf{x} \, \mathrm{d}q.
\end{aligned}$$

We propose to show that $\hat{\sigma}_n^2 \xrightarrow{L_2} \sigma^2$, where $\sigma^2 \equiv \lim_{n\to\infty} |A_n| \times h_{1,n}^2 \times h_{2,n} \times \mathrm{Var}(\hat{\Psi}_n(\mathbf{s}, t))$ and the sample second-order intesity function at lag $(\mathbf{s}, t)$ on $D_{m(r_n)} \times$

$T_{l(n)} + (\mathbf{x}, q)$ is denoted by $\hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}, q)$. Also define

$$\begin{aligned} S_n &\equiv \frac{1}{|D_n^{1-c}| \times |T_n^{1-c}|} \int_{T_n^{1-c}} \int_{D_n^{l-c}} |D_{m(r_n)}| \times |T_{l(n)}| \, h_{1,m(r_n)}^2 \, h_{2,l(n)} \\ &\quad \left(\hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}, q) - \mathrm{E}(\hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}, q))\right)^2 \mathrm{d}\mathbf{x} \, \mathrm{d}q \end{aligned}$$

and

$$\begin{aligned} S_n' &\equiv \frac{1}{|D_n^{1-c}| \times |T_n^{1-c}|} \int_{T_n^{1-c}} \int_{D_n^{1-c}} \sqrt{|D_{m(r_n)}| \times |T_{l(n)}|} \, h_{1,m(r_n)} \sqrt{h_{2,l(n)}} \\ &\quad \left(\hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}, q) - \mathrm{E}(\hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}, q))\right) \mathrm{d}\mathbf{x} \, \mathrm{d}q \end{aligned}$$

Then $\hat{\sigma}^2 = S_n - (S_n')^2$. Therefore it is sufficient to show that (1) $S_n \xrightarrow{L_2} \sigma^2$ and (2) $(S_n')^2 \xrightarrow{L_2} 0$ to see that $\hat{\sigma}_n^2 \xrightarrow{L_2} \sigma^2$.

We first consider S_n: since $\mathrm{E}(S_n) \to \sigma^2$, we only need to show that $\mathrm{Var}(S_n) \to 0$.

$$\begin{aligned} \mathrm{Var}(S_n) &= \frac{1}{|D_n^{1-c}|^2 \times |T_n^{1-c}|^2} \iint\limits_{T_n^{1-c}} \iint\limits_{D_n^{1-c}} \\ &\quad \mathrm{Cov}\Big\{|D_{m(r_n)}||T_{l(n)}|h_{1,m(r_n)}^2 h_{2,l(n)} \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_1, q_1), \\ &\quad |D_{m(r_n)}||T_{l(n)}|h_{1,m(r_n)}^2 h_{2,l(n)} \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_2, q_2)\Big\} \, \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2 \, \mathrm{d}q_1 \, \mathrm{d}q_2 \end{aligned}$$

We define U_n and V_n as follows, where $U_n + V_n = \mathrm{Var}(S_n)$.

$$\begin{aligned} U_n &\equiv \frac{1}{|D_n^{1-c}|^2 \times |T_n^{1-c}|^2} \times \iint\limits_{T_n^{1-c},\, t(q_1,q_2) \le l(n)} \iint\limits_{D_n^{1-c},\, d(\mathbf{x}_1,\mathbf{x}_2) \le m(r_n)} \\ &\quad \mathrm{Cov}\Big\{|D_{m(r_n)}| \, |T_{l(n)}| \, h_{1,m(r_n)}^2 \, h_{2,l(n)} \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_1, q_1), \\ &\quad |D_{m(r_n)}| \, |T_{l(n)}| \, h_{1,m(r_n)}^2 \, h_{2,l(n)} \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_2, q_2)\Big\} \, \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2 \, \mathrm{d}q_1 \, \mathrm{d}q_2, \\ V_n &\equiv \frac{1}{|D_n^{1-c}|^2 \times |T_n^{1-c}|^2} \times \iint\limits_{T_n^{1-c},\, t(q_1,q_2) > l(n)} \iint\limits_{D_n^{1-c},\, d(\mathbf{x}_1,\mathbf{x}_2) > m(r_n)} \\ &\quad \mathrm{Cov}\Big\{|D_{m(r_n)}| \, |T_{l(n)}| \, h_{1,m(r_n)}^2 \, h_{2,l(n)} \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_1, q_1), \\ &\quad |D_{m(r_n)}| \, |T_{l(n)}| \, h_{1,m(r_n)}^2 \, h_{2,l(n)} \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_2, q_2)\Big\} \, \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2 \, \mathrm{d}q_1 \, \mathrm{d}q_2. \end{aligned}$$

$U_n \to 0$ follows from the proof of Theorem 1 in Politis and Sherman (2001): Because $t(q_1, q_2) \le l(n)$ and $d(\mathbf{x}_1, \mathbf{x}_2) \le m(r_n)$, for each $(\mathbf{x}_2, q_2)$, for subshapes of length $c_1 r_n^{\alpha}$ of distance $m(r_n)$ apart in space and of length $c_2 n^{\beta}$ and distance $l(n)$ apart in time, we have that

$$\iint_{t(q_1,q_2) \le l(n),\, d(\mathbf{x}_1,\mathbf{x}_2) \le m(r_n)} \mathrm{d}\mathbf{x}_1 \, \mathrm{d}q_1 \le \left[2m(r_n) + 3c_1 r_n^{\alpha}\right]^d \left[2l(n) + 3c_2 n^{\beta}\right].$$

Therefore,

$$\begin{aligned} U_n &\le \frac{\mathrm{Var}(\hat{\Psi}(\mathbf{x}, q))}{|D_n^{1-c}|^2 |T_n^{1-c}|^2} \iint_{|D_n^{1-c}||T_n^{1-c}|} \iint_{t(q_1,q_2) \le l(n),\, d(\mathbf{x}_1,\mathbf{x}_2) \le m(r_n)} \mathrm{d}\mathbf{x}_1 \, \mathrm{d}q_1 \, \mathrm{d}\mathbf{x}_2 \, \mathrm{d}q_2 \\ &\le \frac{\mathrm{Var}(\hat{\Psi}(\mathbf{x}, q))}{|D_n^{1-c}|^2 |T_n^{1-c}|^2} 5^{d+1} \, c_1^d \, c_2 \, r_n^{d\alpha} \, n^{\beta} \, |D_n^{1-c}| \, |T_n^{1-c}| \\ &\le \frac{\mathrm{Var}(\hat{\Psi}(\mathbf{x}, q))}{|D_n^{1-c}|^2 |T_n^{1-c}|^2} 5^{d+1} \, |D_n| \, |T_n| \, c_1^d \, c_2 \to 0. \end{aligned}$$

To see that $B_n \to 0$, note that for any $(\mathbf{x}_1, q_1)$ and $(\mathbf{x}_2, q_2)$ in the integral defining B_n, we can see an upper bound

$$\begin{aligned} \mathrm{Cov}\Big\{ &|A_{l(n)}| \times h_{1,m(r_n)}^2 \times h_{2,l(n)} \times \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_1, q_1), \\ &|A_{l(n)}| \times h_{1,m(r_n)}^2 \times h_{2,l(n)} \times \hat{\Psi}_{m(r_n),l(n)}(\mathbf{x}_2, q_2) \Big\} \\ &\le C_{\delta} \alpha^{\delta/(2+\delta)}(|A_{l(n)}|; m(r_n); l(n)) \end{aligned}$$

by Minkowski's inequality and condition (2.2). Thus we see that by the third assumption for Theorem 1 and our mixing condition (2.1), $B_n \to 0$. Using the same rationale, $S_n' \xrightarrow{L_2} 0$, and $(S_n')^2 \xrightarrow{L_2} 0$ follows. Therefore $\hat{\sigma_n^2} \xrightarrow{L_2} \sigma^2$.

For the multivariate case, we follow the proof of Theorem 3 in Guan's dissertation [40]. Let $\mathbf{b} \equiv \{b_{\mathbf{t}}, \mathbf{t} \in \Lambda\}$ be a nonzero vector. Also let $S(A_n, \mathbf{b}) \equiv \mathbf{b}' \times (\hat{\mathbf{G}}_n - \mathbf{G})$.

By Minkowski's inequality and condition (2.1), it can be seen that

$$\sup_n E\left[|\sqrt{|A_n|} \times \{S(A_n, \mathbf{b}) - E[S(A_n, \mathbf{b})]\}\,|^{2+\delta}\right] \leq C_\delta$$

for some $\delta > 0$, $C_\delta < \infty$. Then define

$$\begin{aligned} \theta_{\mathbf{b}} &\equiv \lim_{n\to\infty} |A_n| \times \mathrm{Var}(S(A_n, \mathbf{b})) \\ &= \mathbf{b}' \left[\lim_{n\to\infty} |A_n| \mathrm{Cov}(\hat{\mathbf{G}}_n, \hat{\mathbf{G}}_n)\right] \mathbf{b} = \mathbf{b}'\Sigma\mathbf{b}. \end{aligned}$$

Then the subsampling estimator for $\theta_{\mathbf{b}}$ is

$$\hat{\theta}_{\mathbf{b},n} = \sum_{i=1}^{k_n} |D^i_{m(r_n)}||T^i_{l(n)}| \left[S(D^i_{m(r_n)} \times T^i_{l(n)}, \mathbf{b}) - \bar{S}_n\right]^2 / k_n = \mathbf{b}' \,\hat{\Sigma}\, \mathbf{b}.$$

By Politis and Sherman [41], $\hat{\theta}_{\mathbf{b},n} \xrightarrow{L_2} \theta_{\mathbf{b}}$. Thus $\mathbf{b}' \,\hat{\Sigma}\, \mathbf{b} \to \mathbf{b} \xrightarrow{L_2} \mathbf{b}' \,\Sigma\, \mathbf{b}$ for all nonzero $\mathbf{b}$. The L_2 consistency of $\hat{\Sigma}_n$ follows directly. ■

APPENDIX B

TABLES IN CHAPTER III

Table I. Summaries of Explanatory Variables. Explanatory variables used in the logistic and Poisson models. Most of the land parcels classified as water did not hold water for most of the year; some were washes which fill with water only during the monsoon season [29].

Parameter	Mean	Standard Deviation
Percentage Neighbors with bees	2.02%	0.0242
Construction year	1979	18.59
Acreage	0.498	2.59
Distance from "water"	0.270 miles	0.243
Distance from recreational land	0.801 miles	0.875
Distance from natural land	0.397 miles	0.456
Distance from vacant land	0.297 miles	0.396
Distance from agricultural land	2.013 miles	1.319
Distance from transportation land	0.298 miles	0.358
Indicator of pool within 0.5 mile	0.742	0.438
Indicator of residential location	0.828	0.377

Table II. Logistic Model Results. Parameter estimates and corresponding p-values from the logistic regression model. To measure importance, we created column “Factor” to measure the multiplicative factor for the odds when the explanatory variable changes by the amount specified in the column “Change”.

Parameter	Estimate	Standard Error	P-value	Change	Factor
(Intercept)	35.62	3.97	< 0.0001		
Percent neighbors with colonies	24.91	2.35	< 0.0001	2	1.65
Construction year	-0.02027	2.00 e-03	< 0.0001	20	0.67
Land hectares [Ln(acres +0.1)]	0.4403	2.74 e-02	< 0.0001	0.203	1.31
Distance from “water”** (km)	2.000 e-05	1.64 e-05	0.1117	0.8	1.05
Dist from recreational land (km)	9.153 e-06	1.87 e-05	0.3121	0.8	1.02
Dist from natural land (km)	2.411 e-05	1.07 e-05	0.0119	0.8	1.07
Dist from vacant land (km)	-1.075 e-04	1.81 e-05	< 0.0001	0.8	0.75
Dist from agricultural land (km)	-9.247 e-07	4.98 e-06	0.4263	0.8	1.00
Dist from transportation land (km)	1.462 e-05	1.39 e-05	0.1463	0.8	1.04
Indicator of pool within 0.8 km	-6.366 e-03	0.0510	0.4503	1	0.99
Indicator of residential location	1.559 e-01	0.0567	0.0030	1	1.17

Table III. Poisson Model Results. Parameter estimates and corresponding p-values from the Poisson regression model. To measure importance, we created column "Factor" to measure the multiplicative factor for the predicted number of occupancies when the explanatory variable changes by the amount specified in the column "Change".

Parameter	Parameter Estimates	Standard Errors	P-values	Change	Factor
(Intercept)	35.49	5.44	< 0.001		
Percent neighbors with colonies	24.94	3.14	< 0.001	2	1.65
Construction year	-0.02003	2.74 e-03	< 0.001	20	0.67
Land hectares [Ln(acres +0.1)]	0.4898	3.70 e-02	< 0.001	0.203	1.35
Distance from "water"** (km)	1.968 e-05	2.38 e-05	0.4090	0.8	1.05
Dist from recreational land (km)	1.721 e-06	1.99 e-05	0.9312	0.8	1.00
Dist from natural land (km)	2.846 e-05	1.61 e-05	0.0767	0.8	1.08
Dist from vacant land (km)	-1.316 e-04	2.21 e-05	< 0.001	0.8	0.71
Dist from agricultural land (km)	2.410 e-06	7.24 e-06	0.7394	0.8	1.01
Dist from transportation land (km)	2.001 e-05	2.04 e-05	0.3261	0.8	1.05
Indicator of pool within 0.8 km	0.02771	0.0784	0.7237	1	1.03
Indicator of residential location	0.1219	8.33 e-03	< 0.001	1	1.13

VITA

Elizabeth Young Kolodziej recieved her Bachelor of Science degree in statistics from The University of Georgia in 2003. She received a Master of Science from the same school, also in statistics, in 2005, from the Department of Statistics. In the fall of 2005, she entered the doctoral program at Texas A&M University. She received her Doctor of Philosophy from Texas A&M University in August 2010. Elizabeth Kolodziej may be reached at Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, or by email at elizabethkolodziej@gmail.com.