

NONPARAMETRIC METHODS OF ASSESSING SPATIAL ISOTROPY

A Dissertation

by

YONG TAO GUAN

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 2003

Major Subject: Statistics

NONPARAMETRIC METHODS OF ASSESSING SPATIAL ISOTROPY

A Dissertation

by

YONG TAO GUAN

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

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Michael Sherman
(Co-Chair of Committee)

James A. Calvin
(Co-Chair of Committee)

Suojin Wang
(Member)

Sherry I. Bame
(Member)

James A. Calvin
(Head of Department)

August 2003

Major Subject: Statistics

ABSTRACT

Nonparametric Methods of Assessing Spatial Isotropy. (August 2003)

Yong Tao Guan, B.S., Peking University, P.R.China

Co-Chairs of Advisory Committee: Dr. Michael Sherman
Dr. James A. Calvin

A common requirement for spatial analysis is the modeling of the second-order structure. While the assumption of isotropy is often made for this structure, it is not always appropriate. A conventional practice to check for isotropy is to informally assess plots of direction-specific sample second-order properties, e.g., sample variogram or sample second-order intensity function. While a useful diagnostic, these graphical techniques are difficult to assess and open to interpretation. Formal alternatives to graphical diagnostics are valuable, but have been applied to a limited class of models.

In this dissertation, we propose a formal approach testing for isotropy that is both objective and appropriate for a wide class of models. This approach, which is based on the asymptotic joint normality of the sample second-order properties, can be used to compare these properties in multiple directions. An L_2 consistent subsampling estimator for the asymptotic covariance matrix of the sample second-order properties is derived and used to construct the test statistic with a limiting χ^2 distribution under the null hypothesis.

Our testing approach is purely nonparametric and can be applied to both quantitative spatial processes and spatial point processes. For quantitative processes, the results apply to both regularly spaced and irregularly spaced data when the point locations are generated by a homogeneous point process. In addition, the shape of the random field can be quite

irregular. Examples and simulations demonstrate the efficacy of the approach.

To Lian and Michelle

ACKNOWLEDGEMENTS

Having Dr. Sherman and Dr. Calvin as my advisors is something that I will always be proud of and cherish for the rest of my life. I am grateful for their guidance, encouragement, enduring support for my professional development, and for the most part, their friendship. Without their help, I would have never been able to accomplish as much as what I have achieved. Thank you!

Thanks extend to my committee members, Dr. Suojin Wang and Dr. Sherry Bame, for their willingness to serve on my committee, and to Dr. Daniel Sui for serving as a substitute for my defense. I also want to thank Dr. Dahm for his effort recruiting me to this department five years ago and for his trust and friendship over the years.

A special thanks is owed to my parents. In a situation where even having some meat once in a while was considered as a luxury expense, they managed to pay for my tuition and fees from elementary school through college. For this I want to thank them. I am also very grateful for my parents-in-law's support both spiritually and financially when we needed it the most. Our life would not have been as enjoyable as it was without their support.

I don't know how to start to thank Lian since no word would fully describe my emotions. It was still like yesterday when I picked her up at the airport and we were all in tears. Since that moment, my life has been full of sunshine. I appreciate the love, joy, hope and comfort that she has brought to my life and the sacrifice she has made for being with me; I also admire her courage, strength and determination while struggling through all the up and downs together with me. Very often I ask myself why I am so lucky to have such a wonderful wife. Although I don't know what the answer is, I do know that I will always love her. Honey, this is for you!

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
I INTRODUCTION	1
1.1 Quantitative Spatial Processes	1
1.2 Spatial Point Processes	4
1.3 Overview Structure	6
II A TEST OF ISOTROPY FOR QUANTITATIVE SPATIAL PRO- CESSES	7
2.1 Introduction	7
2.2 Definitions and Asymptotic Results	9
2.3 Assessment of Isotropy	15
2.4 A Simulation Study	20
2.5 Applications	25
III A TEST OF ISOTROPY FOR SPATIAL POINT PROCESSES	28
3.1 Introduction	28
3.2 Definitions and Asymptotic Results	30
3.3 Assessment of Isotropy	35

CHAPTER	Page
3.4 A Simulation Study	39
3.5 Applications	43
IV A TEST OF ISOTROPY FOR MARKED-POINT PROCESSES . . .	47
4.1 Introduction	47
4.2 Preliminary Asymptotic Results	49
4.3 A Cross-validation Approach for Bandwidth Selection	53
4.4 A Test for Isotropy	55
4.5 An Application	59
V CONCLUSIONS AND FUTURE RESEARCH	62
REFERENCES	66
APPENDIX A	70
APPENDIX B	89
APPENDIX C	101
VITA	106

LIST OF TABLES

TABLE	Page
1 Simulation results for regularly spaced observations. Each table entry is the percentage of rejections at the 5% nominal level from ten thousand simulations	23
2 Simulation results for irregularly spaced observations. Each entry is the percentage of rejections at the 5% nominal level from one thousand simulations	24
3 Simulation results from one thousand realizations of Poisson clustering processes. Each entry is the percentage of rejections at the 5% nominal level. ρ is the expected number of parents; μ is the expected number of offspring per parent.	42
4 Simulation results from five thousand realizations of marked Poisson cluster processes. Each entry is the percentage of rejections at the 5% nominal level. σ is a spread parameter for the radially symmetric normal distribution; m is a range parameter that defines the correlation strength among marks.	59

LIST OF FIGURES

FIGURE	Page
1 Sample Variograms in Two Directions for the Wind-Speed Data. \times : E-W direction, \circ : N-S direction.	25
2 K Function Plot. The dotted line is the estimated K function assuming isotropy, the solid lines are the upper and lower envelopes from 100 simulations from the best fitting isotropic model.	29
3 Locations of Aggravated Assault in an Area of Downtown Houston.	43
4 Locations of Leukemia Cases and Controls. The first graph plots locations of the cases, while the second plots those of the controls.	45
5 Isotropic Sample Second-order Intensity Function for Leukemia Data. \times : cases, \circ : controls.	46
6 Plot of Cross-validation Sample Mean Squared Errors for Different Bandwidth Balues. The cross-validation procedure is introduced in Section 4.3.	61
7 Partition of the Field for Lemma A.3	72
8 Partition of the Field for Lemma A.4. \times marks boundary subsquares.	77
9 Partition of the Field for Theorem II.1	80

CHAPTER I

INTRODUCTION

Spatial statistics is concerned with data that are observed in a two or higher dimensional space, where observations can be either quantitative measurements recorded at various locations or simply the spatial locations themselves. Based on this very fact, spatial processes can be split into two major branches: quantitative spatial processes, which deal with quantitative spatial observations, and spatial point processes, where the locations of events are the primary interest of study. A unifying characterization of a spatial process is through its second-order characteristics, which are often expressed as functions of relative locations of two observations. In this dissertation, we consider a second-order property, isotropy, for both the quantitative spatial processes and spatial point processes.

1.1 Quantitative Spatial Processes

Quantitative spatial statistics is often originally associated with geostatistics, where Matheron (1963) developed tools to predict ore reserves in a region. Geostatistics emerged in the early 1980s and has seen rapid growth over the past twenty years. Successful applications abound beyond geostatistics in, for example, rainfall data (Ord and Rees 1979), groundwater research (Myers et al. 1982), ozone exposure study (Carroll et al. 1997), and wind-speed prediction (Cressie and Huang 1999).

A distinct characteristic of spatial data, in contrast to independent and identical observations, is that spatially close observations are often correlated. A commonly made

The format and style follow that of the *Journal of the American Statistical Association*.

assumption while modeling the correlation structure is that of second-order stationarity. Consider a spatial process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$, where \mathbf{s} denotes the locations where $Z(\cdot)$ is observed. $Z(\cdot)$ is said to be second-order stationary if

$$E[Z(\mathbf{s})] = \mu \quad \forall \mathbf{s} \in \mathbb{R}^2 \quad \text{and} \quad (1.1)$$

$$\text{Cov}[Z(\mathbf{s} + \mathbf{t}), Z(\mathbf{s})] = C(\mathbf{t}) \quad \forall \mathbf{s}, \mathbf{t} \in \mathbb{R}^2. \quad (1.2)$$

Thus a second-order stationary process has a constant mean structure and the covariance between two observations depends only on their relative locations. The function $C(\cdot)$ defined by (1.2) is known as the covariance function.

A weaker condition than second-order stationarity is called intrinsic stationarity, which specifies that (1.1) holds and that

$$\text{Var}[Z(\mathbf{s} + \mathbf{t}) - Z(\mathbf{s})] = \gamma(\mathbf{t}) \quad \forall \mathbf{s}, \mathbf{t} \in \mathbb{R}^2, \quad (1.3)$$

Spatial processes that are not second-order stationary, e.g., Brownian motions, can still be intrinsically stationary. The function $\gamma(\cdot)$ in (1.3) is called the variogram function. The variogram function is often used in place of the usual covariance function due to its greater generality and the fact that it can be estimated more accurately for a variety of data structures (Cressie, 1991). Observe that if a process is second-order stationary, then $\gamma(\mathbf{t}) = 2C(\mathbf{0}) - 2C(\mathbf{t})$.

The variogram function (or the covariance function) plays an important role in many spatial analyses, particularly in spatial prediction (e.g., kriging). In the wind-speed example detailed in Chapter 2, wind speeds are measured at 289 different locations over a region in the western tropical Pacific Ocean. The goal is to predict the wind speed at a new location in the region, say \mathbf{s}_0 . Assume that wind speeds are generated by an intrinsically stationary process. Then the optimal linear predictor is given in terms of $\gamma(\cdot)$ as (see, e.g., Cressie

1991)

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i),$$

where

$$\begin{aligned} (\lambda_1, \dots, \lambda_n, m)' &= \mathbf{\Gamma}^{-1} \gamma_0, \\ \gamma_0 &\equiv (\gamma(\mathbf{s}_0 - \mathbf{s}_1), \dots, \gamma(\mathbf{s}_0 - \mathbf{s}_n), 1)', \\ \mathbf{\Gamma}_{i,j} &= \begin{cases} \gamma(\mathbf{s}_i - \mathbf{s}_j) & \text{if } i, j = 1, \dots, n, \\ 1 & \text{if } i = n+1, j = 1, \dots, n, \\ 1 & \text{if } j = n+1, i = 1, \dots, n, \\ 0 & \text{if } i = n+1, j = n+1. \end{cases} \end{aligned}$$

and m denote a Lagrange multiplier that ensures $\sum_{i=1}^n \lambda_i = 1$.

Typically the variogram function is unknown and must be estimated. A common practice is to estimate the variogram at a set of observed lags and fit a parametric model, which ensures that the estimated variogram function satisfies a “conditionally negative-definite” property (see Cressie 1991 for a detailed discussion of this property). For example, let $\|\mathbf{t}\| \equiv (\mathbf{t}'\mathbf{t})^{1/2}$ where \mathbf{t} denotes an arbitrary lag, an *Exponential* variogram model is specified as

$$\gamma(\mathbf{t}; \theta) = \theta_1 \{1 - \exp(-\theta_2 \|\mathbf{t}\|)\}. \quad (1.4)$$

In (1.4), $\gamma(\mathbf{t}; \theta)$ is a function of the (Euclidean) length of \mathbf{t} but not of its direction. This direction invariant property is known as isotropy. A spatial process is said to be isotropic if $\gamma(\mathbf{t}_1) = \gamma(\mathbf{t}_2)$ for any \mathbf{t}_1 and \mathbf{t}_2 such that $\|\mathbf{t}_1\| = \|\mathbf{t}_2\|$. Otherwise, it is called anisotropic. In practice, isotropy is often assumed due to its simpler interpretation and ease of estimation. Not assessing the validity of this assumption, however, may lead to inefficient spatial inferences. In an example detailed in Chapter 2, it can be seen that misspecification of an anisotropic model as isotropic can lead to noticeably less efficient spatial prediction.

Chapter 2 and Chapter 4 of this dissertation are devoted to developing a formal approach to test for isotropy for quantitative spatial processes. Chapter 2 considers regularly spaced data and irregularly spaced data whose locations can be modeled by a marked-Poisson process. Chapter 4 extends the results to locations from a more general marked-point process. Our approach is purely nonparametric and can be applied in a variety of settings.

1.2 Spatial Point Processes

The origin of the use of spatial point processes can be traced back to the early development of counting problems, which deal with the numbers of events in intervals or regions of various types. A simple, but important example, among these is the Poisson process, which was first given by Poisson (1837). Spatial point processes have been undergoing rapid development in the last quarter century due to newly discovered applications in many fields (e.g., astrophysics, biology, epidemiology, forestry and image processing). For additional information, see Diggle (1983) who discusses many applications, and Daley and Vere-Jones (2002) who provide a balance between theory and applications.

As with quantitative spatial processes, the analysis of spatial point processes often assumes certain stationarity conditions on the underlying process, among which the most frequently used, perhaps, is second-order stationarity. Consider a spatial point process N . Let dx be an infinitesimal region which contains the point \mathbf{x} , $|dx|$ denote the area of dx , and $N(dx)$ be the random number of points in dx . N is said to be second-order stationary if

$$\lim_{|dx| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(dx)]}{|dx|} \right\} = \nu \quad \forall \mathbf{x} \in \mathbb{R}^2, \quad (1.5)$$

$$\lim_{|dx_1|, |dx_2| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(dx_1) \times N(dx_2)]}{|dx_1| \times |dx_2|} \right\} = \Psi(\mathbf{x}_1 - \mathbf{x}_2) \quad \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2. \quad (1.6)$$

The functions defined by (1.5) and (1.6) are known in point process literature as the first-

and second-order intensity function, respectively. A spatial point process is further said to be isotropic if $\Psi(\mathbf{t}) = \Psi_0(\|\mathbf{t}\|)$ for some function $\Psi_0(\cdot)$, where \mathbf{t} denotes the lag between two points \mathbf{x}_1 and \mathbf{x}_2 , and $\|\mathbf{t}\|$ denotes its Euclidean length. When a spatial point process is isotropic, an alternative approach to characterizing its second-order property is to use the K function. As in Diggle (1983), we define the K function to be

$$K(t) = \nu^{-1} \text{E}(\text{number of further events within distance } t \text{ of an arbitrary event}).$$

A primary interest in analyzing spatial point processes is to model the spatial distribution of the event locations, which can often be achieved by fitting a parametric model (see, Diggle 1983 for more discussions). In the leukemia data detailed in Chapter 3, a cluster model seems to be plausible to describe the locations of leukemia patients, where each residential area can be viewed as a cluster and patients living in a particular area can be viewed as members of the cluster to which that area corresponds. Let θ be a vector of parameters that defines a class of such cluster models, $K(t; \theta)$ be the theoretical K function and $\hat{K}(t)$ be a sample estimate of $K(t; \theta)$ (see, e.g., Diggle 1983). The estimate for θ is given as the value that minimizes the following criterion

$$D(\theta) = \int_0^{t_0} \{[\hat{K}(t)]^c - [K(t; \theta)]^c\} dt,$$

where t_0 and c are two pre-selected “tuning constants” (Diggle 1983).

The above modeling procedure is built upon the assumption that N is isotropic. When a process is not isotropic, this approach is no longer applicable. Thus, assessing isotropy is an important part of the model building process.

Chapter 3 of this dissertation develops a formal approach to test for isotropy in stationary spatial point processes. This approach, which is based on the asymptotic joint normality of the sample second-order intensity function, can be used to compare intensities in multiple directions. The testing approach requires minimal conditions on the underlying process and can be applied in a variety of settings.

1.3 Overview Structure

As previously stated, Chapter 2 and Chapter 4 develop tests for isotropy for quantitative spatial processes, while Chapter 3 develops a test for isotropy for spatial point processes. Applications to four real data examples, wind speed over a region in the western Pacific Ocean, longleaf pine data from Cressie (1991), crime locations in an area of downtown Houston, and leukemia patient locations in the Houston area, are used as illustrative examples for our testing methods. The first two examples are quantitative spatial processes while the latter two are spatial point processes. Simulation results under a variety of settings are presented in these chapters to demonstrate the efficacy of the proposed approaches. Chapter 5 gives final conclusions and considers a future research topic of fitting anisotropic variogram models. Lemmas and proofs of the theorems are detailed in the appendices.

CHAPTER II

A TEST OF ISOTROPY FOR QUANTITATIVE SPATIAL PROCESSES

2.1 Introduction

Data that are spatially close are often correlated. For example, crop yields on adjacent plots in an agricultural study tend to be similar (i.e., positively correlated) because of similar soil types and watering conditions; wind speeds from nearby sites at the same point in time are often alike due to atmospheric circulation and similar geographical conditions. A correlation structure is called isotropic if the correlation between observations at any two sites depends only on the distance between those sites and not on their relative orientation. When a correlation structure is not isotropic, it is said to be anisotropic. The assumption of isotropy is often made in practice due to simpler interpretation and ease of computation.

In many applications, however, isotropy may not be a reasonable assumption. For example, wind plays an important role in transporting pollens for some crops and thus in determining crop yields. Correlation of crop yields might be stronger in the major wind direction than perpendicular to that direction. The correct specification of the correlation structure has two primary benefits. The first is improved scientific interpretation of the model. The second is improved accuracy of prediction. For example, suppose four observations have been made from a stationary random field satisfying the following anisotropic covariance structure:

$$C(x, y) = \exp\{-x^2 - 4y^2\},$$

where $C(x, y)$ denotes the covariance between two observations separated by lag (x, y) . If the goal is to predict the value at an observation equally distant ($=.65$) from the four observed values, then treating the process as isotropic will lead to assigning equal weights

(=0.25 each) to each of the four observations. This choice of weights clearly fails to account for the fact that the correlation in the x direction is much stronger than that in the y direction. If the correct correlation structure is used, then the two extreme points in the x direction receive much greater weight (=0.4997) than the extreme points in the y direction (=0.0003). As a consequence, the isotropic model prediction variance (=0.4938) is appreciably larger than that of the anisotropic model (=0.2814).

Many spatial models (e.g., Diggle 1981, Cressie 1991) utilize isotropy. However, it is clear from the previous example that checking for isotropy before fitting a model is an important part of model building. If the assumption of isotropy is rejected, modification of the model may be necessary with the introduction of an anisotropic model, e.g., Cressie (1991), Zimmerman (1993).

A conventional practice to check for isotropy is to informally assess plots of direction-specific sample (semi)variograms. For example, Isaaks and Srivastava (1989) discuss the use of a rose diagram in detecting anisotropy; Diggle (1981) draws contour plots of the empirical correlation function for a binary mosaic and concludes isotropy. Graphical diagnostics are often difficult to assess and are open to interpretation. In response to this, more formal testing procedures for investigating isotropy have been proposed by Baczkowski and Mardia (1990) and Lu and Zimmerman (2001). Baczkowski and Mardia's method is designed for the specific "doubly geometric" covariance model, and is not appropriate for other covariance models. Lu and Zimmerman's approach is more general and performs well for equally spaced Gaussian processes.

In this chapter, we propose a formal approach to test for isotropy which is both objective and appropriate for a wide class of models. This approach, like that of Lu and Zimmerman, is based on the asymptotic joint normality of the sample variograms and can be used to compare sample variograms in multiple directions. An L_2 consistent subsampling estimator for the asymptotic covariance matrix of the sample variograms is derived

and used to construct a test statistic with a limiting χ^2 distribution under the null hypothesis. Our testing approach is purely nonparametric in that it only requires some mild moment conditions and a weak dependence assumption for the underlying process. In addition, the shape containing the data locations can be quite irregular, and the results apply to both regularly spaced and irregularly spaced data when the point locations are generated by a homogeneous Poisson process.

The rest of the chapter is organized as follows. Section 2.2 demonstrates the asymptotic joint normality of the sample variograms. Section 2.3 illustrates how to use these results to assess the assumption of isotropy. Section 2.4 presents a simulation study conducted to evaluate the performance of our testing approach, while Section 2.5 presents applications to two data sets.

2.2 Definitions and Asymptotic Results

2.2.1 Regularly Spaced Data Case

Consider a strictly stationary random field $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$. Let D be a finite set of lattice points in \mathbb{Z}^2 at which observations are taken. Define the variogram function as

$$\gamma(\mathbf{t}) \equiv \text{Var}\{Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t})\},$$

where $\mathbf{t} [\equiv (t_x, t_y)]$ is an arbitrary lag in \mathbb{R}^2 .

We consider the variogram function in place of the usual covariance function due to its greater generality and the fact that it can be estimated more accurately for a variety of data structures (Cressie, 1991). Note that if two lags have the same length, i.e., they have the same Euclidean distance from the origin, the corresponding values of $\gamma(\mathbf{t})$ will be the same under isotropy. This suggests a test for isotropy may be obtained by comparing variograms at lags with the same length but in different directions. In practice, since the true variograms are typically unknown, we form a test based on estimators of the variograms. The classical

estimator of the variogram, i.e., the sample variogram, is given by

$$\hat{\gamma}(\mathbf{t}) \equiv \frac{1}{|D(\mathbf{t})|} \times \sum \{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2,$$

where the sum is over $D(\mathbf{t}) \equiv \{(i, j) : \mathbf{s}_i, \mathbf{s}_j \in D, \mathbf{s}_i - \mathbf{s}_j = \mathbf{t}\}$ and $|D(\mathbf{t})|$ is the number of distinct elements in $D(\mathbf{t})$.

Let $\mathbf{\Lambda}$ be a set of lags for which we want to calculate and compare the sample variograms. Define $\mathbf{G} \equiv \{\gamma(\mathbf{t}) : \mathbf{t} \in \mathbf{\Lambda}\}$ to be the vector of variograms at lags in $\mathbf{\Lambda}$. Consider a sequence $\{Z(\mathbf{s}) : \mathbf{s} \in D_n\}$ and let $\hat{\gamma}_n(\mathbf{t})$ and $\hat{\mathbf{G}}_n \equiv \{\hat{\gamma}_n(\mathbf{t}) : \mathbf{t} \in \mathbf{\Lambda}\}$ be the estimators of $\gamma(\mathbf{t})$ and \mathbf{G} obtained over D_n , respectively.

To formally state the large sample properties of $\hat{\mathbf{G}}_n$, we need to quantify the strength of dependence in the random field. We do this using a model free mixing condition. Following Rosenblatt (1956), we make use of a particular type of strong mixing coefficients defined by

$$\alpha_p(k) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), |E_i| \leq p, i = 1, 2, d(E_1, E_2) \geq k\},$$

where $|E|$ is the cardinality of the index set E , $\mathcal{F}(E)$ is the σ -algebra generated by the random variables $\{Z(\mathbf{s}) : \mathbf{s} \in E\}$ and $d(E_1, E_2)$ is the minimal ‘‘city block’’ distance between E_1 and E_2 .

If the observations are independent, then $\alpha_p(k) = 0$ for all $k \geq 1$. Here we will need $\alpha_p(k)$ to approach 0 for large k , at some rate depending on the cardinality p . Following Sherman and Carlstein (1994), we assume the following mixing condition

$$\sup_p \frac{\alpha_p(k)}{p} = O(k^{-\epsilon}) \text{ for some } \epsilon > 2. \quad (2.1)$$

Condition (2.1) says that at a fixed distance k , as the cardinality increases, we allow dependence to increase at a rate controlled by p . As the distance increases, the dependence must decrease at a polynomial rate in k . Examples of spatial processes satisfying this condition

can be found in, e.g., Sherman and Carlstein (1994) and Sherman (1996). See also Bradley (1993) for the importance of accounting for cardinality p .

We need to account for the shape of the field from which we observe data. As in Bolthausen (1982), we define the boundary of a set D to be the set $\partial D \equiv \{\mathbf{s} \in D : \text{there exists } \mathbf{s}' \notin D \text{ with } d(\mathbf{s}, \mathbf{s}') = 1\}$, where $d[(s_x, s_y), (s'_x, s'_y)] \equiv \max(|s_x - s'_x|, |s_y - s'_y|)$. Let $|\partial D|$ denote the number of points in ∂D . Assume

$$|D_n| = O(n^2) \text{ and } |\partial D_n| = O(n). \quad (2.2)$$

Condition (2.2) is satisfied by many commonly encountered field sequences. For example, let $A \subset (0, 1] \times (0, 1]$ be the interior of a simple closed curve which is of finite length. Now multiply the set A by n , to obtain the set $A_n \subset (0, n] \times (0, n]$; that is, A_n is the shape A inflated by a factor n . Define $D_n \equiv \{\mathbf{s} : \mathbf{s} \in A_n \cap \mathbb{Z}^2\}$. Then D_n satisfies condition (2.2). This formulation allows for a wide variety of shapes on which the data can be observed, including squares, rectangles, circles and starshapes.

Finally we require the following mild moment condition

$$\sup_n \mathbf{E} \left\{ \left| \sqrt{|D_n|} \times [\hat{\gamma}_n(\mathbf{t}) - \gamma(\mathbf{t})] \right|^{2+\delta} \right\} \leq C_\delta \text{ for some } \delta > 0, C_\delta < \infty. \quad (2.3)$$

Condition (2.3) is only slightly stronger than the existence of the (standardized) asymptotic variance of $\hat{\gamma}_n(\mathbf{t})$. If the random field is m -dependent, it can be shown that the finiteness of $\mathbf{E}(|Z(\mathbf{s})|^{4+2\delta})$ will be sufficient for condition (2.3) to hold.

Theorem II.1. Let $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ be a strictly stationary random field which is observed at lattice points in $D_n \subset \mathbb{Z}^2$ satisfying condition (2.2). Assume

$$\sum_{\mathbf{s} \in \mathbb{Z}^2} \left| \text{Cov} \left\{ [Z(\mathbf{0}) - Z(\mathbf{s}_1)]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{s}_2)]^2 \right\} \right| < \infty \text{ for all finite } \mathbf{s}_1, \mathbf{s}_2. \quad (2.4)$$

Then $\Sigma_{\mathbf{R}} \equiv \lim_{n \rightarrow \infty} |D_n| \times \text{Cov}(\hat{\mathbf{G}}_n, \hat{\mathbf{G}}_n)$ exists, the (i, j) th element of which is

$$\sum_{\mathbf{s} \in \mathbb{Z}^2} \text{Cov} \left\{ [Z(\mathbf{0}) - Z(\mathbf{t}_i)]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t}_j)]^2 \right\}.$$

If we further assume that $\Sigma_{\mathbf{R}}$ is positive definite and that conditions (2.1) and (2.3) hold, then the limiting distribution of $\sqrt{|D_n|} \times (\hat{\mathbf{G}}_n - \mathbf{G})$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathbf{R}}$.

Proof. See Appendix A. □

Condition (2.4) is analogous to the condition $\sum_{\mathbf{s} \in \mathbb{Z}^2} |\text{Cov}\{Z(\mathbf{0}), Z(\mathbf{s})\}| < \infty$, which is often assumed in deriving the asymptotic normality of the univariate sample mean (e.g., Bolthausen, 1982). Any process that is m -dependent with finite fourth moment satisfies this condition. For a Gaussian process, it can be shown that the absolute integrability of its covariance function, i.e., $\int_{\mathbb{R}^2} |R(\mathbf{t})| d\mathbf{t} < \infty$, where $R(\mathbf{t}) \equiv \text{Cov}(Z(\mathbf{0}), Z(\mathbf{t}))$, is sufficient for (2.4) to hold. Many covariance models, e.g., Exponential, Gaussian, Spherical models, can be shown to satisfy the integrability condition and thus satisfy (2.4).

2.2.2 Irregularly Spaced Data Case

Consider a strictly stationary random field $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$. Let $D \subset \mathbb{R}^2$ be the domain of interest in which observations are taken. We view the points at which $Z(\cdot)$ is observed as *random* in number and location; specifically they are generated from a homogeneous two-dimensional Poisson process with intensity parameter ν . Karr (1986) makes a strong case for the plausibility of the Poisson assumption for many practical situations, e.g., data arising from meteorological studies and geological explorations.

In what follows, denote the random point process by N and the random number points of N contained in B by $N(B)$, where B is any given Borel set. We further assume N to be independent of $Z(\cdot)$. To construct a test statistic, an estimate of the variogram is needed. Here we consider one based on kernel smoothing.

In an adaption of the notation in Section 2.2.1, let $|D|$ denote the volume (not the cardinality) of D , ∂D to denote the boundary of D and $|\partial D|$ to denote the length (not the

number of points) of ∂D . Let h be a positive constant and $w(\cdot)$ be a bounded, nonnegative, isotropic density function which takes positive values only on a finite support, C . Here and henceforth, we use $d\mathbf{x}$ to denote an infinitesimally small disc centered at \mathbf{x} . Define $N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) \equiv N(d\mathbf{x}_1)N(d\mathbf{x}_2)I(\mathbf{x}_1 \neq \mathbf{x}_2)$, where $I(\mathbf{x}_1 \neq \mathbf{x}_2) = 1$ if $\mathbf{x}_1 \neq \mathbf{x}_2$ and 0 otherwise. The kernel variogram estimator is given by

$$\hat{\gamma}(\mathbf{t}) = \frac{1}{\nu^2} \int_{\mathbf{x}_1 \in D} \int_{\mathbf{x}_2 \in D} h^{-2} w\left(\frac{\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2}{h}\right) \times \frac{[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2}{|D \cap (D - \mathbf{x}_1 + \mathbf{x}_2)|} N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2).$$

In practice, ν is usually replaced by $N(D)/|D|$, which is a consistent estimator of ν (Stoyan and Stoyan, 1994). We adopt the definitions of $\mathbf{\Lambda}$, \mathbf{G} , $\hat{\gamma}_n(\mathbf{t})$ and $\hat{\mathbf{G}}_n$ in Section 2.2.1, with the understanding that the variogram estimator is now defined by the kernel estimator.

To account for dependence, we modify the mixing condition introduced in Section 2.2.1. Following Politis et al. (1998), we make use of a particular type of strong mixing coefficients defined by

$$\alpha_p(k) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}(E_1), A_2 \in \mathcal{F}(E_2), \\ E_2 = E_1 + \mathbf{s}, \lambda(E_1) = \lambda(E_2) \leq p, d(E_1, E_2) \geq k\},$$

where the supremum is taken over all compact and convex subsets $E_1 \subset \mathbb{R}^2$, and over all $\mathbf{s} \in \mathbb{R}^2$ such that $d(E_1, E_2) \geq k$; in the above, $\mathcal{F}(E)$ denotes the σ -algebra generated by the random variables $\{Z(\mathbf{s}) : \mathbf{s} \in E\}$. We, again, assume the following mixing condition

$$\sup_p \frac{\alpha_p(k)}{p} = \mathbf{O}(k^{-\epsilon}) \text{ for some } \epsilon > 2. \quad (2.1')$$

We also need to account for the shape of the random field and the choice of bandwidth. Consider a sequence of random fields, D_n , and a sequence of bandwidths, h_n . Assume

$$|D_n| = \mathbf{O}(n^2), |\partial D_n| = \mathbf{O}(n), \text{ and } h_n = \mathbf{O}(n^{-\beta}) \text{ for some } \beta \in (0, 1). \quad (2.2')$$

Let $\gamma^{(4)}(\mathbf{t}) \equiv \mathbf{E}\{[Z(\mathbf{t}) - Z(\mathbf{0})]^4\}$. We require the following moment conditions:

$$\gamma(\mathbf{t}), \gamma^{(4)}(\mathbf{t}) \text{ are bounded and continuous, and} \quad (2.3'.a)$$

$$\sup_n \mathbb{E} \left\{ \left| \sqrt{|D_n|} \times h_n \times [\hat{\gamma}_n(\mathbf{t}) - \mathbb{E}(\hat{\gamma}_n(\mathbf{t}))] \right|^{2+\delta} \right\} \leq C_\delta \text{ for some } \delta > 0, C_\delta < \infty. \quad (2.3'.b)$$

The following theorem states that $\hat{\gamma}_n(\mathbf{t})$ is a consistent estimator for $\gamma(\mathbf{t})$ and $\hat{\mathbf{G}}_n$ is asymptotically jointly normal under some mild conditions.

Theorem II.2. Let $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ be a strictly stationary random field observed on a general shaped field D_n , where the points at which $Z(\cdot)$ is observed are generated by a homogeneous Poisson process. Assume condition (2.2'), (2.3'.a) and

$$\int_{\mathbf{s} \in \mathbb{R}^2} \left| \text{Cov} \left\{ [Z(\mathbf{0}) - Z(\mathbf{s}_1)]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{s}_2)]^2 \right\} \right| d\mathbf{s} < \infty \text{ for all finite } \mathbf{s}_1, \mathbf{s}_2, \quad (2.4')$$

then $\mathbb{E}[\hat{\gamma}_n(\mathbf{t})] \rightarrow \gamma(\mathbf{t})$ and

$$\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times \text{Cov}(\hat{\gamma}_n(\mathbf{t}), \hat{\gamma}_n(\mathbf{t}')) = \int_C w(\mathbf{s})^2 d\mathbf{s} \times \gamma^{(4)}(\mathbf{t}) \times I(\mathbf{t} = \pm \mathbf{t}'),$$

where $I(\mathbf{t} = \pm \mathbf{t}') = 1$ if $\mathbf{t} = \pm \mathbf{t}'$ and 0 otherwise. If we further assume $\gamma^{(4)}(\mathbf{t}) > 0$, conditions (2.1') and (2.3'.b) hold, then $\sqrt{|D_n|} \times h_n \times \{\hat{\mathbf{G}}_n - \mathbb{E}(\hat{\mathbf{G}}_n)\}$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathbb{R}}$ the structure of which is given by the above expression.

Proof. See Appendix A. □

An interesting property from the above theorem is that $\hat{\gamma}_n(\mathbf{t})$ and $\hat{\gamma}_n(\mathbf{t}')$, for $\mathbf{t} \neq \pm \mathbf{t}'$, are asymptotically uncorrelated. Observe that for large n and small h_n , very few data points relative to the total number of points on D_n will be used to calculate the sample variogram at a given lag. Because of the randomness of the point process, the chance that the same or even nearby data points are used to calculate $\hat{\gamma}_n(\mathbf{t})$ and $\hat{\gamma}_n(\mathbf{t}')$, where $\mathbf{t} \neq \pm \mathbf{t}'$, becomes small as n becomes large. Thus $\hat{\gamma}_n(\mathbf{t})$ and $\hat{\gamma}_n(\mathbf{t}')$ tend to be uncorrelated due to the assumption of weak dependence on the process.

2.3 Assessment of Isotropy

2.3.1 Choice of Λ

To assess the hypothesis of isotropy, the lag set, Λ , needs to be specified. The choice of Λ depends on a number of factors, including the configuration of a data set, the goal of the study, and the underlying physical/biological phenomenon of interest. There is no unique rule for choosing Λ . Generally speaking, smaller lags are preferable to larger ones. This is due mainly to the following two facts: sample variograms at larger lags are based on fewer observations than estimates at smaller lags and therefore are more variable; and observations at smaller lags are usually more correlated and identification of the correlation among these lags is more important for spatial prediction.

For regularly spaced observations, note that the two components of any lag \mathbf{t} (i.e., t_x and t_y) should both be integers so that $\hat{\gamma}(\mathbf{t})$ can be calculated. Thus only lags with integer components can be included in Λ (e.g., $\Lambda = \{(1, 0), (0, 1), (1, 1), (-1, 1)\}$). For irregularly spaced observations, sample variograms at more lags can be calculated and thus more options for Λ are available. The choice of Λ should mainly rely on the knowledge of the underlying physical/biological process generating the observations. For example, wind might be a suspected source of anisotropy in an air pollution study. Thus a natural choice for Λ is to include lags in the major wind direction and those perpendicular to that direction. In the long-leaf pine data example which is discussed in Section 5, 584 unequally spaced trees were observed in a $200m \times 200m$ square field. The dbh (diameter at breast height) for each of them was measured. A 24×24 grid may be laid over the field such that approximately one point can be assigned to the nearest node of the grid. This gives the distance between adjacent nodes approximately as eight meters, which is a reasonable reflection of distances among neighboring trees. The test for isotropy can be performed by comparing the sample variograms at lags whose lengths are eight meters. This may provide

us with some insights into how environmental conditions affect the growth of trees and how they compete with each other.

2.3.2 Estimating the Covariance Matrix

To formally compare the directional sample variograms, we must take the sampling variation as well as the correlation among them into consideration, i.e., we need the knowledge of $\Sigma_{\mathbf{R}}$ (for regularly spaced data) or $\Sigma_{\mathbf{IR}}$ (for irregularly spaced data). These covariance matrices are usually unknown and thus need to be estimated. While the expression for $\Sigma_{\mathbf{IR}}$ in Section 2.2.2 suggests a plug-in method might be available for irregularly spaced data, it is ambitious to do so for regularly spaced data due to the large number of elements in $\Sigma_{\mathbf{R}}$ and the complex structure for each of them (see Section 2.2.1).

2.3.2.1 Regularly spaced data case

To estimate $\Sigma_{\mathbf{R}}$, we apply a subsampling technique. Subsampling has been widely used to estimate the variance of a general spatial statistic (e.g., Hall 1988, Possolo 1991, Sherman and Carlstein 1994). Heagerty and Lumley (2000) used a subsampling method to estimate the covariance matrix of estimators derived from estimating functions. Here we apply subsampling to estimate the covariance matrix of $\hat{\mathbf{G}}_{\mathbf{n}}$.

Toward this end, we divide the original field D_n into overlapping subblocks. These subblocks are obtained by moving a subsampling window across D_n . If the window is fully contained in D_n , then a subblock is obtained. The subsampling window is chosen to be congruent to D_n both in configuration and orientation (so as to retain the same dependence structure as the original data) but is much smaller than D_n . In practice, we set its cardinality to be of order l_n^2 , where $l(n) = cn^\alpha$ for some $c > 0$ and $\alpha \in (0, 1)$. In what follows, we denote the total number of subblocks by k_n , the i th subblock by $D_{l(n)}^i$, and the vector of

sample variograms calculated on $D_{l(n)}^i$ by $\hat{\mathbf{G}}_{l(n)}^i$. An estimator of $\Sigma_{\mathbf{R}}$ is given by

$$\hat{\Sigma}_{\mathbf{R},n} \equiv \frac{1}{k_n} \times \sum_{i=1}^{k_n} \left\{ |D_{l(n)}^i| (\hat{\mathbf{G}}_{l(n)}^i - \bar{\mathbf{G}}_n) (\hat{\mathbf{G}}_{l(n)}^i - \bar{\mathbf{G}}_n)' \right\} \text{ with } \bar{\mathbf{G}}_n \equiv \sum_{i=1}^{k_n} \hat{\mathbf{G}}_{l(n)}^i / k_n.$$

The following theorem states that under some mild conditions, $\hat{\Sigma}_{\mathbf{R},n}$ is an L_2 consistent estimator for $\Sigma_{\mathbf{R}}$, in the sense that every element of $\hat{\Sigma}_{\mathbf{R},n}$ is an L_2 consistent estimator for its counterpart in $\Sigma_{\mathbf{R}}$.

Theorem II.3. Assume that condition (2.3) holds with $\delta > 2$ and all the remaining conditions in Theorem II.1 hold, then $\hat{\Sigma}_{\mathbf{R},n}$ is an L_2 consistent estimator for $\Sigma_{\mathbf{R}}$.

Proof. See Appendix A. □

In order to apply the subsampling method, one needs to decide on an appropriate, or “optimal” subblock size, i.e., the value of $l(n)$ that best estimates the unknown covariance matrix. In general, this choice depends on the dependence structure of the underlying process and the definition of “optimal”. Sherman (1996) showed that $cn^{1/2}$ for some $c > 0$ is the “optimal” rate for estimating the variance of a statistic on a spatial lattice, where the word optimal therein refers to minimizing the mean squared error. In practice, we suggest choosing a reasonable value of c and setting $l(n) = cn^{1/2}$. We perform a sensitivity study in Section 2.4 to assess how the choice of subblock size affects the testing results.

2.3.2.2 Irregularly spaced data case

Following the asymptotic properties given in Theorem II.2, a natural choice to estimate $\Sigma_{\mathbf{IR}}$ is to use a plug-in method, e.g., to set the off diagonal elements to zero and estimate the diagonal elements by replacing ν by $N(D)/|D|$ and $\gamma^{(4)}(\cdot)$ by a kernel estimator similar to the one used in Section 2.2.2. This method, however, may be overly simplistic because it completely ignores the off diagonal elements of $\Sigma_{\mathbf{IR}}$, which may be nonnegligible in finite sample applications. For example, we used a five-dependent isotropic covariance model

(see Section 2.4) to simulate 20×20 marked-Poisson Gaussian random fields. Five thousands realizations showed that $\text{Cov}(\hat{\gamma}((1, 0)), \hat{\gamma}((0, 1)))$ is about one third of $\text{Var}(\hat{\gamma}((1, 0)))$ using the uniform kernel with bandwidth 0.4. Further, to estimate $\gamma^{(4)}(\cdot)$, one also needs to decide what bandwidth to use, which is not an easy task.

To estimate Σ_{IR} , we again apply the subsampling technique. Politis and Sherman (2001) justified the use of subsampling for scalar variables from a marked-point process. Here we extend their results to the multivariate case.

Toward that end, let $D_{l(n)}$ be the subfield with the same shape as D_n but rescaled, where $l(n)$ is defined in Section 3.2.1. Define $D_{l(n)} + \mathbf{y} \equiv \{\mathbf{s} + \mathbf{y} : \mathbf{s} \in D_{l(n)}\}$ be its shifted copy, where $\mathbf{y} \in D_n^{1-c}$ and $D_n^{1-c} \equiv \{\mathbf{y} \in D_n : D_{l(n)} + \mathbf{y} \subset D_n\}$. Let $h_{l(n)}$ be the bandwidth used to get the sample variograms on $D_{l(n)} + \mathbf{y}$ (denoted by $\hat{\mathbf{G}}(D_{l(n)} + \mathbf{y})$). An estimator of Σ_{IR} , $\hat{\Sigma}_{\text{IR},n}$, is given by

$$\frac{1}{|D_n^{1-c}|} \times \int_{D_n^{1-c}} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times (\hat{\mathbf{G}}(D_{l(n)} + \mathbf{y}) - \bar{\mathbf{G}}_n)(\hat{\mathbf{G}}(D_{l(n)} + \mathbf{y}) - \bar{\mathbf{G}}_n)' \right\} d\mathbf{y}, \quad (2.5)$$

where $\bar{\mathbf{G}}_n \equiv \int_{D_n^{1-c}} \hat{\mathbf{G}}(D_{l(n)} + \mathbf{y}) d\mathbf{y} / |D_n^{1-c}|$.

Theorem II.4. Assume that condition (2.3'.b) holds with $\delta > 2$ and all the remaining conditions in Theorem II.2 hold, then $\hat{\Sigma}_{\text{IR},n}$ is an L_2 consistent estimator for Σ_{IR} .

Proof. See Appendix A. □

In practice, to obtain $\hat{\Sigma}_{\text{IR},n}$ the integral in (2.5) has to be approximated by a finite sum. Politis and Sherman (2001) suggested both a deterministic approximation and a Monte-Carlo or stochastic approximation. In this article, we adopt the deterministic approach, which is discussed further in Section 4.

2.3.3 The Test Statistic

Observe that the hypothesis of isotropy can be expressed as $H_0 : \gamma(\mathbf{t}) = \gamma_0(\|\mathbf{t}\|)$ for some function $\gamma_0(\cdot)$, where $\|\mathbf{t}\| = \sqrt{\mathbf{t}'\mathbf{t}}$. It can be rewritten, in terms of variograms at lags

belonging to Λ , as

$$H_0 : \gamma(\mathbf{t}_1) = \gamma(\mathbf{t}_2), \quad \mathbf{t}_1, \mathbf{t}_2 \in \Lambda, \mathbf{t}_1 \neq \mathbf{t}_2, \text{ but } \|\mathbf{t}_1\| = \|\mathbf{t}_2\|.$$

Thus under the hypothesis of isotropy, there exists a full row rank matrix \mathbf{A} such that $\mathbf{A}\mathbf{G} = \mathbf{0}$ (Lu and Zimmerman, 2001). For example, if $\Lambda = \{(1, 0), (0, 1), (1, 1), (-1, 1)\}$, i.e., $\mathbf{G} = \{\gamma((1, 0)), \gamma((0, 1)), \gamma((1, 1)), \gamma((-1, 1))\}'$, then

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

Here and henceforth, we will use d to denote the row rank of \mathbf{A} . For regularly spaced observations, it follows from Theorem II.1 that if H_0 is true, then

$$|D_n| \times (\mathbf{A}\hat{\mathbf{G}}_n)'(\mathbf{A}\Sigma_{\mathbf{R}}\mathbf{A}')^{-1}(\mathbf{A}\hat{\mathbf{G}}_n) \xrightarrow{D} \chi_d^2 \text{ as } n \rightarrow \infty.$$

We estimate $\Sigma_{\mathbf{R}}$ by the subsampling estimator $\hat{\Sigma}_{\mathbf{R},n}$ which is given in Section 2.3.2.1 and propose the following test statistic:

$$TS_{R,n} \equiv |D_n| \times (\mathbf{A}\hat{\mathbf{G}}_n)'(\mathbf{A}\hat{\Sigma}_{\mathbf{R},n}\mathbf{A}')^{-1}(\mathbf{A}\hat{\mathbf{G}}_n).$$

Since $\hat{\Sigma}_{\mathbf{R},n} \xrightarrow{L_2} \Sigma_{\mathbf{R}}$, $TS_{R,n} \xrightarrow{D} \chi_d^2$ as $n \rightarrow \infty$ by the multivariate Slutsky's theorem (Ferguson 1996). For irregularly spaced observations, $E(\mathbf{A}\hat{\mathbf{G}}) = \mathbf{0}$ under H_0 from the proof of Theorem II.2. We then naturally extend the above test statistic as follows

$$TS_{IR,n} \equiv |D_n| \times h_n^2 \times (\mathbf{A}\hat{\mathbf{G}}_n)'(\mathbf{A}\hat{\Sigma}_{\mathbf{IR},n}\mathbf{A}')^{-1}(\mathbf{A}\hat{\mathbf{G}}_n),$$

where $\hat{\Sigma}_{\mathbf{IR},n}$ is the subsampling estimator given in Section 2.3.2.2. Similarly $TS_{IR,n} \xrightarrow{D} \chi_d^2$ as $n \rightarrow \infty$. Thus an approximate size- α test for isotropy is to reject H_0 if $TS_{R,n}$ (for regularly spaced observations) or $TS_{IR,n}$ (for irregularly spaced observations) is bigger than $\chi_{d,\alpha}^2$, i.e., the upper α percentage point of a χ^2 distribution with d degrees of freedom.

2.4 A Simulation Study

2.4.1 Simulation Design

We consider realizations from a zero-mean, second-order stationary Gaussian random field. Each random field was either isotropic or geometrically anisotropic with the following covariance structure

$$C(r; m) = \begin{cases} \theta(1 - \frac{3r}{2m} + \frac{r^3}{2m^3}) & \text{if } 0 \leq r \leq m \\ 0 & \text{otherwise,} \end{cases}$$

where $r = \sqrt{\mathbf{t}'\mathbf{B}\mathbf{t}}$ and \mathbf{B} is a 2×2 positive definite matrix. The parameter θ is a scale parameter which was set equal to 1.0 throughout the study. The parameter m defines the range and strength of dependence. In the simulation study, m was set to be 2, 5, and 8, denoting weak, moderate and relatively strong spatial dependence, respectively. This setup was considered by Lu and Zimmerman (2001).

The following five matrices of \mathbf{B} were used:

$$\mathbf{B1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{B2} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \quad \mathbf{B3} = \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix},$$

$$\mathbf{B4} = \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}, \quad \mathbf{B5} = \begin{bmatrix} 8.5 & -7.5 \\ -7.5 & 8.5 \end{bmatrix}.$$

Matrix **B1** yields isotropic random fields while **B2**, **B3**, **B4** and **B5** yield geometrically anisotropic random fields. More specifically, **B2** and **B3** yield isocorrelation ellipses whose main axes are aligned with the (x, y) axes; **B4** and **B5** yield isocorrelation ellipses whose main axes are oriented at 45 and 135 degree angles with the x -axis. In addition, the anisotropy ratio, defined as the ratio of the lengths of the main axes, is 2:1 for **B2** and **B4** but 4:1 for **B3** and **B5**.

For the regularly spaced case, ten thousand realizations for each choice of \mathbf{B} and m were generated on a 20×20 square grid and a 20×20 parallelogram grid. For each realization, the proposed test was conducted for $\Lambda = [(1, 0), (0, 1), (1, 1), (-1, 1)]$. For the square field, c was set to be 0.5, 1, 1.5, corresponding to subblocks of size 2×2 , 4×4 and 7×7 , respectively. For the parallelogram field, c was taken to be .75, 1, 1.5.

In the irregularly spaced case, the number of points for each realization was generated according to a Poisson distribution with parameter 400; then point locations were determined by a uniform distribution on a 20×20 square field and a 20×20 parallelogram field. Given these locations, the values of the observations were then generated from a Gaussian process following the above covariance structure. One thousand realizations were simulated for each choice of \mathbf{B} and m . The proposed test was performed for $c = 1$ and $h = 0.4, 0.7, 1$. The same lag set Λ was used as in the regularly spaced case.

2.4.2 Finite Sample Adjustments to the Subsampling Estimator

Although our subsampling estimator is consistent for a wide class of situations, we modify it slightly for better finite sample performance. This modification corrects for edge effects and reduces bias, but does not change our asymptotic results. For regularly spaced data, the (j, k) th element of $\Sigma_{\mathbf{R}, \mathbf{n}}$ is estimated by:

$$\frac{1}{k'_n} \times \sum_{i=1}^{k_n} \left\{ \sqrt{|D_{l(n)}^i(\mathbf{t}_j)| \times |D_{l(n)}^i(\mathbf{t}_k)|} \times (\hat{\gamma}_{l(n)}^i(\mathbf{t}_j) - \bar{\gamma}_n(\mathbf{t}_j)) \times (\hat{\gamma}_{l(n)}^i(\mathbf{t}_k) - \bar{\gamma}_n(\mathbf{t}_k)) \right\},$$

where $\hat{\gamma}_{l(n)}^i(\mathbf{t})$, $|D_{l(n)}^i(\mathbf{t})|$ and $\bar{\gamma}_n(\mathbf{t})$ denote the estimate of the sample variogram at lag \mathbf{t} on $D_{l(n)}^i$, the number of distinct pairs in $D_{l(n)}^i$ used to calculate $\hat{\gamma}_{l(n)}^i(\mathbf{t})$, and the average of the sample variograms at lag \mathbf{t} obtained from all the subfields, respectively.

In the above expression, k'_n is set to be the right standardizing constant that produces an unbiased variance estimator for the sample mean of i.i.d. observations. The choice of k'_n can be obtained via simulation. For example, one thousand realizations of i.i.d. random

variables with variance σ^2 were simulated on a 20×20 square grid in this study. The target parameter σ^2 was estimated by subsampling method during each realization. The value of k'_n was chosen such that the average of these one thousand variance estimates was approximately equal to the target parameter. We obtained $k'_n = 357, 274, 162$ for $c = .5, 1, 1.5$ respectively, corresponding to $k_n = 361, 289$ and 196 . Since the sample variogram estimator (defined in Section 2.1) is a mean of derived variables, k'_n is a reasonable choice for bias correction under weak dependence assumptions.

To calculate the subsampling estimator for the irregularly spaced case, we laid a 20×20 grid on the field during each simulation, with a square grid on the square field, a parallelogram grid on the parallelogram field respectively. A 4×4 ($c = 1$) subblock window (either square or parallelogram) was moved across the field, with its bottom left corner starting from one of the nodes on the grid each time. On each subfield $D_{l(n)}^i$ and for each $\mathbf{t} \in \Lambda$, the following quantity was calculated:

$$\tilde{\gamma}_{l(n)}^i(\mathbf{t}) \equiv \iint h_n^{-1} w\left(\frac{\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2}{h_n}\right) \times \frac{[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2}{\sqrt{|D_{l(n)}^i \cap (D_{l(n)}^i - \mathbf{x}_1 + \mathbf{x}_2)|}} N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2),$$

where the integrals are over $\{\mathbf{x}_1 \in D_{l(n)}^i, \mathbf{x}_2 \in D_{l(n)}^i\}$. The (j, k) th element of $\Sigma_{\text{IR},n}$ is then estimated by:

$$\frac{1}{k'_n} \times \sum_{i=1}^{k_n} (\tilde{\gamma}_{l(n)}^i(\mathbf{t}_j) - \bar{\gamma}_n(\mathbf{t}_j)) \times (\tilde{\gamma}_{l(n)}^i(\mathbf{t}_k) - \bar{\gamma}_n(\mathbf{t}_k)),$$

where $\bar{\gamma}_n(\mathbf{t})$ denote the average of all $\tilde{\gamma}_{l(n)}^i(\mathbf{t})$, $i = 1, \dots, k_n$ and k'_n is as being defined in the regularly spaced case.

2.4.3 Results and Analysis

Table 1 reports the percentages of rejections at the 5% nominal level for the regularly spaced case. Note $c = 1$, regardless of the dependence strength, gives the best results in achieving the nominal size for square grids. This is roughly true also for parallelogram grids. Thus

Table 1. Simulation results for regularly spaced observations. Each table entry is the percentage of rejections at the 5% nominal level from ten thousand simulations

Shape of Grids	m	c	B1	B2	B3	B4	B5
square	2	0.5	0.0268	0.9811	0.9796	0.1038	0.1029
		1.0	0.0486	0.9754	0.9726	0.2190	0.2044
		1.5	0.0659	0.9633	0.9610	0.2560	0.2437
	5	0.5	0.1113	1.0000	1.0000	0.9972	1.0000
		1.0	0.0812	0.9999	1.0000	0.9953	1.0000
		1.5	0.0835	0.9986	0.9999	0.9853	1.0000
	8	0.5	0.1500	1.0000	1.0000	0.9978	1.0000
		1.0	0.0903	1.0000	1.0000	0.9942	1.0000
		1.5	0.0938	0.9978	1.0000	0.9816	0.9999
parallelogram	2	0.75	0.0469	0.9711	0.9700	0.1564	0.1479
		1.0	0.0576	0.9728	0.9685	0.1972	0.1919
		1.5	0.0678	0.9621	0.9572	0.2349	0.2339
	5	0.75	0.1080	0.9999	1.0000	0.9973	1.0000
		1.0	0.0916	0.9998	1.0000	0.9960	1.0000
		1.5	0.0919	0.9988	1.0000	0.9877	1.0000
	8	0.75	0.1305	0.9999	1.0000	0.9980	1.0000
		1.0	0.1017	0.9999	1.0000	0.9968	1.0000
		1.5	0.0960	0.9989	1.0000	0.9864	0.9999

we recommend using $c = 1$ in practice. It is generally true that the empirical sizes of the test are higher than the nominal one, especially when correlations are high. This is because the subsampling method tends to underestimate elements in the covariance matrix and thus to inflate the test statistic. The problem will diminish for larger data grids. Our results are comparable to those in Lu and Zimmerman (2001).

Table 2 reports the percentages of rejections at the 5% nominal level for the irregularly spaced case. An interesting phenomenon is that the empirical sizes of the tests are much closer to the nominal 5% compared to the regularly spaced case. This can be intuitively explained as follows. Consider a $n \times n$ square field D_n . Let $D_n^{i,j}$ denote the (i, j) th cell obtained by overlaying a $n \times n$ grid. Define a new process $\{X_n(i, j) : 1 \leq i \leq n, 1 \leq j \leq$

Table 2. Simulation results for irregularly spaced observations. Each entry is the percentage of rejections at the 5% nominal level from one thousand simulations

Shape of Fields	m	h	B1	B2	B3	B4	B5
square	2	0.4	0.052	0.211	0.123	0.074	0.054
		0.7	0.049	0.232	0.123	0.086	0.047
		1.0	0.061	0.108	0.063	0.087	0.057
	5	0.4	0.052	0.571	0.927	0.502	0.804
		0.7	0.071	0.775	0.964	0.736	0.911
		1.0	0.081	0.740	0.912	0.748	0.889
	8	0.4	0.067	0.619	0.968	0.566	0.945
		0.7	0.074	0.788	0.997	0.777	0.989
		1.0	0.086	0.777	0.996	0.824	0.993
parallelogram	2	0.4	0.054	0.219	0.137	0.061	0.060
		0.7	0.071	0.233	0.146	0.090	0.062
		1.0	0.051	0.110	0.092	0.089	0.076
	5	0.4	0.064	0.591	0.921	0.527	0.823
		0.7	0.079	0.790	0.972	0.738	0.939
		1.0	0.095	0.788	0.917	0.770	0.902
	8	0.4	0.040	0.626	0.963	0.555	0.950
		0.7	0.067	0.810	0.993	0.786	0.989
		1.0	0.095	0.826	0.993	0.826	0.994

$n\}$ where

$$X_n(i, j) \equiv \int_{\mathbf{x}_1 \in D_n} \int_{\mathbf{x}_2 \in D_n^{i,j}} h_n^{-1} w\left(\frac{\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2}{h_n}\right) \times [Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2 N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2).$$

Let $\bar{X}_n \equiv \sum_{i=1}^n \sum_{j=1}^n X_n(i, j)/n^2$. Observe that for small h_n , $|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \approx n^2$ and thus $\sqrt{|D_n|} \times h_n \times \{\hat{\gamma}_n(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_n(\mathbf{t}))\} \approx \sqrt{|D_n|} \times \{\bar{X}_n - \mathbf{E}(\bar{X}_n)\}$. A simple analysis shows the correlation between $X_n(i, j)$ and $X_n(i', j')$ is of order h_n^2 if $i \neq i'$ or $j \neq j'$, i.e., $X_n(i, j)$ and $X_n(i', j')$ tend to be less correlated than in the regularly spaced setting when h_n is small. The subsampling method in general gives reasonably good estimates when correlations among observations are weak. This led to the better empirical sizes in our setting. Note for fixed n , $X_n(i, j)$ and $X_n(i', j')$ tend to be more correlated as h_n increases. This explains the increasing empirical sizes as h increases in Table 2. Comparing the results in Table 1 and Table 2 shows that the test for the irregularly spaced case is less powerful

than that for the regularly spaced case. This can be attributed to the necessary smoothing in the irregularly spaced case.

2.5 Applications

2.5.1 The Wind-Speed Data

The wind-speed data consist of the east-west component of the wind speed (in meters per second) over a region in the tropical western Pacific Ocean (145 deg E-175deg E, 14deg S-16 deg N). The data are given on a regular spatio-temporal grid of 17×17 sites with grid spacing of about 210 km, and every 6 hours for the period November 1992 through February 1993. This gives data at 289 locations and 480 time points. Cressie and Huang (1999) examined the second-order stationary assumption and did not find evidence against it. They further fitted a stationary spatio-temporal variogram to the data with the spatial component being isotropic.

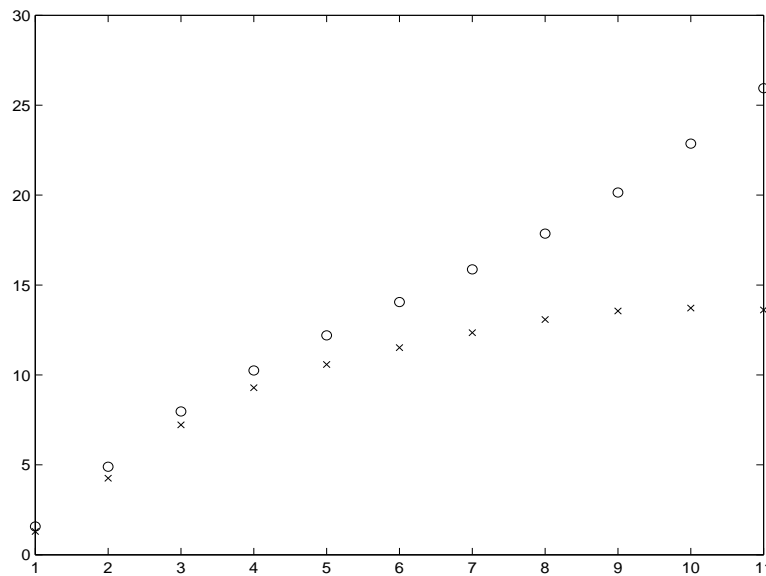


Figure 1. Sample Variograms in Two Directions for the Wind-Speed Data. \times : E-W direction, \circ : N-S direction.

We here assess the validity of the isotropy assumption. We first plot the empirical

variograms at spatial lags from 1 to 11 in the E-W direction and N-S direction (see Figure 1). This is a natural choice of directions to study due to the directional feature of the observations. The empirical variogram at a lag was calculated by averaging all the sample variograms at each time for that lag (480 in total). Some interesting features of the plot include a strong linear trend in the N-S direction, a sill around 13 in the E-W direction, and seemingly close agreement at small lags of these two directions. This closeness is relative to the results at larger lags. At lag distance one, for example, the sample variogram in the north-south direction is 21.24% higher than that in the east-west direction.

We applied our test using $\Lambda = \{(1, 0), (0, 1)\}$ and 4×4 subblocks to estimate $\Sigma_{\mathbf{R}}$ for each of the 480 time points. 22.29% of the p-values are less than 0.01, 37.29% less than 0.05, 47.71% less than 0.1. Thus the isotropic assumption does not appear to be very reasonable for this data set.

2.5.2 *The Long-leaf Pine Data*

The long-leaf pine data consists of locations and dbh (diameter at breast height) of 584 long-leaf pine trees in a $200m \times 200m$ square field. The data set was given by Cressie (1991). Although the locations of the trees exhibit apparent clustering, our test was applied as if they come from a homogeneous Poisson process. We conjecture our asymptotic results will remain valid if the locations of trees are generated by an isotropic, second-order stationary point process. Λ was set to be $[(8, 0), (0, 8), (8/\sqrt{2}, 8/\sqrt{2}), (-8/\sqrt{2}, 8/\sqrt{2})]'$ due to the argument given in Section 2.3.1; \mathbf{A} was as follows

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix}.$$

A 24×24 grid was laid on the field corresponding to approximately one point at each node. For the imposed grid, $c = 1$ gave an approximate subblock size of 5×5 , which translated

back to the original scale yielded $40m \times 40m$ subblocks. Two bandwidths h , equal to four and eight meters, were considered. The p-values for the tests are 0.6448 and 0.4610 respectively, indicating no strong evidence against the isotropy assumption.

CHAPTER III

A TEST OF ISOTROPY FOR SPATIAL POINT PROCESSES

3.1 Introduction

Spatial events in the form of a set of points in space are often irregularly scattered in nature; examples include the locations of trees in a forest, of leukemia patients in a state, or of crime events in a city. A common approach is to model these observations by a stationary spatial point process (e.g., Diggle 1983, Cressie 1991, Stoyan and Stoyan 1994). The development of the second-order structure for a point process, e.g., second-order intensity function, K -function, is critical for spatial modeling (e.g., Diggle 1983, Cressie 1991, Stoyan and Stoyan 1994). A spatial point process is called isotropic if its second-order characteristics are direction invariant. Otherwise the process is said to be anisotropic.

In many applications, it is of great interest to test if a spatial process is isotropic. Consider the example of childhood leukemia study (Section 3.5.2) where it is suspected that exposure to pollution may increase the risk of developing the disease. Since wind is an important media in transporting many of these pollutants, spatial locations of leukemia cases may be more clustered in the major wind direction than perpendicular to that direction. Testing for isotropy in this case can help justify our conception of certain risk factors being influential or rule them out from future studies. In a study of violent crimes in a city (Section 3.5.1), a rejection of isotropy will provide evidence for the existence of some important directional factors, for example, major highways, and to what extent these factors influence the distribution. This in turn can play an important role for future policy making and crime prevention.

Many spatial models utilize isotropy (see, e.g., Cuzick and Edwards 1990, Diggle and Chetwynd 1991) due to its simplicity. Often in practice, however, these models are fitted

without further checking of this assumption. As a consequence, this can lead to inaccurate inferences. For example, consider a Poisson cluster process with (expected) 100 parents and 4 offspring per parent on the unit square. Suppose the positions of the offspring relative to their parents follow the p.d.f.

$$f(x, y) = \frac{1}{\pi\sigma^2} \exp\{-(x^2 + 4y^2)/2\sigma^2\}.$$

The above model yields an anisotropic process where the dependence in the x direction is twice as strong as that in the y direction. Figure 3 plots the estimated (isotropic) K function as well as an upper and lower simulation envelopes obtained from the best fitting isotropic model based on a realization from the above model with $\sigma = 0.02$. Without checking for isotropy, we may wrongly conclude the fitted isotropic model is appropriate for the process since the estimated K function lies inside the envelopes. This example, together with the discussions in the proceeding paragraph, demonstrate that testing isotropy is an important part of model building.

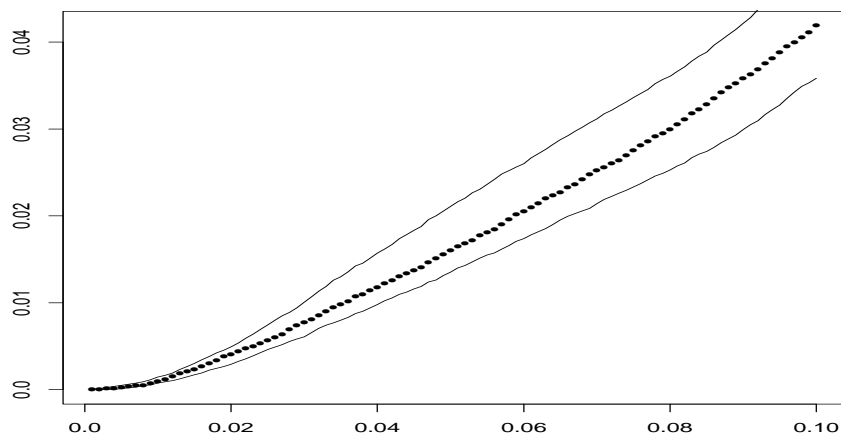


Figure 2. K Function Plot. The dotted line is the estimated K function assuming isotropy, the solid lines are the upper and lower envelopes from 100 simulations from the best fitting isotropic model.

A conventional practice to check for isotropy is to informally assess plots of direction-specific sample second-order intensity function or the K function (e.g., Ohser and Stoyan 1981, Stoyan and Stoyan 1994). While useful, graphical diagnostics are often difficult to assess and are open to interpretation. Progress has been made lately with formally investigating isotropy for quantitative lattice data (e.g., Baczkowski and Mardia 1990, Lu and Zimmerman 2001). Such testing procedures, however, are yet to be developed under the spatial point process set-up.

In this article, we propose a formal approach to test isotropy for a spatial point process. Our approach is based on the asymptotic joint normality of the sample second-order intensity functions and can be used to compare these values in multiple directions. We derive an L_2 consistent subsampling estimator for the asymptotic covariance matrix of the sample second-order intensity function and use that to construct a test statistic with a χ^2 limiting distribution. Our testing approach is purely nonparametric in that it only requires mild moment conditions and a weak dependence assumption for the underlying process. In addition, the shape containing the spatial locations can be quite irregular.

The rest of the article is organized as follows. In Section 3.2, we introduce the definition of the sample second-order intensity function and study its asymptotic properties. Based on these results, we form a test statistic in Section 3.3 and further derive its limiting distribution. Some practical concerns are discussed in Section 3.4. Section 3.5 presents some simulation results, while in Section 3.6 our method is applied to two data sets. Proofs of the theorems are given in Appendix B.

3.2 Definitions and Asymptotic Results

3.2.1 Notations and Set-up

Consider a stationary spatial point process N . A process is said to be *stationary* if all probability statements about the process in any region D of the plane are invariant under

arbitrary translation of D . In what follows, let $N(D)$ denote the random number of points in D , $|D|$ represent the area of D , and $d\mathbf{x}$ be an infinitesimal region which contains the point \mathbf{x} . Define the first and second-order intensity function as

$$\lambda(\mathbf{x}) \equiv \lim_{|d\mathbf{x}| \rightarrow 0} \left\{ \frac{\mathbf{E}[N(d\mathbf{x})]}{|d\mathbf{x}|} \right\},$$

$$\lambda^{(2)}(\mathbf{x}, \mathbf{y}) \equiv \lim_{|d\mathbf{x}|, |d\mathbf{y}| \rightarrow 0} \left\{ \frac{\mathbf{E}[N(d\mathbf{x}) \times N(d\mathbf{y})]}{|d\mathbf{x}| \times |d\mathbf{y}|} \right\}.$$

The value of $\lambda(\mathbf{x})|d\mathbf{x}|$ gives the approximate probability that one point of N falls in $d\mathbf{x}$, while $\lambda^{(2)}(\mathbf{x}, \mathbf{y})|d\mathbf{x}||d\mathbf{y}|$ approximates the probability that there is one observed point in both $d\mathbf{x}$ and $d\mathbf{y}$. For a stationary process, $\lambda(\mathbf{x})$ assumes a constant value, say ν and $\lambda^{(2)}(\mathbf{x}, \mathbf{y}) \equiv \Psi(\mathbf{x} - \mathbf{y})$ for some function $\Psi(\cdot)$.

Let \mathbf{t} denote a lag between two points and let $\|\mathbf{t}\|$ denote the Euclidean distance of \mathbf{t} from the origin. Note that if $\|\mathbf{t}\| = \|\mathbf{t}'\|$, then $\Psi(\mathbf{t}) = \Psi(\mathbf{t}')$ under isotropy. Thus we may form a test for isotropy by comparing second-order intensity functions at lags with the same length but in different directions. In practice, since the true function is typically unknown, we form a test based on an estimator of it.

Let h be a positive constant and $w(\cdot)$ be a bounded, nonnegative, isotropic density function which takes positive values only on a finite support, C . Define $N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) \equiv N(d\mathbf{x}_1)N(d\mathbf{x}_2)I(\mathbf{x}_1 \neq \mathbf{x}_2)$, where $I(\mathbf{x}_1 \neq \mathbf{x}_2) = 1$ if $\mathbf{x}_1 \neq \mathbf{x}_2$ and 0 otherwise. A kernel estimator of $\Psi(\mathbf{t})$ is given by

$$\hat{\Psi}(\mathbf{t}) = \int_{\mathbf{x}_1 \in D} \int_{\mathbf{x}_2 \in D} \frac{w[(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2)/h]}{|D \cap (D - \mathbf{x}_1 + \mathbf{x}_2)| \times h^2} N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2).$$

We base our test on sample second-order intensity functions at a finite number of lags in multiple directions. Let Λ be a user chosen lag set of interest. Define $\mathbf{G} \equiv \{\Psi(\mathbf{t}) : \mathbf{t} \in \Lambda\}$ to be the vector of second-order intensity functions at lags in Λ . Consider a sequence of random fields D_n and let $\hat{\Psi}_n(\mathbf{t})$ and $\hat{\mathbf{G}}_n \equiv \{\hat{\Psi}_n(\mathbf{t}) : \mathbf{t} \in \Lambda\}$ be the estimators of $\Psi(\mathbf{t})$

and \mathbf{G} obtained over D_n , respectively. In the remainder of this section, we investigate the large sample properties of $\hat{\Psi}_n(\mathbf{t})$ and $\hat{\mathbf{G}}_n$.

3.2.2 Asymptotic Bias and Covariance of Sample Second-order Intensity Function

Define the k th-order cumulant function as

$$C_N^{(k)}(\mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_k - \mathbf{x}_1) \equiv \lim_{|\mathbf{d}\mathbf{x}_1|, \dots, |\mathbf{d}\mathbf{x}_k| \rightarrow 0} \left\{ \frac{\text{Cum}[N(\mathbf{d}\mathbf{x}_1), \dots, N(\mathbf{d}\mathbf{x}_k)]}{|\mathbf{d}\mathbf{x}_1| \times \dots \times |\mathbf{d}\mathbf{x}_k|} \right\},$$

where $\text{Cum}(Y_1, \dots, Y_k)$ is given by the coefficient of $i^k t_1 \dots t_k$ in the Taylor series expansion of $\log\{\mathbf{E}[\exp(i \sum_{j=1}^k Y_j t_j)]\}$ about the origin (see, Brillinger 1975). Roughly speaking, the cumulant functions have an interpretation in terms of dependence and independence. For example, if N is Poisson, then all $C_N^{(k)}(\mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_k - \mathbf{x}_1)$ will be equal to zero if any of $\mathbf{x}_j - \mathbf{x}_1 \neq \mathbf{0}$, $j = 2, \dots, k$. Throughout this article, we will assume

$$C_N^{(2)}(\cdot), C_N^{(3)}(\cdot, \cdot) \text{ are bounded, } C_N^{(2)}(\cdot) \text{ is continuous and integrable,} \quad (3.1)$$

$$\int_{\mathbb{R}^2} |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_2)| d\mathbf{u}_1 < \infty, \int_{\mathbb{R}^2} |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_1 + \mathbf{u}_2)| d\mathbf{u}_1 < \infty \text{ for finite } \mathbf{u}_2, \text{ and} \\ \int_{\mathbb{R}^2} |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_3)| d\mathbf{u}_2 < \infty \text{ for finite } \mathbf{u}_1, \mathbf{u}_3. \quad (3.2)$$

All the above conditions are clearly satisfied if N is a homogeneous Poisson process with finite intensity. If a point process is m -dependent with $C_N^{(k)}$ being finite, $k = 1, 2, 3, 4$, then conditions (3.1) and (3.2) also hold. For a Poisson cluster process, these integrability conditions can further be written in terms of the intensity for the parent process, moments of the number of offspring per parent, and the distribution of an offspring's position relative to its parent (see Lemma B.4 in Appendix B). Then (3.1) and (3.2) are satisfied by commonly used models such as the one discussed in Section 3.1, where the positions of the offspring relative to their parents follow a radially symmetric normal distribution.

We need to account for the shape of the field from which we observe data. Let ∂D denote the boundary of D_n and $|\partial D|$ denote the length of ∂D . Assume

$$|D_n| = \mathcal{O}(n^2), \quad |\partial D_n| = \mathcal{O}(n). \quad (3.3)$$

Condition (3.3) allows for a variety of field sequences, e.g., squares, rectangles, circles and starshapes, etc. Throughout the article, we also assume the following condition on the bandwidth, $h = h_n$:

$$h_n = \mathcal{O}(n^{-\beta}) \text{ for some } \beta \in (0, 1). \quad (3.4)$$

The following theorem gives conditions under which the sample second-order intensity function is a consistent estimator for the target function.

Theorem III.1. Let N be a stationary point process observed on domain D_n . Assume conditions (3.1), (3.3) and (3.4), then

$$\mathbb{E}[\hat{\Psi}_n(\mathbf{t})] = \int_C w(\mathbf{x}) \Psi(\mathbf{t} - h_n \mathbf{x}) d\mathbf{x} \rightarrow \Psi(\mathbf{t}).$$

If we further assume condition (3.2), then

$$\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times \text{Cov}[\hat{\Psi}_n(\mathbf{t}), \hat{\Psi}_n(\mathbf{t}')] = \int_C w^2(\mathbf{x}) d\mathbf{x} \times \Psi(\mathbf{t}) \times I(\mathbf{t} = \pm \mathbf{t}'),$$

where $\mathbf{t}, \mathbf{t}' \in \Lambda$, $I(\mathbf{t} = \pm \mathbf{t}') = 1$ if $\mathbf{t} = \pm \mathbf{t}'$ and 0 otherwise.

Proof. See Appendix B. □

3.2.3 Asymptotic Normality of Sample Second-order Intensity Function

To formally state the asymptotic normality of $\hat{\mathbf{G}}_n$, we need to quantify the strength of dependence in the random field. In the nonparametric spirit, we do so using a model free mixing condition. Following Rosenblatt (1956), we make use of a particular type of strong mixing coefficient defined as

$$\alpha_N(p; k) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_N(E_1), A_2 \in \mathcal{F}_N(E_2),$$

$$E_2 = E_1 + \mathbf{s}, |E_1| = |E_2| \leq p, d(E_1, E_2) \geq k\},$$

where the supremum is taken over all compact and convex subsets $E_1 \subset \mathbb{R}^2$, and over all $\mathbf{s} \in \mathbb{R}^2$ such that $d(E_1, E_2) \geq k$; in the above, $\mathcal{F}_N(E)$ is the σ -algebra generated by the random points of the point process N that happen to fall in set E .

If N is Poisson, then $\alpha_N(p; k) = 0$ for all $k > 0$. Here we require $\alpha_N(p; k)$ to approach 0 for large k , at some rate depending on the volume p . Following Sherman and Carlstein (1994), we assume the following mixing condition

$$\sup_p \frac{\alpha_N(p; k)}{p} = O(k^{-\epsilon}) \text{ for some } \epsilon > 2. \quad (3.5)$$

Condition (3.5) says that at a fixed distance k , as the volume increases, we allow dependence to increase at a rate controlled by p . As the distance increases, the dependence must decrease at a polynomial rate in k . Any m -dependent point process satisfies this condition. Jensen (1993a, b) shows that the Strauss point process satisfies $\alpha_N(p; k) \leq Ake^{-Bp}$ for some constants A and B and thus satisfies condition (3.5); more discussion regarding this condition can be found in Politis and Sherman (2001).

Finally we require the following mild moment condition

$$\sup_n \mathbf{E} \left\{ \left| \sqrt{|D_n|} \times h_n \times [\hat{\Psi}_n(\mathbf{t}) - \Psi(\mathbf{t})] \right|^{2+\delta} \right\} \leq C_\delta \text{ for some } \delta > 0, C_\delta < \infty. \quad (3.6)$$

Condition (3.6) is only slightly stronger than the existence of the (standardized) asymptotic variance of $\hat{\Psi}_n(\mathbf{t})$. For an m -dependent point process with bounded cumulant functions up to order eight, it can be shown that (3.6) holds for $\delta = 2$ (see Lemma B.3 in Appendix B).

Theorem III.2. Let N be a stationary point process observed on domain D_n . Assume conditions (3.1)-(3.6), then $\sqrt{|D_n|} \times h_n \times \{\hat{\mathbf{G}}_n - \mathbf{E}(\hat{\mathbf{G}}_n)\}$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix Σ , the elements of which are given in Theorem III.1.

Proof. See Appendix B. □

3.3 Assessment of Isotropy

3.3.1 Estimating the Covariance Matrix

To formally compare the directional sample second-order intensity functions, we need knowledge of Σ . This covariance matrix is usually unknown and thus needs to be estimated. From the asymptotic properties given in Section 3.2, the following plug-in method seems to be a natural choice

$$\hat{\sigma}_{i,j} = \begin{cases} \int_C w(\mathbf{x})^2 d\mathbf{x} \times \hat{\Psi}(\mathbf{t}_i) & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

where $\hat{\sigma}_{i,j}$ denotes the (i, j) th element of the estimated covariance matrix.

The plug-in method, however, is often overly simplistic because the off diagonal elements may be nonnegligible in finite samples. This is especially true in the presence of clustering. For example, we simulated a Poisson cluster process on a 20×20 square field, where the number of parents is assumed to be a random variable from Poisson(100); the expected number of offspring is four and the position of each offspring relative to its parent is determined by the p.d.f. given in Section 3.5 with \mathbf{B} being a 2×2 identity matrix. One thousand realizations showed that $\text{Cov}[\hat{\Psi}(1, 0), \hat{\Psi}(0, 1)]$ is about 35% of $\text{Var}[\hat{\Psi}(1, 0)]$ using a uniform kernel with bandwidth 0.1. It is this potential problem of the plug-in estimator that prompts us to search for an alternative method.

Here we apply a subsampling technique. Subsampling has been widely used to estimate the variance of a general spatial statistic on a regularly spaced grid (e.g., Hall 1988, Possolo 1991, Sherman and Carlstein 1994). Politis and Sherman (2001) consider subsampling based variance estimation for statistics computed from marked-point processes. We adapt their results to the point process case. In particular, to form a test statistic, we require an estimator of the covariance matrix of $\hat{\mathbf{G}}_n$.

Toward this end, let $D_{l(n)}$ be a subshape that is congruent to D_n both in configuration

and orientation but rescaled, where $l(n) = cn^\alpha$ for some positive constant c and $\alpha \in (0, 1)$. Define its shifted copy $D_{l(n)} + \mathbf{x} \equiv \{\mathbf{s} + \mathbf{x} : \mathbf{s} \in D_{l(n)}\}$, where $\mathbf{x} \in D_n^{1-c}$ and $D_n^{1-c} \equiv \{\mathbf{x} \in D_n : D_{l(n)} + \mathbf{x} \subset D_n\}$. Let $\hat{\mathbf{G}}_{l(n)}(\mathbf{x})$ be the sample second-order intensity function on $D_{l(n)} + \mathbf{x}$, $h_{l(n)}$ be the bandwidth used to obtain $\hat{\mathbf{G}}_{l(n)}(\mathbf{x})$. The subsampling estimator (denoted by $\hat{\Sigma}_n$) is given as

$$\frac{1}{|D_n^{1-c}|} \times \int_{D_n^{1-c}} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times (\hat{\mathbf{G}}_{l(n)}(\mathbf{x}) - \overline{\mathbf{G}}_{l(n)}) (\hat{\mathbf{G}}_{l(n)}(\mathbf{x}) - \overline{\mathbf{G}}_{l(n)})' \right\} d\mathbf{x}, \quad (3.7)$$

where $\overline{\mathbf{G}}_{l(n)} \equiv \int_{D_n^{1-c}} \hat{\mathbf{G}}_{l(n)}(\mathbf{x}) d\mathbf{x} / |D_n^{1-c}|$.

The following theorem states that under some mild conditions, $\hat{\Sigma}_n$ is an L_2 consistent estimator for Σ , in the sense that every element of the subsampling estimator is L_2 consistent for its counterpart in the target covariance matrix.

Theorem III.3. Assume that condition (3.6) holds with $\delta > 2$ and all the remaining conditions in Theorem III.1 and III.2 hold, then $\hat{\Sigma}_n$ is an L_2 consistent estimator for Σ .

Proof. See Appendix B. □

In practice, to obtain $\hat{\Sigma}_n$ the integral in (3.7) has to be approximated by a finite sum. This approximation, as well as some finite sample adjustments, will be discussed in Section 3.5.

3.3.2 The Test Statistic

First recall the following result in Theorem III.1:

$$\mathbb{E}[\hat{\Psi}_n(\mathbf{t})] = \int_c w(\mathbf{x}) \Psi(\mathbf{t} - h_n \mathbf{x}) d\mathbf{x} \rightarrow \Psi(\mathbf{t}).$$

Consider two lags $\mathbf{t}, \mathbf{t}' \in \Lambda$, where $\|\mathbf{t}\| = \|\mathbf{t}'\|$. Since $w(\cdot)$ is an isotropic kernel function and $\Psi(\mathbf{t}) = \Psi(\mathbf{t}')$ under isotropy, we conclude that $\mathbb{E}[\hat{\Psi}_n(\mathbf{t})] = \mathbb{E}[\hat{\Psi}_n(\mathbf{t}')]$. Thus the

null hypothesis can be expressed, in terms of the expected sample second-order intensity functions at lags belonging to Λ , as

$$H_0 : E[\hat{\Psi}_n(\mathbf{t})] = E[\hat{\Psi}_n(\mathbf{t}')], \quad \mathbf{t}, \mathbf{t}' \in \Lambda, \mathbf{t} \neq \mathbf{t}', \text{ but } \|\mathbf{t}\| = \|\mathbf{t}'\|.$$

Further we form a contrast based on the above equations in H_0 , i.e., find a full row rank matrix \mathbf{A} such that $\mathbf{A}E(\hat{\mathbf{G}}_n) = \mathbf{0}$. For example, if $\Lambda = \{(1, 0), (0, 1), (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})\}$, then

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Thus instead of assessing isotropy directly, we will test the hypothesis $H_0 : \mathbf{A}E(\hat{\mathbf{G}}_n) = \mathbf{0}$. A similar technique has been applied by Lu and Zimmerman (2001) in testing for spatial isotropy for a quantitative spatial process.

Here and henceforth, we will use d to denote the row rank of \mathbf{A} . Define the following test statistic

$$TS_n \equiv |D_n| h_n^2 (\mathbf{A} \hat{\mathbf{G}}_n)' (\mathbf{A} \hat{\Sigma}_n \mathbf{A}')^{-1} (\mathbf{A} \hat{\mathbf{G}}_n),$$

where $\hat{\Sigma}_n$ is the subsampling estimator of Σ defined in Section 3.3.1.

Since $\hat{\Sigma}_n \xrightarrow{L_2} \Sigma$, $TS_n \xrightarrow{D} \chi_d^2$ as $n \rightarrow \infty$ by the multivariate Slutsky's theorem (Ferguson 1996). Thus an approximate size- α test for isotropy rejects H_0 if TS_n is bigger than $\chi_{d,\alpha}^2$, i.e., the upper α percentage point of a χ^2 distribution with d degrees of freedom.

3.3.3 Choice of Lag Set, Subblock Size, Bandwidth

The test statistic in Section 3.3.2 has the given asymptotic distribution for any choice of Λ , $l(n)$, h_n , assuming that conditions (3.1)-(3.6) are satisfied. Nevertheless, for any data set the performance of the test depends on the choice of these parameters. In this section, we will address these issues.

The choice of the lag set, Λ , plays an important role, especially in determining the power of the test. Generally speaking, there is no unique rule in determining this set. However, we often have some idea/suspicion about what may cause anisotropy, e.g., major highways for crime locations, wind for leukemia patients' locations. Such knowledge may help us in choosing directions to compare. The lengths of lags in Λ should be neither too large nor too small. The second-order intensity function at large lags are often close to a limiting constant, the square of the first-order intensity and thus the difference between them is small. The functions at very small lags, on the other hand, also tend to be very similar. Empirical results presented in Section 3.5 suggest a good choice is to set this length to be between 1/3-1/2 of the dependence range. This range can be gauged by studying an isotropic sample second-order intensity function plot. Specifically, we recommend to use the starting value beyond which the function becomes flat for the range.

The choice of the subblock size depends on a number of issues, including the dependence structure of the underlying point process and the definition of "optimal". Sherman (1996) showed that $cn^{1/2}$ for some $c > 0$ is the "optimal" rate for estimating the variance of a statistic calculated from quantitative lattice data, where the word optimal therein refers to minimizing the mean squared error. We conjecture a similar result will hold in the spatial point process setting. Empirical methods, such as those proposed by Hall and Jing (1996), Politis and Sherman (2001), can be applied in selecting this constant but require very large samples. Simulation results suggest when the first-order intensity is approximately equal to one, $c = 1$ is usually a reasonable choice. In light of this observation, we recommend to choose $c = 1$ in practice.

To select the bandwidth, we further assume that the second-order intensity function is twice continuously differentiable. A direct extension of Theorem III.1 yields

$$E[\hat{\Psi}_n(\mathbf{t})] = \Psi(\mathbf{t}) + \left\{ \int_C (\mathbf{x}'\Psi''(\mathbf{t})\mathbf{x})w^2(\mathbf{x})d\mathbf{x} \right\} \times h_n^2/2 + o(h_n^2).$$

Combining the above result and the variance expression of $\hat{\Psi}_n(\mathbf{t})$ in Section 3.2.2, we obtain the (asymptotic) optimal bandwidth choice in minimizing the mean squared error as

$$h_n = \left[\frac{2 \times \int_C w^2(\mathbf{x}) d\mathbf{x} \times \Psi(\mathbf{t})}{|D_n| \times \left\{ \int_C (\mathbf{x}' \Psi''(\mathbf{t}) \mathbf{x}) w^2(\mathbf{x}) d\mathbf{x} \right\}^2} \right]^{1/6} \equiv c(\mathbf{t}) \times \left[\frac{1}{|D_n|} \right]^{1/6}.$$

For a smaller but still relatively large field, say D_m , the optimal bandwidth h_m is approximately $c(\mathbf{t}) \times |D_m|^{-1/6}$. This suggests if an estimate for h_m , say \hat{h}_m , is available, then \hat{h}_n can be simply calculated as $\{|D_m|/|D_n|\}^{-1/6} \times \hat{h}_m$.

To estimate h_m , we first estimate $\Psi(\mathbf{t})$ by $\hat{\Psi}_n(\mathbf{t})$ using some pilot bandwidth h'_n . Then we split D_n into N_m subblocks of size $|D_m|$. Let $\hat{\Psi}_m^{i,j}(\mathbf{t})$ denote the estimate of the second-order intensity function calculated on subblock D_m^i using a bandwidth h_m^j , where h_m^j is from a pre-chosen bandwidth set, say $H \equiv \{h_m^j : j = 1, \dots, k\}$. Define the mean squared error associated with h_m^j as

$$MSE(m, h_m^j) \equiv \sum_{i=1}^{N_m} \{ \hat{\Psi}_m^{i,j}(\mathbf{t}) - \hat{\Psi}_n(\mathbf{t}) \}^2 / N_m.$$

Then \hat{h}_m is set to be the bandwidth in H that minimizes the above quantity. This procedure can be iterated by setting h'_n equal to \hat{h}_m until certain convergence criterion is satisfied. If the optimal bandwidth for estimating the second-order intensity function at a set of lags (instead of one particular lag) is needed, which is the case for our testing purpose, we simply set \hat{h}_m to be the one that minimizes the sum of the mean squared errors for all the lags.

3.4 A Simulation Study

3.4.1 Simulation Design

We consider realizations from a Poisson cluster process (see, e.g., Diggle 1983) on a 20×20 square field. For each realization, parent locations are generated from a Poisson process

with intensity .25 or 0.125, which correspond to 100 or 50 expected parents. Each parent produces S offspring according to a Poisson distribution with parameter equal to 5 or 10, respectively. Thus, both of these two set-ups lead to approximately 500 observations for each realization. This enables us to evaluate the effect of offspring number per parent on test performance. The position of an offspring relative to its parents follows a radially symmetric normal distribution with p.d.f.

$$f(x, y) = \frac{1}{2\pi\sigma^2 \times |\mathbf{B}|^{-1/2}} \exp\{-(x, y)\mathbf{B}(x, y)'/2\sigma^2\},$$

where \mathbf{B} is a 2×2 positive definite matrix and $|\mathbf{B}|$ here denotes its determinant. The parameter σ in the above expression defines the spread of each cluster; for the isotropic case, $2\sigma^2$ is the mean squared distance of an offspring from its parent. In the simulation, we set σ equal to 0.4 and 0.8 respectively. These are reasonable values in many real applications; see, for example, Cressie's (1991) analysis of long-leaf pine data, and Diggle's (1983) study of redwood seedlings. The \mathbf{B} matrix takes the following three values

$$\mathbf{B1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{B2} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{B3} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}.$$

Matrix $\mathbf{B1}$ yields isotropic point processes while $\mathbf{B2}$ and $\mathbf{B3}$ yield geometrically anisotropic point processes. More specifically, the main anisotropic axes of $\mathbf{B2}$ and $\mathbf{B3}$ are aligned with the (x, y) axes; the anisotropy ratio, defined as the ratio of the lengths of the main axes, is $\sqrt{2} : 1$ for $\mathbf{B2}$ and 2:1 for $\mathbf{B3}$.

One thousand realizations for each choice of \mathbf{B} are simulated. For each realization, sample second-order intensity functions at 0, 45, 90 and 135 degrees are calculated and compared. For $\sigma = .4$, the lengths of lags are set to be 0.4, 0.6, 0.8 and 1.0 while 0.8, 1.2, 1.6 and 2.0 are used for $\sigma = .8$. $l(n)$ is calculated as $20/N(D)^{1/4}$, which is in accordance with the aforementioned rate $n^{1/2}$. h is selected by the data-driven method introduced in Section 3.3.3.

We approximate the integral (3.7) by a finite sum, as suggested by Politis and Sherman (2001). Following the procedures given in Section 2.4.2, a 20×20 grid is laid on the field during each simulation. A square subblock window with the defined size is then moved across the field, with its bottom left corner starting from one of the nodes on the grid. On each subfield $D_{l(n)}^i$ and for each $\mathbf{t} \in \Lambda$, the following quantity is calculated:

$$\tilde{\Psi}_{l(n)}^i(\mathbf{t}) \equiv \iint \frac{w((\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2)/h_n)}{\sqrt{|D_{l(n)}^i \cap (D_{l(n)}^i - \mathbf{x}_1 + \mathbf{x}_2)|} \times h_n} N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2),$$

where the integrals are over $\{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1, \mathbf{x}_2 \in D_{l(n)}^i\}$. The (j, k) th element of Σ is then estimated by:

$$\frac{1}{k'_n} \times \sum_{i=1}^{k_n} [\tilde{\Psi}_{l(n)}^i(\mathbf{t}_j) - \bar{\Psi}_n(\mathbf{t}_j)] \times [\tilde{\Psi}_{l(n)}^i(\mathbf{t}_k) - \bar{\Psi}_n(\mathbf{t}_k)].$$

In the above expression, $\bar{\Psi}_n(\mathbf{t})$ denotes the average of all $\tilde{\Psi}_{l(n)}^i(\mathbf{t})$, $i = 1, \dots, k_n$ and k'_n is calculated as $k_n(1 - l(n)^2/n^2)$. The first modification (using $\tilde{\Psi}$) better estimates the dominant terms in the expression of Σ while the second (using k'_n) reduces bias. Observe that k'_n is the right standardizing constant when nonoverlapping subsampling is applied, for which k_n is approximately equal to $n^2/l(n)^2$. These modifications do not change our asymptotic results.

3.4.2 Results and Analysis

Table 3 reports the percentages of rejection from one thousand simulations at the 5% nominal level. The empirical sizes of the test are all reasonably close to 5%, except for $|\mathbf{t}| = 2.0$, which is almost half of the side length of each subsquare. Thus our test approximately achieves the nominal size for a variety of settings.

For relatively strong clustering ($\sigma = .4$), our testing method has excellent power in detecting anisotropy for different values of ρ and μ when the anisotropy ratio is only 1:2. As the degree of anisotropy increases, the power of the test increases. As σ increases, however,

the power drops significantly. This is because the strength of the clustering becomes weaker as the cluster spread increases and the process becomes more similar to a Poisson process. As a result, the true second-order intensity functions in different directions are closer to a directional independent constant, the square of the first-order intensity. When the true difference becomes smaller, the power of detecting such a difference becomes weaker as well.

Table 3. Simulation results from one thousand realizations of Poisson clustering processes. Each entry is the percentage of rejections at the 5% nominal level. ρ is the expected number of parents; μ is the expected number of offspring per parent.

(ρ, μ)	σ	$ t $	B1	B2	B3
(100, 5)	0.4	0.4	0.068	0.247	0.919
		0.6	0.053	0.510	0.982
		0.8	0.066	0.480	0.925
		1.0	0.069	0.296	0.756
	0.8	0.8	0.062	0.117	0.564
		1.2	0.044	0.180	0.577
		1.6	0.059	0.180	0.462
		2.0	0.079	0.138	0.250
(50, 10)	0.4	0.4	0.072	0.420	0.977
		0.6	0.062	0.676	0.992
		0.8	0.067	0.611	0.972
		1.0	0.074	0.442	0.848
	0.8	0.8	0.047	0.180	0.818
		1.2	0.066	0.289	0.800
		1.6	0.089	0.240	0.623
		2.0	0.103	0.191	0.402

The above paragraph indicates that the power of our testing method is closely related to the strength of clustering (as given by σ). The stronger the clustering effect, the more power the test has to detect anisotropy. The clustering effect also is strong if there are more offspring per parent. This in turn leads to larger power. As shown in Table 1, better powers are typically achieved for $\mu = 10$ than $\mu = 5$.

3.5 Applications

3.5.1 Crime Locations in Houston

Figure 3 plots locations of aggravated assault occurrences in the downtown Houston area during the year 2000. These locations are within a rectangle region (95.31 deg W-95.41 deg W, 29.765 deg N-29.815 deg N), with length roughly equal to 10 km in the east-west direction and width about 5 km in the north-south direction. The northern border of this area is along Highway 610, while Highway 10 crosses the southern part of it in the east-west direction. Two other major highways, Highway 59 and 45, run through this area both at approximately 90° direction, where 0° is defined as the west-east direction.

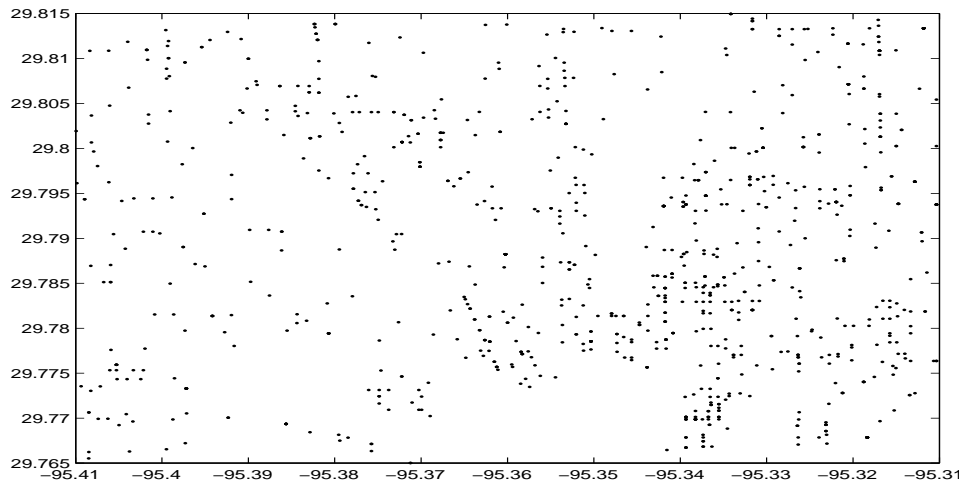


Figure 3. Locations of Aggravated Assault in an Area of Downtown Houston.

A potential source of anisotropy for this example is the direction of major roads since major roads provide convenient transportation for many human activities. For example, Nelson et al. (2001) find approximately 52% of violent crimes in the city centre of Cardiff, Britain, occurred along streets during the year 1993. Observe that roads in this area are mostly in the east-west or the north-south direction, we chose to compare sample second-order intensity functions in the 0° and 90° directions with the 45° and 135° directions. We

apply the proposed testing method to the following two lag sets,

$$\Lambda \equiv \{(1, 0), (0, 1), (\sqrt{2}/2, \sqrt{2}/2), (-\sqrt{2}/2, \sqrt{2}/2)\} \times 0.1 \text{ km},$$

$$\Lambda \equiv \{(1, 0), (0, 1), (\sqrt{2}/2, \sqrt{2}/2), (-\sqrt{2}/2, \sqrt{2}/2)\} \times 0.2 \text{ km},$$

respectively. The matrix \mathbf{A} is set as $[1, 1, -1, -1]'$.

The smoothing parameters are determined as 0.046 km for $\|\mathbf{t}\| = 0.1$ km and 0.092 km for $\|\mathbf{t}\| = 0.2$ km by the algorithm in Section 3.3.3 with $N_m = 16$. The subblock size is 1.57×0.91 km, which gives about 33 nonoverlapping replicates. The two calculated test statistics are 6.6313 and 0.5025, with respective p-values 0.01 and 0.4784. Thus strong anisotropy is detected when $\|\mathbf{t}\| = 0.1$ km, while for a larger distance, 0.2 km, such a pattern is not significant any more. We further examine the estimated second-order intensity functions for $\|\mathbf{t}\| = .1$ km and find that the estimates in the 0° , 90° directions are indeed much larger than those in the 45° and 135° directions. This confirms our conjecture that crime locations are more correlated in the major road directions.

3.5.2 *Childhood Leukemia Data*

The leukemia data consists of locations of leukemia cases diagnosed in the state of Texas between 1990 and 1998. For each case, a control matched on sex and date of birth is randomly selected from all births in Texas during that period. In this article, we will study observations within a rectangular area in the city of Houston. The western side of this rectangle borders with Highway 6 while its eastern, northern and southern sides are all along Highway 8. The Houston Ship Channel, an area of major petroleum refining and petrochemical industries, sits to the east of the region. 545 cases and 566 controls are observed in this region, respectively. Figure 4 plots the transformed locations of leukemia cases and controls in a (approximately) 25×22 field. The actual locations are not given for reasons of confidentiality.

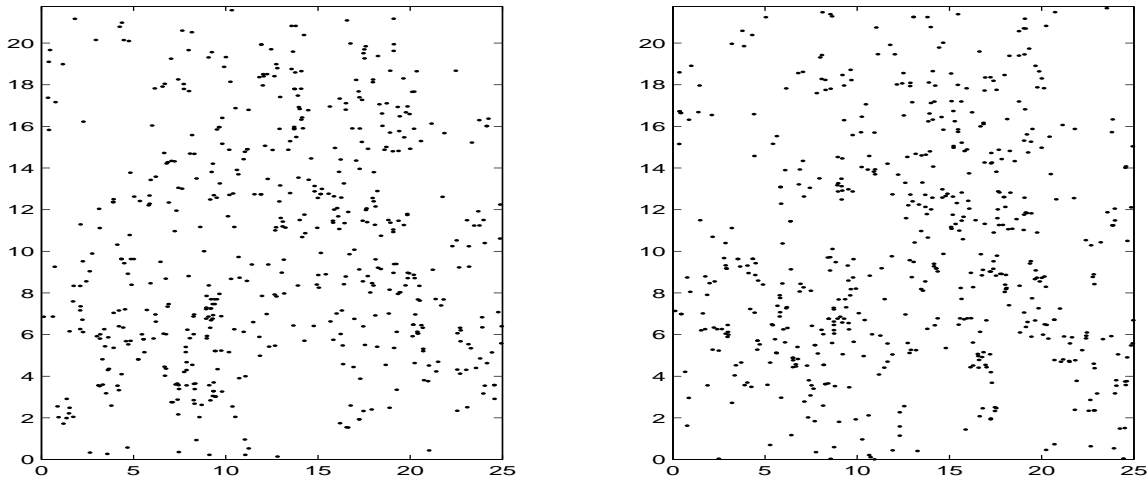


Figure 4. Locations of Leukemia Cases and Controls. The first graph plots locations of the cases, while the second plots those of the controls.

It is suspected that exposure to pollution, e.g., to agricultural pesticides in a rural area, to emissions of chemicals, say Benzene, from traffic and industry in an urban area, may increase the risk of developing leukemia. For our study region, there are two main possible sources of pollutions: local traffic and the ship channel to the east. Since roads in this area are mostly in the east-west or the north-south direction, and in particular, the ship channel is directly to the east of this region, we chose Λ to be

$$\{(0.6, 0), (0, 0.6), (0.6/\sqrt{2}, 0.6/\sqrt{2}), (-0.6/\sqrt{2}, 0.6/\sqrt{2})\}.$$

The length 0.6 is selected for $\|\mathbf{t}\|$ due to the clustering feature for cases (see Figure 5) and our finding that $1/3$ - $1/2$ of the range (≈ 1.5 here, see Figure 5) is a good choice for clustering processes. To compare the spatial distribution of the cases with that of the controls, we further test isotropy for the control group. The same directions are selected but a slightly smaller lag length, 0.5, is used. This is due to the property that the control group seems to have a shorter range than the case group. For both cases and controls, we thus set the matrix $\mathbf{A} = [1, 1, -1, -1]'$.

The bandwidths are chosen to be 0.26 for cases and 0.23 for controls determined by

the algorithm in Section 3.3.3 with $N_m = 9$. The subblock size is approximately 5.2×4.5 , which gives about 23 nonoverlapping replicates. The calculated test statistics are equal to 1.7601 for the cases and 0.2279 for the controls, with respective p-values 0.1846 and 0.6331. Thus we do not reject spatial isotropy for either the cases or the controls.

We further compare the strengths of clustering between these two groups. Cuzick and Edwards (1990) introduce a testing approach based on the so-called k nearest neighbors that utilizes the assumption of isotropy. In light of our testing results, their method can be applied to the leukemia data. In particular, we perform the test using their nearest neighbor approach, i.e., $k = 1$. This yields a test statistic equal to $-.5305$ and a p value equal to 0.7021. Thus we conclude there is no evidence of spatial clustering. This may be due to the fact that the field being studied here is fairly small and thus exposure to some risk factors, e.g., pollution, is relatively homogeneous for the whole region.

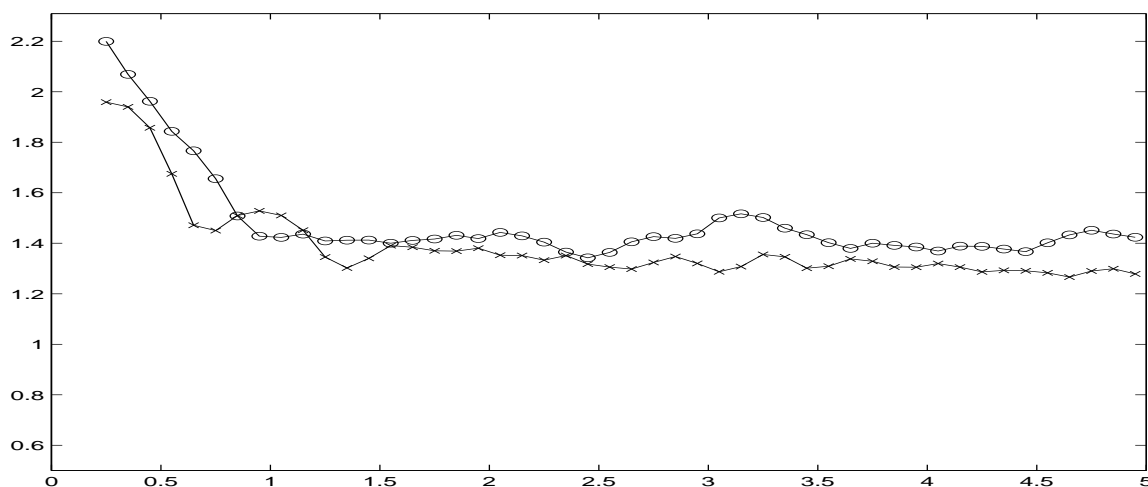


Figure 5. Isotropic Sample Second-order Intensity Function for Leukemia Data. \times : cases, \circ : controls.

CHAPTER IV

A TEST OF ISOTROPY FOR MARKED-POINT PROCESSES

4.1 Introduction

An important component of spatial analysis is to model the second-order structure of the underlying spatial process. A commonly-made assumption while modeling such a structure is that of intrinsic stationarity. Consider a spatial process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$, where \mathbf{s} denotes an arbitrary location where observation can take place. The process is said to be intrinsically stationary if

$$\begin{aligned} E[Z(\mathbf{s})] &= \mu, \quad \forall \mathbf{s} \in \mathbb{R}^2, \\ \text{Var}[Z(\mathbf{s} + \mathbf{t}) - Z(\mathbf{s})] &= \gamma(\mathbf{t}) \quad \forall \mathbf{s}, \mathbf{t} \in \mathbb{R}^2. \end{aligned} \quad (4.1)$$

Thus by definition an intrinsically stationary process has a constant mean structure and the variance between the difference of two observations is a function of their relative locations.

The function $\gamma(\cdot)$ defined by (4.1) is known as the variogram function, which plays an important role in spatial statistics, e.g., spatial prediction (see e.g., Cressie 1991), understanding the underlying process (see, e.g., Dibalasi and Bowman 2001). When observations occur on a regularly spaced grid, say D , the classical nonparametric, method-of-moments estimator of the variogram, i.e., the sample variogram, is defined as

$$\hat{\gamma}(\mathbf{t}) \equiv \frac{1}{|D(\mathbf{t})|} \times \sum [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2,$$

where the sum is over $D(\mathbf{t}) \equiv \{(i, j) : \mathbf{s}_i, \mathbf{s}_j \in D, \mathbf{s}_i - \mathbf{s}_j = \mathbf{t}\}$ and $|D(\mathbf{t})|$ is the number of distinct elements in $D(\mathbf{t})$.

Many spatial locations, however, are often irregularly scattered. Examples include precipitation in towns, heights of trees in a forest, and the size of cancer cells in a tissue.

When observations are irregularly scattered, the sample variogram is modified to

$$\hat{\gamma}(\mathbf{t}) \equiv \text{ave} \left\{ [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 : (i, j) \in D(\mathbf{s}), \mathbf{s} \in T(\mathbf{t}) \right\},$$

where the region $T(\mathbf{t})$ is some specified “tolerance” region around \mathbf{t} and $\text{ave}\{\cdot\}$ denotes a possibly weighted averaged over the elements in $\{\cdot\}$ (see, e.g., Cressie 1991). Observe that $T(\mathbf{t})$ plays essentially the same role as that of the bandwidth in nonparametric regression. The choice of $T(\mathbf{t})$ will directly affect how well a variogram function can be estimated. Journel and Huijbregts (1978) recommend that the number of distinct pairs in $T(\mathbf{t})$ be at least 30. Data-driven methods to select $T(\mathbf{t})$ are yet to be developed.

Once the sample variogram has been calculated, a parametric model can then be fit based on these estimates to ensure the “conditionally negative-definite” property of the variogram (see Cressie 1991 for detailed discussion of this property). Typically a class of isotropic models is used, mainly due to the simpler interpretation and ease of computation that the assumption of isotropy provides. However, many spatial processes are anisotropic (see, e.g., Cressie 1991, Hobert et al. 1997). Misspecification of an anisotropic process as isotropic typically leads to less efficient spatial prediction (see the example in Section 2.1). A conventional approach to check for isotropy is to informally assess plots of direction-specific sample (semi)variogram (see, e.g., Diggle 1981, Isaaks and Srivastava 1989, Cressie 1991). These diagnostics are often difficult to assess and are open to interpretation. This is especially true for irregularly spaced data since the variogram estimate can be sensitive to the “tolerance” region specified (Myers et al. 1982). More formal testing procedures for investigating isotropy have been proposed by Baczkowski and Mardia (1990) and Lu and Zimmerman (2001). Their methods, however, are not appropriate for irregularly spaced data.

In this chapter, we address the selection of $T(\mathbf{t})$ and a test for isotropy in developing the variogram function for irregularly spaced data. Here and henceforth, we make the as-

assumption that $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ arises from a spatial marked-point process, i.e., the index locations can be modeled by a spatial point process. We propose a cross-validation procedure to select the tolerance region. The optimal tolerance region $T(\mathbf{t})$ is defined as the one that gives the smallest value for a sample version of mean squared error. To test for isotropy, we first develop the asymptotic properties of the sample variogram. From these results we construct a test statistic with a limiting χ^2 distribution under the null hypothesis. This limiting distribution is used to compare sample variogram in multiple directions. Our testing approach is purely nonparametric in that no explicit knowledge (except for some mild moment conditions) of the marginal distribution of the process is necessary.

The rest of the chapter is organized as follows. In Section 4.2 we investigate the asymptotic properties of the sample variogram. Section 4.3 deals with bandwidth selection, while Section 4.4 tests for isotropy. In Section 4.5, we apply the proposed procedures to the long-leaf pine data given by Cressie (1991).

4.2 Preliminary Asymptotic Results

4.2.1 Notations and Set-up

Consider a marked-point process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$, where the point locations at which $Z(\cdot)$ is observed are generated from a second-order stationary point process N . We make the further assumption that N is independent of $Z(\cdot)$.

Let D be the domain of interest and ∂D be the boundary of D . In what follows, we use $|D|$ and $|\partial D|$ to denote the volume of the domain and the length of its boundary, respectively. Let h be a positive constant and $w(\cdot)$ be a bounded, nonnegative, isotropic density function which takes positive values on only a finite support, say $C (\subset \mathbb{R}^2)$. Here and henceforth, we use $d\mathbf{x}$ to denote an infinitesimally small disc centered at \mathbf{x} and $N(d\mathbf{x})$ to denote the number of points in $d\mathbf{x}$. Define $N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) \equiv N(d\mathbf{x}_1)N(d\mathbf{x}_2)I(\mathbf{x}_1 \neq \mathbf{x}_2)$, where $I(\mathbf{x}_1 \neq \mathbf{x}_2) = 1$ if $\mathbf{x}_1 \neq \mathbf{x}_2$ and 0 otherwise. The kernel variogram estimator is

given by

$$\hat{\gamma}(\mathbf{t}) = \frac{1}{\Psi(\mathbf{t})} \iint_D h^{-2} w\left(\frac{\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2}{h}\right) \times \frac{[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2}{|D \cap (D - \mathbf{x}_1 + \mathbf{x}_2)|} N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2), \quad (4.2)$$

where

$$\Psi(\mathbf{t}) \equiv \lim_{|d\mathbf{x}|, |d(\mathbf{x} + \mathbf{t})| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(d\mathbf{x}) \times N(d(\mathbf{x} + \mathbf{t}))]}{|d\mathbf{x}| \times |d(\mathbf{x} + \mathbf{t})|} \right\}.$$

In practice, $\Psi(\mathbf{t})$ is usually replaced by its estimate, see, e.g., Diggle (1983) for parametric procedures and Stoyan and Stoyan (1994) for nonparametric procedures to estimate $\Psi(\mathbf{t})$.

In what follows, we assume that $\Psi(\mathbf{t})$ is known. This extra assumption has minor effects on our major results, especially in testing for isotropy when the point process itself is isotropic (see Section 4.4.1).

4.2.2 Asymptotic Bias and Covariance

To establish the asymptotic consistency of the variogram estimator, we need to account for the shape of the random field and the choice of bandwidth. Consider a sequence of random fields, D_n , and a sequence of constants, h_n , where

$$|D_n| = O(n^2), |\partial D_n| = O(n), \text{ and} \quad (4.3)$$

$$h_n = O(n^{-\beta}) \text{ for some } \beta \in (0, 1). \quad (4.4)$$

Practically, condition (4.3) requires D_n to grow in all directions. This condition allows for a variety of field sequences, e.g., squares, circles and starshapes. Condition (4.4) was chosen to ensure sufficient averaging for each $\hat{\gamma}(\mathbf{t})$.

Let $\gamma^{(4)}(\mathbf{t}) \equiv \mathbb{E}\{[Z(\mathbf{t}) - Z(\mathbf{0})]^4\}$. As in Chapter 2, we require the following conditions on the mark process, $Z(\cdot)$.

$$\gamma(\mathbf{t}) \text{ and } \gamma^{(4)}(\mathbf{t}) \text{ are bounded and continuous,} \quad (4.5)$$

$$\int_{\mathbf{s} \in \mathbb{R}^2} \left| \text{Cov}\left\{ [Z(\mathbf{0}) - Z(\mathbf{s}_1)]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{s}_2)]^2 \right\} \right| ds < \infty \text{ for all finite } \mathbf{s}_1, \mathbf{s}_2, \quad (4.6)$$

Condition (4.6) is a weak dependence assumption. Any process that is m -dependent (i.e., observations separated by a distance larger than m are independent) with finite fourth moment satisfies this condition. For a Gaussian process, it can be shown that the absolute integrability of its covariance function, i.e., $\int_{\mathbb{R}^2} |R(\mathbf{t})| d\mathbf{t} < \infty$ is sufficient for (4.6) to hold. Many covariance models, e.g., Exponential, Gaussian, Spherical models, can be shown to satisfy the integrability condition and thus satisfy (4.6).

In addition to the above assumptions, some conditions on the point process N are also needed. Define the k th-order cumulant function as

$$C_N^{(k)}(\mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_k - \mathbf{x}_1) \equiv \lim_{|d\mathbf{x}_1|, \dots, |d\mathbf{x}_k| \rightarrow 0} \left\{ \frac{\text{Cum}[N(d\mathbf{x}_1), \dots, N(d\mathbf{x}_k)]}{|d\mathbf{x}_1| \times \dots \times |d\mathbf{x}_k|} \right\},$$

where $\text{Cum}(Y_1, \dots, Y_k)$ is given by the coefficient of $i^k t_1 \dots t_k$ in the Taylor series expansion of $\log\{E[\exp(i \sum_{j=1}^k Y_j t_j)]\}$ about the origin (see, Brillinger 1975). As in Masry (1983), we assume

$$C_N^{(2)}(\cdot), C_N^{(3)}(\cdot, \cdot) \text{ are bounded, } C_N^{(2)}(\cdot) \text{ is continuous and integrable,} \quad (4.7)$$

$$\int_{\mathbb{R}^2} |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_2)| d\mathbf{u}_1 < \infty, \int_{\mathbb{R}^2} |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_1 + \mathbf{u}_2)| d\mathbf{u}_1 < \infty \text{ for finite } \mathbf{u}_2, \text{ and} \\ \int_{\mathbb{R}^2} |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_3)| d\mathbf{u}_2 < \infty \text{ for finite } \mathbf{u}_1, \mathbf{u}_3. \quad (4.8)$$

The integrability conditions specified by (4.7) and (4.8) require the point process to be weakly dependent. These conditions are clearly satisfied for a homogeneous Poisson process, in which case $C_N^{(k)}(\mathbf{u}_1, \dots, \mathbf{u}_{k-1}) = 0$, $k \geq 2$, if $\mathbf{u}_i \neq \mathbf{0}$ for any $1 \leq i \leq k-1$. If a point process is m -dependent and $C_N^{(k)}$ is finite up to order four, then (4.7) and (4.8) also hold. It has been shown in Chapter 3 that these conditions are satisfied by commonly used Poisson cluster models such as the one to be discussed in Section 4.4.

The following theorem states under the above conditions, the sample variogram is a consistent estimator for the target variogram function.

Theorem IV.1. Consider a stationary marked-point process $\{Z(\mathbf{s}) : \mathbf{s} \in N \cap \mathbb{R}^2\}$.

Assume conditions (4.3)-(4.8), then

$$\mathbb{E}[\hat{\gamma}_n(\mathbf{t})] = \frac{1}{\Psi(\mathbf{t})} \int_C w(\mathbf{v}) \gamma(\mathbf{t} - h_n \mathbf{v}) \Psi(\mathbf{t} - h_n \mathbf{v}) d\mathbf{v} \rightarrow \gamma(\mathbf{t}) \text{ and}$$

$$\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times \text{Cov}[\hat{\gamma}_n(\mathbf{t}_1), \hat{\gamma}_n(\mathbf{t}_2)] = \int_C w(\mathbf{v})^2 d\mathbf{v} \times \frac{I(\mathbf{t}_1 = \pm \mathbf{t}_2) \times \gamma^{(4)}(\mathbf{t}_1)}{\Psi(\mathbf{t}_1)},$$

where $I(\mathbf{t}_1 = \pm \mathbf{t}_2) = 1$ if $\mathbf{t}_1 = \pm \mathbf{t}_2$ and 0 otherwise.

Proof. See Appendix C. □

4.2.3 Asymptotic Normality

Following Rosenblatt (1956), we can quantify the strength of dependence in the random field through the following strong mixing coefficients

$$\alpha_Z(p; k) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_Z(E_1), A_2 \in \mathcal{F}_Z(E_2),$$

$$E_2 = E_1 + \mathbf{s}, |E_1| = |E_2| \leq p, d(E_1, E_2) \geq k\}, \text{ and}$$

$$\alpha_N(p; k) \equiv \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_N(E_1), A_2 \in \mathcal{F}_N(E_2),$$

$$E_2 = E_1 + \mathbf{s}, |E_1| = |E_2| \leq p, d(E_1, E_2) \geq k\},$$

where the suprema are taken over all compact and convex subsets $E_1 \subset \mathbb{R}^2$, and over all $\mathbf{s} \in \mathbb{R}^2$ such that $d(E_1, E_2) \geq k$. In the above, $\mathcal{F}_Z(E)$ and $\mathcal{F}_N(E)$ denote the σ -algebras generated by the random variables $\{Z(\mathbf{s}) : \mathbf{s} \in E\}$ and $\{\mathbf{s} : \mathbf{s} \in N \cap E\}$, respectively.

Consider the following mixing conditions

$$\sup_p \frac{\alpha_Z(p; k)}{p} = \mathcal{O}(k^{-\epsilon}) \text{ for some } \epsilon > 2, \text{ and} \quad (4.9)$$

$$\sup_p \frac{\alpha_N(p; k)}{p} = \mathcal{O}(k^{-\epsilon}) \text{ for some } \epsilon > 2. \quad (4.10)$$

Conditions (4.9) and (4.10) require that at a fixed distance k , as the volume p increases, we allow the dependence to increase at a rate controlled by p . As the distance increases, the

dependence must decrease at a polynomial rate in k . Similar conditions have been used, e.g., in Sherman (1996), Heagerty and Lumley (2000), Politis and Sherman (2001).

We also require the following mild moment condition:

$$\sup_n \mathbf{E} \left\{ \left| \sqrt{|D_n|} \times h_n \times \left[\hat{\gamma}_n(\mathbf{t}) - \mathbf{E}\{\hat{\gamma}_n(\mathbf{t})\} \right] \right|^{2+\delta} \right\} \leq C_\delta \text{ for some } \delta > 0, C_\delta < \infty. \quad (4.11)$$

Condition (4.11) is only slightly stronger than the existence of the (standardized) asymptotic variance of $\hat{\gamma}_n(\mathbf{t})$. If the random field is m -dependent and N is Poisson, it can be shown that a finite eighth moment of $Z(\cdot)$ will be sufficient for condition (4.11) to hold.

Theorem IV.2. Consider a stationary marked-point process $\{Z(\mathbf{s}) : \mathbf{s} \in N \cap \mathbb{R}^2\}$. Assume conditions (4.3)-(4.11), then

$$\sqrt{|D_n|} \times h_n \times \{\hat{\gamma}_n(\mathbf{t}_1) - \mathbf{E}[\hat{\gamma}_n(\mathbf{t}_1)], \dots, \hat{\gamma}_n(\mathbf{t}_k) - \mathbf{E}[\hat{\gamma}_n(\mathbf{t}_k)]\}' \xrightarrow{D} N(\mathbf{0}, \Sigma)$$

where Σ is as given in Theorem IV.1.

Proof. See Appendix C. □

4.3 A Cross-validation Approach for Bandwidth Selection

The two theorems in Section 4.2 hold for any bandwidth h_n satisfying condition 4.4. Nevertheless, in any application the user needs to make a specific choice. The basic idea of cross-validation (see Stone 1974) is to build a model from one part of the data and assess the model using the remaining data. For any given model, an average prediction error over all the predicted data points can be computed and used as a criterion to evaluate the goodness-of-fit of the model. It is a very useful tool in selecting the smoothing parameter in the nonparametric regression setting, see, e.g. Hart (1997).

A common form of cross-validation is the so-called “leave-one-out” version. Suppose $(X_i, Y_i), i = 1, \dots, l$ represent observations from

$$Y_i = F(X_i) + \epsilon_i, \quad i = 1, \dots, l, \quad (4.12)$$

where $F(\cdot)$ is an unknown function and the ϵ_i are independent errors. Let $\hat{F}(\cdot; h)$ be a nonparametric estimator, e.g., a kernel estimator, of $F(\cdot)$ using smoothing parameter h ; $\hat{F}^{(i)}(\cdot; h)$ be the new estimator of the same type as $\hat{F}(\cdot; h)$ excluding (X_i, Y_i) . Define the cross-validation criterion $CV(h)$ as follows

$$CV(h) = \frac{1}{l} \sum_{i=1}^l [Y_i - \hat{F}^{(i)}(X_i; h)]^2. \quad (4.13)$$

The cross-validation choice of the smoothing parameter is the value of h that minimizes the above criterion.

In the case of estimating the variogram function, $F(\cdot)$ denotes the variogram function $\gamma(\cdot)$, while $(\mathbf{x}_1 - \mathbf{x}_2, [Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2)$ for $\mathbf{x}_1, \mathbf{x}_2 \in D$ are the observations, i.e., (X_i, Y_i) . To choose a bandwidth h for lag \mathbf{t} , we modify the cross-validation criterion in (4.13) as follows

$$CV(h) = \text{ave} \left\{ \left([Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2 - \hat{\gamma}^{(i)}(\mathbf{x}_1 - \mathbf{x}_2; h) \right)^2 : \mathbf{x}_1 - \mathbf{x}_2 \in T'(\mathbf{t}) \right\}, \quad (4.14)$$

where $\hat{\gamma}^{(i)}(\cdot; h)$ is the “leave-one-out” version of $\hat{\gamma}(\cdot)$ and $T'(\mathbf{t})$ denotes a new “tolerance” region that defines the pairs to be cross-validated. The bandwidth is then chosen as the value of h that minimizes (4.14).

The motivation for the method is as follows. Observe that

$$\text{Cov}\{[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2, \hat{\gamma}^{(i)}(X_i; h) | N\} \approx 0$$

due to the assumed weak dependence of the process and stationarity of the point process.

Thus $\text{E}\{CV(h) | N\}$ is approximately equal to the sum of

$$\begin{aligned} & \text{ave} \left\{ \text{Var} \{ [Z(\mathbf{x}_1) - Z(\mathbf{x}_2)]^2 \} : \mathbf{x}_1 - \mathbf{x}_2 \in T'(\mathbf{t}) \right\}, \text{ and} \\ & \text{ave} \left\{ \text{E} \{ [\hat{\gamma}^{(i)}(\mathbf{x}_1 - \mathbf{x}_2; h) - \gamma(\mathbf{x}_1 - \mathbf{x}_2)]^2 | N \} : \mathbf{x}_1 - \mathbf{x}_2 \in T'(\mathbf{t}) \right\}. \end{aligned} \quad (4.15)$$

When the total number of observations is large, (4.15) is approximately equal to

$$MASE(h) \equiv \text{ave} \left\{ \text{E} \{ [\hat{\gamma}(\mathbf{x}_1 - \mathbf{x}_2; h) - \gamma(\mathbf{x}_1 - \mathbf{x}_2)]^2 | N \} : \mathbf{x}_1 - \mathbf{x}_2 \in T'(\mathbf{t}) \right\}.$$

Thus the cross-validation bandwidth approximately minimizes $MASE(h)$. By setting $T'(\mathbf{t})$ to be a small neighborhood around \mathbf{t} , the chosen bandwidth is expected to be a good choice for estimating the variogram at any lag in $T'(\mathbf{t})$ which also includes \mathbf{t} . In practice, the cross-validation procedure is typically not sensitive to the choice of $T'(\mathbf{t})$ (see Section 4.5 for an application).

4.4 A Test for Isotropy

4.4.1 Test Statistic

In this section, we assume the point process is isotropic. This assumption can be tested by applying the techniques presented in Chapter 3. Recall from Theorem IV.1 that

$$E[\hat{\gamma}_n(\mathbf{t})] = \int_c w(\mathbf{x})\gamma(\mathbf{t} - h_n\mathbf{x})\Psi(\mathbf{t} - h_n\mathbf{x})d\mathbf{x}.$$

Consider two lags such that $\mathbf{t}_1 \neq \mathbf{t}_2$, but $\|\mathbf{t}_1\| = \|\mathbf{t}_2\|$. Since $w(\cdot)$ and $\Psi(\cdot)$ are both isotropic functions, $E[\hat{\gamma}_n(\mathbf{t}_1)] = E[\hat{\gamma}_n(\mathbf{t}_2)]$ under isotropy. This suggests the testable null hypothesis

$$H_0 : E[\hat{\gamma}_n(\mathbf{t}_1)] = E[\hat{\gamma}_n(\mathbf{t}_2)], \text{ if } \|\mathbf{t}_1\| = \|\mathbf{t}_2\|.$$

Let $\{\mathbf{t}_i : i = 1, \dots, k\}$ denote a set of user-chosen lags, $\mathbf{G}_n \equiv \{E[\hat{\gamma}_n(\mathbf{t}_i)] : i = 1, \dots, k\}'$ and $\hat{\mathbf{G}}_n \equiv \{\hat{\gamma}_n(\mathbf{t}_i) : i = 1, \dots, k\}'$. We then form appropriate contrasts $\mathbf{A}\mathbf{G}_n = \mathbf{0}$ and test the hypothesis $H_0 : \mathbf{A}\mathbf{G}_n = \mathbf{0}$ based on the sample contrasts $\mathbf{A}\hat{\mathbf{G}}_n$. The idea of testing isotropy in this manner is given by Lu and Zimmerman (2001) who consider equally spaced observations, and is also used in Chapter 2 where a simpler, smaller class of locations, i.e., Poisson process, is considered.

To formally compare $\mathbf{A}\hat{\mathbf{G}}_n$ with $\mathbf{0}$, we require knowledge of the covariance matrix of $\hat{\mathbf{G}}_n$. This matrix is typically unknown and thus needs to be estimated. We apply a subsampling technique to estimate it. Subsampling has been widely used to estimate the variance of a general spatial statistic on a regularly spaced grid (e.g., Hall 1988, Possolo

1991, Sherman and Carlstein 1994). Politis and Sherman (2001) extended the methodology to estimate the variance of statistics computed from marked-point processes. Here we consider a more general multivariate application of this method.

Let $D_{l(n)}$ be a subfield with the same shape as D_n , where $l(n) = cn^{1/2}$ for some positive constant c defines the size of a subfield. Sherman (1996) showed that $n^{1/2}$ is the “optimal” rate (in the sense of minimizing mean squared error) for estimating the variance of a statistic calculated from spatial lattice data. We conjecture a similar result holds in our setting.

In what follows, define $D_{l(n)}(\mathbf{x}) \equiv \{\mathbf{s} + \mathbf{x} : \mathbf{s} \in D_{l(n)}\}$ as the shifted copy of the $D_{l(n)}$. Then $\mathbf{x} \in D_n^{1-c}$ and $D_n^{1-c} \equiv \{\mathbf{x} \in D_n : D_{l(n)}(\mathbf{x}) \subset D_n\}$. Let $\hat{\mathbf{G}}_{l(n)}(\mathbf{x})$ be the sample variogram calculated on $D_{l(n)}(\mathbf{x})$ using a bandwidth $h_{l(n)}$. A subsampling estimator of $\Sigma \equiv |D_n|h_n^2 \text{Cov}(\hat{\mathbf{G}}_n, \hat{\mathbf{G}}_n)$ is then given by

$$\hat{\Sigma}_n \equiv \frac{1}{|D_n^{1-c}|} \int_{D_n^{1-c}} \left\{ |D_{l(n)}| h_{l(n)}^2 (\hat{\mathbf{G}}_{l(n)}(\mathbf{x}) - \bar{\mathbf{G}}_{l(n)}) (\hat{\mathbf{G}}_{l(n)}(\mathbf{x}) - \bar{\mathbf{G}}_{l(n)})' \right\} d\mathbf{x}, \quad (4.16)$$

where $\bar{\mathbf{G}}_{l(n)} \equiv \int_{D_n^{1-c}} \hat{\mathbf{G}}_{l(n)}(\mathbf{x}) d\mathbf{x} / |D_n^{1-c}|$.

Theorem IV.3. Assume that condition (4.11) holds with $\delta > 2$ and all the remaining conditions in Theorems IV.1 and IV.2 hold, then $\hat{\Sigma}_n$ is an L_2 consistent estimator for Σ , in that every element of $\hat{\Sigma}_n$ is L_2 consistent for its counterpart in Σ .

Proof. See Appendix C. □

In light of Theorem IV.3, we propose the following test statistic

$$TS_n \equiv |D_n| h_n^2 (\mathbf{A} \hat{\mathbf{G}}_n)' (\mathbf{A} \hat{\Sigma}_n \mathbf{A}')^{-1} (\mathbf{A} \hat{\mathbf{G}}_n).$$

$TS_n \xrightarrow{D} \chi_d^2$ as $n \rightarrow \infty$ by the multivariate Slutsky’s theorem (Ferguson 1996), where d denotes the row rank of \mathbf{A} . Thus an approximate size α test for isotropy is to reject H_0 if TS_n is bigger than $\chi_{d,\alpha}^2$, i.e., the upper α percentage point of a χ^2 distribution with d degrees of freedom.

4.4.2 Simulation Experiment

We consider realizations from a zero mean, second-order stationary Gaussian process on a 20×20 field. Note that the 20×20 field used in our simulation could be viewed by different users as, for example, a unit square or a 200×200 square. Our simulation results will remain unchanged as long as all pertinent quantities (e.g., subblock size, bandwidth) are transformed accordingly.

We first generated the observation locations by an isotropic Poisson cluster process. Specifically, cluster centers are determined by a Poisson process with intensity 0.2, the number of members per cluster is a Poisson random variable with parameter equal to 4, and the position of each member relative to the cluster center follows the probability density function

$$f(\mathbf{t}) = \frac{1}{2\pi\sigma^2} \exp\{-(t_x^2 + t_y^2)/2\sigma^2\}. \quad (4.17)$$

The above design yields (expected) 400 points on the field in each realization. The parameter σ in (4.17) defines the spread of each cluster in that $2\sigma^2$ is the mean squared distance of an offspring from its parent. In the simulation we set σ equal to 0.4 and 0.8 respectively. These values are similar to the estimates of σ in Cressie's (1991) analysis of long-leaf pine data and Diggle's (1983) study of redwood seedlings, respectively.

Given these simulated point locations, the values of the observations were then generated from a zero mean Gaussian process following the covariance structure

$$C(r; m) = \begin{cases} \theta(1 - \frac{3r}{2m} + \frac{r^3}{2m^3}) & \text{if } 0 \leq r \leq m \\ 0 & \text{otherwise,} \end{cases}$$

where $r = \sqrt{\mathbf{t}'\mathbf{B}\mathbf{t}}$ and \mathbf{B} is a 2×2 positive definite matrix. The parameter m defines the range and strength of dependence. In the simulation study, m was set to be 2, 5, and 8, which, relative to the size of the field, denote weak, moderate and strong spatial depen-

dence, respectively. The following three matrices of \mathbf{B} are used:

$$\mathbf{B1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{B2} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \quad \mathbf{B3} = \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix},$$

Matrix $\mathbf{B1}$ yields isotropic random fields while $\mathbf{B2}$ and $\mathbf{B3}$ yield geometrically anisotropic random fields. More specifically, the anisotropy ratio, defined as the ratio of the lengths of the main axes, is 2:1 for $\mathbf{B2}$ but 4:1 for $\mathbf{B3}$.

One thousand realizations were simulated for each choice of \mathbf{B} , m and σ . For each realization, we compared sample variogram at lags with unit length in four directions (0° , 45° , 90° and 135° , respectively). Three different bandwidths, 0.4, 0.7 and 1, were used to assess how sensitive the test is to different values of h . The bandwidth selection procedure introduced in Section 4.3 was not applied in the simulation for computational reasons. We will demonstrate its application while analyzing the long-leaf pine data in Section 4.5.

The subsampling window is set to be a 4×4 square window, with corresponding c value approximately equal to 0.9. Chapter 2 performs a sensitivity analysis regarding the subblock choice and find $c = 0.9$ performs reasonably well. The integral in (4.16) is then approximated in the same manner therein.

Table 4 presents the percentages of rejections at the 5% nominal level from one thousand simulations, where the column starting with $\mathbf{B1}$ gives the empirical sizes of the test. Observe that regardless of the strength of dependence, clustering strength and choice of bandwidth, the nominal size is approximately achieved. An interesting phenomenon is that for both $m = 5$ and $m = 8$ the empirical powers increase by 15%-75% when the value of σ changes from 0.4 to 0.8. Such a pattern is not seen for $m = 2$ due to the weak correlation and the consequent small differences among the variograms being compared. This observation indicates that strong clustering may result in less powerful testing results.

Table 4. Simulation results from five thousand realizations of marked Poisson cluster processes. Each entry is the percentage of rejections at the 5% nominal level. σ is a spread parameter for the radially symmetric normal distribution; m is a range parameter that defines the correlation strength among marks.

σ	m	h	B1	B2	B3
0.4	2	0.4	0.060	0.158	0.156
		0.7	0.052	0.128	0.133
		1.0	0.067	0.109	0.167
	5	0.4	0.073	0.318	0.732
		0.7	0.059	0.350	0.701
		1.0	0.062	0.283	0.462
	8	0.4	0.043	0.343	0.802
		0.7	0.050	0.391	0.791
		1.0	0.052	0.337	0.655
0.8	2	0.4	0.053	0.222	0.175
		0.7	0.054	0.173	0.127
		1.0	0.061	0.106	0.151
	5	0.4	0.057	0.475	0.908
		0.7	0.060	0.565	0.892
		1.0	0.068	0.495	0.725
	8	0.4	0.069	0.497	0.924
		0.7	0.076	0.602	0.949
		1.0	0.077	0.531	0.917

4.5 An Application

The long-leaf pine data consists of locations and diameter at breast height (dbh) of 584 long-leaf pine trees in a $200m \times 200m$ square field. The data set is given and analyzed by Cressie (1991), who detects spatial clustering among locations of trees. Cressie (1991) further fits an isotropic Neyman-Scott process, which is a stationary process, to these locations and obtains a reasonably good fit. Thus our test approach appears to be appropriate for this data set.

As in Chapter 2, we choose the lags to be compared as $(8, 0)$, $(0, 8)$, $(8/\sqrt{2}, 8/\sqrt{2})$

and $(-8/\sqrt{2}, 8/\sqrt{2})$. \mathbf{A} is thus as follows

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Using the cross-validation procedure in Section 4.3, the bandwidth was chosen to be 1.99, where the “tolerance region” $T'(t)$ is defined to be within a two-meter distance of the four selected lags (see Figure 6). We have also used $T'(t)$ to be within three and four-meter distance of the selected lags and obtained the same choice of bandwidth. To select the subblock size, we transform the original field to be a 24×24 square such that the intensity is approximately equal to one. In light of the simulation result, we then set the subshape to be a 4.4×4.4 square window, where the value 4.4 is calculated from $0.9 \times \sqrt{24}$ (i.e., $c = .9$, $n = 24$). This gives approximately thirty-one nonoverlapping subreplicates. The calculated test statistic is 4.9954. By comparing this value to a χ^2 distribution with three degrees of freedom, we obtain the p-value as 0.1721. Thus there is no strong evidence against the null hypothesis of spatial isotropy.

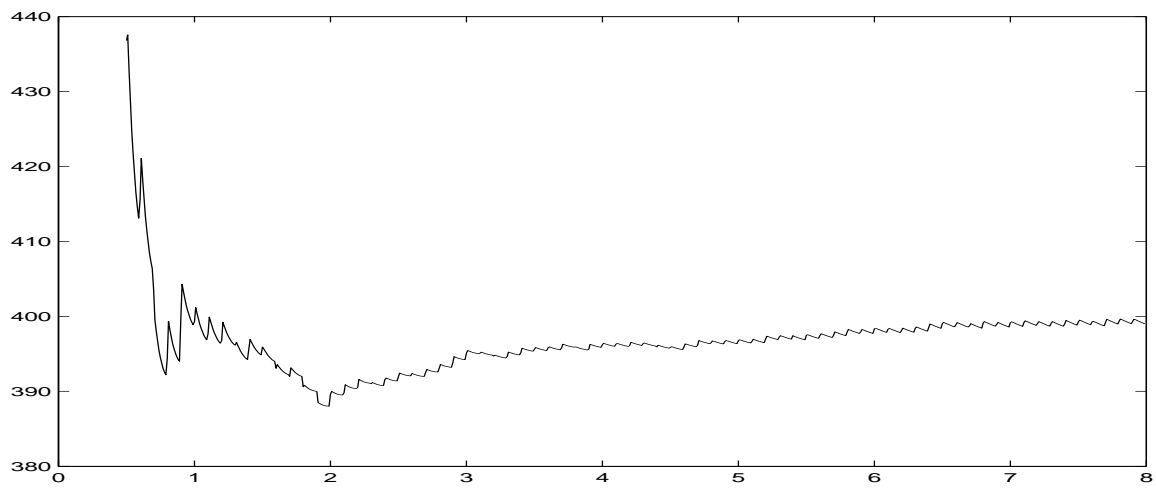


Figure 6. Plot of Cross-validation Sample Mean Squared Errors for Different Bandwidth Values. The cross-validation procedure is introduced in Section 4.3.

CHAPTER V

CONCLUSIONS AND FUTURE RESEARCH

In this dissertation, we have presented methodologies of testing for isotropy for both quantitative spatial processes and spatial point processes. The testing approach is based on the extent to which a sample second-order property, i.e., the sample variogram for quantitative spatial processes or the sample second-order intensity function for spatial point processes, satisfies a linear constraint implied by the null hypothesis. As an intermediate step, we established the asymptotic properties of the estimators of the second-order functions. An L_2 consistent subsampling estimator for the asymptotic covariance matrix of the estimators was derived and used to construct the test statistic with a limiting χ^2 distribution under the null hypothesis.

Our testing approach requires only very mild moment and weak dependence conditions on the underlying process and thus can be applied in a variety of settings. In addition, the results are appropriate for both regularly spaced and irregularly spaced data when the point locations are generated by a stationary point process. Four real data examples, wind speed measured over a region in the western tropical Pacific Ocean, long-leaf pine data from Cressie (1991), and crime data and leukemia patient locations, both in Houston, TX, have been analyzed to illustrate the testing approach. We have also conducted simulation studies for different settings that further demonstrate the efficacy of our approach.

If isotropy is rejected, fitting a parametric anisotropic model to the process is often of interest. Zimmerman (1993) provides a number of anisotropic models. In the future, we plan to focus on fitting geometrically anisotropic models, a class of anisotropic models that often provides a useful representation in many applications. For examples of these see Isaaks and Srivastava (1989) or Hobert et al. (1997). To obtain such a model, we

parameterize the matrix \mathbf{B} defined in Chapter 2 and Chapter 4 as follows

$$\mathbf{B} = \begin{bmatrix} 1 & \theta_1 \\ \theta_1 & \theta_2 \end{bmatrix} \quad \text{with } \theta_2 - \theta_1^2 \geq 0.$$

The geometrically anisotropic exponential model without a nugget effect (i.e., continuous at the origin) can be written as

$$\gamma(\mathbf{t}) = c_e \{1 - \exp[-a_e(\mathbf{t}'\mathbf{B}\mathbf{t})^{1/2}]\},$$

where $a_e > 0$, $c_e \geq 0$ (Zimmerman 1993). In what follows, we will denote the variogram at \mathbf{t} by $\gamma(\mathbf{t}; \theta)$, where $\theta \equiv (a_e, c_e, \theta_1, \theta_2)$.

Cressie (1985) proposed a weighted least squares method to fit parametric variogram models to regularly spaced spatial data. This method minimizes $W_{1,R}(\theta)$ with respect to θ , where

$$W_{1,R}(\theta) \equiv \sum_{i=1}^l \frac{|D(\mathbf{t}_i)|}{\gamma(\mathbf{t}_i; \theta)^2} [\hat{\gamma}(\mathbf{t}_i) - \gamma(\mathbf{t}_i; \theta)]^2,$$

and l denotes the number of sample variogram lags used to fit a model.

When observations are irregularly spaced, the above criterion needs to be modified. In light of our asymptotic results, we propose to minimize

$$W_{1,I}(\theta) \equiv \sum_{i=1}^l \frac{|D(\mathbf{t}_i)| h_i^2 \Psi(\mathbf{t}_i)}{\gamma^{(4)}(\mathbf{t}_i; \theta)^2} [\hat{\gamma}(\mathbf{t}_i) - \gamma(\mathbf{t}_i; \theta)]^2, \quad (5.1)$$

where $D(\mathbf{t}_i) \equiv D \cap D - \mathbf{t}_i$ and h_i is the bandwidth used to calculate $\hat{\gamma}(\mathbf{t}_i)$. The values of $\gamma^{(4)}(\mathbf{t}_i; \theta)$ in (5.1) can be estimated by a nonparametric procedure such as the one used to estimate $\gamma(\cdot)$ in Chapter 4. For a Gaussian process, it can also be replaced by $3\gamma(\mathbf{t}_i; \theta)^2$ since $\gamma^{(4)}(\mathbf{t}_i; \theta) = 3\gamma(\mathbf{t}_i; \theta)^2$ in that case.

In order to apply the modified least squares approach given by (5.1), one needs to decide on the set of lags to be used. In addition, the bandwidth used to calculate the sample variogram also has to be chosen. A weighted least square criterion which avoids such

choices can be defined as

$$W_2(\theta) \equiv \sum_{i=1}^n \sum_{j>i} \frac{\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 - \gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)\}^2}{\text{Var}\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2\}}, \quad (5.2)$$

where n denotes the number of observations that have been made. If the process is Gaussian, (5.2) can be rewritten as

$$W_2^G(\theta) \equiv \sum_{i=1}^n \sum_{j>i} \frac{\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 - \gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)\}^2}{\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)^2}.$$

The values of $\text{Var}\{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2\}$ and $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ can be estimated by $\hat{\gamma}^{(4)}(\mathbf{s}_i - \mathbf{s}_j) - \hat{\gamma}(\mathbf{s}_i - \mathbf{s}_j)^2$ and $\hat{\gamma}(\mathbf{s}_i - \mathbf{s}_j)$, respectively. The weighted least squares estimate of θ is the θ that minimizes (5.2).

The idea of fitting variogram models using squared differences of individual pairs is not entirely new; Curriero and Lele (1999) form a composite likelihood using such differences and obtain estimates of parameters for variogram models by maximizing the formed likelihood. The approach we propose here lies between those of Curriero and Lele (1999) and Cressie (1985), in that it utilizes information from all pairs but is still in the spirit of least squares.

For regularly spaced Gaussian data, we demonstrate that this method is essentially equivalent to Cressie's (1985) weighted least squares approach. To see that, we first differentiate $W_2^G(\theta)$ by treating the denominator as known since it would otherwise lead to an inconsistent estimator as pointed out by Curriero and Lele (1999)

$$\frac{\partial W_2^G(\theta)}{\partial \theta} = -2 \sum_{i=1}^n \sum_{j>i} \frac{[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 - \gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)}{\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)^2} \frac{\partial \gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)}{\partial \theta}.$$

If we focus on only the pairs that are separated by a lag in $\{\mathbf{t}_i : i = 1, \dots, l\}$, the above quantity can be rewritten as

$$-2 \sum_{i=1}^l \frac{|D(\mathbf{t}_i)| [\hat{\gamma}(\mathbf{t}_i) - \gamma(\mathbf{t}_i; \theta)]}{\gamma(\mathbf{t}_i; \theta)^2} \frac{\partial \gamma(\mathbf{t}_i; \theta)}{\partial \theta},$$

which is identical to $\frac{\partial W_{1,R}(\theta)}{\partial \theta}$. Thus these two methods will yield the same estimates. For irregularly spaced data, the newly proposed approach is expected to better capture directional differences, if there are any, due to the fact that no aggregation of variogram in different directions is applied.

REFERENCES

- Baczkowski, A. J. and Mardia, K. V. (1990), "A Test of Spatial Symmetry with General Application", *Communications in Statistics-Theory and Methods*, 19, 555–572.
- Bolthausen, E. (1982), "On the Central Limit Theorem for Stationary Mixing Random Fields", *The Annals of Probability*, 10, 1047–1050.
- Bradley, R. C. (1993), "Some Examples of Mixing Random Fields", *Rocky Mountain Journal of Mathematics*, 23, 495–519.
- Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*, New York: Holt, Rinehart & Winston.
- Carroll, R. J., Chen, R., George, E. I., Li, T. H., Newton, H. J., Schmiediche, H. and Wang, N. (1997), "Ozone Exposure and Population Density in Harris County, Texas", *Journal of the American Statistical Association*, 92, 392–404.
- Cressie, N. A. C. (1985), "Fitting Variogram Models by Weighted Least Squares", *Journal of the International Association for Mathematical Geology*, 17, 563–586.
- Cressie, N. A. C. (1991), *Statistics for Spatial Data*, New York: Wiley.
- Cressie, N. A. C. and Huang H. (1999), "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions", *Journal of the American Statistical Association*, 94, 1330–1340.
- Curriero, F. C. and Lele, S. (1999), "A Composite Likelihood Approach to Semivariogram Estimation", *Journal of Agricultural, Biological and Environmental Statistics*, 4, 9–28.
- Cuzick, J. and Edwards, R. (1990), "Spatial Clustering for Inhomogeneous Populations", *Journal of the Royal Statistical Society ser. B*, 19, 555–572.
- Daley, D. J. and Vere-Jones, D. (2002), *An Introduction to the Theory of Point Processes*,

Second Edition, New York: Springer.

- Dibiasi, A. and Bowman, A. W. (2001), “On the Use of the Variogram in Checking for Independence in Spatial Data”, *Biometrics*, 57, 211–218.
- Diggle, P. J. (1981), “Binary Mosaics and the Spatial Pattern of Heather”, *Biometrics*, 37, 531–539.
- Diggle, P. J. (1983), *Statistical Analysis of Spatial Point Patterns*, New York: Wiley.
- Diggle, P. J. and Chetwynd, A. G. (1991), “Second-order Analysis of Spatial Clustering for Inhomogeneous Populations”, *Biometrics*, 47, 1155–1163.
- Doukhan, P. (1994), *Mixing: Properties and Examples*, New York: Springer-Verlag.
- Ferguson, T. (1996), *A Course in Large Sample Theory*, Boca Raton: Chapman and Hall.
- Hall, P. (1988), “On Confidence Intervals for Spatial Parameters Estimated from Nonreplicated Data”, *Biometrics*, 44, 271–277.
- Hall, P. and Jing, B. (1996), “On Sample Reuse Methods for Dependent Data”, *Journal of the Royal Statistical Society ser. B*, 58, 727–737.
- Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer.
- Heagerty, P. J. and Lumley, T. (2000), “Window Subsampling of Estimating Functions with Application to Regression Models”, *Journal of the American Statistical Association*, 92, 846–854.
- Hobert, J. P., Altman, N. S. and Schofield, C. L. (1997), “Analyses of Fish Species Richness with Spatial Covariate”, *Journal of the American Statistical Association*, 95, 197–211.
- Ibragimov, I. A. and Linnik, Y. V. (1971), *Independent and Stationary Sequences of Random Variables*, Groningen: Wolters-Noordhoff.
- Isaaks, E. H. and Srivastava, R. M. (1989), *An Introduction to Applied Geostatistics*, Oxford: Oxford University Press.
- Jensen, J. L. (1993a) “A Note on Asymptotic Expansions for Sums Over a Weakly Dependent Random Field with Application to the Poisson and Strauss Processes”, *Annals of*

- the Institute of Statistical Mathematics*, 45, 353-360.
- Jensen, J. L. (1993b) "Asymptotic Normality of Estimates in Spatial Point Processes", *Scandinavian Journal of Statistics*, 20, 97-109.
- Journel, A. G. and Huijbregts, C. J. (1978), *Mining Geostatistics*, London: Academic Press.
- Karr, A. F. (1986), "Inference for Stationary Random Fields Given Poisson Samples", *Advanced Applied Probability*, 18, 406-422.
- Lu, H. and Zimmerman D. L. (2001), "Testing for Isotropy and Other Directional Symmetry Properties of Spatial Correlation" (personal collection, Y. Guan).
- Masry, E. (1983), "Nonparametric Covariance Estimation from Irregularly Spaced Data", *Advances in Applied Probability*, 15, 113-132.
- Matheron, G. (1963), "Principles of Geostatistics", *Economic Geology*, 58, 1246-1266.
- McCullagh, P. (1987), *Tensor Methods in Statistics*, New York: Chapman and Hall.
- Myers, D. E., Begovich, C. L., Butz, T. R. and Kane, V. E. (1982), "Variogram Models for Regional Groundwater Geochemical Data", *Mathematical Geology*, 14, 629-644.
- Nelson A. L., Bromley R. D. F. and Thomas C. J. (2001), "Identifying Micro and Temporal Patterns of Violent Crime and Disorder in the British City Center", *Applied Geography*, 21, 249-274.
- Ohser, J. and Stoyan, D. (1981), "On the Second-order and Orientation Analysis of Planar Stationary Point Processes", *Biometrical Journal*, 23, 523-533.
- Ord, J. K. and Rees, M. (1979), "Spatial Processes: Recent Developments with Applications to Hydrology", in *The Mathematics of Hydrology and Water Resources*, 95-118, eds. E. H. Lloyd, T. O'Donnell, and J. C. Wilkinson, London: Academic Press.
- Poisson, S. D. (1837), *Recherches sur la Probabilite des Jugements en Matiere Criminelle et en Matiere Civile, Precedees des Regles Generales du Calcul des Probabilites*, Paris: Bachelier.
- Politis, D. N., Paparoditis, E. and Romano, J. P. (1998), "Large Sample Inference for Irreg-

- ularly Spaced Dependent Observations Based on Subsampling”, *Sankhya*, ser. A, 60, 274–292.
- Politis, D. N. and Sherman, M. (2001), “Moment Estimation for Statistics From Marked Point Processes”, *Journal of the Royal Statistical Society*, ser. B, 63, 261–275.
- Possolo, A. (1991), “Subsampling a Random Field”, *Spatial Statistics and Imaging*, IMS Lecture Notes Monograph Series, Vol. 20, ed. A. Possolo, Hayward, CA: Institute of Mathematical Statistics, pp. 286–294.
- Ripley, B. D. (1976), “The Second-order Analysis of Stationary Point Processes”, *Journal of Applied Probability*, 13, 255–266.
- Rosenblatt, M. (1956), “A Central Limit Theorem and a Strong Mixing Condition”, *Biometrics*, 37, 531–539.
- Sherman, M. (1996), “Variance Estimation for Statistic Computed from Spatial Lattice Data”, *Journal of the Royal Statistical Society*, ser. B, 58, 509–523.
- Sherman, M. and Carlstein, E. (1994), “Nonparametric Estimation of the Moment of a General Statistic Computed from Spatial Data”, *Journal of the American Statistical Association*, 89, 496–500.
- Stone, M. (1974), “Cross-validatory Choice and Assessment of Statistical Predictions (with Discussion)”, *Journal of the Royal Statistical Society*, ser. B, 36, 111–147.
- Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes and Point Fields*, New York: Wiley.
- Zimmerman D. L. (1993), “Another Look at Anisotropy in Geostatistics”, *Mathematical Geology*, Vol. 25, No. 4, 453–470.

APPENDIX A

LEMMAS AND PROOF OF THEOREMS IN CHAPTER II

A.1 Lemmas

Lemma A.1. For a Gaussian process, the integrability of its covariance function, i.e., $\int_{\mathbf{s} \in \mathbb{R}^2} |R(\mathbf{s})| d\mathbf{s} < \infty$, is sufficient for (2.4) and (2.4') to hold.

Proof. Note the following results for Gaussian processes (Cressie, 1991)

$$\text{Var}\{[Z(\mathbf{x}) - Z(\mathbf{y})]^2\} = 2 \times \gamma(\mathbf{x} - \mathbf{y})^2,$$

$$\text{Corr}\{[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)]^2, [Z(\mathbf{y}_2) - Z(\mathbf{y}_1)]^2\} = \left\{ \text{Corr}[Z(\mathbf{x}_2) - Z(\mathbf{x}_1), Z(\mathbf{y}_2) - Z(\mathbf{y}_1)] \right\}^2.$$

Thus

$$\begin{aligned} & \text{Cov}\{[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)]^2, [Z(\mathbf{y}_2) - Z(\mathbf{y}_1)]^2\} \\ &= \left\{ \text{Corr}[Z(\mathbf{x}_2) - Z(\mathbf{x}_1), Z(\mathbf{y}_2) - Z(\mathbf{y}_1)] \right\}^2 \times \sqrt{2 \times \gamma(\mathbf{x}_2 - \mathbf{x}_1)^2 \times 2 \times \gamma(\mathbf{y}_2 - \mathbf{y}_1)^2} \\ &= 2 \times \gamma(\mathbf{x}_2 - \mathbf{x}_1) \times \gamma(\mathbf{y}_2 - \mathbf{y}_1) \times \left\{ \frac{\text{Cov}[Z(\mathbf{x}_2) - Z(\mathbf{x}_1), Z(\mathbf{y}_2) - Z(\mathbf{y}_1)]}{\sqrt{\text{Var}[Z(\mathbf{x}_2) - Z(\mathbf{x}_1)] \times \text{Var}[Z(\mathbf{y}_2) - Z(\mathbf{y}_1)]}} \right\}^2 \\ &= 2 \times \gamma(\mathbf{x}_2 - \mathbf{x}_1) \times \gamma(\mathbf{y}_2 - \mathbf{y}_1) \times \left\{ \frac{R(\mathbf{x}_2 - \mathbf{y}_2) + R(\mathbf{x}_1 - \mathbf{y}_1) - R(\mathbf{x}_2 - \mathbf{y}_1) - R(\mathbf{x}_1 - \mathbf{y}_2)}{\sqrt{\gamma(\mathbf{x}_2 - \mathbf{x}_1) \times \gamma(\mathbf{y}_2 - \mathbf{y}_1)}} \right\}^2 \\ &= 2 \times [R(\mathbf{x}_2 - \mathbf{y}_2) + R(\mathbf{x}_1 - \mathbf{y}_1) - R(\mathbf{x}_2 - \mathbf{y}_1) - R(\mathbf{x}_1 - \mathbf{y}_2)]^2. \end{aligned}$$

Thus

$$\begin{aligned} & \int_{\mathbf{s} \in \mathbb{R}^2} \left| \text{Cov}\{[Z(\mathbf{0}) - Z(\mathbf{s}_1)]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{s}_2)]^2\} \right| d\mathbf{s} \\ &= 2 \int_{\mathbf{s} \in \mathbb{R}^2} [R(\mathbf{s}) + R(\mathbf{s}_1 - \mathbf{s} - \mathbf{s}_2) - R(\mathbf{s} + \mathbf{s}_2) - R(\mathbf{s}_1 - \mathbf{s})]^2 d\mathbf{s} \\ &\leq 8R(\mathbf{0}) \int_{\mathbf{s} \in \mathbb{R}^2} [|R(\mathbf{s})| + |R(\mathbf{s}_1 - \mathbf{s} - \mathbf{s}_2)| + |R(\mathbf{s} + \mathbf{s}_2)| + |R(\mathbf{s}_1 - \mathbf{s})|] d\mathbf{s} \\ &< \infty \text{ for all finite } \mathbf{s}_1, \mathbf{s}_2, \end{aligned}$$

Thus condition (2.4') holds. Replacing the integral with summation, we have that condition (2.4) holds also. \square

Lemma A.2. Minkowski's inequality: Let X and Y be any two random variables, then for $1 \leq p < \infty$, $[\mathbf{E}|X + Y|^p]^{1/p} \leq [\mathbf{E}|X|^p]^{1/p} + [\mathbf{E}|Y|^p]^{1/p}$, i.e., $\mathbf{E}|X + Y|^p \leq \{[\mathbf{E}|X|^p]^{1/p} + [\mathbf{E}|Y|^p]^{1/p}\}^p$.

Lemma A.3. If D_n is a $n \times n$ square field and $Z(\cdot)$ is m -dependent with finite eighth moment, then condition (2.3'.b) holds for $\delta = 2$.

Proof. Split the original $n \times n$ square D_n into $[n/m']^2$ subsquares, where m' is a fixed value which is bigger than the sum of m and the length of lag t and $[n/m']$ denotes the largest integer that is smaller than or equal to n/m' . We use $l(n, m')$ to denote the length of each subsquare here and henceforth; $l(n, m')$ is finite and at least as big as m' . Group these subsquares into disjoint "blocks" of four, each block being $2l(n, m') \times 2l(n, m')$; label the four subsquares within a block (1,2,3,4), beginning with "1" in the lower left and proceeding clockwise through the block. Let k_n^j denote the number of subsquares with label j ($j = 1, 2, 3, 4$), and denote by $D_{l(n, m')}^{i, j}$, $i = 1, 2, \dots, k_n^j$, the i th subsquare with label j (see Figure 7).

By definition, we have

$$\begin{aligned}
 \hat{\gamma}_n(\mathbf{t}) &= \frac{1}{n^2} \times \iint_{D_n} [Z(\mathbf{x}) - Z(\mathbf{y})]^2 \times w_n(t - \mathbf{x} + \mathbf{y}) \times N^{(2)}(d\mathbf{x}, d\mathbf{y}) \\
 &= \frac{1}{n^2} \times \sum_{j=1}^4 \sum_{i=1}^{k_n^j} \int_{x \in D_{l(n, m')}^{i, j}} \int_{y \in D_n} [Z(\mathbf{x}) - Z(\mathbf{y})]^2 \times w_n(t - \mathbf{x} + \mathbf{y}) \times N^{(2)}(d\mathbf{x}, d\mathbf{y}) \\
 &= \underbrace{\sum_{j=1}^4 \frac{1}{n^2} \times \sum_{i=1}^{k_n^j} \int_{x \in D_{l(n, m')}^{i, j}} \int_{y \in D_n} [Z(\mathbf{x}) - Z(\mathbf{y})]^2 \times w_n(t - \mathbf{x} + \mathbf{y}) \times N^{(2)}(d\mathbf{x}, d\mathbf{y})}_{\hat{\gamma}_n^j(\mathbf{t})}
 \end{aligned}$$

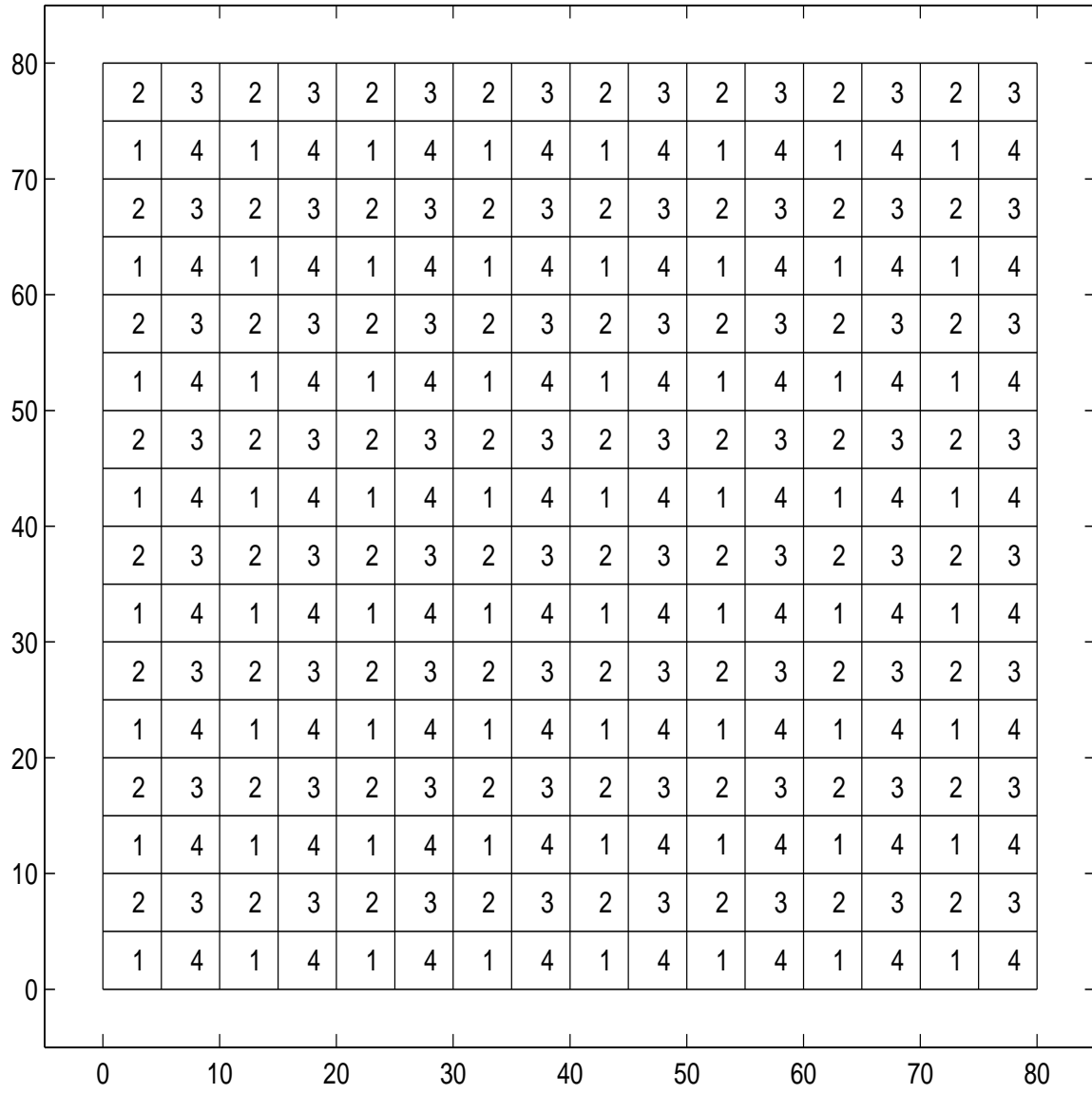


Figure 7. Partition of the Field for Lemma A.3

and $\mathbf{E}(\hat{\gamma}_n(\mathbf{t})) = \sum_{j=1}^4 \mathbf{E}(\hat{\gamma}_n^j(\mathbf{t}))$. To show (2.3'.b) holds, we only need to show

$$\sup_n \mathbf{E} \left\{ \left| n \times h_n \times \left[\hat{\gamma}_n^j(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_n^j(\mathbf{t})) \right] \right|^4 \right\} < \infty \text{ for each } j. \quad (\text{A.1.1})$$

due to Minkowski's inequality. Define $\hat{\gamma}_n^{i,j}(\mathbf{t}) \equiv \frac{1}{l(n,m')^2} \times \int_{\mathbf{x} \in D_{l(n,m')}^{i,j}} \int_{\mathbf{y} \in D_n} [Z(\mathbf{x}) - Z(\mathbf{y})]^2 \times w_n(t - \mathbf{x} + \mathbf{y}) \times N^{(2)}(d\mathbf{x}, d\mathbf{y})$. Then

$$n \times h_n \times \hat{\gamma}_n^j(\mathbf{t}) = l(n, m')^2 \times \frac{1}{n} \times \sum_{i=1}^{k_n^j} \{ \hat{\gamma}_n^{i,j}(\mathbf{t}) \times h_n \},$$

and

$$n \times h_n \times \left[\hat{\gamma}_n^j(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_n^j(\mathbf{t})) \right] = l(n, m')^2 \times \frac{1}{n} \times \sum_{i=1}^{k_n^j} \left\{ \left[\hat{\gamma}_n^{i,j}(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_n^{i,j}(\mathbf{t})) \right] \times h_n \right\}.$$

Since $n^2/k_n^j \rightarrow 4(m')^2$ and (m') and $l(n, m')$ are finite, (A.1.1) is equivalent to that

$$\sup_n \mathbf{E} \left\{ \left| \sqrt{\frac{1}{k_n^j}} \times \sum_{i=1}^{k_n^j} \underbrace{\left[\hat{\gamma}_n^{i,j}(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_n^{i,j}(\mathbf{t})) \right]}_{X_n^{i,j}} \times h_n \right|^4 \right\} < \infty \text{ for each } j. \quad (\text{A.1.2})$$

Since $Z(\cdot)$ is m -dependent and $l(n, m')$ is bigger than the sum of m and the length of lag t , we conclude that $X_n^{i,j}$ s are independent. Notice also $\mathbf{E}(X_n^{i,j})=0$. Thus

$$\begin{aligned} & \mathbf{E} \left\{ \left| \sqrt{\frac{1}{k_n^j}} \times \sum_{i=1}^{k_n^j} \left\{ \left[\hat{\gamma}_n^{i,j}(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_n^{i,j}(\mathbf{t})) \right] \times h_n \right\} \right|^4 \right\} \\ &= \mathbf{E} \left\{ \left| \sqrt{\frac{1}{k_n^j}} \times \sum_{i=1}^{k_n^j} X_n^{i,j} \right|^4 \right\} \\ &= \left(\frac{1}{k_n^j} \right)^2 \times \left\{ \sum_{i=1}^{k_n^j} \mathbf{E} \left[(X_n^{i,j})^4 \right] + \sum_{i_1 \neq i_2} \sum \mathbf{E} \left[(X_n^{i_1,j})^3 \times X_n^{i_2,j} \right] + \sum_{i_1 \neq i_2} \sum \mathbf{E} \left[(X_n^{i_1,j})^2 \times (X_n^{i_2,j})^2 \right] \right. \\ & \quad \left. + \sum_{i_1 \neq i_2 \neq i_3} \sum \sum \mathbf{E} \left[(X_n^{i_1,j})^2 \times X_n^{i_2,j} \times X_n^{i_3,j} \right] + \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \sum \sum \sum \sum \mathbf{E} \left[X_n^{i_1,j} \times X_n^{i_2,j} \times X_n^{i_3,j} \times X_n^{i_4,j} \right] \right\} \\ &= \left(\frac{1}{k_n^j} \right)^2 \times \left\{ \sum_{i=1}^{k_n^j} \mathbf{E} \left[(X_n^{i,j})^4 \right] + \sum_{i_1 \neq i_2} \sum \mathbf{E} \left[(X_n^{i_1,j})^2 \right] \times \mathbf{E} \left[(X_n^{i_2,j})^2 \right] \right\} \\ &= \underbrace{\left(\frac{1}{k_n^j} \right)^2 \times \sum_{i=1}^{k_n^j} \mathbf{E} \left[(X_n^{i,j})^4 \right]}_{(A)} + \underbrace{\left(\frac{1}{k_n^j} \right)^2 \times \sum_{i_1 \neq i_2} \sum \mathbf{E} \left[(X_n^{i_1,j})^2 \right] \times \mathbf{E} \left[(X_n^{i_2,j})^2 \right]}_{(B)}. \end{aligned}$$

Since $\mathbb{E}\left[(X_n^{i_1,j})^2\right]$ is finite, $\mathbb{E}\left[(X_n^{i_1,j})^2\right] \times \mathbb{E}\left[(X_n^{i_2,j})^2\right]$ is finite too. Notice there are $k_n^j \times (k_n^j - 1)$ terms in the summand (B), thus (B) is finite. To show (A) is finite, we only need to show for each i , $\mathbb{E}\left[(X_n^{i,j})^4\right] = o(k_n^j)$, i.e., $\mathbb{E}\left[(X_n^{i,j})^4\right] = o(n^2)$. Since $n^2 h_n^2 \rightarrow \infty$, $\frac{1}{h_n^2} = o(n^2)$. Thus $\mathbb{E}\left[(X_n^{i,j})^4\right] = O\left(\frac{1}{h_n^2}\right)$ will suffice for $\mathbb{E}\left[(X_n^{i,j})^4\right] = o(n^2)$ to hold and thus for (A) to be finite. By the definition of $X_n^{i,j}$ and Minkowski's inequality, we have

$$\begin{aligned} \mathbb{E}\left[(X_n^{i,j})^4\right] &= \mathbb{E}\left[\left(\hat{\gamma}_n^{i,j}(\mathbf{t}) - \mathbb{E}\left[\hat{\gamma}_n^{i,j}(\mathbf{t})\right]\right)^4 \times h_n^4\right] \\ &\leq h_n^4 \times \left\{ \left[\mathbb{E}\left(\hat{\gamma}_n^{i,j}(\mathbf{t})^4\right) \right]^{1/4} + \mathbb{E}\left(\hat{\gamma}_n^{i,j}(\mathbf{t})\right) \right\}^4. \end{aligned}$$

Thus we only need to show $h_n^4 \times \mathbb{E}\left(\hat{\gamma}_n^{i,j}(\mathbf{t})^4\right)$ and $h_n^4 \times \left[\mathbb{E}\left(\hat{\gamma}_n^{i,j}(\mathbf{t})\right)\right]^4$ are of order not higher than $\frac{1}{h_n^2}$. By definition, we have

$$\begin{aligned} \mathbb{E}\left(\hat{\gamma}_n^{i,j}(\mathbf{t})^4\right) &= \frac{1}{l(n, m')^8} \times \iiint\limits_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in D_{l(n, m')}} \iiint\limits_{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4 \in D_n} \\ &\mathbb{E}\left\{ \left[Z(\mathbf{x}_1) - Z(\mathbf{y}_1) \right]^2 \times \left[Z(\mathbf{x}_2) - Z(\mathbf{y}_2) \right]^2 \times \left[Z(\mathbf{x}_3) - Z(\mathbf{y}_3) \right]^2 \times \left[Z(\mathbf{x}_4) - Z(\mathbf{y}_4) \right]^2 \right\} \\ &\times w_n(t - \mathbf{x}_1 + \mathbf{y}_1) \times w_n(t - \mathbf{x}_2 + \mathbf{y}_2) \times w_n(t - \mathbf{x}_3 + \mathbf{y}_3) \times w_n(t - \mathbf{x}_4 + \mathbf{y}_4) \\ &\times \mathbb{E}\left[N^{(2)}(d\mathbf{x}_1, d\mathbf{y}_1) N^{(2)}(d\mathbf{x}_2, d\mathbf{y}_2) N^{(2)}(d\mathbf{x}_3, d\mathbf{y}_3) N^{(2)}(d\mathbf{x}_4, d\mathbf{y}_4) \right]. \end{aligned}$$

Notice the $\mathbb{E}\left[N^{(2)}(d\mathbf{x}_1, d\mathbf{y}_1) N^{(2)}(d\mathbf{x}_2, d\mathbf{y}_2) N^{(2)}(d\mathbf{x}_3, d\mathbf{y}_3) N^{(2)}(d\mathbf{x}_4, d\mathbf{y}_4) \right]$ term can be split into a number of terms as follows

$$\begin{aligned} & d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{x}_4 d\mathbf{y}_1 d\mathbf{y}_2 d\mathbf{y}_3 d\mathbf{y}_4 \\ &+ d\mathbf{x}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{x}_2) d\mathbf{x}_3 d\mathbf{x}_4 d\mathbf{y}_1 d\mathbf{y}_2 d\mathbf{y}_3 d\mathbf{y}_4 \\ &+ \dots \\ &+ d\mathbf{x}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{x}_2, d\mathbf{x}_3, d\mathbf{x}_4) d\mathbf{y}_1 \epsilon_{\mathbf{y}_1}(d\mathbf{y}_2, d\mathbf{y}_3) d\mathbf{y}_4 \\ &+ d\mathbf{x}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{x}_2, d\mathbf{x}_3, d\mathbf{x}_4) d\mathbf{y}_1 \epsilon_{\mathbf{y}_1}(d\mathbf{y}_2, d\mathbf{y}_3, d\mathbf{y}_4) \\ &+ d\mathbf{x}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2, d\mathbf{x}_3, d\mathbf{x}_4) d\mathbf{y}_1 \epsilon_{\mathbf{y}_1}(d\mathbf{x}_2, d\mathbf{y}_3, d\mathbf{y}_4) \end{aligned}$$

$$\begin{aligned}
& + \dots \\
& + dx_{\mathbf{1}} \epsilon_{\mathbf{x}_1}(dy_2, dy_3, dx_4) dy_{\mathbf{1}} \epsilon_{\mathbf{y}_1}(dx_2, dx_3, dy_4) \\
& + \dots \\
& + dx_{\mathbf{1}} \epsilon_{\mathbf{x}_1}(dy_2, dy_3, dy_4) dy_{\mathbf{1}} \epsilon_{\mathbf{y}_1}(dx_2, dx_3, dx_4).
\end{aligned}$$

The dominant terms of $h_n^4 \times \mathbb{E}(\hat{\gamma}_n^{i,j}(\mathbf{t})^4)$ are the ones including and after $dx_{\mathbf{1}} \epsilon_{\mathbf{x}_1}(dx_2, dx_3, dx_4) dy_{\mathbf{1}} \epsilon_{\mathbf{y}_1}(dy_2, dy_3, dy_4)$. Each of these terms is of order $\frac{1}{h_n^2}$ and all the rest are of order one or less. For example, the term $dx_{\mathbf{1}} \epsilon_{\mathbf{x}_1}(dx_2, dx_3, dx_4) dy_{\mathbf{1}} \epsilon_{\mathbf{y}_1}(dy_2, dy_3, dy_4)$ gives

$$\begin{aligned}
& h_n^4 \times \int_{\mathbf{x}_1 \in D_{l(n,m')}^{i,j}} \int_{\mathbf{y}_1 \in D_n} \mathbb{E}\left\{ [Z(\mathbf{x}_1) - Z(\mathbf{y}_1)]^8 \right\} \times [w_n(t - \mathbf{x}_1 + \mathbf{y}_1)]^4 dx_{\mathbf{1}} dy_{\mathbf{1}} \\
& = \frac{1}{h_n^4} \times \int_{\mathbf{x}_1 \in D_{l(n,m')}^{i,j}} \int_{\mathbf{y}_1 \in D_n} \mathbb{E}\left\{ [Z(\mathbf{x}_1) - Z(\mathbf{y}_1)]^8 \right\} \times \left[w\left(\frac{t - \mathbf{x}_1 + \mathbf{y}_1}{h_n}\right) \right]^4 dx_{\mathbf{1}} dy_{\mathbf{1}}.
\end{aligned}$$

Since $|D_{l(n,m')}^{i,j}|$ is finite, the eighth moment of $Z(\cdot)$ exists and $h_n \rightarrow 0$ (and thus $1/h_n \rightarrow \infty$), a simple change of variable gives that the above expression is of order $\frac{1}{h_n^2}$. Now consider a different term, $dx_{\mathbf{1}} \epsilon_{\mathbf{x}_1}(dx_2, dx_3, dx_4) dy_{\mathbf{1}} \epsilon_{\mathbf{y}_1}(dy_2, dy_3) dy_4$,

$$\begin{aligned}
& h_n^4 \times \int_{x \in D_{l(n,m')}^{i,j}} \int_{\mathbf{y}_1 \in D_n} \int_{\mathbf{y}_4 \in D_n} \mathbb{E}\left\{ [Z(\mathbf{x}_1) - Z(\mathbf{y}_1)]^6 \times [Z(\mathbf{x}_1) - Z(\mathbf{y}_4)]^2 \right\} \\
& \times [w_n(t - \mathbf{x}_1 + \mathbf{y}_1)]^3 \times w_n(t - \mathbf{x}_1 + \mathbf{y}_4) dx_{\mathbf{1}} dy_{\mathbf{1}} dy_4 \\
& = \frac{1}{h_n^4} \times \int_{x \in D_{l(n,m')}^{i,j}} \int_{\mathbf{y}_1 \in D_n} \int_{\mathbf{y}_4 \in D_n} \mathbb{E}\left\{ [Z(\mathbf{x}_1) - Z(\mathbf{y}_1)]^6 \times [Z(\mathbf{x}_1) - Z(\mathbf{y}_4)]^2 \right\} \\
& \times \left[w\left(\frac{t - \mathbf{x}_1 + \mathbf{y}_1}{h_n}\right) \right]^3 \times w\left(\frac{t - \mathbf{x}_1 + \mathbf{y}_4}{h_n}\right) dx_{\mathbf{1}} dy_{\mathbf{1}} dy_4.
\end{aligned}$$

A simple change of variable shows the above expression is of order one. Thus $h_n^4 \times \mathbb{E}(\hat{\gamma}_n^{i,j}(\mathbf{t})^4)$ is of order $\frac{1}{h_n^2}$. Similarly we may show that $h_n^4 \times \left[\mathbb{E}(\hat{\gamma}_n^{i,j}(\mathbf{t})) \right]^4$ is of order h_n^4 . Thus we conclude (A) is finite and (A.1.2) holds, i.e., condition (2.3'.b) holds. \square

Lemma A.4. Assume condition (2.2), we have $|D^{m(n)}|/|D_n| \rightarrow 1$ as $n \rightarrow \infty$, where $D^{m(n)}$ is as defined in the proof for Theorem II.1 and II.2 in Section A.2.

Proof. Introduce the following notations,

$$D_{n/l(n)} : \text{field in } D_n \text{ but not in } D^{l(n)},$$

$$D_{l(n)/m(n)} : \text{field in } D^{l(n)} \text{ but not in } D^{m(n)},$$

$k'(n)$: minimal number of extra $l(n) \times l(n)$ subsquares needed to cover the whole D_n ,

$D_{n'}$: union of all extra $l(n) \times l(n)$ subsquares needed to cover the whole D_n .

Since

$$|D^{l(n)}| = k_n |D_{l(n)}^i| = k_n l_n^2 = k_n n^{2\alpha},$$

$$|D^{m(n)}| = k_n |D_{m(n)}^i| = k_n m_n^2 = k_n \{n^{2\alpha} + n^{2\beta} - 2n^{\alpha+\beta}\},$$

$$k_n n^{2\beta} = o(k_n n^{2\alpha}), \quad k_n n^{\alpha+\beta} = o(k_n n^{2\alpha}) \text{ due to } \beta < \alpha,$$

we have $|D_{l(n)/m(n)}| = |D^{l(n)}| - |D^{m(n)}|$ is of order $o(|D^{l(n)}|)$.

Now we want to show $|D_{n/l(n)}|$ is of order no larger than $O(nl(n))$. Notice $D_{n/l(n)} \subset D_{n'}$, i.e., $|D_{n/l(n)}| < |D_{n'}|$, thus $|D_{n'}|$ is of order no larger than $O(nl(n))$ will be sufficient for $|D_{n/l(n)}|$ to be of order no larger than $O(nl(n))$. Since $|D_{n'}| = k'_n \times l_n^2$, to show $|D_{n'}|$ is of order no larger than $O(nl(n))$ is equivalent to show k'_n is of order no larger than $O(n/l(n))$.

To show this, we split the boundary of D_n into lines whose lengths are all $l(n)$. Here we use $L_{n,i}$ to denote the i th line and k''_n denote the number of all the lines available. k''_n is of order $O(n/l(n))$ due to condition (2.2). For every $L_{n,i}$, we may form a $l(n) \times l(n)$ square (denoted by $SQ_{l(n),i}$) which fully contains $L_{n,i}$; in addition to that, we form a $3l(n) \times 3l(n)$ square (denoted as $SQ_{3l(n),i}$) which has the same center as $SQ_{l(n),i}$ (see Figure 8).

Notice any $l(n) \times l(n)$ square that intersects with $L_{n,i}$ (and thus with $SQ_{l(n),i}$) will be fully contained in $SQ_{3l(n),i}$. Since every square in $D_{n'}$ intersects with the boundary and thus with at least one of these $L_{n,i}$ s, it will be fully contained in one of the $SQ_{3l(n),i}$ s. Notice

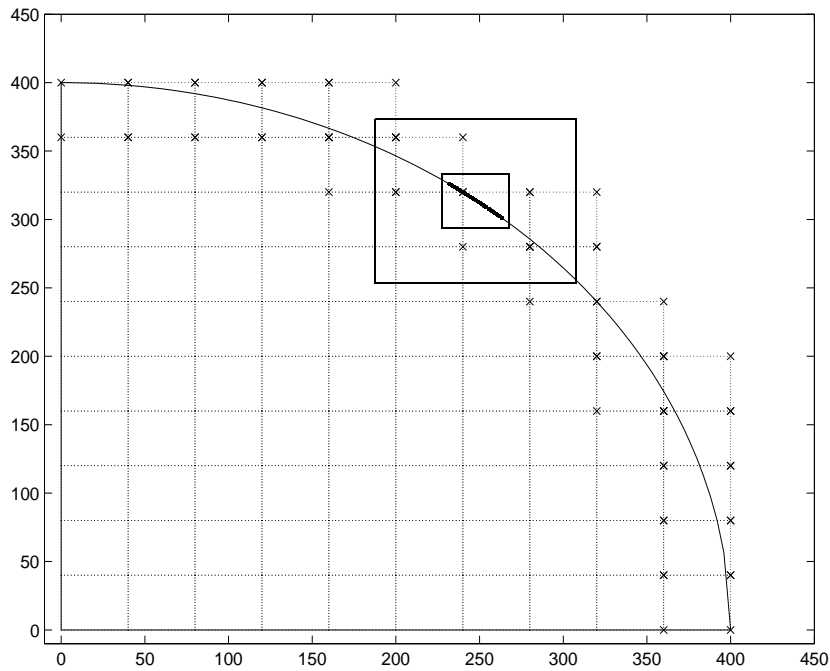


Figure 8. Partition of the Field for Lemma A.4. \times marks boundary subsquares.

these $l(n) \times l(n)$ squares in D_n' do not intersect with each other except at the boundary. The maximum number of such squares that intersect with $L_{n,i}$ can not be larger than 9. Since size of k_n'' is of order $O(n/l(n))$, we conclude k_n' is of order no larger than $O(n/l(n))$ due to $k_n' < 9 \times k_n''$. Thus $|D_{n/l(n)}|$ is of order no larger than $O(nl(n))$. Then

$$\begin{aligned}
 & |D_n| - |D^{m(n)}| \\
 &= [|D_n| - |D^{l(n)}|] + [|D^{l(n)}| - |D^{m(n)}|] \\
 &= |D_{n/l(n)}| + |D_{l(n)/m(n)}| \\
 &= O(nl(n)) + o(|D^{l(n)}|) \\
 &= o(n^2) + o(|D_n|) \text{ (since } D^{l(n)} \subset D_n) \\
 &= o(|D_n|) \text{ (since } |D_n| = O(n^2)).
 \end{aligned}$$

Thus $|D^{m(n)}|/|D_n| \rightarrow 1$ as $n \rightarrow \infty$. □

Lemma A.5. Consider two compact and convex sets U, V in R^2 such that $|U| = |V| \leq p$ and $d(U, V) \geq k$. Let X and Y be measurable random variables with respect to $\mathcal{F}(U)$ and $\mathcal{F}(V)$ such that $|X| < C_1$ and $|Y| < C_2$. Then

$$\text{Cov}(X, Y) \leq 4C_1C_2\alpha_p(k).$$

Proof. The following proof is analogous to that of Theorem 17.2.1 in Ibragimov and Linnik (1971 p. 306).

$$\begin{aligned} |\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)| &= |\mathbf{E}\{\mathbf{E}(XY|\mathcal{F}(U))\} - \mathbf{E}(X)\mathbf{E}(Y)| \\ &= |\mathbf{E}\{X[\mathbf{E}(Y|\mathcal{F}(U)) - \mathbf{E}(Y)]\}| \leq C_1\mathbf{E}|\mathbf{E}(Y|\mathcal{F}(U)) - \mathbf{E}(Y)| \\ &\equiv C_1\mathbf{E}\{u[\mathbf{E}(Y|\mathcal{F}(U)) - \mathbf{E}(Y)]\}, \end{aligned}$$

where $u = \text{sign}\{\mathbf{E}(Y|\mathcal{F}(U)) - \mathbf{E}(Y)\}$. Clearly u is measurable with respect to $\mathcal{F}(U)$, and therefore

$$|\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)| \leq C_1|\mathbf{E}(uY) - \mathbf{E}(u)\mathbf{E}(Y)|.$$

Now we consider $|\mathbf{E}(uY) - \mathbf{E}(u)\mathbf{E}(Y)|$. Since u and Y are measurable random variables with respect to $\mathcal{F}(U)$ and $\mathcal{F}(V)$, we may apply the above arguments to get

$$|\mathbf{E}(uY) - \mathbf{E}(u)\mathbf{E}(Y)| \leq C_2|\mathbf{E}(uv) - \mathbf{E}(u)\mathbf{E}(v)|,$$

where $v = \text{sign}\{\mathbf{E}(u|\mathcal{F}(V)) - \mathbf{E}(u)\}$. And thus

$$|\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)| \leq C_1C_2|\mathbf{E}(uv) - \mathbf{E}(u)\mathbf{E}(v)|.$$

Introducing the events:

$$\begin{aligned} A &\equiv \{u = 1\} \in \mathcal{F}(U), \bar{A} \equiv \{u = -1\} \in \mathcal{F}(U), \\ B &\equiv \{v = 1\} \in \mathcal{F}(V), \bar{B} \equiv \{v = -1\} \in \mathcal{F}(V). \end{aligned}$$

The definition of the mixing coefficient gives

$$\begin{aligned}
|\mathbf{E}(uv) - \mathbf{E}(u)\mathbf{E}(v)| &= |P(AB) + P(\bar{A}\bar{B}) - P(\bar{A}B) - P(A\bar{B}) \\
&\quad - P(A)P(B) - P(\bar{A})P(\bar{B}) + P(\bar{A})P(B) + P(A)P(\bar{B})| \\
&\leq |P(AB) - P(A)P(B)| + |P(\bar{A}\bar{B}) - P(\bar{A})P(\bar{B})| \\
&\quad + |P(\bar{A}B) - P(\bar{A})P(B)| + |P(A\bar{B}) - P(A)P(\bar{B})| \\
&\leq 4\alpha_p(k).
\end{aligned}$$

And thus $\text{Cov}(X, Y) \leq 4C_1C_2\alpha_p(k)$ □

Note: if the variables X, Y are complex, then separating the real and imaginary parts, we again arrive at the same expression, with 4 replaced by 16.

Lemma A.6. Lyapounov's Theorem: suppose $\{X_{n,i}\}$ is a independent triangular array satisfying $\mathbf{E}(X_{n,i}) = 0$ and for some $\delta > 0$,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} \frac{\mathbf{E}(|X_{n,i}|^{2+\delta})}{\sigma_n^{2+\delta}} = 0,$$

where r_n is number of elements in the n th row of the triangular array and $\sigma_n^2 = \sum_{k=1}^{r_n} \text{var}(X_{n,i})$.

Then $\sum_{k=1}^{r_n} X_{n,i}/\sigma_n \rightarrow N(0, 1)$.

A.2 Proof of Theorems

A.2.1 Proof of Theorem II.1

Proof. Consider the covariance term, $\text{Cov}[\hat{\gamma}_n(\mathbf{t}_i), \hat{\gamma}_n(\mathbf{t}_j)]$, where $\mathbf{t}_i, \mathbf{t}_j \in \Lambda$.

$$\begin{aligned}
&\frac{1}{|D_n(\mathbf{t}_i)| \times |D_n(\mathbf{t}_j)|} \sum_{D_n(\mathbf{t}_i)} \sum_{D_n(\mathbf{t}_j)} \text{Cov} \left\{ \left[Z(\mathbf{s}_1) - Z(\mathbf{s}_1 + \mathbf{t}_i) \right]^2, \left[Z(\mathbf{s}_2) - Z(\mathbf{s}_2 + \mathbf{t}_j) \right]^2 \right\} \\
&= \frac{1}{|D_n(\mathbf{t}_i)| \times |D_n(\mathbf{t}_j)|} \sum_{D_n(\mathbf{t}_j) - D_n(\mathbf{t}_i)} \sum_{D_n(\mathbf{t}_i) \cap (D_n(\mathbf{t}_j) - \mathbf{s})} \text{Cov} \left\{ \left[Z(0) - Z(\mathbf{t}_i) \right]^2, \left[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t}_j) \right]^2 \right\} \\
&= \sum_{D_n(\mathbf{t}_i) - D_n(\mathbf{t}_j)} \text{Cov} \left\{ \left[Z(0) - Z(\mathbf{t}_i) \right]^2, \left[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t}_j) \right]^2 \right\} \times \frac{|D_n(\mathbf{t}_i) \cap (D_n(\mathbf{t}_j) - \mathbf{s})|}{|D_n(\mathbf{t}_i)| \times |D_n(\mathbf{t}_j)|}
\end{aligned}$$

Applying conditions (2.2), (2.4) and the Kronecker's lemma, we conclude

$$|D_n| \times \text{Cov}[\hat{\gamma}_n(\mathbf{t}_i), \hat{\gamma}_n(\mathbf{t}_j)] \rightarrow \sum_{\mathbf{s} \in \mathbb{Z}^2} \text{Cov}\{[Z(\mathbf{0}) - Z(\mathbf{t}_i)]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t}_j)]^2\}.$$

Let $\sigma^2 \equiv \sum_{\mathbf{s} \in \mathbb{Z}^2} \text{Cov}\{[Z(\mathbf{0}) - Z(\mathbf{t})]^2, [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t})]^2\}$, $S_n \equiv \sqrt{|D_n|} \times [\hat{\gamma}_n(\mathbf{t}) - \gamma(\mathbf{t})]$.

Now we prove $S_n \xrightarrow{D} N(0, \sigma^2)$. To do so, we apply a blocking technique (e.g., Ibragimov and Linnik, 1971) in conjunction with the mixing condition (2.1).

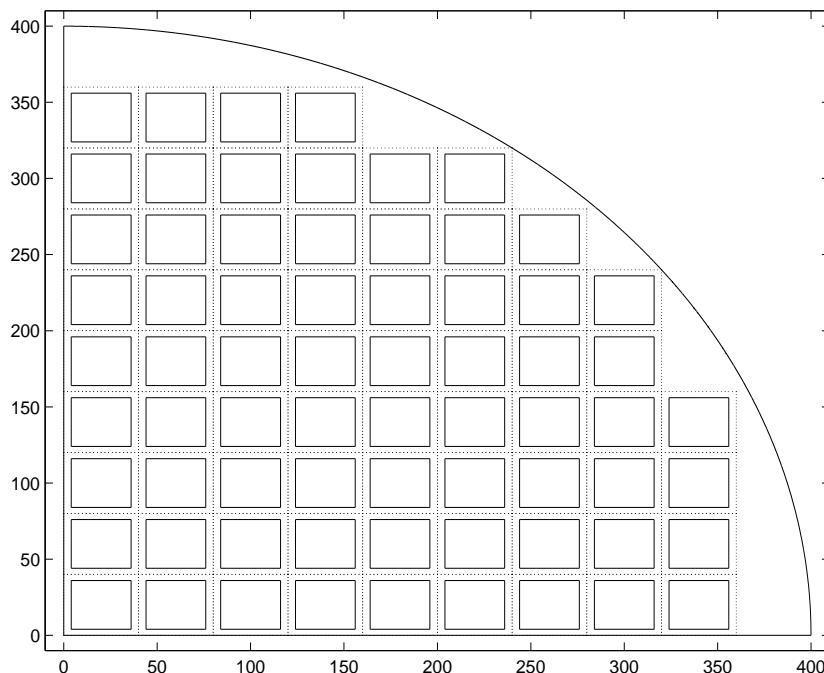


Figure 9. Partition of the Field for Theorem II.1

Let $l(n) = n^\alpha$, $m(n) = n^\alpha - n^\eta$ for some $4/(2 + \epsilon) < \eta < \alpha < 1$. Divide the original field D_n into nonoverlapping $l(n) \times l(n)$ subsquares, $D_{l(n)}^i$, $i = 1, \dots, k_n$; within each subsquare, further obtain $D_{m(n)}^i$ which shares the same center as $D_{l(n)}^i$ (see Figure 9). Thus $d(D_{m(n)}^i, D_{m(n)}^j) \geq n^\eta$ for $i \neq j$. Let $\hat{\gamma}_{m(n)}^i(\mathbf{t})$ denote the sample variogram obtained from $D_{m(n)}^i$. Let $s_n \equiv \sum_{i=1}^{k_n} s_n^i / \sqrt{k_n}$, $s'_n \equiv \sum_{i=1}^{k_n} (s_n^i)' / \sqrt{k_n}$, where $s_n^i \equiv m(n) \times [\hat{\gamma}_{m(n)}^i(\mathbf{t}) - \gamma(\mathbf{t})]$ and $(s_n^i)'$ have the same marginal distributions as s_n^i but are independent. Let $\phi'_n(x)$ and $\phi_n(x)$ be the characteristic functions of s'_n and s_n respectively. The proof consists of

the following three steps.

$$\text{S1 } S_n - s_n \xrightarrow{p} 0;$$

$$\text{S2 } \phi'_n(x) - \phi_n(x) \rightarrow 0;$$

$$\text{S3 } s'_n \xrightarrow{D} N(0, \sigma^2).$$

Proof of S1: Since $E(S_n - s_n) = 0$, it suffices to show $\text{Var}(S_n - s_n) \rightarrow 0$. Let $D^{m(n)}$ denote the union of all $D_{m(n)}^i$. Observe that

$$\begin{aligned} s_n &= k_n^{-\frac{1}{2}} \times \sum_{i=1}^{k_n} s_n^i \\ &= k_n^{-\frac{1}{2}} \times \sum_{i=1}^{k_n} \left\{ m(n) \times [\hat{\gamma}_{m(n)}^i(\mathbf{t}) - \gamma(\mathbf{t})] \right\} \\ &= k_n^{-\frac{1}{2}} \times m(n) \times \sum_{i=1}^{k_n} \frac{1}{|D_{m(n)}^i|} \times \sum_{s \in D_{m(n)}^i} \left\{ [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t})]^2 - \gamma(\mathbf{t}) \right\} \\ &= \frac{m(n)}{\sqrt{k_n} \times |D_{m(n)}^i(\mathbf{t})|} \times \sum_{i=1}^{k_n} \sum_{s \in D_{m(n)}^i} \left\{ [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t})]^2 - \gamma(\mathbf{t}) \right\} \\ &= \frac{m(n)}{\sqrt{k_n} \times |D_{m(n)}^i(\mathbf{t})|} \times \sum_{s \in D^{m(n)}} \left\{ [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t})]^2 - \gamma(\mathbf{t}) \right\} \\ &= \frac{\sqrt{k_n} \times m_n}{\sqrt{|D^{m(n)}(\mathbf{t})|}} \times \sum_{s \in D^{m(n)}} \left\{ [Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t})]^2 - \gamma(\mathbf{t}) \right\} \\ &= \sqrt{k_n} \times m_n \times \left\{ \hat{\gamma}_{D^{m(n)}}(\mathbf{t}) - \gamma(\mathbf{t}) \right\} \\ &= \sqrt{|D_{m(n)}|} \times \left\{ \hat{\gamma}_{D^{m(n)}}(\mathbf{t}) - \gamma(\mathbf{t}) \right\}. \end{aligned}$$

Thus $\text{Var}(s_n) \rightarrow \sigma^2$ from the proof for the covariance term and $\text{Cov}(S_n, s_n)$ is equal to

$$\frac{\sqrt{|D_n|} \times |D^{m(n)}|}{|D_n(\mathbf{t})| \times |D^{m(n)}(\mathbf{t})|} \times \sum_{D_n(\mathbf{t})} \sum_{D^{m(n)}(\mathbf{t})} \text{Cov} \left\{ \left[Z(\mathbf{s}_1) - Z(\mathbf{s}_1 + \mathbf{t}) \right]^2, \left[Z(\mathbf{s}_2) - Z(\mathbf{s}_2 + \mathbf{t}) \right]^2 \right\},$$

where the summation can be rewritten as

$$\sum_{D^{m(n)}(\mathbf{t}) - D_n(\mathbf{t})} \text{Cov} \left\{ \left[Z(0) - Z(\mathbf{t}) \right]^2, \left[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{t}) \right]^2 \right\} \times |D_n(\mathbf{t}) \cap (D^{m(n)}(\mathbf{t}) - \mathbf{s})|.$$

Further observe that

$$\frac{\sqrt{|D_n| \times |D^{m(n)}| \times |D_n(\mathbf{t}) \cap (D^{m(n)}(\mathbf{t}) - \mathbf{s})|}}{|D_n(\mathbf{t})| \times |D^{m(n)}(\mathbf{t})|} \rightarrow 1$$

for any fixed \mathbf{s} due to Lemma A.4. Thus $\text{Cov}(S_n, s_n) \rightarrow \sigma^2$ and $\text{Var}(S_n - s_n) \rightarrow 0$.

Proof of S2: In what follows, we use I (instead of the commonly used notation i) to denote the imaginary number. By definition,

$$\begin{aligned}\phi_n(x) &= \mathbf{E}\{\exp(Ix s_n)\} = \mathbf{E}\left\{\exp\left(Ix \sum_{i=1}^{k_n} \frac{s_n^i}{\sqrt{k_n}}\right)\right\}, \\ \phi'_n(x) &= \mathbf{E}\{\exp(Ix s'_n)\} = \mathbf{E}\left\{\exp\left(Ix \sum_{i=1}^{k_n} \frac{(s_n^i)'}{\sqrt{k_n}}\right)\right\}.\end{aligned}$$

Since $(s_n^i)'$, $i = 1, 2, \dots, k_n$, are independent and have the same marginal distribution as s_n^i , $i = 1, 2, \dots, k_n$, $\phi'_n(x)$ can be rewritten as $\prod_{i=1}^{k_n} \mathbf{E}\left\{\exp\left(Ix \frac{s_n^i}{\sqrt{k_n}}\right)\right\}$. Define

$$U_i \equiv \exp\left(Ix \frac{s_n^i}{\sqrt{k_n}}\right),$$

Then

$$\begin{aligned}& |\phi_n(x) - \phi'_n(x)| \\ &= \left| \mathbf{E}\left\{\prod_{i=1}^{k_n} U_i\right\} - \prod_{i=1}^{k_n} \mathbf{E}\{U_i\} \right| \\ &= \left| \mathbf{E}\left\{\prod_{i=1}^{k_n} U_i\right\} - \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} \times \mathbf{E}\{U_{k_n}\} + \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} \times \mathbf{E}\{U_{k_n}\} - \prod_{i=1}^{k_n} \mathbf{E}\{U_i\} \right| \\ &\leq \left| \mathbf{E}\left\{\prod_{i=1}^{k_n} U_i\right\} - \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} \times \mathbf{E}\{U_{k_n}\} \right| + \left| \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} \times \mathbf{E}\{U_{k_n}\} - \prod_{i=1}^{k_n} \mathbf{E}\{U_i\} \right| \\ &\leq \left| \mathbf{E}\left\{\prod_{i=1}^{k_n} U_i\right\} - \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} \times \mathbf{E}\{U_{k_n}\} \right| + \left| \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} - \prod_{i=1}^{k_n-1} \mathbf{E}\{U_i\} \right| \times |\mathbf{E}\{U_{k_n}\}| \\ &\leq \left| \mathbf{E}\left\{\prod_{i=1}^{k_n} U_i\right\} - \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} \times \mathbf{E}\{U_{k_n}\} \right| + \left| \mathbf{E}\left\{\prod_{i=1}^{k_n-1} U_i\right\} - \prod_{i=1}^{k_n-1} \mathbf{E}\{U_i\} \right|.\end{aligned}$$

The last inequality in the above expression is due to $|\mathbf{E}\{U_{k_n}\}| \leq 1$. By induction, we have

$$|\phi_n(x) - \phi'_{t,n}(x)| \leq \sum_{j=1}^{k_n-1} \left| \mathbf{E}\left\{\prod_{i=1}^{j+1} U_i\right\} - \mathbf{E}\left\{\prod_{i=1}^j U_i\right\} \times \mathbf{E}\{U_{j+1}\} \right| = \sum_{j=1}^{k_n-1} \left| \text{Cov}\left\{\prod_{i=1}^j U_i, U_{j+1}\right\} \right|.$$

Define

$$X_j = \prod_{i=1}^j U_i, \quad Y_j = U_{j+1}.$$

Since X_j is measurable with respect to $\mathcal{F}(\bigcup_{i=1}^j D_{m(n)}^i)$, Y_j is measurable with respect to $\mathcal{F}(D_{m(n)}^{j+1})$, $|X_j| \leq 1$, $|Y_j| \leq 1$ and $|D_{m(n)}^{j+1}| \leq |\bigcup_{i=1}^j D_{m(n)}^i| = j \times m(n)^2$, we have

$$\text{Cov}(X_j, Y_j) \leq 16\alpha_{j \times m(n)^2}(n^\eta) \leq 16j \times m(n)^2 \times n^{-\eta\epsilon} = 16j \times \{n^{2\alpha} + n^{2\eta} - 2n^{\alpha+\eta}\} \times n^{-\eta\epsilon}$$

by the mixing condition. Thus

$$|\phi_{t,n}(\lambda) - \phi'_{t,n}(\lambda)| \leq \sum_{j=1}^{k_n-1} \left\{ 16j \times n^{2\alpha} \times n^{-\eta\epsilon} \right\} = \mathcal{O}(n^{2\alpha-\eta\epsilon} \times k_n^2) = \mathcal{O}(n^{4-2\alpha-\eta\epsilon}).$$

The last equality in the above expression is due to $\mathcal{O}(k_n) = \mathcal{O}(n^2/l(n)^2) = \mathcal{O}(n^{2-2\alpha})$. Since $4/(2+\epsilon) < \eta < \alpha < 1$, we have $4 - 2\alpha - \eta\epsilon < 4 - 2\eta - \eta\epsilon < 4 - (2+\epsilon) \times 4/(2+\epsilon) = 0$.

Thus $|\phi'_n(x) - \phi_n(x)| \rightarrow 0$ as $n \rightarrow \infty$.

Proof of S3: Observe that $\mathbb{E}(|(s_n^i)'|^{2+\delta}) < C_\delta$ for a constant C_δ . Since $(s_n^i)'$ are i.i.d.,

$$\text{Var} \left[\sum_{i=1}^{k_n} (s_n^i)' \right] = k_n \times \text{Var}((s_n^i)').$$

Define $\sigma_n^2 \equiv \text{Var}((s_n^i)'),$ we have $\sigma_n^2 \rightarrow \sigma^2$ by the proof of S1. Thus

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \frac{\mathbb{E}(|(s_n^i)'|^{2+\delta})}{\left[\sqrt{\text{Var}(\sum_{i=1}^{k_n} (s_n^i)')} \right]^{2+\delta}} \leq \lim_{n \rightarrow \infty} C_\delta \times \frac{k_n}{[k_n \sigma_n^2]^{(2+\delta)/2}} = 0.$$

Thus apply Lyapounov's Theorem, we have

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

To prove the joint normality, we apply the Cramer-Wold device. Consider the case where $\mathbf{\Lambda} = \{\mathbf{t}_1, \mathbf{t}_2\}$, the more general case follows in a similar way.

Let a, b be two arbitrary real numbers such that $|a| \leq 1, |b| \leq 1$ and at least one of them is not zero. Define

$$S_n \equiv a\sqrt{|D_n|} \times (\hat{\gamma}_n(\mathbf{t}_1) - \gamma(\mathbf{t}_1)) + b\sqrt{|D_n|} \times (\hat{\gamma}_n(\mathbf{t}_2) - \gamma(\mathbf{t}_2)) \equiv S_n(\mathbf{t}_1) + S_n(\mathbf{t}_2),$$

$$s_n = a s_n(\mathbf{t}_1) + b s_n(\mathbf{t}_2), \quad s'_n = a s_n(\mathbf{t}_1)' + b s_n(\mathbf{t}_2)',$$

where $s_n(\mathbf{t}_i)$ and $s_n(\mathbf{t}_i)'$, $i = 1, 2$ are as being defined in the univariate case. Let $\phi_n(x)$, $\phi'_n(x)$ be the characteristic functions of s_n and s'_n , respectively. Define

$$\sigma^2 = \lim_{n \rightarrow \infty} |D_n| \times \{a^2 \text{Var}(\hat{\gamma}_n(\mathbf{t}_1)) + 2ab \text{Cov}(\hat{\gamma}_n(\mathbf{t}_1), \hat{\gamma}_n(\mathbf{t}_2)) + b^2 \text{Var}(\hat{\gamma}_n(\mathbf{t}_2))\}.$$

The proof of this theorem consists of the following three steps

S1 $S_n - s_n \xrightarrow{p} 0$;

S2 $\phi'_n(x) - \phi_n(x) \rightarrow 0$;

S3 $s'_n \xrightarrow{d} N(0, \sigma^2)$.

These three steps combined together suggest $S_n \xrightarrow{d} N(0, \sigma^2)$. Since this is true for arbitrary a, b , we conclude that $(S_n(\mathbf{t}_1), S_n(\mathbf{t}_2))$ is asymptotically jointly normal.

Proof of S1 is trivial because $[S_n(\mathbf{t}_1) - s_n(\mathbf{t}_1)]$ and $[S_n(\mathbf{t}_2) - s_n(\mathbf{t}_2)]$ converge to zero in probability as n increases.

To prove S2, note that

$$\begin{aligned} \phi_n(x) &= \mathbf{E} \left\{ \exp(Ix s_n) \right\} = \mathbf{E} \left\{ \exp \left(Ix \sum_{i=1}^{k_n} \frac{a s_n^i(\mathbf{t}_1) + b s_n^i(\mathbf{t}_2)}{\sqrt{k_n}} \right) \right\}, \\ \phi'_n(x) &= \mathbf{E} \left\{ \exp(Ix s'_n(t)) \right\} = \mathbf{E} \left\{ \exp \left(Ix \sum_{i=1}^{k_n} \frac{a s_n^i(\mathbf{t}_1)' + b s_n^i(\mathbf{t}_2)'}{\sqrt{k_n}} \right) \right\}, \end{aligned}$$

where $(s_n^i(\mathbf{t}_1)', s_n^i(\mathbf{t}_2)')$ have the same distributions as $(s_n^i(\mathbf{t}_1), s_n^i(\mathbf{t}_2))$ but are independent. Define $s_n^i \equiv a s_n^i(\mathbf{t}_1) + b s_n^i(\mathbf{t}_2)$ and $(s_n^i)' \equiv a s_n^i(\mathbf{t}_1)' + b s_n^i(\mathbf{t}_2)'$, $i = 1, 2, \dots, k_n$. Then $(s_n^i)'$ have the same marginal distributions as s_n^i but are independent. The proof will then follow in a similar way as the univariate case.

To prove S3, first note that

$$\mathbf{E}(|s_n^i|^{2+\delta}) \leq \left\{ a[\mathbf{E}(|s_n^i(\mathbf{t}_1)|^{2+\delta})]^{\frac{1}{2+\delta}} + b[\mathbf{E}(|s_n^i(\mathbf{t}_2)|^{2+\delta})]^{\frac{1}{2+\delta}} \right\}^{2+\delta} < \infty.$$

Define $\sigma_n^2 \equiv \text{Var}(as_n^i(\mathbf{t}_1) + bs_n^i(\mathbf{t}_2))$. Since $(s_n^i)'$ have the same marginal distribution as s_n^i and are i.i.d., we have

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \frac{\mathbb{E}(|(s_n^i)'|^{2+\delta})}{\left(\sqrt{\text{Var}(\sum_{i=1}^{k_n} (s_n^i)')}\right)^{2+\delta}} = \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \frac{\mathbb{E}(|(s_n^i)'|^{2+\delta})}{(k_n \sigma_n^2)^{(2+\delta)/2}} = 0.$$

Applying Lyapounov's Theorem again, we prove S3 and thus the joint normality. \square

A.2.2 Proof of Theorem II.2

Proof. Let $w_n(\mathbf{x}) \equiv w(\mathbf{x}/h_n)$. For large n such that $C \in D_n - D_n$,

$$\begin{aligned} \mathbb{E}(\hat{\gamma}_n(\mathbf{t})) &= \int_{D_n} \int_{D_n} w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times \frac{\gamma(\mathbf{x}_2 - \mathbf{x}_1)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)|} d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \int_{D_n - D_n} w_n(\mathbf{t} + \mathbf{u}) \gamma(\mathbf{u}) d\mathbf{u} \\ &= \int_C w(\mathbf{v}) \gamma(\mathbf{t} - h_n \mathbf{v}) d\mathbf{v} \\ &\rightarrow \gamma(\mathbf{t}). \end{aligned}$$

The derivation for the variance follows similarly as in Karr (1986). Specifically, consider two lags, \mathbf{t} and \mathbf{t}' , where $\mathbf{t}, \mathbf{t}' \in \Lambda$. First

$$\begin{aligned} &\mathbb{E}[\hat{\gamma}_n(\mathbf{t}) \times \hat{\gamma}_n(\mathbf{t}')] \\ &= \frac{1}{\nu^4} \times \iiint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\ &\times \mathbb{E}\left\{ [Z(\mathbf{x}_2) - Z(\mathbf{x}_1)]^2 [Z(\mathbf{y}_2) - Z(\mathbf{y}_1)]^2 \right\} \times \mathbb{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)]. \end{aligned}$$

Observe that

$$\begin{aligned} &\mathbb{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \\ &= \nu^4 d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 + \nu^3 d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) d\mathbf{y}_2 + \nu^3 d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \\ &+ \nu^3 d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) d\mathbf{y}_2 + \nu^3 d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) \\ &+ \nu^2 d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) \epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) + \nu^2 d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2), \end{aligned}$$

where $\epsilon_{\mathbf{x}}(\cdot)$ denotes a point measure. Thus $E(\hat{\gamma}_n(\mathbf{t}) \times \hat{\gamma}_n(\mathbf{t}'))$ can be decomposed into seven terms, say (A1) – (A7), respectively. Ignoring ν , (A1) is

$$\begin{aligned}
& \iiint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\
& \quad \times \Gamma_*(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 \\
& = \iiint_{D_n - D_n} \frac{|D_n \cap (D_n - \mathbf{u}_1) \cap (D_n - \mathbf{u}_2) \cap (D_n - \mathbf{u}_3)|}{|D_n \cap (D_n + \mathbf{u}_1)| \times |D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times w_n(\mathbf{t} + \mathbf{u}_1) \\
& \quad \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2) \times \Gamma_*(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
& \leq \iiint_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2)}{|D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times |\Gamma_*(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
& \leq \iiint_{\mathbb{R}^2} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_4)}{|D_n \cap (D_n + \mathbf{u}_4)|} \times |\Gamma_*(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_4)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_4 \\
& \leq C_1 \times \iint_{\mathbb{R}^2} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_4)}{|D_n \cap (D_n + \mathbf{u}_4)|} d\mathbf{u}_1 d\mathbf{u}_4 \\
& = O\left(\frac{1}{|D_n|}\right).
\end{aligned}$$

Now consider the second term (A2). Ignoring ν , (A2) is

$$\begin{aligned}
& \iiint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{x}_1 + \mathbf{y}_2) \times \Gamma(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{0}, \mathbf{y}_2 - \mathbf{x}_1)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{x}_1 + \mathbf{y}_2)|} d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_2 \\
& \leq C_1 \iint_{D_n - D_n} \frac{|D_n \cap (D_n - \mathbf{u}_1) \cap (D_n - \mathbf{u}_2)|}{|D_n \cap (D_n + \mathbf{u}_1)| \times |D_n \cap (D_n + \mathbf{u}_2)|} \times w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_2) d\mathbf{u}_1 d\mathbf{u}_2 \\
& \leq \iiint_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2)}{|D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times |\Gamma_*(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
& = O\left(\frac{1}{|D_n|}\right).
\end{aligned}$$

(A3), (A4) and (A5) can be shown of order $\frac{1}{|D_n|}$ similarly. The sixth term can be written as

$$\begin{aligned}
& \frac{1}{\nu^2} \times \iint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{x}_1 + \mathbf{x}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)|^2} \times \gamma^{(4)}(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\
& = \frac{1}{\nu^2} \times \int_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}) \times w_n(\mathbf{t}' + \mathbf{u})}{|D_n \cap (D_n + \mathbf{u})|} \times \gamma^{(4)}(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

$$= \frac{1}{\nu^2} \times \int_C \frac{w(\mathbf{v}) \times w(\mathbf{v} + (\mathbf{t}' - \mathbf{t})/h_n)}{|D_n \cap (D_n + h_n \mathbf{v} - \mathbf{t})| \times h_n^2} \times \gamma^{(4)}(h_n \mathbf{v} - \mathbf{t}) d\mathbf{v}$$

Similarly the seventh term can be written as

$$\frac{1}{\nu^2} \times \int_C \frac{w(\mathbf{v}) \times w(\mathbf{v} - (\mathbf{t}' + \mathbf{t})/h_n)}{|D_n \cap (D_n + h_n \mathbf{v} - \mathbf{t})| \times h_n^2} \times \gamma^{(4)}(h_n \mathbf{v} - \mathbf{t}) d\mathbf{v}.$$

Thus

$$|D_n| \times h_n^2 \times \text{Cov}(\hat{\gamma}_n(\mathbf{t}), \hat{\gamma}_n(\mathbf{t}')) \rightarrow \frac{1}{\nu^2} \times \int_C w(\mathbf{v})^2 d\mathbf{v} \times \gamma^{(4)}(\mathbf{t}) \times I(\mathbf{t} = \pm \mathbf{t}').$$

Let $\sigma^2 \equiv \int_C w(\mathbf{v})^2 d\mathbf{v} \times \gamma^{(4)}(\mathbf{t})/\nu^2$, $S_n \equiv \sqrt{|D_n|} \times h_n \times \{\hat{\gamma}_n(\mathbf{t}) - \mathbf{E}[\hat{\gamma}_n(\mathbf{t})]\}$. To show that $S_n \xrightarrow{D} N(0, \sigma^2)$, we again apply the blocking technique. Choose α such that $n^\alpha h_n \rightarrow \infty$. Then divide D_n as in the proof of the previous theorem. We adopt the notations therein with the understanding that now $s_n^i \equiv m(n) \times h_n \times [\hat{\gamma}_{m(n)}^i(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_{m(n)}^i(\mathbf{t}))]$. We need to show

$$\text{S1 } S_n - s_n \xrightarrow{p} 0;$$

$$\text{S2 } \phi'_n(x) - \phi_n(x) \rightarrow 0;$$

$$\text{S3 } s'_n \xrightarrow{d} N(0, \sigma^2).$$

Proof of S1 is analogous to that of the previous theorem. Observe that s_n can be written as $\sqrt{|D^{m(n)}|} \times h_n \times \{\hat{\gamma}_{D^{m(n)}}(\mathbf{t}) - \mathbf{E}(\hat{\gamma}_{D^{m(n)}}(\mathbf{t}))\}$. Similarly we can show

$$\lim_{n \rightarrow \infty} \text{Var}(S_n) = \lim_{n \rightarrow \infty} \text{Var}(s_n) = \lim_{n \rightarrow \infty} \text{Cov}(S_n, s_n) = \sigma^2 \Rightarrow \text{Var}(S_n - s_n) \rightarrow 0.$$

To prove S2, define

$$\mathbf{E}^N(X_j) \equiv \mathbf{E}(X_j|N), \quad \mathbf{E}^N(Y_j) \equiv \mathbf{E}(Y_j|N),$$

$$\text{Cov}^N(X_j, Y_j) \equiv \text{Cov}(X_j, Y_j|N),$$

then

$$\text{Cov}(X_j, Y_j) = \mathbf{E}\left\{\text{Cov}^N(X_j, Y_j)\right\} + \text{Cov}\left\{\mathbf{E}^N(X_j), \mathbf{E}^N(Y_j)\right\}.$$

Since N is a homogeneous Poisson process and X_j, Y_j are random variables defined on two disjoint random fields, we have

$$\text{Cov}\left\{\mathbf{E}^N(X_j), \mathbf{E}^N(Y_j)\right\} = 0.$$

For given N , $X_j|N$ is measurable with respect to $\mathcal{F}(\bigcup_{i=1}^j D_{m(n)}^i)$ and $Y_j|N$ is measurable with respect to $\mathcal{F}(D_{m(n)}^{j+1})$. Then we have

$$|\phi_n(x) - \phi'_n(x)| = \mathbf{O}(n^{4-2\alpha-\eta\epsilon}) \rightarrow 0.$$

Proof of S3 is completely analogous to the previous theorem. \square

A.2.3 Proof of Theorem II.3 and II.4

Proof. Let $\mathbf{b} \equiv \{b_{\mathbf{t}}, \mathbf{t} \in \Lambda\}$ be a nonzero vector. Define

$$S(D_n, \mathbf{b}) = \mathbf{b}' \times (\hat{\mathbf{G}}_n - \mathbf{G}).$$

By condition (2.3) and Minkowski's inequality, we conclude

$$\sup_n \mathbf{E} \left\{ \left| \sqrt{|D_n|} \times [S(D_n, \mathbf{b}) - \mathbf{E}(S(D_n, \mathbf{b}))] \right|^{4+(\delta-2)} \right\} \leq C_2 \text{ for some } \delta > 2, C_2 < \infty.$$

Define

$$\theta_{\mathbf{b}} \equiv \lim_{n \rightarrow \infty} |D_n| \times \text{Var}(S(D_n, \mathbf{b})) = \mathbf{b}' \left\{ \lim_{n \rightarrow \infty} |D_n| \text{Cov}(\hat{\mathbf{G}}_n, \hat{\mathbf{G}}_n) \right\} \mathbf{b} = \mathbf{b}' \Sigma_{\mathbf{R}} \mathbf{b}.$$

The subsampling estimator for $\theta_{\mathbf{b}}$ is

$$\hat{\theta}_{\mathbf{b},n} = \frac{\sum_{i=1}^{k_n} |D_{l(n)}^i| \{S(D_{l(n)}^i, \mathbf{b}) - \bar{S}_n\}^2}{k_n} = \mathbf{b}' \hat{\Sigma}_{\mathbf{R},n} \mathbf{b}.$$

$\hat{\theta}_{\mathbf{b},n} \xrightarrow{L_2} \theta_{\mathbf{b}}$ due to Sherman (1996). Thus $\mathbf{b}' \hat{\Sigma}_{\mathbf{R},n} \mathbf{b} \xrightarrow{L_2} \mathbf{b}' \Sigma_{\mathbf{R}} \mathbf{b}$ for all $\mathbf{b} \neq \mathbf{0}$. The L_2 consistency of $\hat{\Sigma}_{\mathbf{R},n}$ then follows directly for Theorem II.3.

Theorem II.4 is a direct result from Theorem 2 in Politis and Sherman (2001) and the above proof. \square

APPENDIX B

LEMMAS AND PROOF OF THEOREMS IN CHAPTER III

B.1 Lemmas

Lemma B.1. Assume the intensity functions of the point process exist up to order four.

Then for $\mathbf{x}_1 \neq \mathbf{x}_2$ and $\mathbf{y}_1 \neq \mathbf{y}_2$,

$$\begin{aligned}
& \mathbb{E}(N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) \times N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)) \\
&= \Psi^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 \\
&+ \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) + \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \\
&+ \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_2 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) + \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) \\
&+ \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) \epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) + \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \epsilon_{\mathbf{x}_2}(d\mathbf{y}_1),
\end{aligned}$$

where $\epsilon_x(\cdot)$ is a point measure, $\Psi^{(k)}$ denotes the k th order intensity function, $k = 3, 4$.

Proof. Let $I_B(\mathbf{s}) = 1$ if $\mathbf{s} \in B$ and zero otherwise. Define the factorial k th moment measure $\alpha^{(k)}$ as

$$\alpha^{(k)}(B_1, B_2, \dots, B_k) = \mathbb{E} \left\{ \sum_{\mathbf{s}_1 \neq \mathbf{s}_2 \neq \dots \neq \mathbf{s}_k, \mathbf{s}_i \in N} I_{B_1}(\mathbf{s}_1) I_{B_2}(\mathbf{s}_2) \cdots I_{B_k}(\mathbf{s}_k) \right\}$$

Observe that

$$\begin{aligned}
& \mathbb{E}(N(d\mathbf{x}_1) \times N(d\mathbf{x}_2) \times N(d\mathbf{y}_1) \times N(d\mathbf{y}_2)) \\
&= \mathbb{E} \left\{ \sum_{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4 \in N} I_{d\mathbf{x}_1}(\mathbf{s}_1) I_{d\mathbf{x}_2}(\mathbf{s}_2) I_{d\mathbf{y}_1}(\mathbf{s}_3) I_{d\mathbf{y}_2}(\mathbf{s}_4) \right\}
\end{aligned}$$

and

$$\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4 \in N\}$$

$$\begin{aligned}
&= \{s_1 \neq s_2 \neq s_3 \neq s_4\} \cup \{s_1 = s_2 \neq s_3 \neq s_4\} \cup \{s_1 = s_3 \neq s_2 \neq s_4\} \cup \{s_1 = s_4 \neq s_2 \neq s_3\} \\
&\cup \{s_2 = s_3 \neq s_1 \neq s_4\} \cup \{s_2 = s_4 \neq s_1 \neq s_3\} \cup \{s_3 = s_4 \neq s_1 \neq s_2\} \cup \{s_1 = s_2 = s_3 \neq s_4\} \\
&\cup \{s_1 = s_2 = s_4 \neq s_3\} \cup \{s_1 = s_3 = s_4 \neq s_2\} \cup \{s_2 = s_3 = s_4 \neq s_1\} \cup \{s_1 = s_2 \neq s_3 = s_4\} \\
&\cup \{s_1 = s_3 \neq s_2 = s_4\} \cup \{s_1 = s_4 \neq s_2 = s_3\} \cup \{s_1 = s_2 = s_3 = s_4\}.
\end{aligned}$$

Thus $E(N(dx_1) \times N(dx_2) \times N(dy_1) \times N(dy_2))$ can be written as fifteen terms. By definition, the first term

$$E\left\{\sum_{s_1 \neq s_2 \neq s_3 \neq s_4} I_{dx_1}(s_1)I_{dx_2}(s_2)I_{dy_1}(s_3)I_{dy_2}(s_4)\right\} = \alpha^{(4)}(dx_1, dx_2, dy_1, dy_2).$$

Consider the second term:

$$\begin{aligned}
&E\left\{\sum_{s_1 = s_2 \neq s_3 \neq s_4} I_{dx_1}(s_1)I_{dx_2}(s_2)I_{dy_1}(s_3)I_{dy_2}(s_4)\right\} \\
&= E\left\{\sum_{s_1 \neq s_3 \neq s_4} I_{dx_1 \cap dx_2}(s_1)I_{dy_1}(s_3)I_{dy_2}(s_4)\right\} \\
&= \alpha^{(3)}(dx_1, dy_1, dy_2)\epsilon_{x_1}(dx_2)
\end{aligned}$$

The second equality comes from $dx_1 \cap dx_2$ is empty unless $x_1 = x_2$ (i.e. two infinitesimally small discs centered at x_1 and x_2 are disjoint if $x_1 \neq x_2$); and the definition of $\alpha^{(3)}(\cdot, \cdot, \cdot)$.

By working out the remaining terms in a similar fashion, we obtain

$$\begin{aligned}
&E(N(dx_1) \times N(dx_2) \times N(dy_1) \times N(dy_2)) \\
&= \alpha^{(4)}(dx_1, dx_2, dy_1, dy_2) + \alpha^{(3)}(dx_1, dy_1, dy_2)\epsilon_{x_1}(dx_2) + \alpha^{(3)}(dx_1, dx_2, dy_2)\epsilon_{x_1}(dy_1) \\
&+ \alpha^{(3)}(dx_1, dx_2, dy_1)\epsilon_{x_1}(dy_2) + \alpha^{(3)}(dx_1, dx_2, dy_2)\epsilon_{x_2}(dy_1) + \alpha^{(3)}(dx_1, dx_2, dy_1)\epsilon_{x_2}(dy_2) \\
&+ \alpha^{(3)}(dx_1, dx_2, dy_1)\epsilon_{y_1}(dy_2) + \alpha^{(2)}(dx_1, dy_2)\epsilon_{x_1}(dx_2, dy_1) + \alpha^{(2)}(dx_1, dy_1)\epsilon_{x_1}(dx_2, dy_2) \\
&+ \alpha^{(2)}(dx_1, dx_2)\epsilon_{x_1}(dy_1, dy_2) + \alpha^{(2)}(dx_1, dx_2)\epsilon_{x_2}(dy_1, dy_2) + \alpha^{(2)}(dx_1, dy_1)\epsilon_{x_1}(dx_2)\epsilon_{y_1}(dy_2) \\
&+ \alpha^{(2)}(dx_1, dx_2)\epsilon_{x_1}(dy_1)\epsilon_{x_2}(dy_2) + \alpha^{(2)}(dx_1, dx_2)\epsilon_{x_1}(dy_2)\epsilon_{x_2}(dy_1) + \nu dx_1 \epsilon_{x_1}(dx_2, dy_1, dy_2)
\end{aligned}$$

Further imposing the condition that $x_1 \neq x_2$ and $y_1 \neq y_2$, we obtain

$$E(N^{(2)}(dx_1, dx_2) \times N^{(2)}(dy_1, dy_2))$$

$$\begin{aligned}
&= \alpha^{(4)}(d\mathbf{x}_1, d\mathbf{x}_2, d\mathbf{y}_1, d\mathbf{y}_2) \\
&+ \alpha^{(3)}(d\mathbf{x}_1, d\mathbf{x}_2, d\mathbf{y}_2)\epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) + \alpha^{(3)}(d\mathbf{x}_1, d\mathbf{x}_2, d\mathbf{y}_1)\epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \\
&+ \alpha^{(3)}(d\mathbf{x}_1, d\mathbf{x}_2, d\mathbf{y}_2)\epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) + \alpha^{(3)}(d\mathbf{x}_1, d\mathbf{x}_2, d\mathbf{y}_1)\epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) \\
&+ \alpha^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)\epsilon_{\mathbf{x}_1}(d\mathbf{y}_1)\epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) + \alpha^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)\epsilon_{\mathbf{x}_1}(d\mathbf{y}_2)\epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) \\
&= \Psi^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2d\mathbf{y}_1d\mathbf{y}_2 \\
&+ \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2d\mathbf{y}_2\epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) + \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2d\mathbf{y}_1\epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \\
&+ \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2d\mathbf{y}_2\epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) + \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2d\mathbf{y}_1\epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) \\
&+ \Psi(\mathbf{x}_2 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2\epsilon_{\mathbf{x}_1}(d\mathbf{y}_1)\epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) + \Psi(\mathbf{x}_2 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2\epsilon_{\mathbf{x}_1}(d\mathbf{y}_2)\epsilon_{\mathbf{x}_2}(d\mathbf{y}_1).
\end{aligned}$$

□

Lemma B.2. Assume the intensity functions of the point process exist up to order four.

Then $\Psi(\mathbf{x}_2 - \mathbf{x}_1) = C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1) + \nu^2$ and

$$\begin{aligned}
&\Psi^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \\
&= C_N^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \\
&+ \nu C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_2, \mathbf{y}_2 - \mathbf{x}_2) + C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1)C_N^{(2)}(\mathbf{y}_2 - \mathbf{y}_1) \\
&+ C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1)C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2) + C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1)C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2) + \nu^2 C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1) + \nu^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1) \\
&+ \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1) + \nu^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2) + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2) + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{y}_1) + \nu^4
\end{aligned}$$

Proof. We repeatedly use the relationship between moments and cumulants (e.g., McCullagh 1987).

$$\begin{aligned}
\mathbb{E}(N(d\mathbf{x}_1)N(d\mathbf{x}_2)) &= \text{cum}(N(d\mathbf{x}_1), N(d\mathbf{x}_2)) + \text{cum}(N(d\mathbf{x}_1))\text{cum}(N(d\mathbf{x}_2)) \\
&= C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1)d\mathbf{x}_1d\mathbf{x}_2 + \nu^2d\mathbf{x}_1d\mathbf{x}_2
\end{aligned}$$

$$\mathbb{E}(N(d\mathbf{x}_1)N(d\mathbf{x}_2)N(d\mathbf{y}_1)N(d\mathbf{y}_2))$$

$$\begin{aligned}
&= \text{cum}(N(dx_1), N(dx_2), N(dy_1), N(dy_2)) \\
&+ \text{cum}(N(dx_1), N(dx_2), N(dy_1))\text{cum}(N(dy_2)) + \text{cum}(N(dx_1), N(dx_2), N(dy_2))\text{cum}(N(dy_1)) + \\
&\quad \text{cum}(N(dx_1), N(dy_1), N(dy_2))\text{cum}(N(dx_2)) + \text{cum}(N(dx_2), N(dy_1), N(dy_2))\text{cum}(N(dx_1)) \\
&+ \text{cum}(N(dx_1), N(dx_2))\text{cum}(N(dy_1), N(dy_2)) + \text{cum}(N(dx_1), N(dy_1))\text{cum}(N(dx_2), N(dy_2)) + \\
&\quad \text{cum}(N(dx_1), N(dy_2))\text{cum}(N(dx_2), N(dy_1)) \\
&+ \text{cum}(N(dx_1), N(dx_2))\text{cum}(N(dy_1))\text{cum}(N(dy_2)) + \\
&\quad \text{cum}(N(dx_1), N(dy_1))\text{cum}(N(dx_2))\text{cum}(N(dy_2)) + \\
&\quad \text{cum}(N(dx_1), N(dy_2))\text{cum}(N(dx_2))\text{cum}(N(dy_2)) + \\
&\quad \text{cum}(N(dx_2), N(dy_1))\text{cum}(N(dx_1))\text{cum}(N(dy_2)) + \\
&\quad \text{cum}(N(dx_2), N(dy_2))\text{cum}(N(dx_1))\text{cum}(N(dy_1)) + \\
&\quad \text{cum}(N(dy_1), N(dy_2))\text{cum}(N(dx_1))\text{cum}(N(dx_2)) \\
&+ \text{cum}(N(dx_1))\text{cum}(N(dx_2))\text{cum}(N(dy_1))\text{cum}(N(dy_2)).
\end{aligned}$$

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

Lemma B.3. If D_n is a $n \times n$ square field and N is m -dependent with bounded cumulant functions up to order eight, then (3.6) holds for $\delta = 2$.

Proof. The proof here follows in the same way as Lemma A.3 in the marked-Poisson case, except for replacing $[Z(\mathbf{x}) - Z(\mathbf{y})]^2$ therein with one and the variogram and sample variogram function with the second-order intensity and sample second-order intensity function, respectively. \square

Lemma B.4. Consider a Poisson cluster process N . Let ρ denote the intensity for the parent process, S represent the number of offspring per parent and $f(\mathbf{x})$ be the p.d.f. of an offspring's position relative to its parent. Then the integrability conditions introduced in

(3.1) and (3.2) can be rewritten as

$$\begin{aligned} \rho \mathbf{E}[S(S-1)] \iint_{\mathbb{R}^2} f(\mathbf{x})f(\mathbf{x}-\mathbf{u})d\mathbf{x}d\mathbf{u} &< \infty, \\ \rho \mathbf{E}[S(S-1)(S-2)] \iint_{\mathbb{R}^2} f(\mathbf{x})f(\mathbf{x}-\mathbf{u}_1)f(\mathbf{x}-\mathbf{u}_2)d\mathbf{x}d\mathbf{u}_1 &< \infty, \\ \rho \mathbf{E}[S(S-1)(S-2)(S-3)] \iint_{\mathbb{R}^2} f(\mathbf{x})f(\mathbf{x}-\mathbf{u}_1)f(\mathbf{x}-\mathbf{u}_2)f(\mathbf{x}-\mathbf{u}_2-\mathbf{u}_3)d\mathbf{x}d\mathbf{u}_2 &< \infty. \end{aligned}$$

Proof. Observe that $C_N^{(2)}(\mathbf{u}) = \rho \mathbf{E}[S(S-1)] \int f(\mathbf{x})f(\mathbf{x}-\mathbf{u})d\mathbf{x}$ from Diggle (1983), p.55. Thus the integrability condition of $C_N^{(2)}(\mathbf{u})$ can be written as the first expression presented above.

Now we consider $C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_2)$. For notation purpose, let N_p and $\alpha_o^{(k)}(|\mathbf{x})$ denote the parent process and the k th factorial measure of the offspring process with center at \mathbf{x} . We here consider all possible ways in which three distinct points from the superposition of clusters could fall into the product set $A \times B \times C$, where A , B and C are disjoint. There are three possibilities: all three points come from the same parent, two of them come from the same parent or none of them comes from the same parent. Incorporating all cases, we obtain

$$\begin{aligned} \mathbf{E}[N^{(3)}(A \times B \times C | N_p)] &= \int \alpha_o^{(3)}(A \times B \times C | \mathbf{x}) N_p(d\mathbf{x}) \\ &+ \iint \alpha_o^{(2)}(A \times B | \mathbf{x}_1) \alpha_o^{(1)}(C | \mathbf{x}_2) N_p^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) |_{(3)} \\ &+ \iiint \alpha_o^{(1)}(A | \mathbf{x}_1) \alpha_o^{(1)}(B | \mathbf{x}_2) \alpha_o^{(1)}(C | \mathbf{x}_3) N_p^{(3)}(d\mathbf{x}_1, d\mathbf{x}_2, d\mathbf{x}_3), \end{aligned}$$

where $|_{(3)}$ represents three terms, each a rearrangement of A , B and C . Thus

$$\begin{aligned} \alpha^{(3)}[A \times B \times C] &= \rho \int \alpha_o^{(3)}(A \times B \times C | \mathbf{x}) d\mathbf{x} \\ &+ \rho^2 \iint \alpha_o^{(2)}(A \times B | \mathbf{x}_1) \alpha_o^{(1)}(C | \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 |_{(3)} \\ &+ \rho^3 \iiint \alpha_o^{(1)}(A | \mathbf{x}_1) \alpha_o^{(1)}(B | \mathbf{x}_2) \alpha_o^{(1)}(C | \mathbf{x}_3) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3. \end{aligned}$$

Observe that

$$\begin{aligned}
\int \alpha_o^{(2)}[A \times B | \mathbf{x}] d\mathbf{x} &= \int_A \int_B \left[\int \mathbf{E}[S(S-1)] f(\mathbf{u} - \mathbf{x}) f(\mathbf{v} - \mathbf{x}) d\mathbf{x} \right] d\mathbf{u} d\mathbf{v} \\
&= \int_A \int_B \left[\int \mathbf{E}[S(S-1)] f(\mathbf{x}) f(\mathbf{x} - \mathbf{u} + \mathbf{v}) d\mathbf{x} \right] d\mathbf{u} d\mathbf{v} \\
&= \int_A \int_B C_N^{(2)}(\mathbf{u} - \mathbf{v}) d\mathbf{u} d\mathbf{v} = C_N^{(2)}[A \times B],
\end{aligned}$$

$$\int \alpha_o^{(1)}[A | \mathbf{x}] d\mathbf{x} = \int_A \left[\int \mathbf{E}(S) f(\mathbf{u} - \mathbf{x}) d\mathbf{x} \right] d\mathbf{u} = \int_A \left[\int \mathbf{E}(S) f(\mathbf{x}) d\mathbf{x} \right] d\mathbf{u} = C_N^{(1)}[A].$$

Thus we obtain, due to the relationship between cumulants and moments, that

$$\begin{aligned}
&C_N^{(3)}[A \times B \times C] \\
&= \rho \int \alpha_o^{(3)}(A \times B \times C | \mathbf{x}) d\mathbf{x} \\
&= \int_A \int_B \int_C \left[\int \mathbf{E}[S(S-1)(S-2)] f(\mathbf{u} - \mathbf{x}) f(\mathbf{v} - \mathbf{x}) f(\mathbf{w} - \mathbf{x}) d\mathbf{x} \right] d\mathbf{u} d\mathbf{v} d\mathbf{w} \\
&= \int_A \int_B \int_C \left[\int \mathbf{E}[S(S-1)(S-2)] f(\mathbf{x}) f(\mathbf{x} + \mathbf{u} - \mathbf{v}) f(\mathbf{x} + \mathbf{u} - \mathbf{w}) d\mathbf{x} \right] d\mathbf{u} d\mathbf{v} d\mathbf{w}
\end{aligned}$$

Therefore $C_N^{(3)}[\mathbf{u}_1, \mathbf{u}_2] = \int \mathbf{E}[S(S-1)(S-2)] f(\mathbf{x}) f(\mathbf{x} - \mathbf{u}_1) f(\mathbf{x} - \mathbf{u}_2) d\mathbf{x}$. Similarly

$$C_N^{(4)}[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = \int \mathbf{E}[S(S-1)(S-2)(S-3)] f(\mathbf{x}) f(\mathbf{x} - \mathbf{u}_1) f(\mathbf{x} - \mathbf{u}_2) f(\mathbf{x} - \mathbf{u}_3) d\mathbf{x}.$$

Thus the lemma is proved. \square

B.2 Proof of Theorems

B.2.1 Proof of Theorem III.1

Proof. Let $w_n(\mathbf{x}) \equiv h_n^{-2} w(\mathbf{x}/h_n)$. For large n such that $C \in D_n - D_n$,

$$\begin{aligned}
\mathbf{E}(\hat{\gamma}_n(\mathbf{t})) &= \int_{D_n} \int_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)|} \times \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \int_{D_n - D_n} w_n(\mathbf{t} + \mathbf{u}) \Psi(\mathbf{u}) d\mathbf{u}
\end{aligned}$$

$$\begin{aligned}
&= \int_C w(\mathbf{v}) \Psi(\mathbf{t} - h_n \mathbf{v}) d\mathbf{v} \\
&\rightarrow \Psi(\mathbf{t}).
\end{aligned}$$

The derivation of the variance follows similarly as in Masry (1983). Specifically, consider two lags, \mathbf{t} and \mathbf{t}' , where $\mathbf{t}, \mathbf{t}' \in \Lambda$. $\text{Cov}(\hat{\Psi}_n(\mathbf{t}), \hat{\Psi}_n(\mathbf{t}'))$ can be written as

$$\begin{aligned}
&\mathbf{E}(\hat{\Psi}_n(\mathbf{t}) \times \hat{\Psi}_n(\mathbf{t}')) - \mathbf{E}(\hat{\Psi}_n(\mathbf{t})) \times \mathbf{E}(\hat{\Psi}_n(\mathbf{t}')) \\
&= \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \times \mathbf{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \\
&- \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \times \mathbf{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)]\mathbf{E}[N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \\
&= \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \times \left\{ \mathbf{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \right. \\
&\quad \left. - \mathbf{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)]\mathbf{E}[N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \right\}
\end{aligned}$$

From the results of Lemma B.1, we obtain

$$\begin{aligned}
&\mathbf{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] - \mathbf{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)]\mathbf{E}[N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \\
&= \{ \Psi^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) - \Psi(\mathbf{x}_2 - \mathbf{x}_1)\Psi(\mathbf{y}_2 - \mathbf{y}_1) \} d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 \\
&+ \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) + \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \\
&+ \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_2 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_1) + \Psi^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 \epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) \\
&+ \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_1) \epsilon_{\mathbf{x}_2}(d\mathbf{y}_2) + \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \epsilon_{\mathbf{x}_1}(d\mathbf{y}_2) \epsilon_{\mathbf{x}_2}(d\mathbf{y}_1).
\end{aligned}$$

Thus the covariance can be written in seven terms. In turn, we denote them as term 1-7, which correspond to the above seven terms respectively. From the results of Lemma B.2, we further have

$$\begin{aligned}
&\Psi^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) - \Psi(\mathbf{x}_2 - \mathbf{x}_1)\Psi(\mathbf{y}_2 - \mathbf{y}_1) \\
&= C_N^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1)
\end{aligned}$$

$$\begin{aligned}
& + \nu C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_2, \mathbf{y}_2 - \mathbf{x}_2) \\
& + C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1)C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2) + C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1)C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2) + \nu^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1) \\
& + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1) + \nu^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2) + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2).
\end{aligned}$$

Similarly we denote terms in the covariance resulted from the above expression as terms (1.1)-(1.11). We need to show that all eleven terms are of order $\frac{1}{|D_n|}$ except the sixth or the seventh term. Here and henceforth, we will simply assume that $\nu = 1$. First consider (1.1).

$$\begin{aligned}
& \iiint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\
& \quad \times C_N^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 \\
= & \iiint_{D_n - D_n} \frac{|D_n \cap (D_n - \mathbf{u}_1) \cap (D_n - \mathbf{u}_2) \cap (D_n - \mathbf{u}_3)|}{|D_n \cap (D_n + \mathbf{u}_1)| \times |D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times w_n(\mathbf{t} + \mathbf{u}_1) \\
& \quad \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2) \times C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
\leq & \iiint_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2)}{|D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
\leq & \iiint_{\mathbb{R}^2} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_4)}{|D_n \cap (D_n + \mathbf{u}_4)|} \times |C_N^{(4)}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_2 + \mathbf{u}_4)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_4 \\
& \text{(by setting } \mathbf{u}_4 = \mathbf{u}_3 - \mathbf{u}_2) \\
\leq & C_1 \times \iiint_{\mathbb{R}^2} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_4)}{|D_n \cap (D_n + \mathbf{u}_4)|} d\mathbf{u}_1 d\mathbf{u}_4 \\
= & \mathcal{O}\left(\frac{1}{|D_n|}\right).
\end{aligned}$$

First we look at (1.2).

$$\begin{aligned}
& \iiint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\
& \quad \times C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 \\
= & \iiint_{D_n - D_n} \frac{|D_n \cap (D_n - \mathbf{u}_1) \cap (D_n - \mathbf{u}_2) \cap (D_n - \mathbf{u}_3)|}{|D_n \cap (D_n + \mathbf{u}_1)| \times |D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times w_n(\mathbf{t} + \mathbf{u}_1) \\
& \quad \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2) \times C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_2) d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3
\end{aligned}$$

$$\begin{aligned}
&\leq \iiint_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2)}{|D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times |C_N^{(3)}(\mathbf{u}_1, \mathbf{u}_2)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
&\leq C_2 \times \iint_{\mathbb{R}^2} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_4)}{|D_n \cap (D_n + \mathbf{u}_4)|} d\mathbf{u}_1 d\mathbf{u}_4 \\
&= \mathcal{O}\left(\frac{1}{|D_n|}\right).
\end{aligned}$$

Similarly we can prove that terms (1.3)-(1.5) are all of order $\frac{1}{|D_n|}$. Now let's consider (1.6).

$$\begin{aligned}
&\iiint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\
&\quad \times C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1) \times C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1 d\mathbf{y}_2 \\
&\leq \iiint_{D_n - D_n} \frac{|D_n \cap (D_n - \mathbf{u}_1) \cap (D_n - \mathbf{u}_2) \cap (D_n - \mathbf{u}_3)|}{|D_n \cap (D_n + \mathbf{u}_1)| \times |D_n \cap (D_n + \mathbf{u}_3 - \mathbf{u}_2)|} \times w_n(\mathbf{t} + \mathbf{u}_1) \\
&\quad \times w_n(\mathbf{t}' + \mathbf{u}_3 - \mathbf{u}_2) \times |C_N^{(2)}(\mathbf{u}_2)| d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \\
&\leq C_3 \times \iint_{\mathbb{R}^2} \frac{w_n(\mathbf{t} + \mathbf{u}_1) \times w_n(\mathbf{t}' + \mathbf{u}_4)}{|D_n \cap (D_n + \mathbf{u}_4)|} d\mathbf{u}_1 d\mathbf{u}_4 \\
&= \mathcal{O}\left(\frac{1}{|D_n|}\right).
\end{aligned}$$

Similarly we can prove that terms (1.6)-(1.11) are all of order $\frac{1}{|D_n|}$. Terms 2-5 can be shown all of order $\frac{1}{|D_n|}$ due to that $\Psi^{(3)}(\cdot, \cdot)$ is finite. The proof follows similarly as that in the marked-Poisson case. Now let's consider the sixth term.

$$\begin{aligned}
&\iint_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{x}_1 + \mathbf{x}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)|^2} \times \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \int_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}) \times w_n(\mathbf{t}' + \mathbf{u})}{|D_n \cap (D_n - \mathbf{u})|} \times \Psi(\mathbf{u}) d\mathbf{u} \\
&= \int_{D_n - D_n} \frac{w(\mathbf{v}) \times w(\mathbf{v} + (\mathbf{t}' - \mathbf{t})/h_n)}{|D_n \cap (D_n + \mathbf{t} - h_n \mathbf{v})| \times h_n^2} \times \Psi(h_n \mathbf{v} - \mathbf{t}) d\mathbf{v}.
\end{aligned}$$

Thus $\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times (6) = \int_C w(\mathbf{v})^2 d\mathbf{v} \times \Psi(\mathbf{t}) \times I(\mathbf{t} = \mathbf{t}')$. Similarly we can show $\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times (7) = \int_C w(\mathbf{v})^2 d\mathbf{v} \times \Psi(\mathbf{t}) \times I(\mathbf{t} = -\mathbf{t}')$. Thus we prove Theorem

III.1. \square

B.2.2 Proof of Theorem III.2

Proof. Let $\sigma^2 \equiv \int_C w(\mathbf{v})^2 d\mathbf{v} \times \Psi(\mathbf{t})$, $S_n \equiv \sqrt{|D_n|} \times h_n \times \{\hat{\Psi}_n(\mathbf{t}) - \mathbb{E}[\hat{\Psi}_n(\mathbf{t})]\}$. Now we prove $S_n \xrightarrow{D} N(0, \sigma^2)$. To do so, we apply a blocking technique (e.g., Ibragimov and Linnik, 1971) in conjunction with the mixing condition (3.5).

Let $m(n) = n^\alpha$, $m(n)' = n^\alpha - n^\eta$ for some $4/(2 + \epsilon) < \eta < \alpha < 1$. Divide the original field D_n into nonoverlapping $m(n) \times m(n)$ subsquares, $D_{m(n)}^i$, $i = 1, \dots, k_n$; within each subsquare, further obtain $D_{m(n)'}^i$ which shares the same center as $D_{m(n)}^i$. Thus $d(D_{m(n)'}^i, D_{m(n)'}^j) \geq n^\eta$ for $i \neq j$. Let $\hat{\Psi}_{m(n)'}^i(\mathbf{t})$ denote the sample second-order intensity function obtained from $D_{m(n)'}^i$. Let $s_n \equiv \sum_{i=1}^{k_n} s_n^i / \sqrt{k_n}$, $s_n' \equiv \sum_{i=1}^{k_n} (s_n^i)' / \sqrt{k_n}$, where $s_n^i \equiv m(n)' \times h_n \times \{\hat{\Psi}_{m(n)'}^i(\mathbf{t}) - \mathbb{E}[\hat{\Psi}_{m(n)'}^i(\mathbf{t})]\}$ and $(s_n^i)'$ have the same marginal distributions as s_n^i but are independent. Let $\phi_n'(x)$ and $\phi_n(x)$ be the characteristic functions of s_n' and s_n respectively. The proof consists of the following three steps.

$$\text{S1 } S_n - s_n \xrightarrow{p} 0;$$

$$\text{S2 } \phi_n'(x) - \phi_n(x) \rightarrow 0;$$

$$\text{S3 } s_n' \xrightarrow{D} N(0, \sigma^2).$$

Proof of S1:

Since $\mathbb{E}(S_n - s_n) = 0$, it suffices to show $\text{Var}(S_n - s_n) \rightarrow 0$. Observe $\text{Var}(S_n) \rightarrow \sigma^2$ as $n \rightarrow \infty$. Let $D^{m(n)'}$ denote the union of all $D_{m(n)'}^i$. Simple analysis shows s_n can be written as $\sqrt{|D^{m(n)'}|} \times h_n \times \{\hat{\Psi}_{D^{m(n)'}}(\mathbf{t}) - \mathbb{E}[\hat{\Psi}_{D^{m(n)'}}(\mathbf{t})]\}$ (similar to the marked-Poisson case). Since $D^{m(n)'}$ is the union of a set of disjoint squares whose sizes tend to infinity, it satisfies condition (3.3). By this property and conditions (3.1) and (3.2), $\text{Var}(s_n) \rightarrow \sigma^2$ as $n \rightarrow \infty$ by Theorem III.1. Since $\text{Var}(S_n - s_n) = \text{Var}(S_n) + \text{Var}(s_n) - 2 \times \text{Cov}(S_n, s_n)$, we only need to show $\text{Cov}(S_n, s_n) \rightarrow \sigma^2$ in order to show $\text{Var}(S_n - s_n) \rightarrow 0$. Notice $D^{m(n)'} \subset D_n$ and $|D^{m(n)'}|/|D_n| \rightarrow 1$ (by Lemma A.4), therefore from the proof of Theorem III.1, we

conclude

$$\text{Cov}(S_n, s_n) \rightarrow \sigma^2 \Rightarrow \text{Var}(S_n - s_n) \rightarrow 0 \Rightarrow S_n - s_n \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

Proof of S2:

$|\phi_n(x) - \phi'_n(x)| \leq 16k_n \alpha_{n^2}(n^n) \leq C_1 n^{4-2\alpha-\epsilon\eta} \rightarrow 0$. The first inequality follows from a natural extension of theorem 17.2.1 of Ibragimov and Linnik (1971) and from their “telescoping” argument (p. 338). The second follows from the mixing condition (3.5).

Proof of S3:

This follows directly by applying the Lyapounov central limit theorem.

Proof of the joint normality:

This follows directly by applying the Cramer-Wold device. \square

B.2.3 Proof of Theorem III.3

Proof. We first consider the univariate case, i.e., \mathbf{G} and $\hat{\mathbf{G}}_n$ are both scalars, $\Psi(\mathbf{t})$ and $\hat{\Psi}_n(\mathbf{t})$ respectively, say. Thus (3.7) becomes

$$\hat{\sigma}_n^2 = \frac{1}{|D_n^{1-c}|} \times \int_{D_n^{1-c}} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times [\hat{\Psi}_{l(n)}(\mathbf{x}) - \bar{\Psi}_{l(n)}]^2 \right\} d\mathbf{x}.$$

Define $\sigma^2 \equiv \lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times \text{Var}(\hat{\Psi}_n(\mathbf{t}))$. Our goal is to show that $\hat{\sigma}_n^2 \xrightarrow{L_2} \sigma^2$. Denote the sample second-order intensity function at lag \mathbf{t} on $D_{l(n)} + \mathbf{x}$ by $\hat{\Psi}_{l(n)}(\mathbf{x})$. Define

$$S_n \equiv \frac{1}{|D_n^{1-c}|} \times \int_{D_n^{1-c}} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times [\hat{\Psi}_{l(n)}(\mathbf{x}) - \mathbf{E}(\hat{\Psi}_{l(n)}(\mathbf{x}))]^2 \right\} d\mathbf{x}.$$

and

$$S'_n \equiv \frac{1}{|D_n^{1-c}|} \times \int_{D_n^{1-c}} \left\{ \sqrt{|D_{l(n)}|} \times h_{l(n)} \times [\hat{\Psi}_{l(n)}(\mathbf{x}) - \mathbf{E}(\hat{\Psi}_{l(n)}(\mathbf{x}))] \right\} d\mathbf{x}.$$

Observe that $\hat{\sigma}_n^2$ is equal to $S_n - (S'_n)^2$. Thus $S_n \xrightarrow{L_2} \sigma^2$ and $(S'_n)^2 \xrightarrow{L_2} 0$ will be sufficient for $\hat{\sigma}_n^2 \xrightarrow{L_2} \sigma^2$. For the first term, since $\mathbf{E}(S_n) \rightarrow \sigma^2$, we only need to prove that $\text{Var}(S_n) \rightarrow 0$.

$$\text{Var}(S_n) = \frac{1}{|D_n^{1-c}|^2} \times \iint_{D_n^{1-c}} \text{Cov} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{x}), |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{y}) \right\} d\mathbf{x} d\mathbf{y}.$$

We write the above expression in two terms, namely

$$A_n \equiv \frac{1}{|D_n^{1-c}|^2} \times \iint_{D_n^{1-c}, d(\mathbf{x}, \mathbf{y}) \leq l(n)} \text{Cov} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{x}), |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{y}) \right\} d\mathbf{x}d\mathbf{y},$$

$$B_n \equiv \frac{1}{|D_n^{1-c}|^2} \times \iint_{D_n^{1-c}, d(\mathbf{x}, \mathbf{y}) > l(n)} \text{Cov} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{x}), |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{y}) \right\} d\mathbf{x}d\mathbf{y}.$$

$A_n \rightarrow 0$ follows directly from the proof of Theorem 1 in Politis and Sherman (2001). For any \mathbf{x} and \mathbf{y} in the integral defining B_n , we have

$$\text{Cov} \left\{ |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{x}), |D_{l(n)}| \times h_{l(n)}^2 \times \hat{\Psi}_{l(n)}(\mathbf{y}) \right\} \leq C_\delta \alpha^{\delta/(2+\delta)} (|D_{l(n)}|; l(n))$$

by a covariance inequality in Doukhan (1994) and condition (3.6). Thus $B_n \rightarrow 0$ due to condition (3.3) and (3.5) and thus $\text{Var}(S_n) \rightarrow 0$.

Applying a same technique, we can show $S'_n \xrightarrow{L_2} 0$. $(S'_n)^2 \xrightarrow{L_2} 0$ then follows directly from Sherman (1996). Thus $\hat{\sigma}_n^2 \xrightarrow{L_2} \sigma^2$.

The more general multivariate case can be proved by applying the above univariate case result and the proof of Theorem 2 in Guan et al. (2002). \square

APPENDIX C

PROOF OF THEOREMS IN CHAPTER IV

C.1 Proof of Theorem IV.1

Proof. Let $w_n(\mathbf{x}) \equiv h_n^{-2}w(\mathbf{x}/h_n)$. For large n such that $C \in D_n - D_n$,

$$\begin{aligned}
\mathbb{E}(\hat{\gamma}_n(\mathbf{t})) &= \frac{1}{\Psi(\mathbf{t})} \int_{D_n} \int_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)|} \times \gamma(\mathbf{x}_2 - \mathbf{x}_1) \times \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \frac{1}{\Psi(\mathbf{t})} \int_{D_n - D_n} w_n(\mathbf{t} + \mathbf{u}) \gamma(\mathbf{u}) \Psi(\mathbf{u}) d\mathbf{u} \\
&= \frac{1}{\Psi(\mathbf{t})} \int_C w(\mathbf{v}) \gamma(\mathbf{t} - h_n \mathbf{v}) \Psi(\mathbf{t} - h_n \mathbf{v}) d\mathbf{v} \\
&\rightarrow \gamma(\mathbf{t}).
\end{aligned}$$

For the covariance term, first introduce the following notation:

$$\begin{aligned}
\Gamma(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) &\equiv \mathbb{E} \left\{ \left[Z(\mathbf{x}_2) - Z(\mathbf{x}_1) \right]^2 \left[Z(\mathbf{y}_2) - Z(\mathbf{y}_1) \right]^2 \right\} \\
\Gamma_*(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) &\equiv \text{Cov} \left\{ \left[Z(\mathbf{x}_2) - Z(\mathbf{x}_1) \right]^2, \left[Z(\mathbf{y}_2) - Z(\mathbf{y}_1) \right]^2 \right\}
\end{aligned}$$

Consider two lags, \mathbf{t} and \mathbf{t}' , where $\mathbf{t}, \mathbf{t}' \in \Lambda$. The covariance of two lags $(\mathbf{t}, \mathbf{t}')$ times $[\Psi(\mathbf{t})\Psi(\mathbf{t}')] can be written as:$

$$\begin{aligned}
&\mathbb{E}[\hat{\gamma}_n(\mathbf{t}) \times \hat{\gamma}_n(\mathbf{t}')] - \mathbb{E}[\hat{\gamma}_n(\mathbf{t})] \times \mathbb{E}[\hat{\gamma}_n(\mathbf{t}')] \\
&= \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2) \times \Gamma(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\
&\quad \times \mathbb{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2) N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)] \\
&- \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2) \times \gamma(\mathbf{x}_2 - \mathbf{x}_1) \times \gamma(\mathbf{y}_2 - \mathbf{y}_1)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} \\
&\quad \times \mathbb{E}[N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)] \mathbb{E}[N^{(2)}(d\mathbf{y}_1, d\mathbf{y}_2)]
\end{aligned}$$

Thus as in the point process case (see Appendix B), the above expression can be written as seven terms, where the second to the seventh terms depend only on $\mathbb{E}[\hat{\gamma}_n(\mathbf{t}) \times \hat{\gamma}_n(\mathbf{t}')]$.

Following the proof therein, the second to the fifth terms can be shown of order $1/|D_n|$ due to the fact that $\Gamma(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)$ is finite. Define

$$\begin{aligned}
& F_1(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \\
&= \Gamma(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \left\{ C_N^{(4)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \right. \\
&+ \nu C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \\
&+ \nu C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) + \nu C_N^{(3)}(\mathbf{y}_1 - \mathbf{x}_2, \mathbf{y}_2 - \mathbf{x}_2) \\
&+ C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1) C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2) + C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1) C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2) \\
&\left. + \nu^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_1) + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_1) + \nu^2 C_N^{(2)}(\mathbf{y}_1 - \mathbf{x}_2) + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{x}_2) \right\}
\end{aligned}$$

and

$$\begin{aligned}
& F_2(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \\
&= \Gamma_*(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1) \left\{ \nu^2 C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1) + \nu^2 C_N^{(2)}(\mathbf{y}_2 - \mathbf{y}_1) \right. \\
&\left. + C_N^{(2)}(\mathbf{x}_2 - \mathbf{x}_1) C_N^{(2)}(\mathbf{y}_2 - \mathbf{y}_1) + \nu^4 \right\}.
\end{aligned}$$

The first term can be written as the sum of the following two terms

$$\begin{aligned}
& \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2) \times F_1(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1, d\mathbf{y}_2, \\
& \iiint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{y}_1 + \mathbf{y}_2) \times F_2(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)| \times |D_n \cap (D_n - \mathbf{y}_1 + \mathbf{y}_2)|} d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{y}_1, d\mathbf{y}_2.
\end{aligned}$$

The first quantity is of order $1/|D_n|$ due to the proof for the point process case and that $\Gamma(\mathbf{x}_2 - \mathbf{x}_1, \mathbf{y}_1 - \mathbf{x}_1, \mathbf{y}_2 - \mathbf{x}_1)$ is finite. The second quantity is of order $1/|D_n|$ due to the proof for the marked-Poisson case (see Appendix A) and that the cumulant functions are finite.

Consider the sixth term, which can be written as

$$\iint\limits_{D_n} \frac{w_n(\mathbf{t} - \mathbf{x}_1 + \mathbf{x}_2) \times w_n(\mathbf{t}' - \mathbf{x}_1 + \mathbf{x}_2)}{|D_n \cap (D_n - \mathbf{x}_1 + \mathbf{x}_2)|^2} \times \gamma(\mathbf{x}_2 - \mathbf{x}_1) \times \Psi(\mathbf{x}_2 - \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2$$

$$\begin{aligned}
&= \int_{D_n - D_n} \frac{w_n(\mathbf{t} + \mathbf{u}) \times w_n(\mathbf{t}' + \mathbf{u})}{|D_n \cap (D_n - \mathbf{u})|} \times \gamma(\mathbf{u}) \times \Psi(\mathbf{u}) d\mathbf{u} \\
&= \int_{D_n - D_n} \frac{w(\mathbf{v}) \times w(\mathbf{v} + (\mathbf{t}' - \mathbf{t})/h_n)}{|D_n \cap (D_n + \mathbf{t} - h_n \mathbf{v})| \times h_n^2} \times \gamma(h_n \mathbf{v} - \mathbf{t}) \times \Psi(h_n \mathbf{v} - \mathbf{t}) d\mathbf{v}.
\end{aligned}$$

Thus

$$\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times \text{the sixth term} = \int_C w(\mathbf{v})^2 d\mathbf{v} \times \gamma(\mathbf{t}) \times \Psi(\mathbf{t}) \times I(\mathbf{t} = \mathbf{t}')$$

Similarly we can show

$$\lim_{n \rightarrow \infty} |D_n| \times h_n^2 \times \text{the seventh term} = \int_C w(\mathbf{v})^2 d\mathbf{v} \times \gamma(\mathbf{t}) \times \Psi(\mathbf{t}) \times I(\mathbf{t} = -\mathbf{t}').$$

Thus the theorem is proved. \square

C.2 Proof of Theorem IV.2

Proof. Let $\sigma^2 = \int_C w(\mathbf{v})^2 d\mathbf{v} \times \gamma^{(4)}(\mathbf{t})/\nu^2$, $S_n \equiv \sqrt{|D_n|} \times h_n \times \{\hat{\gamma}_n(\mathbf{t}) - \mathbf{E}[\hat{\gamma}_n(\mathbf{t})]\}$. To show that $S_n \xrightarrow{D} N(0, \sigma^2)$, we again apply the blocking technique. Choose α such that $n^\alpha h_n \rightarrow \infty$. Then divide D_n as in marked-Poisson case. We adopt the notations therein with the understanding that we are studying a general marked-point process. We need to show

$$\text{S1 } S_n - s_n \xrightarrow{p} 0;$$

$$\text{S2 } \phi'_n(x) - \phi_n(x) \rightarrow 0;$$

$$\text{S3 } s'_n \xrightarrow{d} N(0, \sigma^2).$$

The second and the third step indicate that $s_n \xrightarrow{d} N(0, \sigma^2)$. Since $S_n - s_n \xrightarrow{p} 0$, we can conclude that $S_n \xrightarrow{d} N(0, \sigma^2)$ as well.

Proof of S1: The proof is the same as that of the Poisson case.

Proof of S2: In what follows, we use I (instead of the commonly used notation i) to denote the imaginary number. By definition,

$$\begin{aligned}\phi_n(x) &= \mathbf{E}\{\exp(Ixs_n)\} = \mathbf{E}\left\{\exp\left(Ix \sum_{i=1}^{k_n} \frac{s_n^i}{\sqrt{k_n}}\right)\right\}, \\ \phi'_n(x) &= \mathbf{E}\{\exp(Ixs'_n)\} = \mathbf{E}\left\{\exp\left(Ix \sum_{i=1}^{k_n} \frac{(s_n^i)'}{\sqrt{k_n}}\right)\right\}.\end{aligned}$$

Since $(s_n^i)'$, $i = 1, 2, \dots, k_n$, are independent and have the same marginal distribution as s_n^i , $i = 1, 2, \dots, k_n$, $\phi'_n(x)$ can be rewritten as $\prod_{i=1}^{k_n} \mathbf{E}\left\{\exp\left(Ix \frac{s_n^i}{\sqrt{k_n}}\right)\right\}$. Define

$$U_i \equiv \exp\left(Ix \frac{s_n^i}{\sqrt{k_n}}\right),$$

Then

$$\phi_n(x) = \mathbf{E}\left\{\prod_{i=1}^{k_n} U_i\right\}, \quad \phi'_n(x) = \prod_{i=1}^{k_n} \mathbf{E}\{U_i\}.$$

From the proof of S2 in the Poisson case, we have

$$|\phi_n(x) - \phi'_n(x)| \leq \sum_{j=1}^{k_n-1} \underbrace{\left| \mathbf{E}\left\{\prod_{i=1}^{j+1} U_i\right\} - \mathbf{E}\left\{\prod_{i=1}^j U_i\right\} \times \mathbf{E}\{U_{j+1}\} \right|}_{(A)}.$$

Define

$$\begin{aligned}X_j &= \prod_{i=1}^j U_i, \quad Y_j = U_{j+1}, \\ \mathbf{E}^N(X_j) &\equiv \mathbf{E}(X_j|N), \quad \mathbf{E}^N(Y_j) \equiv \mathbf{E}(Y_j|N), \\ \text{Cov}^N(X_j, Y_j) &\equiv \text{Cov}(X_j, Y_j|N),\end{aligned}$$

then

$$(A) = \text{Cov}(X_j, Y_j) = \mathbf{E}\left\{\text{Cov}^N(X_j, Y_j)\right\} + \text{Cov}\left\{\mathbf{E}^N(X_j), \mathbf{E}^N(Y_j)\right\}.$$

For given N , $X_j|N$ is measurable with respect to $\mathcal{F}(\bigcup_{i=1}^j D_{m(n)}^i)$ and $Y_j|N$ is measurable with respect to $\mathcal{F}(D_{m(n)}^{j+1})$. Since $|X_j| \leq 1$, $|Y_j| \leq 1$ and $|D_{m(n)}^{j+1}| \leq |\bigcup_{i=1}^j D_{m(n)}^i| = j \times m(n)^2$, we have

$$\text{Cov}^N(X_j, Y_j) \leq 16j \times \{n^{2\alpha} + n^{2\eta} - 2n^{\alpha+\eta}\} \times n^{-\eta\epsilon}$$

due to the assumed mixing condition for the mark process and the proof for the marked-Poisson case (see Appendix A). Note $E^N(X_j)$ is measurable with respect to $\mathcal{F}_N(\bigcup_{i=1}^j D_{m(n)}^i)$ and $E^N(Y_j)$ is measurable with respect to $\mathcal{F}_N(D_{m(n)}^{j+1})$; $|E^N(X_j)| \leq 1$, $|E^N(Y_j)| \leq 1$, we have

$$\text{Cov}\{E^N(X_j), E^N(Y_j)\} \leq 16j \times \{n^{2\alpha} + n^{2\eta} - 2n^{\alpha+\eta}\} \times n^{-\eta\epsilon}$$

due to the assumed mixing condition for the point process. Combining the above results, we have

$$|\phi_n(x) - \phi'_n(x)| \leq \mathbf{O}(n^{4-2\alpha-\eta\epsilon}) \rightarrow 0.$$

Proof of S3: the proof is analogous to the marked-Poisson case. □

C.3 Proof of Theorem IV.3

Proof. This is a direct result from Theorem 2 in Politis and Sherman (2001) and the proof of Theorem II.3. □

VITA

Yong Tao Guan was born in Pingdu, China. He received a Bachelor of Science degree in Probability and Statistics in 1998 from the Peking University. That same year, he was admitted to the Ph.D. program in the Department of Statistics at Texas A&M University. He received his Ph.D. degree in August 2003. His permanent address is

Wang Xian Zhuang, Men Cun

Ping Du, Shan Dong 266708

People's Republic of China