Spring 2019

Spatio-Temporal Cluster Detection and Local Moran Statistics of Point Processes

Jennifer L. Matthews
Old Dominion University

SPATIO-TEMPORAL CLUSTER DETECTION AND LOCAL MORAN STATISTICS OF POINT PROCESSES

by

Jennifer L. Matthews
Commander, United States Navy
B.S. May 2000, United States Naval Academy
M.S. March 2006, National University of Singapore
M.S. June 2006, Naval Postgraduate School

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
May 2019

Approved by:

Norou Diawara (Director)

N. Rao Chaganty (Member)

Kayoung Park (Member)

Bryan Porter (Member)

ABSTRACT

Approved for public release; distribution is unlimited

SPATIO-TEMPORAL CLUSTER DETECTION
AND LOCAL MORAN STATISTICS OF POINT PROCESSES

Jennifer L. Matthews
Commander, United States Navy
Old Dominion University, 2019
Director: Dr. Norou Diawara

Moran's index is a statistic that measures spatial dependence, quantifying the degree of dispersion or clustering of point processes and events in some location/area. Recognizing that a single Moran's index may not give a sufficient summary of the spatial autocorrelation measure, a local indicator of spatial association (LISA) has gained popularity. Accordingly, we propose extending LISAs to time after partitioning the area and computing a Moran-type statistic for each subarea. Patterns between the local neighbors are unveiled that would not otherwise be apparent. We consider the measures of Moran statistics while incorporating a time factor under simulated multilevel Palm distribution, a generalized Poisson phenomenon where the clusters and dependence among the subareas are captured by the rate of increase of the process over time. Event propagation is built under spatial nested sequences over time. The Palm parameters, Moran statistics and convergence criteria are calculated from an explicit algorithm in a Markov chain Monte Carlo simulation setting and further analyzed in two real datasets.

To my loving husband

Chuck

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my committee chair Dr. Norou Diawara, who continually conveyed excitement and adventure in regard to research and scholarship. This dissertation would not have been possible without his supervision and persistent guidance.

I would like to thank Dr. N. Rao Chaganty who supported me in beginning my doctoral studies part-time. I also thank him for his review of this dissertation as a committee member.

I greatly appreciate the contributions of my other committee members, Dr. Kayoung Park and Dr. Bryan Porter, who have been very gracious with their time and improved this work with their input.

I also thank Nhan Bui for his contributions in assisting with coding for the simulations and for his work on a published manuscript.

I thank the following publishers: Susan Rivers' Cultural Institute for permission to include material in Subsection 5.1.1, which was originally published in the Journal of Probability and Statistical Science; Canadian Center of Science and Education, for permission to include material in Section 4.2 and Subsection 5.1.2, which was originally published in the International Journal of Statistics and Probability; and Taylor & Francis Group for permission to include material in Chapter 5, which was originally published in Communications in Statistics - Simulation and Computation.

Finally, I would like to thank the U.S. Navy and the U.S. Naval Academy for financial support throughout my studies and research. Material contained herein is solely the responsibility of the author and its contents do not necessarily reflect the views of the Department of the Navy or the Department of Defense.

NOMENCLATURE

$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{R}_+$	Set of non-negative real numbers
$\mathbb{Z}$, $\mathbb{Z}_+$	Set of integers in $\mathbb{R}$ or $\mathbb{R}_+$
$A \stackrel{d}{=} B$	A is equal in distribution to B
δ	Dirac-delta function
$\mathbb{E}$	Expectation
$\mathbb{E}_R$	Expectation under the assumption of randomization
$\mathcal{H}$	History
L	Likelihood
ℓ	Log-likelihood
$\mathbb{L}$	Law
$\mathcal{L}$	Laplace transform (or functional)
P	Probability
$\mathbb{P}$	Palm distribution
PL	Pseudo-likelihood
$\mathbb{Q}$	Reduced Palm distribution
$\mathbf{1}$	Indicator function
$\mathcal{M}(S)$	Class of σ-finite measures on a Borel space S
$\mathcal{M}_{\mathcal{S}}^{\#}$	Boundedly finite measures on $\mathcal{S}$
$\mathcal{N}(S)$	Class of locally finite counting measures on a Borel space S

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In this dissertation, we consider spatio-temporal Moran-type statistics of point processes, capturing a measure of dependence or correlation of events. Motivations are based on simulations of an infectious disease that spreads over continuous space (spatial) and discrete time (temporal). It has applications in high dimensional data under asymptotic theory of continuous space and discrete time. Moran measures of autocorrelation, properties and comparisons with other statistics are made. We describe a branching process with time and space evolution of events that reproduce according to an offspring phenomena. The evolution is captured by a conditional distribution of Palm measure kind under a Markov process. The distribution of events that occur within a region surrounding an event occurrence from the previous time period is captured. Instead of assuming overall area based estimation assumed in so-called convenient space with challenging computations, here we propose subarea partition-based spatial analysis over time. Simulations are performed along with real data analysis. This work mimics real world phenomena much closer than global trend analysis.

1.1 MOTIVATION AND BACKGROUND

The geographical spread of infectious disease is of interest to health care professionals, government officials, and the general population. Diseases spread throughout a region (globally or locally) over time. Survival analysis makes use of hazard rates to describe spread of disease over time. Disease spread may be a diffusion process in that "the location of the earliest infected may be random, but those locations exert influence on the locations of later infected" (Odland, 1988). Moreover, the spread is usually irregular in space and time factors.

Some data collection includes only the spatial (space) or temporal (time) aspect such as John Snow's disease mapping of the cholera outbreak in Soho, London in 1854. However, spatio-temporal data have been collected for hundreds of years. The *Lienzo de Quauhquechollan* in Figure 1 tells the story of the Spanish conquest of current day Guatemala from

1527 to 1530 (Cressie & Wikle, 2011). A close look at the tapestry (Figure 1) reveals that it consists of 15 individual pieces (space element) that are stitched together to form the time sequence of the events. Statistical analysis of space-time data has become a field of interest, especially over the past 25 years due to advancements in technology. In many cases, the space and time components of the models are estimated independently, when in fact, interaction between those two factors is dynamic (Li, Calder, & Cressie, 2007). Wikle and Hooten (2010) provide a framework for predictive models in dynamic space and time. As they suggested, the implemented statistical modeling requires developing more optimal methods.

Figure 1: The *Lienzo de Quauhquechollan* shows both space and time elements of the Spanish conquest of Guatemala from 1527 to 1530.

Much has been done to analyze spatial data; however, the addition of time is still an area of development. As the amount of data collected grows, there is a greater demand for new statistical models to analyze this data that often includes space and time components. As mentioned in Møller and Waagepeterson (2007) and Lindgren and Rue (2015), the challenges are to develop new tools and tractable iterative methods and other approximation techniques. Even though there have been major developments in recent years, more is expected in the years to come.

In this dissertation, we extend the spatial dependency and include concepts of space and time inferences of a measure of correlation first developed by P. A. P. Moran in 1950. Martin and Oeppen (1975) studied regional forecasting models using spatial correlation over time, and Stoyan and Stoyan (1998) presented applications in forestry. More recent examples of

modeling and studying spatial correlation is evidenced in: Møller and Díaz-Avalos (2010) modeling forest fires; Vaillant, Puggioni, Waller, and Daugrois (2011) studying the spread of sugarcane yellow leaf virus; and Meddens and Hicke (2014) studying tree mortality caused by a mountain pine beetle outbreak in Colorado.

1.2 CONTRIBUTIONS

We make use of topics in measure theory and random processes to develop a model for the spatio-temporal spread of events and then conduct simulations as ways to provide more accurate description and validation of our model. The Palm distribution is used to condition the point pattern on an event that occurred during the previous time period. We use this construct to develop a spatio-temporal Palm distribution and a spatio-temporal Palm likelihood. Analysis of simulated data is conducted within uniform space and under extension to Voronoi cells. Clusters are revealed and time-dependent estimates are obtained from Moran and Palm distribution functions and are then compared with other statistics. We conduct analysis of the recursive estimates of Moran values and Palm parameters through spatio-temporal autocorrelation measures. We also conduct analysis of two real data sets, seat belt use and brain image data.

1.3 ORGANIZATION

We first present a general overview of spatial point processes, Palm distribution, and associated properties in Chapter 2. We also describe process spread with irregular shaped areas. Then in Chapter 3, these concepts are extended to include time while checking for existence and identifiability. Chapter 4 presents statistical inference through the likelihood and autocorrelation measure. Then in Chapter 5, we present simulated and real data examples of the models and statistics introduced. We conclude with a summary and future work in Chapter 6.

CHAPTER 2

SPATIAL MODELS

Point processes arise in many fields of research including epidemiology (e.g. disease spread), meteorology (e.g. lightning strikes) and astronomy (e.g. patterns of stars). Applications can also be motivated in neuroscience where the brain activity and connectivity are captured by neuron firings in one location or another in response to some stimulus . Other areas of interest include human behavior factors over space and time.

In a one dimensional setting, events or actions associated to a point process could describe the times of occurrence of some disease or event of interest. These examples would most commonly be associated with a counting point process using the standard Lebesgue measure. The counting point process can be extended to two and three dimensions with an area measure and volume measure in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. The designated point process may then be thought of as a random point process in a field with each point representing the location where an event of interest is captured. However, the distribution of points is not always known, nor is their spatial dependence as there is no natural ordering of points in space (see Baddeley, 2007).

Motivated by the above applications, we will work with selected sequences of point processes to validate simulation results and to apply them to real data. We start with the description of a point process and then provide a summary of principles of marked point processes, their spatial relationship and properties. Thinning is applied and its design is studied through the Poisson process. Generalization of the Poisson process through the Palm distribution is presented.

2.1 POINT PROCESSES

Point processes are described in a space S and by a collection of random points $x_1, x_2, \ldots$ pairwise distinct in S. We consider D, a measurable subset of S and $\mathcal{D}$ the σ-field associated with $x_1, x_2, \ldots, x_n$, for $n \in \mathbb{N}$. Given the probability space $(S, \mathcal{D}, P)$, a probability on $D \in \mathcal{D}$ is defined as $P(D)$, where P is the probability measure with unit total mass (Kallenberg, 2002).

Definition 2.1.1. Let $N(D)$ represent the count or number of points x_i in the subarea D. To $N(D)$, we associate a measure called a point measure, also called Lebesgue measure, of the process, defined as:

$$N(D) = \sum_{x \in D} \delta_x, \quad \text{where} \tag{1}$$

- the sum is finite; and
- δ_x is a Dirac-delta function defined as:

$$\delta_x(D) = \begin{cases} 1 & \text{if } x \in D, \\ 0 & \text{otherwise.} \end{cases}$$

If there is no occurrence/point in D then δ is the zero measure in D. The function $N(D)$ is called a point process. The point process is called finite if $P(N(D) < \infty) = 1$ for every bounded subset D of S. Two point processes N and N' are called equivalent if $N(D) = N'(D)$ for all $D \in \mathcal{D}$ where $\mathcal{D}$ is the sigma-field of subsets of S.

Let Q be the probability on D (with probability distribution function q) and suppose $x_1, x_2, \ldots, x_n$ are independent random points/events in D with distribution Q. Then,

$$P(N(D) = k) = \binom{m}{k} Q(D)^k (1 - Q(D))^{m-k}, \ k = 0, 1, \ldots, m,$$

and N is referred to as a binomial process with sample size m and sampling distribution Q.

Every point process has an associated intensity measure μ defined on the subsets of S and described as $\mu(D) = E(N(D))$, the expected number of points.

$$\begin{aligned} \mu : \mathcal{D} &\longrightarrow [0, \infty) \\ D &\longmapsto \mu(D). \end{aligned}$$

Let N be an integer value random variable and $N_1, N_2, \ldots$ be a sequence of independent and identically distributed (*i.i.d*) Bernoulli random variables with parameter $p \in [0, 1]$. Then

$$X = \sum_{i=1}^{N} N_i$$

defines a conditional distribution of X given N and is called a p-thinning, i.e. given $N =$

$n,\ X \sim Bin(n,p)$.

To the set of measurable subsets of S, we define the Laplace transform (or characteristic function) of a point process N on S with the conditional distribution of X given N. The Laplace transform is uniquely defined as

$$\begin{aligned}\mathcal{L}_N :&\mathcal{D} \longrightarrow [0,1] \\ &f \longrightarrow \mathcal{L}_N(f) := \mathbb{E}\left(\exp\left\{-\int f(x)N(dx)\right\}\right), \end{aligned} \tag{2}$$

with $\mathcal{D} :=$ set of measurable functions $f : D \to \mathbb{R}$. Recall from measure theory that we can define f^+ and f^- and for any other measure ν,

$$\begin{aligned}\int f d\nu &= \int f^+ d\nu - \int f^- d\nu, \\ \nu(f) &= \int f(x)\nu(dx) \\ \text{and } \nu(D) &= \nu(\mathbf{1}_D).\end{aligned}$$

If N follows a binomial process then

$$\begin{aligned}\mathcal{L}_N(f) &= \mathbb{E}e^{-\sum_{k=0}^m f(X_k)} = \mathbb{E}\prod_{k=0}^m e^{-f(X_k)} \\ &= \prod_{k=0}^m \mathbb{E}e^{-f(X_k)} = \left[\int e^{-f(x)}Q(dx)\right]^{m+1},\end{aligned}$$

where Q is the distribution function of the random variable X associated with random point x in D. Summary principles require familiarity with some statistical methods, in particular the characteristic function or Laplace transform.

The distribution of the random measure N is uniquely defined by its Laplace transform. For any two point process N and N' on D, we have the following equivalence:

1. $N \stackrel{d}{=} N'$ (i.e. N and N' have the same distribution).
2. $\mathcal{L}_N(f) = \mathcal{L}_{N'}(f),\ \forall\ f \in \mathcal{D}$.
3. $\int f(x)N(dx) \stackrel{d}{=} \int f(x)N'(dx),\ \forall\ f \in \mathcal{D}$.

This class of point processes is quite large. Moreover, if D is a Borel subspace of a

complete separable metric space, then any locally finite point process will define a proper point process (Last & Penrose, 2017). We will focus on three types of point processes: Poisson, Cox, and Markov, after giving a review of kernel density.

2.1.1 KERNEL DENSITY

In the general context of spatial statistics, kernel density has been proposed to build estimations of parameters and density estimation based on covariates and marks (see Diggle, 1985; Parzen, 1962; Rosenblatt, 1956). The mark of a point from a spatial point process carries information such as the radius of the disc or the clustering size of the spatial data. A point process can then be described as a stochastic process where the random variable represents events in space, denoted as ξ. We say that ξ, instead of N as above, is a point process on S if ξD is an integer-valued random variable for every bounded set D in $\mathbb{R}^d$, $d \geq 1$, or $D \subset S$ (where usually $d = 1$, 2, or 3 in applications). Alternatively, ξ can be thought of as a random element in the space $\mathcal{N}(S) \subset \mathcal{M}(S)$ the measure space of all σ-finite, integer-valued measures on S (see Kallenberg, 2002).

The following discussion will be developed within the context of a fixed arbitrary basic probability space $(\Omega, \mathcal{A}, P)$, where P is the probability measure having a total mass of 1 on Ω. $\mathcal{A}$ is also assumed to be a locally compact second countable Hausdorff space (*lcscH*), meaning that the neighborhood of every point in $\mathcal{A}$ is compact, there is a countable base for $\mathcal{A}$, and disjoint neighborhoods can be used to sparate distinct points (Kallenberg, 1983).

If $\mathcal{S}$ is a σ-field of subsets of a space S, then $(S, \mathcal{S})$ is said to be a measurable space. A measure on $(S, \mathcal{S})$ is defined as a mapping $\mu : \mathcal{S} \to \mathbb{R}_+ = [0, +\infty)$ with $\mu(\emptyset) = 0$ and $\mu(\bigcup_{k<\infty} D_k) = \sum_{k\leq\infty} \mu(D_k)$, $D_1, D_2, \ldots$ disjoint in $\mathcal{S}$ (i.e. countably additive) (Casella & Berger 2002). D_k denotes the k^{th} (non-overlapping) partitioned subregion of the larger region D. A measure is σ-finite if there exists sets D_k in $\mathcal{S}$ whose union is S and such that $\mu(D_k) \leq \infty$. The *measure space* $\mathcal{M}(S)$ is the class of σ-finite (or locally finite) measures on S.

Given the triple $(S, \mathcal{S}, \mu)$, known as a measure space, a function $f : S \to \mathbb{R}$ is measurable if $f^{-1}(D) = \{s \in S : f(s) \in D\} \in \mathcal{S}$, $\forall\ D \in S$. All continuous functions are measurable functions. In this case, since f is a function from a sample space S into the real numbers, it is also called a random variable. Indeed $D \in S$, then $\{\xi \in D\} = \xi^{-1}D \in \mathcal{A}$, and the

associated probabilities can be shown as

$$P\{\xi \in D\} = P(\xi^{-1}D) = (P \circ \xi^{-1})D = \mathbb{L}_\xi D, \quad D \in S,$$

where $\mathbb{L}_\xi$ is called the distribution or law of ξ and is a probability measure on the range space S of ξ.

Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two measurable spaces. Then a (probability) *kernel density* from S to T is a function $\nu : S \times \mathcal{T} \to \mathbb{R}_+ = [0, +\infty]$ such that:

1. $\nu_s(B) = \nu(s, B)$ is $\mathcal{S}$-measurable in $s \in S$ for fixed $B \in \mathcal{T}$.
2. $\nu_s(B) = \nu(s, B)$ is a (probability) measure in $B \in \mathcal{T}$ for fixed $s \in S$.

Kernel densities provide tools to quantify spatial variation of event occurrence. If $\mu_s(T) = \mu(s, T) = 1$ for each $s \in S$ then μ is called a *probability kernel*. For any function $f : T \to \mathbb{R}$, a kernel ν defines an integral operator such that $\nu f(s) = \int_S \nu(s, dt) f(t)$ (Kallenberg, 2002).

Kernels are used to estimate the distribution function of the events of the point process and characterize the patterns of spatial distributions. We will use this for the autocorrelation measure defined in Section 4.3 and for the sample of points $x_1, \ldots, x_n$. Kernel density is a nonparametric density estimation method, with inference based on a finite data sample and smoothing of data determined by a bandwidth coefficient. The wider the bandwidth, the smoother the estimate of density but bias may increase. Reich, Hodges, and Zadnik (2006) studied the effects of residual smoothing in the case of spatial randomness on fixed effects and posterior variance (which could be infinite if smoothing is ignored). We will first test/validate such results by getting estimates without spatial smoothing and checking for over-dispersion. To implement smoothing techniques and bandwidth selections we will refer to R packages that have been developed (see Guidoum, 2015; Langrené & Warin, 2017). Kernel smoothing has also been used to estimate functions in survival analysis (Hess & Gentleman, 2015). Commonly used kernels are shown in Table 1 (Rizzo, 2008).

Table 1: Commonly used kernel functions for density estimation

Kernel	$\nu(t)$	Support
Uniform	$\frac{1}{2}$	$\lvert t\rvert < 1$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)$	$\mathbb{R}$
Epanechnikov	$\frac{3}{4}(1-t^2)$	$\lvert t\rvert < 1$
Rectangular	$\frac{1}{2}$	$\lvert t\rvert < 1$
Triangular	$1 - \lvert t\rvert$	$\lvert t\rvert < 1$
Biweight	$\frac{15}{16}(1-t^2)^2$	$\lvert t\rvert < 1$
Cosine	$\frac{\pi}{4} cos \frac{\pi}{2} t$	$\mathbb{R}$

Since we are interested in local intensity of the coverage of events say per unit area centered around a point, we use a locally finite kernel density estimate (KDE). KDE transforms the points/events into local intensity of counts. The locally finite kernel version of a point process ξ from the basic probability space into $\mathbb{R}^d$ is called a random measure. Locally finite here means that there exists a partition $D_1, D_2, \ldots \in \mathcal{S}$ of S such that $\xi D_k < \infty$ for all k (see Kallenberg 2002, Ch. 12). Note that $\xi D = \xi(\cdot, D)$ is a random variable in $[0, +\infty]$ for $D \in S$. To add more to the basic notions of random measure theory, we think of ξ as a random element and write $\xi f = \int f d\xi,\ \forall f \geq 0$, measurable functions and

$$\xi f(s) = \int_D \xi(s, dx) f(x). \tag{3}$$

The intensity of ξ is then defined as $\mathbb{E}\xi D = \mathbb{E}(\xi D) = \mu D$. The basic uniqueness criteria for random measures follows (Kallenberg, 2002).

Let ξ and η be random measures on S. Then $\xi \stackrel{d}{=} \eta$ under each of these conditions:

i. $(\xi D_1, \ldots, \xi D_n) \stackrel{d}{=} (\eta D_1, \ldots, \eta D_n)$ for any $D_1, \ldots, D_n \in S,\ n \in \mathbb{N}$;

ii. $\xi f \stackrel{d}{=} \eta f$ for any measurable function $f \geq 0$ on S.

If the random variables $\xi D_1, \ldots, \xi D_n$ are independent for any disjoint sets $D_1, \ldots, D_n \in S$ then the random measure ξ is said to have *independent increments* and Equation (3) can be written as:

$$\xi f = \sum_{k=1}^{n} \xi f \mathbf{1}_{D_k}(s) = \sum_{k=1}^{n} \int_{D_k} \xi(s, dx) f(x).$$

A distribution is said to be stable if a linear combination of two independent random variables with this distribution has the same distribution up to location and scale parameters. Given X_1, X_2 that are independent copies from X, if for any constants $a > 0$ and $b > 0$ the random variable $(aX_1 + bX_2)$ has the same distribution as $(cX + d)$ for suitable constants $c > 0$ and d, then X is stable (Kallenberg, 2002). The dynamics of the process are such that the choice of the kernels will be of the stable family of distributions.

The generalized family of stable distributions first described by Paul Lévy (1925) are the fundamental case of Markov processes and are expressed in terms of a characteristic triple, (a, b, ν), where a is the diffusion rate, b is the drift coefficient, and ν is the Lévy measure that determines the rates for jumps of different sizes.

Definition 2.1.2. (Kallenberg, 2005) The basic properties that define a Lévy process X are:

(i) right continuous with left-hand limits (*rcll*) almost surely;

(ii) X has independent increments;

(iii) X has stationary increments, $\forall\ t, s \geq 0, \forall\ h > 0, \quad X_{t+h} - X_{s+h} \stackrel{d}{=} X_t - X_s$;

(iv) $X_0 = 0$ almost surely;

(v) X_t is stochastically continuous, $X_{t+h} \xrightarrow{P} X_s$ as $h \to 0$.

Randomness in stochastic calculus is introduced through the change associated with a process. Itô's formula is similar to the fundamental theorem of calculus, and extends to the second derivative. To describe the structure of general Lévy processes, we use the Lévy-Itô representation that describes a Lévy process as the sum of three independent components, a linear drift, a Brownian motion, and a Lévy jump process that is a stochastic sum of independent Poisson random variables.

The Lévy-Itô representation with Lévy measure ν is

$$\begin{aligned} X_t &= bt + W_A(t) + \int_{|y|<1} y(N(t, dy) - t\nu dy) + \int_{|y|\geq 1} yN(t, dy) \\ &= bt + W_A(t) + \int_{\mathbb{R}} y\left\{N(t, dy) - \mathbf{1}_{[-1,1]} t\nu(dy)\right\}, \quad t \geq 0, \end{aligned} \tag{4}$$

where W is a Brownian motion with covariance matrix A and N is a Poisson process on $\mathbb{R}_+ \times (\mathbb{R}^d \setminus \{0\})$, independent of W and with mean measure $t\nu(dy)$ (Applebaum, 2004).

The following definitions are important to understanding properties associated with continuous time Lévy processes.

Definition 2.1.3. A random vector is *infinitely divisible* if for every $n \in \mathbb{N}$ there exists some i.i.d. array of random vectors $\xi_{n1}, \ldots, \xi_{nn}$ with $\sum_k \xi_{nk} \stackrel{d}{=} \xi$.

Definition 2.1.4. A process X is *exchangeable* if $(X_{k_1}, \ldots, X_{k_m}) \stackrel{d}{=} (X_1, \ldots, X_m)$ for any sequence $k_1, \ldots, k_m$ in the index of X.

Definition 2.1.5. A process X is *contractable* if $(X_{k_1}, \ldots, X_{k_m}) \stackrel{d}{=} (X_1, \ldots, X_m)$ for any $k_1 < \cdots < k_m$.

Theorem 2.1. (Kallenberg, 2002) *If X is an $\mathbb{R}^d$-valued process with $X_0 = 0$, then the following conditions are equivalent:*

(i) X is exchangeable.

(ii) X is contractable.

(iii) X has conditionally i.i.d increments.

Additionally, $X_t - X_s \sim \mathcal{P}(t-s)$, where $\mathcal{P}$ represents a class of infinitely divisible distributions. A random vector ξ or its distribution is defined as infinitely divisible if for every $n \geq 2$, $\xi \stackrel{d}{=} X_1 + X_2 + \cdots + X_n$ for some *i.i.d* random vectors $X_1, \ldots, X_n$.

Theorem 2.2. (Kallenberg, 2002) *The following conditions are equivalent for any random vector ξ in $\mathbb{R}^d$:*

(i) ξ is infinitely divisible.

(ii) $\sum_j X_j \stackrel{d}{\to} \xi$ for some i.i.d. array (X_j).

(iii) $\xi \stackrel{d}{=} X_1$ for some Lévy process in $\mathbb{R}^d$.

Under these conditions, $\mathbb{L}(X)$ is determined by $\mathbb{L}(\xi) = \mathbb{L}(X_1)$. Infinite divisibility is an important property of stable probability distributions. Stable distributions are scalable, which is of great use since a small portion of the distribution looks like the whole distribution

(shares the same characteristics). The Poisson and Gaussian distributions are examples of infinitely divisible distributions.

The probability kernel ν from S to some measurable space T is used to define the ν-*randomization* ζ of an arbitrary point process ξ on S. Assume ξ is nonrandom and equal to $\mu = \sum_k \delta_{s_k}$. Then we can take $\zeta = \sum_k \delta_{s_k,\gamma_k}$, where γ_k are independent random elements in T with distributions $\nu(s_k, \cdot)$ (Kallenberg, 2002). The resulting distribution P_μ of ζ only depends on μ. There are two special cases of ν-randomization that are of interest. When $T = \{0, 1\}$ and $\nu(s, \{0\}) \equiv p \in [0, 1]$, the point process $\xi_p = \zeta(\cdot \times \{0\})$ on S is referred to as a *p-thinning* of ξ. The other special case of interest arises when $S = \{0\}$, $\xi = \kappa\delta_0$, and $\nu = \mu/\mu T$ for some $\mu \in \mathcal{M}(T)$ with $\mu T \in (0, \infty)$. In this case, ζ is referred to as a *mixed binomial process* directed by μ and κ. Conditional on κ, ζD is binomially distributed with parameters νD and κ.

2.1.2 POISSON POINT PROCESS

Spatial point process models can be defined by suggesting a constant or deterministic intensity function. Models of this class are presented under Poisson processes (Kingman, 1993).

Definition 2.1.6. The point process N on domain S defines a *Poisson process* with intensity measure λ since

1. For $D_1, D_2, \ldots$ disjoint sets $N(D_1), N(D_2), \ldots$ are independent.
2. For D fixed set, $N(D)$ is a Poisson distribution with parameter $\lambda(D)$, where λ is the intensity measure of N.

The *homogeneous Poisson point process* (PPP) is a basic point process that satisfies Definition 2.1.2. Indeed it has the following properties

(i) Independence: for $D_1, D_2, \ldots$ disjoint sets in S, $N(D_1), N(D_2), \ldots$ are independent; i.e complete randomness;

(ii) Poisson distribution: for a fixed set, $N(D)$ its expected value defines a measure called an intensity measure denoted as λ. In other words, $\mathbb{E}(N(D)) = \lambda D$. If

$N(D)$ is a Poisson distribution of the form

$$P(N(D) = n) = \frac{e^{-\lambda D}(\lambda D)^n}{n!}, \tag{5}$$

where λ is the intensity measure of N;

(iii) Homogeneity: $\mathbb{E}(N(D_k)) = Var(N(D_k)) = \lambda D_k$, for $D_k \in S$ where $k = 1, \ldots$.

The intensity is indeed a measure because $\lambda(D) := \mathbb{E}(N(D))$ satisfies properties of a measure. If the measure N is a binomial process, then

$$\lambda(D) = \mathbb{E}\left(\sum_{k=1}^{m} \mathbf{1}(X_k \in D)\right) = \sum_{k=1}^{m} P(X_k \in D) = m \cdot Q(D), \text{ for } m \geq 1,$$

where Q is such that $D \sim Q$.

The homogeneous PPP or *complete spatial randomness* has the following important properties (Baddeley, Rubak, and Turner, 2016):

(i) Thinning property: given a homogeneous Poisson process with intensity λ, then applying complete random thinning (deleting some points of the point pattern) with retention probability p, results in the remaining points that form a homogeneous Poisson process with intensity $p\lambda$;

(ii) Superposition property: given X_1 and X_2 independent homogeneous Poisson processes with intensities λ_1 and λ_2, combining the points from both processes results in homogeneous Poisson process with intensity of the form $\lambda_1 + \lambda_2$.

The *nonhomogeneous Poisson point process* has varying intensities across its domain (usually space or time) λD with the main assumption that points are independent of each other. The nonhomogeneous Poisson process has the following important properties (Baddeley et al., 2016):

(i) Conditional property: given exactly n points in a region D, these points are independent and each point has the same probability distribution over D, with probability density $f(x) = \lambda(x)/\mu$, where $\mu = \int_D \lambda(x)dx$.

(ii) Random thinning property: Suppose that $N \sim Poisson(\lambda)$, and that $X_1, X_2, \ldots, X_n$ are *i.i.d* multinomial random variables with distribution $Multinomial\ (p_1, p_2, \ldots, p_n)$, that is $P\{X_i = k\} = p_k$ for $k = 1, 2, \ldots, m$. Then the random variables $N_1, N_2, \ldots, N_m$ defined by $N_k = \sum_i \mathbf{1}\{X_i = k\}$ are independent Poisson random variables with parameters $E(N_k) = (\lambda p_k)$.

(iii) Superposition property: If $Y_1, Y_2, \ldots, Y_n$ are independent Poisson random variables with means $\mathbb{E}(Y_i) = \lambda_i$ then $\sum_i Y_i \sim Poisson(\sum \lambda_i)$.

The Laplace transform of a Poisson process N is

$$\mathcal{L}_N(f) = \mathbb{E}\left\{\exp\left[-\int_D (1 - e^{-f(x)})\mu(dx)\right]\right\} \tag{6}$$

. The form of the Laplace functional follows from the properties of the Poisson process. Given x is a Poisson random variable with mean μ, then

$$\mathbb{E}\ e^{-cx} = e^{-\mu} \sum_{k\geq 0} (\mu e^{-c})^k \Big/ k!, \quad c \in \mathbb{R}^2. \tag{7}$$

Let $f = \sum_{k\leq m} c_k \mathbf{1}_{D_k}$, where $c_k \in \mathbb{R}_+$ and the sets $D_k \in \mathcal{S}$ are disjoint with $\mu D_k < \infty$. Then

$$\begin{aligned}
\mathbb{E}e^{-\xi f} &= \mathbb{E}\exp\left\{-\sum_k c_k \xi D_k\right\} = \prod_k \mathbb{E}e^{-c_k \xi D_k} \\
&= \prod_k \exp\left\{-\mu D_k (1 - e^{-c_k})\right\} \\
&= \exp\left\{-\sum_k \mu D_k (1 - e^{-c_k})\right\}.
\end{aligned} \tag{8}$$

The point process ξ on $(S, \mathcal{S})$ is said to be Poisson with intensity μ (σ-finite) if the random variables ξD are independent for disjoint $D \in \mathcal{T}$ and Poisson distributed with mean μD and Laplace transform:

$$\mathbb{E}e^{-\xi f} = \exp\{-\mu(1 - e^{-f})\}. \tag{9}$$

In the more general case of $f \geq 0$, if we take simple functions $f_n \geq 0$ with $f_n \uparrow f$, then by monotone convergence $\xi f_n \to \xi f$ and $\mu(1 - e^{-f_n}) \to \mu(1 - e^{-f})$. Then by dominated convergence, Equation (9) holds for any $f \geq 0$.

If ξ_p is a p-thinning of ξ, then

$$\mathbb{E}e^{-\xi_p f - \xi g} = \mathbb{E}\exp\left\{-\xi\left(g - \log\{1 - p(1 - e^{-f})\}\right)\right\}, \tag{10}$$

for all f and g non-negative measurable functions.

Equivalence of two Poisson point processes: The following are equivalent for point process N and N' on D,

1. $N \stackrel{d}{=} N'$ (i.e. N and N' have the same distribution).
2. $\big(N(D_1), \ldots, N(D_m)\big) \stackrel{d}{=} \big(N'(D_1), \ldots, N'(D_m)\big)$, $\forall\, m \in \mathbb{N}$ and all pairwise disjoint sets $D_1, D_2, \ldots, D_m \in \mathcal{D}$.
3. $\mathcal{L}_N(f) = \mathcal{L}_{N'}(f),\ \forall\ f \in \mathcal{D}$.
4. $\forall\ f \in \mathcal{D},\ N(f) \stackrel{d}{=} N'(f)$.

For any finite measure on D, there exists a Poisson point process on D with intensity μ (Last & Penrose, 2017). If there is a countable sum of finite measure on D, then D is referred to as s-finite (Last & Penrose, 2017; Norris, 2018). We construct the Poisson point process on the same probability space. The distribution modeling has been applied in Hastie and Tibshirani (1990) and in Elith, Leathwick, and Hastie (2008). The spread/occurrence is not just determined by the count (modeled after a logistic as in McCullagh & Nelder, 1989), but also by including the location where occurrence/event has been noticed.

2.1.3 COX PROCESSES

In 1955, David Cox introduced the Cox process or doubly stochastic point process where the intensity is also random and independent of the underlying Poisson process (Stoyan, Kendall, & Mecke, 1995). Cox processes are suggested for modeling processes that are driven by the environment, not those driven by interaction between points as discussed for Markov processes. For example, a Cox process could be used to model a noninfectious disease, whereas a Markov process would be better suited for modeling an infectious disease.

A point process ξ is a *Cox process* directed by η, any random measure on D, if it is conditionally Poisson given η, where $\mathbb{E}(\xi|\eta) = \eta$ almost surely. Another special case of the

Cox process is the *mixed Poisson process* based on μ and α that may be formed by choosing $\eta = \alpha\mu$ for some measure $\mu \in \mathcal{M}(D)$ and random variable $\alpha \geq 0$ (see Kallenberg, 2002). In other words, the mixed Poisson process is conditionally Poisson given the random variable α. From Equation (5), then

$$P(n, D|\alpha) = \frac{e^{-\alpha\mu D}(\alpha\mu D)^n}{n!}, \quad \text{then}$$

$$P(n, D) = \int_D \frac{e^{-\alpha\mu D}(\alpha\mu D)^n}{n!} dG(\alpha), \tag{11}$$

where α is a random variable with cumulative distribution $G(\alpha)$. Using Bayesian analysis, the posterior distribution can be determined given the prior distribution $G(\alpha)$.

In the case where α follows a Gamma distribution, $N(D)$ follows the negative binomial distribution and Equation (11) becomes:

$$\begin{aligned} P(n, D) &= \int_0^\infty P(n, D|\alpha) g(\alpha) d(\alpha) \\ &= \int_0^\infty \frac{e^{-\alpha\mu D}(\alpha\mu D)^n}{n!} \frac{a^b}{\Gamma(b)} \alpha^{b-1} e^{-a\alpha} d\alpha \\ &= \int_0^\infty \frac{a^b}{n!\Gamma(b)} \alpha^{n+b-1} (\mu D)^n e^{-(1-a)\alpha} e^{-\mu D} d\alpha \\ &= \frac{a^b}{n!\Gamma(b)} \frac{\Gamma(n+a)}{(b+1)^{n+a}} \frac{(\mu D)^n}{e^{-\mu D}} \int_0^\infty \frac{(b+1)^{n+a}}{\Gamma(n+a)} \alpha^{n+b-1} e^{-(1+a)\alpha} d\alpha. \end{aligned} \tag{12}$$

It follows from Equation (6) that the Laplace transform of a Cox process ξ directed by η, is

$$\mathcal{L}_\xi(f) = \mathbb{E}\left\{ \exp\left[-\int_D (1 - e^{-f(x)}) \eta(dx) \right] \right\}.$$

Theorem 2.3. (Kallenberg, 2002) *For every $n \in \mathbb{N}$, let xi_n be a p_n-thinning of some point process η_n on S, where S is lcscH and $p_n \to 0$. Then $\xi_n \xrightarrow{d}$ some η iff $p_n\eta_n \xrightarrow{d}$ some η, in which case ξ is distributed as a Cox process directed by η.*

This convergence of thinnings leads to an interesting characterization of Cox processes.

Theorem 2.4. (Kallenberg, 2002) *Let ξ be a point process on S. Then ξ is Cox iff for every $p \in (0, 1)$ there exists a point process ξ_p such that ξ is distributed as a p-thinning of ξ_p.*

Theorem 2.5. (Kallenberg, 2002) *Exchangeability applies to a process X on $[0, 1]$ iff it can*

be represented as:

$$X_t = \alpha t + \sigma B_t + \sum_j j\beta_j(\mathbf{1}\{\tau_j \leq t\} - t), \quad t \in [0,1],$$

for some Brownian bridge B, some independent i.i.d $U(0,1)$ random variables $\tau_1, \tau_2, \ldots$, and some independent set of coefficients α, σ, and $\beta_1, \beta_2, \ldots$ such that $\sum_j \beta_j^2 < \infty$ almost surely. Then, the sum $\sum_j j\beta_j(\mathbf{1}\{\tau_j \leq t\} - t)$ converges in probability, uniformly on $[0,1]$, toward an rcll limit.

This is similar to the representation in Equation (4).

The log-Gaussian process on $\mathbb{R}^d$ is a Cox process whose driving intensity is of the form $\Lambda(x) = \exp\{Y(x)\}$, where Y is a Gaussian process, i.e. every point in the process is associated with a normal distributed random variable and the log of the intensity is a Gaussian process (Møller, Syversveen, & Waagepeterson, 1998). The benefit of using an underlying Gaussian process is that it is uniquely identified by its mean function and covariance function. The log-Gaussian Cox process (LGCP) can then be defined as $\Lambda(x) = \lambda(x)\exp\{Y(x)\}$, where $\lambda(x)$ is a deterministic process and $Y(x)$ is a stationary Gaussian process (Brix & Diggle, 2001). Using the logarithmic scale to build models ensures that intensity values are positive. The LGCP is flexible, easily managed, and useful for spatial prediction rather than simple hypothesis testing (Diggle, Moraga, Rowlingson, & Taylor, 2013).

Shot-noise Cox process (SNCP) is a Cox process whose driving intensity is of the form $\Lambda(x) = \sum_{Y \in N} K(Y, x)$, for a Poisson point process $N(\cdot)$ and kernel $K(\cdot, \cdot)$. SNCP can be viewed as a cluster process that is formed by first generating a point process Y of "parent" points, and then generating a point process around each "parent" point giving rise to a random pattern of "offspring" that replace the "parent." The collection of all "offspring" points forms the cluster point process (Baddeley et al., 2016). Reproduction can continue over time through multiple generations, with each new generation replacing the one before. The "parent" points can be generated from a homogeneous or nonhomogeneous point process. Each cluster of "offspring" from a homogeneous "parent" can have the same (homogeneous) or varying (nonhomogeneous) intensity.

Matérn-thinned Cox process is most useful for models that exhibit regularity on a local scale, but exhibit clustering on a global scale. In other words, a point is deleted if it lies (arbitrarily) too close to its nearest neighbor (Baddeley et al., 2016). The driving intensity

is of the form $\Lambda(x) = \sum_{(\gamma,c)\in N} \gamma K(c,x)$, where N is a Poisson cluster process, c is the cluster center, γ is the cluster intensity, and $K(c,\cdot)$ is a kernel. Matérn-thinned Cox process can be viewed as a Poisson cluster process, since the distribution of $X|N$ is the superposition of independent Poisson processes $X_{(c,\gamma)},\ (c,\gamma) \in N$, with intensity function $\gamma K(c,\cdot)$ (Anderson & Hahn, 2015).

2.1.4 MARKOV PROCESS

Markov (also called *Gibbs*) *point processes* appeared first in physics of spatial point processes mainly to model mark system interactions. It can be thought of as a Lévy process with induced filtration and transition kernel as introduced in Subsection 2.1.1. The Markov point process validates discovery of small scale point-to-point interactions and is often used in statistical physics to model repulsion and attraction in particle systems.

The homogeneous Markov process of n points in the bounded region D has density

$$f(x_1,\ldots,x_n) = c \cdot \exp\left\{-\sum_{i<j} \theta\ \|x_i - x_j\|\right\},\ x_i \in D, \tag{13}$$

where c is a normalizing constant that cannot be directly calculated and θ is a pair potential (measure of potential energy or reaction) function that models the interaction between points, and $\|\cdot\|$ denotes the L_1-norm in D.

Modeling short range interaction between the points can be challenging as the choice of model parameters is not easy to suggest. Stoyan and Stoyan (1998) and Baddeley and van Lieshout (1995) recommend that covariates be observed at each point within a known interaction range. Additionally, Markov processes are useful for models that exhibit regular patterns; however, they are not adequate for strong clustering models (Chiu, Stoyan, Kendall, & Mecke, 2013). While the Poisson process is completely specified by its intensity, one must look at the conditional intensity as a modeling tool for the Markov process (Baddeley et al., 2016). The Poisson process is a special case of the Markov process where $\lambda(x) = e^{\theta(x)}$.

2.2 PALM DISTRIBUTION

Swedish teletrafficist Conny Palm applied stochastic processes to telephone exchanges in 1943 and *Palm calculus* was later named after him for the study of the relationship between

probabilities conditioned on a specified event occurrence. We are interested in studying the interval properties and space of the process at a chosen point. To do so, the conditional probabilities must be calculated conditioned on there being a point (event) of the process at a specified location in a two dimensional space (see Baddeley, 2007). As such, the conditional distribution provides a flexible tool in modeling spatial patterns.

Using this concept, the *Palm distribution* at location s, denoted as $\mathbb{P}_s$ is defined as:

$$\mathbb{P}_s(B) = \frac{\mathbb{E}[\xi(ds); \gamma \in B]}{\mathbb{E}[\xi(ds)]}, \quad s \in S,\ B \in \mathcal{T}, \tag{14}$$

where ξ is a random measure on the measurable space $(S, \mathcal{S})$, γ is a random element in a measurable space $(T, \mathcal{T})$, and B is a random event in the spatial region T. The Poisson, Cox, and Markov (Gibbs) point processes can all be generalized to the Palm distribution. ξ can be the intensity function for a 2-dimensional point process such as Poisson and in that case

$$\mathbb{E}(n(X \cap A)\mathbf{1}_B) = \int_A \mathbb{P}_s(B)\xi s ds,$$

where $A \subseteq S$ is a spatial region. Denoting the restriction of X to A as $X_A := X \cap A$, local finiteness of X means that X_A is finite almost surely when A is bounded (Couerjolly, Møller, & Waagpeterson 2017). If the process can be described as originating at point s, $\mathbb{P}_s(B)$ is the conditional intensity of area B given realization s, thus mirroring the Papangelou conditional intensity defined as the conditional intensity of a process generated from a point (see Coeurjolly, Møller, & Waagpeterson, 2016; Daley & Vere-Jones, 2008; Kallenberg, 1983, 2005).

To understand the above representation, we make use of a formula for moments first developed by Campbell in 1909:

$$\mathbb{E}\left[\int g(x)N(dx)\right] = \int g(x)M(dx). \tag{15}$$

The term Campbell measure was later adopted for the underlying concept of moments and Palm distributions. Assume ξ is σ-finite then Equation (15) can be written as a disintegration formula (Daley & Vere-Jones, 2008).

$$Cf = \mathbb{E}\int f(s,\gamma)\xi(ds) = \int \mathbb{E}\xi(ds) \int f(s,t)(dt) \quad f \geq 0,\ f \in (S \times T). \tag{16}$$

Given the measurable space $(S, \mathcal{S})$, the associated Campbell measure C on $S \times \mathcal{M}(S)$, is defined by

$$Cf = \mathbb{E} \int f(s, \xi)\xi(ds), \quad f \in (S \times \mathcal{M}(S))_+, \tag{17}$$

where $(S \times \mathcal{M}(S)_+)$ denotes the class of measurable functions $f \geq 0$ on $S \times \mathcal{M}(S)$ (Kallenberg, 2005)

It follows that C is also σ-finite, i.e. $C(\sum f\mathbf{1}_{D_k}) = \sum C(f\mathbf{1}_{D_k})$ for D_k partition in S, so f can be chosen to be strictly positive with $Cf < \infty$. In this case, the projection $\nu = (f \cdot C)(\cdot \times \mathcal{M})$ is bounded and $\nu D = 0$ *iff* $\xi D = 0$ almost surely for every $D \in S$. Choosing $\nu = \mathbb{E}\xi$ when the latter is σ-finite simplifies computations, and ν is called the *supporting measure* of ξ.

If S is Borel, i.e. $\mathcal{S} = \mathcal{B}(S)$ the σ-algebra of all open sets of S, then $\mathcal{M}(S)$, the space of probability measures on S, is also Borel, and there exists a kernel family $\mathbb{P} = (\mathbb{P}_s)$, $s \in S$, the *Palm measures* of ξ, from S to $\mathcal{M}(S)$ satisfying the disintegration formula (see Theorem 6.4, Kallenberg, 2002). Then Equation (17) can be expressed as:

$$\begin{aligned}
C(f\mathbf{1}_D) &= \mathbb{E} \int_D f(s, \xi)\xi(ds) = \mathbb{E}\left[\xi D \cdot \int_D f(s, \xi)\frac{\xi(ds)}{\xi D}\right] \\
&= \mathbb{E}\left[\xi D \cdot \int_D f(s, \xi)P[s \in ds|\xi]\right] = \mathbb{E}\left[\xi D \cdot \mathbb{E}[f(s, \xi)|\xi]\right] \\
&= \mathbb{E}\left[\mathbb{E}[\xi D \cdot f(s, \xi)|\xi]\right] = \mathbb{E}\left[\xi D \cdot f(s, \xi)\right] \\
&= \mathbb{E}\left[\mathbb{E}[\xi D \cdot f(s, \xi)|s]; s \in D\right] \quad \text{for } s \notin D \text{ when } \xi D = 0 \\
&= \int \nu(ds) \int f(s, \mu)\mathbb{P}_s(d\mu), \quad f \in (S \times \mathcal{M}(S))_+,
\end{aligned} \tag{18}$$

and we denote this simply as $C = \nu \otimes \mathbb{P}$. The Palm measures $\mathbb{P}_s$ of ξ can be chosen to be probability measures and are then referred to as *Palm distributions*.

Since ξ is a point process, the reduced Campbell measure C' on $S \times \mathcal{N}(S)$ at location s can also be defined from Equation (17) as

$$C'_s f = \mathbb{E} \int f(s, \xi - \delta_s)\xi(ds), \quad f \in (S \times \mathcal{N}(S))_+. \tag{19}$$

Then disintegration of C'_s leads to the reduced Palm measures (denoted $\mathbb{Q}_s$):

$$C'_s f = \int \nu(ds) \int f(s,\mu)\mathbb{Q}_s(d\mu), \quad f \in (S \times \mathcal{N}(S))_+, \tag{20}$$

and we denote this simply as $C' = \nu \otimes \mathbb{Q}$. When compared with Equation (18), we see that

$$\begin{aligned} C'_s &= \int \nu(ds) \int f(\mu)\mathbb{Q}_s(d\mu) \\ &= \int \left(\int f(\mu)\mathbb{Q}_s(d\mu) \right) \nu(ds) = \mathbb{E} \int_D f(\xi - \delta_s)\xi(ds) \\ &= \int \nu(ds) \int f(\mu - \delta_s)\mathbb{P}_s(d\mu). \end{aligned} \tag{21}$$

This implies $\mathbb{Q}_s = \mathbb{P}_s \circ (\mu - \delta_s)^{-1}$, therefore the reduced Palm distribution $\mathbb{Q}_s$ can be described as the conditional distribution of the point process ξ_s by omitting the point at s, given $\xi(s) > 0$.

The above discussion makes use of the Palm measure/distribution and now we will define associated properties. Consider ζ to be a random measure on $D \in \mathbb{R}^d$ and X to be a measurable/counting random process, with $\mathcal{D}$ the σ-field of subsets of D generated by all ξ in the above form values. Then the pair $(D, \mathcal{D})$ is a measurable space.

The pair (X, ζ) is called jointly stationary as defined in Kallenberg (2002) if

$$\theta_s(X, \zeta) \equiv (\theta_s X, \theta_s \zeta) \equiv (X + s, \zeta) \stackrel{d}{=} (X, \zeta),$$

for any $s \in D$, where θ_s denotes the shift operator. In other words, the pair are translation invariant, meaning that the likelihood of an event occurring is the same throughout space. Stationarity of the random process and the (Lebesgue) measure are needed to define the *intensity* of ζ, $\mathbb{E}\zeta = c\lambda^d$, for some constant $c \in [0, \infty]$, where d is the dimension of the space, $\mathbb{R}^d$ and λ is the Lebesgue measure. Intensity is analogous to the expected value of a random variable, and denotes the expected number of events per unit area (Baddeley, 2007). It is referred to as an *intensity function* λ when $\mathbb{E}\zeta = \int_D \lambda(x)dx$ for some $\lambda(x)$, and some small region $dx \subset \mathbb{R}^d$.

If the pair (X, ζ) is jointly stationary, and ζ is of finite positive intensity on D, then the

Palm measure at the point $s \in D$, with respect to ζ, can be defined as

$$\mathbb{P}_s(f) = \frac{1}{\mathbb{E}\zeta D}\mathbb{E}\int_D f(\theta_s(X,\zeta))\zeta(ds), \tag{22}$$

where

1. D is any measurable set, i.e $D \in \mathcal{D}$;
2. $\zeta(D) \in \mathbb{R}_+$; and
3. f is any nonnegative measurable function on D.

Choosing ζ to be probability measures gives rise to the Palm distribution. The Palm distribution can be extended to the multivariate case $\mathbb{P}_{s_1}, \ldots, \mathbb{P}_{s_n}$ for $s_i \in D$. There can be small regions within the area where these measurements are different from the rest of the area, referred to as pockets of nonstationarity, where the covariance function of these pockets is stationary and is given by $c(h) = cov(\zeta(s+h), \zeta(s))$ (Cressie, 1993).

The relationship between the Campbell measure and the first moment measure $M(\cdot)$ is important to understanding the Palm distribution. This relationship can be seen through the definition of the Palm kernel. When $M(\cdot)$ exists as a boundedly finite measure,

$$\int_D \mathbb{P}_s(U)M(ds) = C_{\mathbb{P}}(D \times U), \tag{23}$$

for each fixed $U \in \mathcal{B}(\mathcal{M}_{\mathcal{S}}^{\#})$ (Borel sigma-field of boundedly finite measures on $\mathcal{S}$). Here, each $\mathbb{P}_s(U)$ is a local Palm distribution for ξ. The family of these measures $\{\mathbb{P}_s(U)\}$ is called a Palm kernel if the measures satisfy the following conditions (Daley & Vere-Jones, 2008):

1. for each fixed U, $\mathbb{P}_s(U)$ is a measurable function of s that is fM-integrable on bounded subsets of $\mathcal{S}$; and
2. for each fixed $s \in \mathcal{S}$, $\mathbb{P}_s(U)$ is a probability measure on $\mathcal{B}(\mathcal{M}_{\mathcal{S}}^{\#})$.

The family of Laplace functionals $\{\mathcal{L}_x[f]\}$ derived from the associated Palm kernels is defined as

$$\mathcal{L}_x[f] = \int_{\mathcal{M}_{\mathcal{D}}^{\#}} \exp\left[-\int_{\mathcal{D}} f(y)\xi(dy)\right]\mathbb{P}_x(d\xi), \tag{24}$$

where ξ is a random measure with distribution $\mathbb{P}$ and $\mathcal{M}_{\mathcal{D}}^{\#}$ indicates the boundedly finite measures on $\mathcal{D}$ (Daley & Vere-Jones 2008). The relationship between the Laplace functional associated with the original random measure $\mathbb{L}[f]$ and the Laplace functional of the Palm kernel $\mathcal{L}_x[f]$ for a Poisson process is given by:

$$\mathcal{L}_x[f] = e^{-f(x)}\mathcal{L}[f] = \mathcal{L}_{\delta_x}[f]\mathcal{L}[f], \tag{25}$$

where $\mathcal{L}_{\delta_x}$ denotes the Laplace functional of a degenerate random measure (Daley & Vere-Jones, 2008).

Some examples of Palm measures include the Poisson process, the log Gaussian Cox process and the finite Markov process. To the theory of point process, Palm measure processes and concepts are of prime importance (see Cressie, 1993). They can be described as the distribution conditioned on the original point process (Stoyan et al., 1995). Under a propagation model, even with normal or Poisson conditioning, efficient algorithms can then be built. Propagation involves invariance or exchangeable properties in the locations and their neighborhood areas and densities.

2.2.1 INVARIANCE

The Palm distribution is never stationary (translation invariant) because the underlying point process must always include the origin point (Chiu et al., 2013); however, if the point process is motion-invariant then its Palm distribution is isotropic (rotation invariant) so that the scaling does not depend on the coordinate system used, which allows for flexibility as the spread may tend toward any direction. Let ξ be a mixed Poisson process on S as in Equation (11). Given a fixed σ-finite measure λ and a random measure $\alpha > 0$, the process is directed by $\alpha\mu$. Then

$$P[\xi D = 0] = \psi(\mu D) = \mathbb{E}e^{-\alpha\mu D}, \quad D \subset S, \tag{26}$$

where $\psi(t) = \mathbb{E}e^{-t\alpha}$ denotes the Laplace transform of α. Equation (26) also holds for a mixed binomial process based on the probability measure $\lambda/\lambda S$ and a $\mathbb{Z}_+$-valued random variable κ when $\mu S \in (0, \infty)$ and $\psi(f)$ is chosen to be $\mathbb{E}(1 - t/\lambda S)^{\kappa}$ for $t \in [0, \lambda S]$. In either case, the Laplace functional of ξ is given by

$$\mathbb{E}e^{-\xi f} = \psi\{-\lambda(1 - e^{-f})\} = \mathbb{E}\exp\{-\alpha\lambda(1 - e^{-f})\}, \quad f \geq 0. \tag{27}$$

Therefore, the distribution of ξ is uniquely given by (λ, ψ), and we can write $\mathbb{L}(\xi) = M(\lambda, \psi)$ (Kallenberg, 2005). The Hausdorff-Bernstein Theorem can also be extended by choosing ψ to be any completely monotone function on $0 < \mu S < \infty$.

The mixed Poisson and binomial processes described above can be characterized in terms of their reduced Palm measures as shown in Kallenberg (2005). This reduced Palm measure is translation invariant since the origin has been omitted. The following characterization illustrates the invariance of Palm measures.

Theorem 2.6. (Kallenberg, 2005) *Let ξ be a point process on a Borel space S with reduced Palm measures $\mathbb{Q}_s$, $s \in S$. We can choose $\mathbb{Q}_s$ to be independent of s iff ξ is a mixed Poisson or binomial process, in which case its distribution $\mathcal{L}(\xi) = M(\lambda, \psi_\xi)$ iff $\mathbb{Q}_s = M(\lambda, -\psi_\xi)$ almost everywhere λ, where λ is a σ-finite measure and ψ denotes the Laplace transform described above.*

Proof. Suppose that $\mathcal{L}(\xi) = M(\lambda, \psi)$, where λ is a supporting measure for ξ. Given a measurable function $f \geq 0$ on S with $\lambda f > 0$ and a set $D \in \mathcal{S}$ with $\mu D < \infty$, then

$$\mathbb{E}e^{-\xi f - t\xi D} = \psi(\lambda(1 - e^{-f-t\mathbf{1}_D})), \quad t \geq 0.$$

Then by dominated convergence we can take the right derivatives at $t = 0$, which yields

$$\mathbb{E}\xi D e^{-\xi f} = -\psi'(\lambda(1 - e^{-f}))\lambda(\mathbf{1}_D e^{-f}).$$

Making use of the Palm measures $\mathbb{P}$ of ξ associated with the supporting measure λ as in Equation (14), we obtain

$$\int_D \lambda(ds) \int e^{-\mu f} \mathbb{P}_s(d\mu) = -\psi'(\lambda(1 - e^{-f})) \int_D e^{-f(s)} \lambda(ds). \tag{28}$$

Since the choice of D was arbitrary, for $s \in S$ almost everywhere λ, Equation (28) becomes

$$\int e^{-\mu f} \mathbb{P}_s(d\mu) = -\psi'(\lambda(1 - e^{-f}))e^{-f(s)}, \tag{29}$$

which implies

$$\int e^{-\mu f} \mathbb{Q}_s(d\mu) = \int e^{-(\mu f - \delta_s)} \mathbb{P}_s(d\mu) = -\psi'(\lambda(1 - e^{-f})). \tag{30}$$

By monotone convergence, this extends to arbitrary $f \geq 0$. Thus we conclude that $\mathbb{Q} = M(\lambda, -\psi')$ almost everywhere λ and $\mathbb{Q}$ is independent of $s \in S$ almost everywhere λ. □

Such property is not true for all point process. Ginibre point processes have intermediate properties that do not satisfy both translation and scale invariance (Osada and Shirai, 2015).

2.2.2 SYMMETRIES

We say that a random measure ξ is *symmetrically distributed* with regard to λ or simply λ-*symmetric* if $\xi f^{-1} \stackrel{d}{=} \xi$ for every measurable mapping f with $\lambda f^{-1} = \lambda$ (Kallenberg, 1983).

Exchangeability can be defined by thinking of a point process as a random element Consider the finite or infinite sequence of random elements $\xi = (\xi_1, \xi_2, \dots)$ in the measurable space $(S, \mathcal{S})$. ξ is said to be *exchangeable* if

$$(\xi_{k_1}, \dots, \xi_{k_m}) \stackrel{d}{=} (\xi_1, \dots, \xi_m)$$

for any sequence $k_1, \dots, k_m$ in the index set of ξ. Furthermore, ξ is *contractable* if the same distributional equivalence holds for $k_1 < \cdots < k_m$. The sequence ξ is exchangeable if an only if it is contractable. When S is Borel, ξ is also conditionally *i.i.d* (Kallenberg, 2005).

LEMMA **2.7.** *Now let ξ be a marked point process on $S \times \mathbb{R}_+$, where S is Borel. Then ξ is contractable iff it is a Cox process directed by $\nu \otimes \lambda$ for some random measure ν on S (Kallenberg 2005).*

Proof. (Kallenberg, 2005) Suppose that ξ is contractable, as defined in Section 2.1.1. We see that ξ has stationary, independent increments over $\mathbb{R}_+$. Hence, we conclude that ξ is conditionally Poisson with intensity measure $\nu \otimes \lambda$. Thus, ξ is a Cox process directed by $\nu \otimes \lambda$. □

LEMMA **2.8.** *Let ξ be a marked point process on $S \times [0,1]$, where S is Borel. Then ξ is exchangeable iff it is a uniform randomization of the point process $\beta = \xi(\cdot \times [0,1])$ (see Kallenberg, 2005 for proof).*

Theorem 2.9. (Peng, 2008) *Let ξ be a marked point process on Borel space $S \times T$, where $T = \mathbb{R}_+$ or $[0,1]$. Assume $\mathbb{E}\xi = \nu \otimes \lambda$ for some σ-finite measure ν on S. The following two conditions are equivalent:*

i. ξ *is exchangeable.*

ii. $\mathbb{Q}_{s,t}$ *has a version that is independent of* t*.*

Note that under the invariance of supporting measure, the reduced Palm measure is invariant *iff* the reduced Campbell measure is invariant, i.e.:

$$\begin{aligned} C'f &= \mathbb{E}\int \xi(ds) f(s, \xi - \delta_s) \\ &= \mathbb{E}\mathbb{E}\left[\int \xi(ds) f(s, \xi - \delta_s)|\beta\right] \\ &= \mathbb{E}\mathbb{E}\left[\int \xi(dsdt) f((s,t), \xi - \delta_{s,t})|\beta\right] \end{aligned} \tag{31}$$

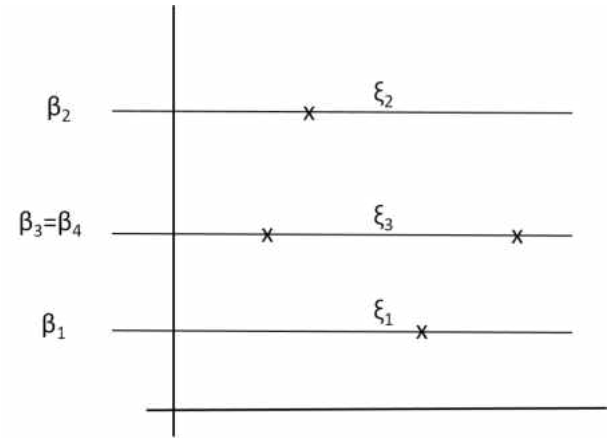


Figure 2: Independent binomial processes: $\xi_1, \xi_2, \ldots$.

$\mathbb{Q}_{s,k} = \mathbb{L}(\xi_1, \ldots, \xi_{k-1}, \xi'_k, \xi_{k+1}, \ldots)$, where ξ'_k is like ξ_k but with one point missing.

Table 2: Conditioning of Palm measures

Level	Point process	Multi-plicity	Red. Palm dist level k	Multiplicity of Red. Palm level k
1	ξ_1	η_1	ξ_1	η_1
2	ξ_2	η_2	ξ_2	η_2
3	ξ_3	η_3	ξ_3	η_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	ξ_k	η_k	ξ'_k	$\eta_k - 1$
$k+1$	ξ_{k+1}	η_{k+1}	ξ_{k+1}	η_{k+1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

2.2.3 EXPECTED VALUE

Point processes are statistical models of data recorded at irregular scattered locations that are determined by the largest areas of independently generated spatial points which will form a partition of the whole space. These partitioned spaces can be described under planar tessellation, or spatial tessellations, in a number of different manners, such as regular grids, or irregular grids based on Voronoi cells, Delaunay tessellations, random-line mosaic, and tilings. We will focus on Voronoi cells. In general, the Voronoi cell is a convex polygon around a point s that contains the entire space within the window that is closer to s than any other point in the region. Cheilaris, Khramtcova, Langerman, and Papadopoulou (2014) describe Hausdorff Voronoi cells in which each site is a cluster of points and the structure subdivides the area into sequentially finer subareas. Kallenberg (2002) further restricts the Voronoi subarea by defining the largest circle that is contained within the convex polygon, defined as:

$$V_{\xi_s} = \{s \in \mathbb{R}^d;\ \mu(S_{|s|} + s) = 0\}, \quad \mu \in \mathcal{N}(\mathbb{R}^d), \tag{32}$$

where $\mathcal{N}(\mathbb{R}^d)$ is the class of locally finite measures on $\mathbb{R}_+$ and S_r denotes the ball of radius r around s. Under such framework, nontrivial measures whenever possible can be defined. V_ξ in Equation (32) represents a stationary random field whose boundary is such that the measure of any point occurring on the boundary is 0; i.e. observations lie at the interior grid of V_ξ. Within V_ξ, we define $X = \{(X_i)\}$ a sequence of stationary point processes from a space of functions that are right continuous and have left limit (*rcll*) endowed with Skorohod

embedding for deriving criteria for a sum of independent random processes (Last, Penrose, Schulte, & Thaele, 2014; Kallenberg, 2002).

Another method of creating the domain area based on usual distance is to construct the ball such that the diameter of D is defined as the distance between the farthest two points in D, i.e. $diam(D) = max\{\rho(x, y),\ x, y \in D\}$. Then the Hausdorff outer measure of dimension d bounded by δ, can be written as

$$H_\delta^d(D) = inf\left\{\sum_{i=1}^{\infty}(diam\ U_i)^d : \cup_{i=1}^{\infty} U_i \supseteq D,\ diam\ U_i < \delta\right\}$$

$$\lim_{\delta \to 0} H_\delta^d(D) = H^d(D) = \text{Hausdorff outer measure of dim } d.$$

We will use the Voronoi construction described in Equation (32) and further generalize the circles of maximum radius to the case of maximum coverage utilizing the entire Voronoi cell.

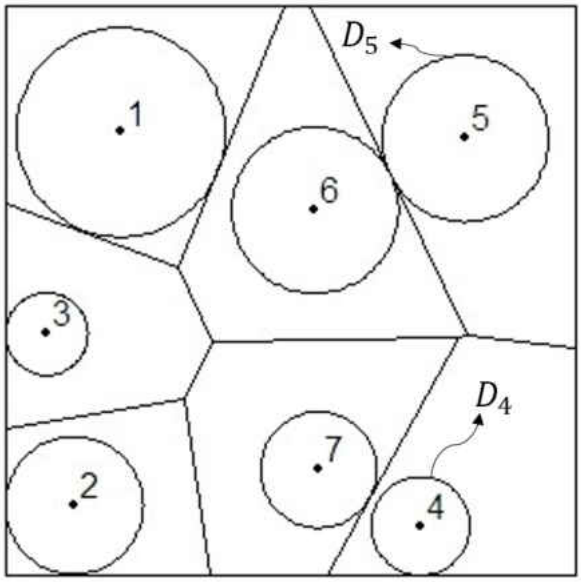


Figure 3: Example of Voronoi cells and largest circle contained within each convex polygon.

Let $\xi = (\xi_1, \xi_2)$ be a random measure in a Borel space $(S^2, \mathcal{S}^2)$, where ξ_1 and ξ_2 represent the (x, y) coordinates for the location and $\mathcal{S}^2$ is the set $\{A \times B\}$. We consider the conditional distribution of the occurrence of events within the partitioned space as a Palm distribution

defined as follows from Equation (14):

$$\mathbb{P}(s \in \cdot \mid \xi) = \frac{\mathbf{1}_{C_1 \times C_2} \cdot \xi}{\xi(C_1 \times C_2)} \tag{33}$$

And the balls (circles) $D_k = C_k^1 \times C_k^2$ depicted in Figure 3 are defined as the partition of the space S^2 of the convex compact balls such that

$$P_s(D_k) = \sup\{r > 0 : B(s_k, r_k) \subseteq D_k,\ D_k \cap D_{k'} = \emptyset, \forall k' \neq k\}.$$

$B(s, r)$ denotes the closed ball with regard to the Euclidean space centered at s, radius r, with

$$\mathbb{E}\Big[f(\xi, s) \mid \xi D_k; s\Big] = \int_D f(\mu, s) q(s, \xi D_k, d\mu), \tag{34}$$

where $q(s, \cdot, \cdot) = \mathbb{P}_s[\,\cdot \mid \mu D_k]$ is the Palm distribution at the point s, and for any non-negative measurable function f.

Let $M \subseteq D$, $M \in \mathcal{S}^2$, then

$$\begin{aligned} P\big((\xi - \delta_s) \in M \mid \xi S < \infty\big) &= \mathbb{P}_s\big[\mu - \delta_s \in M \mid \mu S < \infty\big] \\ P\big((\xi(s); \xi - \xi(s)\delta_s \in \cdot \mid \xi S < \infty, s \in M\big) &= \mathbb{P}_s\big[\mu(s); \mu - \mu(s)\delta_s \cdot \mid \mu S < \infty\big]. \end{aligned} \tag{35}$$

We assume that the marks are generated under a Poisson process and the distribution within the balls follows a Palm distribution. We assume independence almost surely on the set $\{\xi D > 0\}$ for $s \neq s'$ and that the marks themselves are independent and the marks are independent of the points generated, i.e.

$$s \perp\!\!\!\perp \big(\xi(s), \xi - \xi(s)\delta_s\big) = \ \mathbb{P}_s\big[\mu(s); \mu - \mu(s)\delta_s \in \cdot \mid \mu S < \infty\big]. \tag{36}$$

This can be thought of as translation and rotational invariance of the distribution. Then it follows that for a given $s \in D$ with $\xi D > \infty$, δ_s is degenerate and

$$\mathbb{E}(\xi D) = \int_D \mathbb{P}_s(\mu, \cdot, \cdot)\mu(ds) = \ \mathbb{P}_{s'}(\cdot, \xi C, d\mu)\mu(B) \ \text{ with } B \subseteq C. \tag{37}$$

Therefore, $\mathbb{E}(\xi D) = \mathbb{E}(\mathbb{P}_s) \propto \mathbb{P}_{s'}$ with $s' \in D$, i.e. the mean of the Palm distribution is a probability measure that does not depend on the choice of $s \in D$ and the expected value of the Palm distribution is proportional to a Palm distribution at any other point within that convex compact ball D.

Let ξ be a simple point process, i.e. no two points of the process are coincident, whose intensity function exists. Then the statistics in this paper are defined by the Palm distribution satisfying the Campbell-Mecke formula as defined in Baddeley (2007) which states, for any measurable function $f \geq 0$,

$$\mathbb{E}\left[\sum_{s \in D} f(s, D_k)\right] = \int_D \mathbb{E}\big[f(s, D_k)\nu(ds)\big]. \tag{38}$$

Poisson processes, Markov processes, and LGCPs are special cases of point process that share the property that their Palm distributions of any order are again the respective process. Table 3 shows the characteristics for Poisson, Markov, and Cox processes (Coeurjolly et al., 2016).

Table 3: Point process characteristics for Poisson, Markov, and Cox processes when state space S is bounded

Characteristic	Poisson	Markov	Cox
Density $f(x)$	$z_S \prod_{v \in x} \rho(v)$	$\propto \exp\left\{-\sum_{\varnothing \neq y \subseteq x} \Phi(y)\right\}$	$\mathbb{E} f(x\|\Lambda)$
Papangelou conditional intensity $\lambda(u, x)$	$\rho(u)$	$\exp\left\{-\sum_{y \subseteq x \cup \{u\} : u \in y} \Phi(y)\right\}$	$\dfrac{\mathbb{E} f(x \cup \{u\}\|\Lambda)}{\mathbb{E} f(x\|\Lambda)}$
Joint intensity $\rho^{(n)}(x_1, \ldots, x_n)$	$\prod_{i=1}^n \rho(x_i)$	$\mathbb{E} f(\{x_1, \ldots, x_n\} \cup Z)$	$\mathbb{E} \prod_{v \in x} \Lambda(v)$
One-point Palm density $f_u(x)$	$z_S \prod_{v \in x} \rho(v)$	$\propto \exp\left\{-\sum_{\varnothing \neq y \subseteq x \cup \{u\}} \Phi(y)\right\}$	$\mathbb{E} f(x\|\Lambda_u)$
One-point Palm intensity $\rho_v(u)$	$\rho(u)$	$\dfrac{\mathbb{E} f(\{u, v\} \cup Z)}{\mathbb{E} f(\{v\} \cup Z)}$	$\dfrac{\mathbb{E}\{\Lambda(u)\Lambda(v)\}}{\mathbb{E}\Lambda(v)}$

NOTE: For the Poisson process, $\rho(\cdot)$ denotes the intensity function and $z_S = \exp(|S| - \int_S \rho(v)dv)$ the normalizing constant. For the Markov process, Φ denotes the potential function, Z is the unit rate Poisson process on S and the normalizing constants of the density and one-point Palm density and the expectations for the n-th order joint intensity and the one-point Palm intensity are in general intractable. For the Cox process, Λ denotes the random intensity function and $f(\cdot|\Lambda)$ is a Poisson process density.

2.3 PROPERTIES OF POINT PROCESSES

Under certain mathematical conditions for both the original point process and the random displacement, it has been shown via limit theorems that if the points of a point process are repeatedly displaced in a random and independent manner, then the finite distribution of the point process will converge (weakly) to that of a Poisson point process (see Daley & Vere-Jones, 2008 for proof).

Theorem 2.10. Convergence of Superpositions (Kallenberg, 2002) *For (ξ_j) a null array of point processes on a lcscH space S, consider Poisson process ξ with $\mathbb{E}\xi = \mu$. Then $\sum_j \xi_j \xrightarrow{d} \xi$ if and only if (iff) the following conditions hold:*

1. $\sum_j P(\{\xi_j D > 0\} \to \mu D$ *for all* $D \in \mathcal{S}$;
2. $\sum_j P\{\xi_j D > 1\} \to 0$ *for all* $D \in \mathcal{S}$;

The Poisson point process has some convenient properties to investigate algorithms and assumptions. Among these properties, we will consider the isotropic property, mapping property, and marking property.

2.3.1 ISOTROPIC PROPERTY

A point process is said to be isotropic if its statistical properties are unaffected by rotating the point process. In other words, we can analyze the point process from any orientation and the statistical properties would not change. The homogeneous Poisson process is stationary (translation invariant) and isotropic, whereas the nonhomogeneous Poisson process is isotropic and non-stationary (Baddeley et al., 2016).

2.3.2 MAPPING PROPERTY

Given two spaces S and T and a Poisson point process with intensity μ, $N = \sum_k \delta_{X_k}$, $X_k \in S$. Consider f a function $f : S \to T$. Let $Y_k = f(X_k)$, $M = \sum_k \delta_{Y_k}$. Then the new point process M is (again) a Poisson point process in T with intensity $\nu := \mu \circ f^{-1}$, where $\nu(D) = \mu(f^{-1}(D)) = \mu\{s \in S : f(s) \in D\}$, $D \subset T$ (Kingman 1993).

Resnick (2002) provides a more formal definition of the mapping property Suppose $N = \sum_k \delta_{X_k}$ is a Poisson process with intensity μ and state space S. Suppose f is some transformation with domain S and range T, i.e. $f : S \to T$. Now, f^{-1} defines a set mapping

of subsets of T defined as $f^{-1}(D) = \{s \in S : f(e) \in D\}$ for $D' \subset T$. Given the measures N and μ defined on S, we may use f to define induced measures N' and μ' on subsets of T such that

$$\begin{aligned} N'(D) &= N(f^{-1}(D)) = \sum_k \delta_{X_k}(f^{-1}(D)) \\ \text{and} \quad \mu'(D) &= \mu(f^{-1}(D)) \\ \text{then} \quad N'(D) &= \sum_k \varepsilon_{X_k}(f^{-1}(D)) = \sum_k \mathbf{1}_{\{X_k \in f^{-1}(D)\}} \\ &= \sum_k \mathbf{1}_{\{f(X_k) \in D\}} \\ &= \sum_k \delta_{f(X_k)}(D). \end{aligned}$$

2.3.3 MARKING PROPERTY

Recall the notion of a basic point process ξ in a space S and a collection of random points $x_1, x_2, \ldots$. For every point X_n in the Poisson process with intensity λ, we introduce an independent *mark* T_n with a distribution μ. Then the pairs (X_n, T_n) form a marked Poisson process M in the product space $S \times M$ with intensity $\lambda \times \mu$.

$$\begin{aligned} \mathbb{E}N(A) &= \lambda(A), \\ \mu(B) &= P[T_N \in B], \\ \text{and } (\lambda \times \mu) &= \lambda(A) \cdot \mu(B) \end{aligned}$$

A more formal definition of the marking property is given by Resnick (2002). Suppose $\{X_n\}$ are random elements of an Euclidean space S_1 such that $\sum_n \delta_{X_n}$ is a Poisson process with intensity λ. Suppose $\{T_n\}$ are *i.i.d* random elements of a second Euclidean space S_2 with common probability distribution μ. And suppose the Poisson process and the sequence $\{T_n\}$ are defined on the same probability space and are independent. Then $\sum_n \delta_{(X_n, T_n)}$ on $S_1 \times S_2$ is a Poisson process with intensity measure $\lambda \times \mu$ meaning

$$\begin{aligned} (\lambda \times \mu)(A_1 \times A_2) &= (\lambda \times \mu)(\{(s_1, s_2) : s_1 \in A_1 \subset S_1, s_2 \in A_2 \subset S_2\}) \\ &= \lambda(A_1) \cdot \mu(A_2). \end{aligned}$$

Often this procedure is described by saying that we give to point X_n the mark T_n.

(39)

Note:

$$\begin{aligned}\mathbb{E}\sum_n \delta_{(X_n,T_n)}(A_1 \times A_2) &= \sum_n P[(X_n, T_n) \in A_1 \times A_2] \\ &= \sum_n P[X_n \in A_1]P[T_n \in A_2] \end{aligned} \tag{40}$$

because $\{X_n\}$ and $\{T_n\}$ are independent. And since $\{T_n\}$ are iid, Equation (39) can be written as:

$$\begin{aligned}\mathbb{E}\sum_n \delta_{(X_n,T_n)}(A_1 \times A_2) &= \sum_n P[X_n \in A_1]P[T_1 \in A_2] \\ &= \mathbb{E}\left(\sum_n \delta_{X_n}(A_1)\right) P[T_1 \in A_2] \\ &= \lambda(A_1)P[T_1 \in A_2].\end{aligned}$$

This is a way to generalize the spatial point process by combining a stochastic model for attributes, time in this case (marked process), along with the stochastic model for event locations.

Given the measurable space $(S, \mathcal{S}, \mu)$ and the simple point process ξ in S, the marked point process ζ in $S \times M$ is the random sequence $\{(t)\} = \{t_i = (s_k, m_k)\}$, $i \in \mathcal{M} = \{1, \ldots, n\}$ such that s_k are a point process in S and m_k are the marks corresponding to each s_k. Assume that the mark space has a finite reference measure ν on the Borel σ-algebra $\mathcal{B}(\mathcal{M})$. In the special case that $\mathcal{M}$ is finite, ζ can be seen as a multivariate point process $(\zeta_1, \ldots, \zeta_a)$, where ζ_i contains the points marked $i \in \mathcal{M} = \{1, \ldots, a\}$. Then the intensity measure Λ of ζ is defined on the product sets $U = D_k \times T \in \mathcal{B}(\mathcal{S} \times \mathcal{M})$ by

$$\Lambda(U) = \mathbb{E}\zeta(U) = \mathbb{E}N_\zeta(D_k \times T),$$

the expected number of points in D_k with marks in $T \subset M$ (Cronie & van Lieshout 2016).

For the marked point processes ζ, Campbell's formula in Equation (16) takes the form

$$\mathbb{E}\left[\sum_{(s,m)\in\zeta} f(s,m)\right] = \int_{S\times M}\left[f(s,m)\nu(ds,dm)\right] = \int_{\mathcal{S}}\int_{\mathcal{M}} \mathbb{E}f(s,m)\nu(ds,dm), \tag{41}$$

where $f : S \times M \to \mathbb{R}_+$ is a measurable function (Cronie & van Leishout, 2016).

When ζ is stationary, its distributions are invariant under translations of the location, and the reference measure on $\mathcal{M}$ is the mark distribution ν_M, the Palm measures with respect to arbitrary mark sets can be defined as

$$\mathbb{P}_T = \frac{1}{\nu(T)}\int_T \mathbb{P}_{(s,m)}d\nu(m), \quad k = 1,\ldots,n_i, n_i \geq 2, \tag{42}$$

where $T \in \mathcal{B}(\mathcal{M})$, such that $\nu(T) = \nu_M(T) > 0$. This follows from Equations (22) and (23). Then $\mathbb{P}_T$ does not depend on the choice of $s \in D_k \subset S$ and defines a probability measure. The Palm measure can be interpreted as the distribution of ζ given that ζ places a point at s with mark in T (see Cronie & van Leishout, 2016). The properties of invariance and exchangeability introduced in Section 2.2 also hold for the Palm measure defined here.

CHAPTER 3

SPATIO-TEMPORAL MODELS

In this chapter, characterization of the spatial problem is broadened into time dependence. In fact, in many areas such as geostatistics, the modeling description must include a temporal component. A usual method of introducing spatio-temporal models is then through the use of point processes. We present a spatio-temporal Palm distribution process in the context of Poisson, Cox, and Markov processes. To extend the nomenclature from D_s, we add time such that the point process is defined on $D_s \times [0, T]$. We describe properties of the model such as existence and identifiability.

3.1 POINT PROCESSES

For a general point process, consider the number of occurrences at consecutive time periods $t_1 < t_2 < \cdots < t_r < \cdots < t_m$ for some fixed $m \in \mathbb{N}$ in state space D defined as process X^t and the location of event occurrence $\underset{\sim}{X}^t$ to be increasing with time. The times are associated with locations which are sequences of nested events $k(t_1), k(t_2), \ldots, k(t_m)$, where $1 \leq k(t_1) \leq n(t_1)$; $1 \leq k(t_2) \leq n(t_2); \ldots; 1 \leq k(t_r) \leq n(t_r); \ldots; 1 \leq k(t_m) \leq n(t_m)$, $k(t_r) = 1, 2, \ldots, n(t_r)$, and $n(t_r)$ represents the total number of points generated at time t_r. This is an extension of Equation (1) by adding time.

The process follows a discrete time Markov chain as defined in Resnick (2002) if s_i^t for $i = 1, 2, \ldots$ are locations at consecutive times t with $t_1 < t_2 < \cdots < t_{r-2} < t_{r-1} < t_r$:

$$P\Big(\underset{\sim}{X}^{t_r} = s_j^{t_r} | \underset{\sim}{X}^{t_1} = s_1^{t_1}, \ldots, \underset{\sim}{X}^{t_{r-2}} = \ s_i^{t_{r-2}}, \underset{\sim}{X}^{t_{r-1}} = s_i^{t_{r-1}}\Big) = P\Big(\underset{\sim}{X}^{t_r} = s_j^{t_r} | \underset{\sim}{X}^{t_{r-1}} = s_i^{t_{r-1}}\Big), \tag{43}$$

where the holding time of the process X^t in state s_i^t is exponentially distributed with parameter $\lambda(s_i^t)$ some function of propagation from location s_i^t and subarea containing points/locations s_i^t and s_j^t in time interval $[t_{r-1}, t_r)$ and denote $\Delta_r t = t_r - t_{r-1}$. The sequence $\{X^{t_r}\}$, also denoted $\{X(t_r)\}$, follows the Markov property; i.e. given the current state of the system, we can make predictions about the future state without regard for previous states.

This is a discrete finite Markov chain. To simplify notation, we will drop the time from the superscript: $s_i^t \equiv s_i$.

The conditional probability

$$P\left(X^{t_r} = s_j | X^{t_{r-1}} = s_i\right) =: P_{ij}(\Delta_r),\ s_i, s_j \in S,\ t_{r-1}, t_r \geq 0$$

is the transition probability from location s_i to location s_j in interval time Δ_r. This means the process $\{X^{t_r}, r \geq 0\}$ of occurrences at times t_r, has stationary transition probabilities.

$$P_{ij}(\Delta_r) = \begin{cases} p & \text{for } s_j \text{ propogation of } s_i \\ & \text{in successive times } t_{r-1} \text{ and } t_r, \\ 0 & \text{otherwise.} \end{cases}$$

With this transition matrix is associated the probabilities $Q = (Q_{ij})$ constructed from the transition probabilities, $P_{ij}(\Delta_r t)$, where Q_{ij} is the (s_i, s_j) entry p_{ij} from consecutive time periods $t_1, t_2, \ldots, t_T$ adjusted with subarea containing locations i and j, and $Q_{ii} = 0,\ \forall\ s_i \in S$, at each time period.

Following Resnick (2002), we can define on subarea $D_{k(t)}^t$ and each fixed interval time $[t_{r-1}, t_r)$, a sequence $\{E_{k(t_r)}^{t_r}\}_{k \geq 0}$ of *i.i.d* exponential random variables such that $E_{k(t_r)}^{t_r}$ is independent of X^t on $D_{k(t)}^t,\ 1 \leq k(t) \leq n(t_r)$ and set $X_{k(t)}^t$ equal to the count within $D_{k(t)}^t$ for interval time $[t_{r-1}, t_r)$. The exponential distribution is chosen as it describes continuous time Markov chains. Now define $W(D_{k(t)}^t) = \dfrac{E_{k(t)}^t}{\lambda |D_{k(t)}^t|}$, where $|D_{k(t)}^t|$ is the area (Lebesgue) measure of $D_{k(t)}^t$ and λ being the Poisson intensity in $D_{k(t)}^t$. Then

$$P\left(W(D_{k(t)}^t) > u | \underset{\sim}{X}_{k(t)}^{t-1} = s_i \in D_{k(t)}^{t-1}\right) = e^{-\lambda |D_{k(t)}^t| u}, \tag{44}$$

for $u > 0,\ 1 \leq k(t) \leq n(t)$, and at location i. To simplify notation, we write X^{t_r} as X^t and $D_{k(t_r)}^{t_r}$ as $D_{k(t)}^t$.

By discretizing time, we then define two finite sequences $\{(X^t), (t)\}$, where $\mu_t = \sum_{k(t)=1}^{n(t)} X_{k(t)}^t$ with locations $\underset{\sim}{X}_{k(t)}^t$ for $t_r < t < t_{r+1}$ and $\{t_r\}$ is the sequence of times when the process is observed.

3.1.1 KERNEL DENSITY

The temporal extension of the traditional two dimensional kernel density can be expressed using a spatial dimension s that consists of coordinate pairs $s = s(x, y)$ and a temporal dimension $t_{r-1} \leq t < t_r$ as in Subsection 2.1.1. This space-time kernel density can be used to identify spatio-temporal patterns in data sets, which is especially useful when applied to the spread of infectious disease (Saule, Panchananam, Hohl, Tang, & Delmelle, 2017). Besides time dependence, the kernel parameters may be spatially dependent. This is along the same concept as a redistributed time-space kernel presented in Cressie and Wikle (2011). The space-time kernel density can be leveraged as the product of two kernels, the spatial component $K_s(x, y)$, and the temporal component $K_t(t)$ in the following form:

$$\hat{f}(x, y, t) = \frac{1}{nh_s^2 h_t} \sum_i \mathbf{1}(s_i < h_s, t < h_t) K_s\left(\frac{x - x_i}{h_s}, \frac{y - y_i}{h_s}\right) K_t\left(\frac{t - t_i}{h_t}\right), \tag{45}$$

where $(x_i; y_i; t_i)$ is the position and time of events in the vicinity, and can be simplified as:

$$\hat{f}(s, t) = \frac{1}{nh_s^2 h_t} \sum_i \mathbf{1}(s_i < h_s, t_i < h_t) K_s\left(\frac{s - s_i}{h_s}\right) K_t\left(\frac{t - t_i}{h_t}\right), \tag{46}$$

where h_s and h_t are the spatial and temporal bandwidth, respectively.

Consider the point process $\xi = \sum_i \delta_{s_i}$ on some space D. ξ can be extended for a suitable measure in the class of $\mathcal{M}_D$ of measures on D, where the event occurrences are captured alongside their local intensities which could be the number of events over a unit area centered at the point of some other covariates. We consider the probability kernel ν from D to $\mathcal{M}_D$. Let $s_1^t, s_2^t, \ldots$ be the locations of occurrences of events at time t. Then define

$$\xi_t = \sum_i \delta_{s_i^t},$$

and $\xi = (\xi_t, t \geq 0)$, defines a spatial branching/birth process, set $\nu_s = \xi_{t+s} - \xi_t$ and serves as a temporal version of an infinitely divisible process describe in Definition 2.1.3. Let

$$P_{\mu_t}(\xi_t \in \cdot) := \text{ distribution of } \xi_t, \text{ with initial measure } \mu_t.$$

and expectation $\mathbb{E}\mu_t$. Then the Laplace functional

$$\mathbb{E}_{\mu_t} e^{-\xi_t f}, \forall f \in \mathbb{R}_+(X), \text{ i.e. } f : \mathbb{R} \to \mathbb{R}_+$$

satisfies

$$\mathbb{E}_{\mu_t}\left[e^{-\xi_t f}|\xi_s\right] = e^{(-\xi_s \circ \nu_{t-s})f},$$

where $(\nu_t(x), t \geq 0, x \in \mathbb{R}^d)$ is a probability kernel from D to $\mathcal{M}_D$ (see He, 2013).

3.1.2 POISSON PROCESS

Under a Poisson process, the number of occurrences in time interval $[t, t+\tau)$ follows a Poisson distribution with associated intensity parameter $\lambda\tau$ defined as

$$P\left(X^{t+\tau} - X^t = n\right) = e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!} \tag{47}$$

for $n = 0, 1, 2, \ldots$ where $X^{t+\tau} - X^t$ is the number of events in time $[t, t+\tau)$. This is an extension of the spatial model in Equation (5). Extension of the Laplace transform in Equation (6) to times leads to the Laplace transform of the spatio-temporal Poisson process:

$$\mathcal{L}_\mu(f) = \exp\left[-\int\left(1 - e^{-f(x)}\right)\mu(dx)\right], \ \forall f \in \mathbb{R}_+. \tag{48}$$

If λ is not a function of time, the Poisson process X^t is a homogeneous Poisson process. If λ in Equation (47) is such that $\lambda = \lambda(t)$, i.e. under a general rate function say $\lambda(a, b) = \int_a^b \lambda(t)dt$ or $\lambda(t)$, the process $\{X^t\}$ defines a nonhomogeneous Poisson process giving rise to

$$P\left(X^{t+\tau} - X^t = n\right) = e^{-\lambda(\tau)}\frac{(\lambda(\tau))^n}{n!}. \tag{49}$$

The sequence X^t is the total count of occurrence within interval time $[t_{r-1}, t_r)$ in a subset of space D such that $t_r - t_{r-1}$ is conditionally independent and exponential given $X^{t_{r-1}}$. More precisely, if we capture the associated spatio-temporal locations $D^t_{k(t)}$, $t = t_0, t_1, \ldots, t_m$ and $1 \leq k(t) \leq n(t)$, the sequence $\{X^t_{k(t)}\}$, is defined from $n(t_r)$ the number of subareas generated within $[t_{r-1}, t_r)$ based on X^{t_r} and $W(D^t_{k(t)}) = \dfrac{E^t_{k(t)}}{\lambda|D^t_{k(t)}|}$, where $E^t_{k(t)}$ is a sequence of *iid* exponentially distributed random variables with unit mean.

We will explore two methods of capturing the sequence X^t. First, we focus on the propagation at each time step for the entire region of interest D, then we focus on the propagation within a subarea over time.

When we focus on the propagation at each time step, the sequence $\{X^t\}$ is the cumulative count of occurrence within interval time $[t_{r-1}, t_r)$ in a subset of space D such that $t_r - t_{r-1}$ is conditionally independent and exponential given $X^{t_{r-1}}$. More precisely, the sequence $\{X^t_{k(t)}\}$ is defined as follows:

$$
\begin{aligned}
X_1^{t_1} =& X^{t_0}\mathbf{1}_{D_1^{t_1}} + W(D_1^{t_1}),\\
X_2^{t_1} =& X^{t_0}\mathbf{1}_{D_2^{t_1}} + W(D_2^{t_1}),\\
&\vdots\\
X_{k(t_1)}^{t_1} =& X^{t_0}\mathbf{1}_{D_{k(t_1)}^{t_1}} + W(D_{k(t_1)}^{t_1}),\\
&\vdots\\
X_{n(t_1)}^{t_1} =& X^{t_0}\mathbf{1}_{D_{n(t_1)}^{t_1}} + W(D_{n(t_1)}^{t_1}),\\
\text{and } \mu_{\mathbf{t_1}} =& \sum_{\mathbf{k(t_1)=1}}^{\mathbf{n(t_1)}} \mathbf{X_{k(t_1)}^{t_1}};\\
&\vdots\\
X_1^{t_r} =& X^{t_{r-1}}\mathbf{1}_{D_1^{t_r}} + W(D_1^{t_r}),\\
&\vdots\\
X_{n(t_r)}^{t_r} =& X^{t_{r-1}}\mathbf{1}_{D_{n(t_r)}^{t_r}} + W(D_{n(t_r)}^{t_r}),\\
\text{and } \mu_{\mathbf{t_r}} =& \sum_{\mathbf{k(t_r)=1}}^{\mathbf{n(t_r)}} \mathbf{X_{k(t_r)}^{t_r}} \ ;\\
&\vdots\\
X_1^{t_m} =& X^{t_{m-1}}\mathbf{1}_{D_1^{t_r}} + W(D_1^{t_m}),\\
&\vdots\\
X_{n(t_m)}^{t_m} =& X^{t_{m-1}}\mathbf{1}_{D_{n(t_m)}^{t_m}} + W(D_{n(t_m)}^{t_m}),\\
\text{and } \mu_{\mathbf{t_m}} =& \sum_{\mathbf{k(t_m)=1}}^{\mathbf{n(t_m)}} \mathbf{X_{k(t_m)}^{t_r}},
\end{aligned}
\tag{50}
$$

where $n(t_r)$ is the total number of points generated at time t_r.

Figure 4 shows the nested sequence of infinitely divisible propogation studied in the context of arbitrary space-time dependent coefficients on bounded domains as described in Equation (50). The cumulative count is captured by the propagation at the next time step, where $E^t_{k(t)}$ equals the count at time t for area $D^t_{k(t)}$, with $1 \leq k(t) \leq n(t)$, and $n(t)$ represents the number of points generated at time t. The relationship between $E^t_{k(t)}$ and μ_t is also shown.

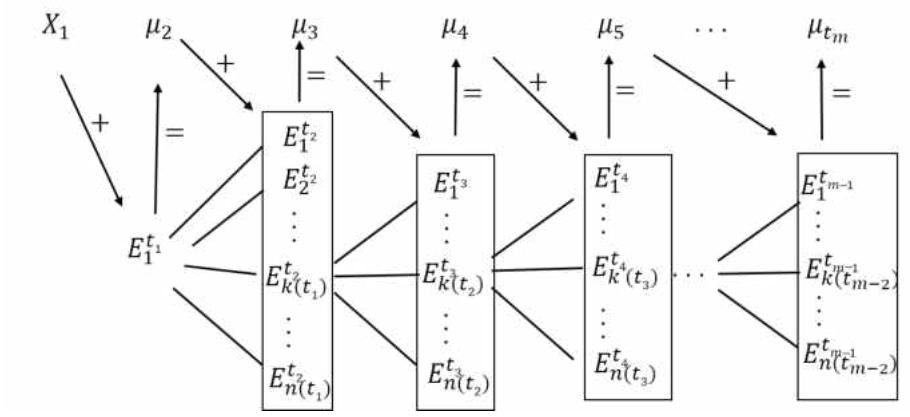


Figure 4: Markov property.

Focusing on the propagation within nested subareas over time, X^t the total count for $t_{r-1} \leq t < t_r,\ 1 \leq r \leq m$ in partitioned area $D = \perp\!\!\!\perp D^{t_r}_{k(t_r)}$ is shown below.

$$\begin{aligned} X^{t_r}_1 =& X^{t_{r-1}} \mathbf{1}_{D^t_1} + W(D^t_1), \\ X^{t_r}_2 =& X^{t_{r-1}} \mathbf{1}_{D^{t_r}_2} + W(D^{t_r}_2), \\ &\vdots \\ X^t_{n(t)} =& X^{t_{r-1}} \mathbf{1}_{D^{t_r}_{n(t)}} + W(D^{t_r}_{n(t)}). \end{aligned} \tag{51}$$

Setting $t_\infty = \lim_{r\to\infty} t_r$ the process X^t becomes explosive. However,

$$\mu_{t_r} = \sum_{k(t)=0}^{n(t)} X^t_{k(t)} \mathbf{1}_{[t_{r-1},t_r)} = \mu_{t_{r-1}} + \sum_{k(t_r)=0}^{\infty} W(D^{t_r}_{k(t_r)}),\ 1 \leq r \leq m.$$

The pair sequence $\{X^t, t_r\}$ has the following two properties (see Resnick, 2002):

1. $\{t_r - t_{r-1} = \Delta_r, 1 \leq r \leq m\}$ are conditionally independent and exponentially distributed given $\{X_n\}$. For $u > 0$,

$$\begin{aligned}
&P(\mu_{t_r} - \mu_{t_{r-1}} > u,\ 1 \leq r \leq m | \underset{\sim}{X}_0^t = s_0, \ldots, \underset{\sim}{X}^{t_{r-1}} = s_i) \\
&\quad = P\left(\sum_{k(t_r)=1}^{t_r} W(D_{k(t_r)}^{t_r}) > u | \underset{\sim}{X}^{t_{r-1}} = s_i \right) \\
&\quad = \prod_{k(t_r)=1}^{n(t_r)} P\left(\frac{E_{k(t_r)}^{t_r}}{\lambda |D_{k(t_r)}^{t_r}|} > u_{k(t_r)} \right), \text{ for } u_1 + \cdots + u_{n(t_r)} > u, \\
&\quad = \prod_{k(t_r)=1}^{n(t_r)} e^{-\lambda |D_{k(t_r)}^{t_r}| u_{k(t_r)}} \\
&\quad = \exp\left\{ - \sum_{k(t_r)=1}^{t_r} \lambda |D_{k(t_r)}^{t_r}| \cdot u_{k(t_r)} \right\}, \text{ for } k(t_r) = 1, 2, \ldots, n(t_r). \qquad (52)
\end{aligned}$$

2. The distributional structure of $\{\mu_{t_r}, (t_r)\}$ is a function of the transition matrix from states s_i to s_j and the limiting function. For $u > 0,\ s_i, s_j \in S$,

$$\begin{aligned}
&P\Big(X_{k(t_r)}^{t_r} = s_j;\ \mu_{t_r} - \mu_{t_{r-1}} > u \mid \underset{\sim}{X}^{t_{r-1}} = s_i \Big) \\
&\qquad = Q^t \cdot \exp\left\{ - \sum_{k(t)=1}^{n(t_r)} \lambda \big| D_{k(t)}^t \big| u \right\}, \quad t \in [t_{r-1}, t_r), \qquad (53)
\end{aligned}$$

where Q is the (s_i, s_j) transition probability of count from locations s_i to s_j between times t_{r-1} sand t_r. When s_j is not a propagation of s_i, Equation (53) is equal to 0.

The special case of transition probabilities following a Poisson process and where the $\lambda(D_{k(t)}^t)$ are proportional to time plays an important role as it naturally occurs in many statistical phenomena. For

$$X^{t_r} = X^{t_{r-1}} + \sum_{k=1}^{n(t_r)} W(D_k^{t_r}),$$

suppose$X^1, X^2, \ldots$ is a sequence of random variables and assume

$$\eta_{t_r} := \sum_{k=1}^{n(t_r)} W(D_k^{t_r})$$

are *iid*, then η_{t_r} defines (staying with discs) a sequence of balls

$$B(s_i^{t_r}) : \{s_i \in \mathbb{R}^2, \|s\| = \|y_i - x_i\| < r\},$$

such that $\cup_{i=1}^{\infty} B(s_i^{t_r}, r)$ defines a partition of S on the set of nonzero measures μ. Last and Penrose (2017) called these closed balls grains, and this has been defined in stochastic geometry as the spherical Boolean model.

3.1.3 COX PROCESSES

A spatio-temporal Cox process is characterized by $\{X^t : t \in \mathbb{Z}\}$ with discrete time $t \in \mathbb{Z}$ and X^t point process. The driving intensity is given by $\eta = \{\eta_t : t \in \mathbb{Z}\}$, where each $\eta_t = \{\lambda(s,t) : s \in \mathbb{R}^2\}$ is a locally integrable non-negative stochastic process. Then $X^t|\eta$ are mutually independent Poisson processes with intensity functions η_t (Møller & Díaz-Avalos, 2010).

Assuming multiplicative decomposition of random intensities,

$$\lambda(s,t) = \lambda_1(s)\lambda_2(t)X(s,t), \quad \mathbb{E}X(s,t) = 1, \quad (s,t) \in \mathbb{R}^2 \times \mathbb{Z}, \tag{54}$$

where $\lambda_1(s)$ and $\lambda_2(t)$ are non-negative deterministic functions and $X(s,t)$ is a spatio-temporal process with unit mean.

When $X(s,t)$ is log-Gaussian, this is referred to as a spatio-temporal log-Gaussian Cox process. The spatio-temporal shot-noise Cox process is formed when

$$X(s,t) = \delta \sum_{\tau=-\infty}^{\infty} \sum_{s' \in \Phi_\tau} \phi(s - s', t - \tau), \tag{55}$$

where Φ_τ is a stationary Poisson point process with some intensity $\lambda > 0$ (not depending on $\tau \in \mathbb{Z}$), ϕ is the joint density on $\mathbb{R}^2 \times \mathbb{Z}$ with regard to the product measure of the Lebesgue measure on $\mathbb{R}^2$ and the counting measure on $\mathbb{Z}$ (Møller & Diaz-Avalos, 2008).

3.1.4 MARKOV PROCESS

We incorporate time into the Markov process described in the previous chapter to develop

a spatio-temporal Markov process (following Baddeley et al., 2016). To incorporate time, we express a Markov process as a hereditary point process in which points at a given time t are only possible as a result of a point generated at time $t-1$. Doing so, we extend a nested process for interaction within $[t_{r-1}, t)$.

$$\mathbb{E}[h(s_t)] = \sum_{k(t)=1}^{n(t)} e^{-|D^t_{k(t)}|} \left[h(s_{t-1}) f(s_{t-1}) + \sum_{k(t)=1, s_t \in D_{k(t)}}^{\infty} \frac{1}{k(t)!} \prod \int h(s_t) f(s_t) ds_t \right], \quad (56)$$

where $\underset{\sim}{s} = ((s_1), (s_2), \ldots, (s_t))$ is a sequence of sequences of lengths $(n(1), n(2), \ldots, n(t))$.

3.2 PALM DISTRIBUTION

Using the theory reviewed in Section 2.2, we propose to extend the Palm distributions over disjoint sets D_k, where k represents the number of subareas, and incorporate time, t_r, partitioned equally into $r = 1, \ldots, m$ periods, each period with its corresponding subdivision partition. The process then depicts a representation of a general decomposition with stationary independent increments (Lévy) process of Brownian components as in Definition 2.1.2. The number of subareas can be fixed from one time to the next or can be dynamic and change over time as determined by the points generated at the previous time point. The subareas will be referred to as $D^t_{k(t)}$, where $k = 1, 2, \ldots, n$ or $k = 1, 2, \ldots, n(t_r)$ for $r = 1, \ldots, m$ for fixed and dynamic subareas, respectively. The spatio-temporal version of Palm distribution described in Equation (14) will be annotated as $\mathbb{P}_{t,s}$ and is defined as:

$$\mathbb{P}_{t,s_{t-1}}(D^t_{k(t)}) = \frac{\mathbb{E}[\xi_t(ds); \gamma_t \in D^t_{k(t)}]}{\mathbb{E}[\xi_t(ds)]}, \quad s \in D^t_{k(t)},\ D^t_{k(t)} \in \mathcal{T}, \quad (57)$$

where ξ_t is a random measure at time t on the measurable space $(S, \mathcal{S})$ and γ_t is a random element at t in a measurable space $(T, \mathcal{T})$.

The Campbell measure presented in Equation (17) can be extended to our construct of disjoint $D^t_{k(t)}$ in $\mathcal{S}$ by following Kallenberg's (2002, 2005) construction:

$$C_{t,s}(f_t \mathbf{1}_{D^t_{k(t)}}) = \mathbb{E} \int_{D^t_{k(t)}} f_t(s, \xi_t) \xi_t(ds) = \mathbb{E}\left[\xi_t D^t_{k(t)} \cdot \int_{D^t_{k(t)}} f_t(s, \xi_t) \frac{\xi_t(ds)}{\xi_t D^t_{k(t)}} \right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\xi_t D^t_{k(t)} \cdot \int_{D^t_{k(t)}} f_t(s,\xi_t) P[s \in ds|\xi_t]\right] = \mathbb{E}\left[\xi_t D^t_{k(t)} \cdot \mathbb{E}[f_t(s,\xi_t)|\xi_t]\right] \\
&= \mathbb{E}\left[\mathbb{E}[\xi_t D^t_{k(t)} \cdot f_t(s,\xi_t)|\xi_t]\right] = \mathbb{E}[\xi_t D^t_{k(t)} \cdot f_t(s,\xi_t)] \\
&= \mathbb{E}\left[\mathbb{E}[\xi_t D^t_{k(t)} \cdot f_t(s,\xi_t)|s]; s \in D^t_{k(t)}\right] \quad \text{and for } s \notin D^t_{k(t)} \text{ when } \xi_t D^t_{k(t)} = 0 \\
&= \int \nu^t_k(ds) \int f_t(s,\mu)\mathbb{P}_s(d\mu), \qquad (58)
\end{aligned}
$$

where ν^t is the supporting measure of ξ_t and $f_t \in (S \times \mathcal{M}(S))_+$, or $C_t = \nu^t_k \otimes \mathbb{P}_t$ for each $D^t_{k(t)}$ sharing the same f_t and ξ_t.

The reduced Campbell measure follows as in Equation (19) and has the form:

$$
C'_{t,s_{t-1}} f_t = \mathbb{E} \int f_t(s, \xi_t - \delta_{s_{t-1}}) \xi_t(ds), \qquad (59)
$$

where for some $s = s^t_{k(t)} \in D^t_{k(t)}$, $k = 1, \ldots, n(t)$, and $t = t_1 < t_2 < \cdots < t_r < \cdots < t_m$. Using disintegration theory as in Equation (21), this can be expressed as:

$$
C'_{t,s_{t-1}} f_t = \int \nu^t_k(ds) \int f_t(s,\mu) \mathbb{Q}_{t,s_{t-1}}(d\mu), \qquad (60)
$$

where $s = s^t_k$ and $\mathbb{Q}_{t,s}$ denotes the reduced Palm measure $\mathbb{P}^!_{t,s} = \mathbb{P}_{t,s} \circ (\mu - \delta_s)^{-1}$ and $f_t \in (S \times \mathcal{N}(S))_+$ or $C'_{t,s} = \nu^t_k \otimes \mathbb{Q}_{t,s}$. Or in general as:

$$
\begin{aligned}
C'_{t,s_{t-1}} &= \int \nu^t_k(ds) \int f_t(\mu) \mathbb{Q}_{t,s_{t-1}}(d\mu) \\
&= \int \left(\int f_t(\mu) \mathbb{Q}_{t,s_{t-1}}(d\mu) \right) \nu^t_k(ds) \\
&= \mathbb{E} \int_{D^t_k} f_t(s, \xi_t - \delta_{s_{t-1}}) \xi_t(ds) \\
&= \int_{D^t_k} \nu^t_k(ds) \int f_t(\mu - \delta_{s_{t-1}}) \mathbb{Q}_{t,s_{t-1}}(d\mu). \qquad (61)
\end{aligned}
$$

The spatio-temporal version of the reduced Palm distribution can be understood as the conditional distribution of the point process $\xi_{t,s}$ by omitting the point at $s = s^{t-1}_k$ given

$\xi_t(s) > 0$. Similar to Equation (57), the reduced Palm can be written as:

$$\begin{aligned}\mathbb{Q}_{t,s_{k(t-1)}}(s^t) &= \frac{\mathbb{E}(\xi_t(ds), \gamma_t \in D_{k(t)})}{\mathbb{E}(\xi_t(ds))} \\ &= \frac{P(N(ds^t) = 1, N(s_{k(t-1)}) = 1)}{P(N(s_{k(t-1)}) = 1)} \\ &= \frac{P(N(ds^t) = 1)}{P(N(s_{k(t-1)} = 1)},\end{aligned} \tag{62}$$

where ξ_t is a random measure at time t on the measurable space $(S, \mathcal{S})$ and γ_t is a random element at t in a measurable space $(T, \mathcal{T})$. This representation is possible since the random element γ_t is assumed to have already existed once we move to the next time interval.

The reduced Palm distribution can also be described by the Palm intensity function or conditional intensity function or second order intensity at time t.

$$\mathbb{Q}_{s_{k(t-1)}}(s)ds = P\{N(ds) = 1 | n(\{s_{k(t-1)}\}) = 1\},$$

where $s \in D_{k(t)}$ and $\mathbb{Q}_{s_k^{t-1}}(s)$ depends only on $D_{k(t)}$ and on ξ_t. As mentioned in Tanaka, Ogata and Stoyan (2008), the best known second order characteristic of stationary and isotropic point processes is the Ripley k-function (Ripley, 1976) transformed into a pair correlation function $g(r)$ and Palm intensity $\lambda_s(r)$ such that $\lambda_s(r) = \lambda g(r)$, where λ is the intensity and can be a function of time.

The regions $D^t_{k(t)}$ are defined by Voronoi cells (or discs/circles) under regular conditions defined as follows:

1. Points that are coincident are the same (i.e. simple point process).
2. Cells are defined such that the areas will represent a partition of D for the points where events occur.
3. Each polygon or Voronoi cell (or disc/circle)as in Figure 3 will be constructed around a center point or "original" point (from the previous time step) with radius/max coverage to include all areas that are closest to that point than any other center point or "original point".

Extending the definition of the reduced Palm measure to our construct of D_k disjoint in

$\mathcal{S}$ and defining the marks as discrete time points yields

$$\mathbb{Q}_{t_r} = \frac{1}{\nu(t_r)} \sum_{k(t_r)=1}^{n(t_r)} \mathbb{Q}_{(s_{k(t_{r-1})})(D^{t_r}_{k(t_r)})}, \tag{63}$$

where s_k is the marked point within the time interval $\left(t_{i-1}, t_{i-1} + \frac{1}{2}(t_i - t_{i-1})\right] = \left(t_{i-1}, \frac{1}{2}(t_i + t_{i-1})\right]$, with disc area generated by $s_{k(t_{r-1})}$. And since $\mathbb{Q}_{t_r}$ does not depend on the choice of s_k, Equation (63) can be further simplified to

$$\mathbb{Q}_{t_r} = \frac{1}{\nu(t_r)} \sum_{k(t_r)=1}^{n(t_r)} \mathbb{Q}_{(D^{t_r}_{k(t_r)})}. \tag{64}$$

3.3 PROPERTIES

3.3.1 MAPPING PROPERTY

A condition to establish interpretable probability of the Palm distributions for point processes is given in the following theorem.

Theorem 3.1. (Daley & Vere-Jones, 2008) *Let (D_t) be a nested sequence of bounded sets from D with non-empty interiors satisfying*

$$diam(D_t) \to 0 \text{ as } t \to \infty,$$

Then

$$\frac{1}{\lambda(D_t)} P\left(N(D_t) > 0\right) \to \mathbb{E}\left(N(D)\right).$$

The aim is to define a family of mappings that links the points of the process over time. Under these mappings, the invariance and exchangeability properties provide characterization of the Palm distribution (Daley & Vere-Jones, 2008).

We establish a thinning for the point process version in space and time similar to that described in subsection 2.1.1.

Theorem 3.2. (Kallenberg, 2002) *For any random process ξ^t in D^t_k,*

(i) ξ_t *is infinitely divisible.*

(ii) $\sum_j \xi^t_{n_j} \xrightarrow{d} \xi^t$ *for any i.i.d. thinning* $(\xi^t_{n_j})$ *of* ξ^t.

(iii) $\xi \stackrel{d}{=} X$ *for some Lévy process* X *in* $\mathbb{R}^d$.

3.3.2 IDENTIFIABILITY

The computational difficulties of our approach are due to the lack of identifiability and hence inferential estimations and stability of any algorithm must be flexible to allow a wide variety of kernel shapes. Indeed, computations can be intractable for space and time processes (see Daniels, Zhou, & Zou, 2006). We will propose a different approach in the inference. Tierney, Kass, and Kadane (1989) illustrated treatment of expected mean and variance in the Bayesian construct with application in Bayesian inference context. The goal here is to identify a system through its design and parameters. If the system is not locally identifiable, the parameters associated with the distribution may not be unique. Dealing with time-space in modeling processes involves manipulation of algebraic functions whose computations are still challenging but accessible under simulations. To estimate the unobserved feature occurrences (state) at time t_{r+1} from the observed points at time t_r, nonlinear model of Laplace Poisson filtering is utilized under time for each subarea of D. Suppose $\{x_{t_r}\}$, $r = 1, 2, \ldots, m$ is a sequence of observations/points (for simplicity, this will be denoted as x_t) directed by Markov chain with predictive distribution $p_D(x_t)$ and let y_t be the expected value generated by $\{x_t\}$, $\mathbb{E}(x_{t-1}) = x_t$, where $x_{t-1} \equiv x_{1:t-1}$. Using Markov property, with initial density $p(x_1)$ and transition $p(y_{t+1}|y_t)$; under Bayes' rule, we obtain

$$p(y_t|x_t) = \frac{p(x_t|y_t)p(y_t|x_{t-1})}{\displaystyle\int p(x_t|y_t)p(y_t|x_{t-1})dx_t}, \tag{65}$$

where

$$p(y_t|x_{t-1}) = \int p(y_t|y_{t-1})p(y_{t-1}|x_{t-1})dx_{t-1}$$

is the predictive distribution. Investigating Equation (65) and its behavior reduces to asymptotic analysis of the Laplace integral.

Let p and $\hat{p}$ denote the true and approximate density of the sequential and nested random variables of point distribution, respectively. Asymptotic approximation allows understanding

of parameters and their precision. The system is locally identifiable if for each subset/subarea D, $\hat{p}(y_t|x_{t-1}) \cong p(y_t|x_{t-1})$. To establish such concept, we adapt symbolic manipulation presented in Wojdylo (2006) and check that for a given sequence at time t_r, the implemented processes generated at time t_{r+1} will be computationally tractable on the basis of stationary processes. The main results are (i) existence and (ii) uniqueness; hence equivalence of the expected values. We introduce some notation. Let $h(y_t)$ be

$$h(y_t) = -\frac{1}{\gamma} \log p(x_t|y_t)p(y_t|x_{t-1}) \tag{66}$$

with the following five regularity conditions (Wojdylo, 2006).

1. $h(y_t)$ is a constant-order function of γ, an expansion parameter that measures concentration of integrand about its peak, as $\gamma \to \infty$, and is five-times differentiable with respect to x_t.

2. $h(y_t)$ has an unique interior minimum, its second derivative is positive.

3. $p(y_{t-1}|y_t)$ is four-times differentiable with respect to y_t.

4. The integral $\int p(y_{t+1}|y_t)e^{-\gamma h(y_t)}dy_t$ exists and is finite, where $p(y_{t+1}|y_t)$ is called the amplitude and $e^{-\gamma h(y_t)}$ is called the phase.

5. Derivatives of $h(y_t)$ up to the fifth order and those $p(y_{t+1}|y_t)$ with respect to y_t up to the third order are bounded uniformly across time.

Then, following Koyama, Prez-Bolde, Shalizi, and Kass (2010), the α-order Laplace Gaussian filtering approximates the predictive distribution as

$$\hat{p}(y_t|x_{t-1}) = p(y_t|x_{t-1}) + O(\gamma^{-\beta}), \tag{67}$$

where $\beta = 1$ for $\alpha = 1$ and $\beta = 2$ for $\alpha \geq 2$. The proof follows inductively from Proposition 2 below since the approximated predictive distribution is $O(\gamma^{-\beta})$ and uniformly bounded by $t \in \mathbb{N}$. For occurrence of events following the Palm distribution as described in Subsection 2.2, we consider the Laplace approximation of the predictive distribution as $\hat{\mathbb{Q}}_{\mathbb{D}}(y_t|x_{t-1}) \cong \mathbb{Q}_D(y_t|x_{t-1})$. To prove this statement, we adjust the results of Koyama et al. (2010) and set the following.

LEMMA **3.3.** *Let*

$$\hat{h}(y_t) = -\frac{1}{\gamma}\log \mathbb{Q}_D(x_t|y_t)\hat{\mathbb{Q}}_{\mathbb{D}}(y_t|x_{t-1}), \tag{68}$$

and $\hat{h}^{(\ell)} \equiv \frac{\partial^{(\ell)}}{\partial y_t}\hat{h}(y_t)$. *Then the* α*-order Laplace approximation of the posterior mean and variance have series expansions as*

$$\tilde{y}_t = \sum_{j=0}^{\alpha-1} A_j(\hat{h}_t^{(\ell)})\gamma^{-j}; \tag{69}$$

and

$$\tilde{\nu}_t = \sum_{j=1}^{\alpha-1} B_j(\hat{h}_t^{(\ell)})\gamma^{-j}; \tag{70}$$

where A_j *and* B_j *are functions of* $\{\hat{h}_t^{(\ell)}\}$. *From Tierney et al. (1989),* $\int b(\theta)\exp^{-nh(\theta)}\,d\theta$ *has Laplace approximation given as*

$$\int b(\theta)\exp^{-nh(\theta)}\,d\theta \cong \sigma\sqrt{2\pi}\exp\left(-n\hat{h}\right)\cdot\left[\hat{b} + \frac{1}{2n}(\sigma^2\hat{b}'' - \sigma^4\hat{b}'\hat{h}''' + \frac{5\sigma^6}{12}\hat{b}(\hat{h}''')^2 - \frac{\sigma^4}{4}\hat{b}\hat{h}^{(4)})\right], \tag{71}$$

where $\hat{b} = b(\theta)$ *and* $\sigma^2 = \left[h'(\hat{\theta})\right]^{-1}$.

Proof. The expectation of a function of $y(t)$, say $g(y_t)$, with respect to the approximated posterior distribution is

$$\hat{\mathbb{E}}(g(y_t)|x_t) = \frac{\int g(y_t)\cdot e^{-\gamma\hat{h}(y_t)}dy_t}{\int e^{-\gamma\hat{h}(y_t)}dy_t},$$

where $g(y_t) = y_t$ for the mean and $g(y_t) = y_t^2$ for the second moment. □

Proposition 2.

$$\mathbb{Q}_D(y_t|x_{t-1}) \cong \mathbb{Q}_D(y_t|x_{t-1}) + \sum_{j=\nu}^{N}\varepsilon_{t,j}(y_t)\gamma^{-j},\ 0 < \nu < N, \tag{72}$$

where $\varepsilon_{t,j}(y_t)$ are constant order functions of γ.

Proof. The predictive distribution at time $t+1$ is

(i) $\mathbb{Q}_D(y_{t+1}|x_t) = \dfrac{\int \hat{\mathbb{Q}}_D(y_{t+1}|y_t)\cdot e^{-\gamma h(y_t)}dy_t}{\int e^{-\gamma h(y_t)}dy_t}$ and approximation under Laplace's

method is expanded as

$$\mathbb{Q}_D(y_{t+1}|x_t) \cong \frac{\sum_{s=0}^{N} \Gamma(s+\frac{1}{2})(\frac{2}{h''})^s c_{2s}^* \gamma^{-s}}{\sum_{s=0}^{N} \Gamma(s+\frac{1}{2})(\frac{2}{h''})^s \overline{c}_{2s}^* \gamma^{-s}},$$

where N is the number of observations or sample size in area D,

$$c_s^* = \sum_{i=0}^{s} \frac{q^{s-i}(y_{t+1})}{(s-1)!} \sum_{j=0}^{i} \binom{-(s+1)/2}{j} C_{i,j}(A_1, \dots),$$

$$\overline{c}_s^* = \binom{-(s+1)/2}{j} C_{s,j}(A_1, \dots),$$

and $C_{s,j}(A_1, \dots)$ is a partial ordinary Bell polynomial. Hence

$$\mathbb{Q}(y_{t+1}|x_t) = \mathbb{Q}(y_{t+1}) + \sum_{j=1}^{N} C_j(x_{t+1})\gamma^{-j} + O(\gamma^{-(N+1)}).$$

(ii) Now consider the approximated predictive distribution of time $t+1$

$$\hat{\mathbb{Q}}_D(y_{t+1}|x_{t+1}) = \int \mathbb{Q}_D(y_{t+1}|y_t)\hat{\mathbb{Q}}_D(y_t|x_t)dy_t, \tag{73}$$

where $\hat{\mathbb{Q}}_D(y_t|x_t)$ is the Palm approximation of kernel distribution whose mean and variance are given by Equations 69 and 70. Equation (73) can be rewritten as

$$\hat{\mathbb{Q}}_D(y_{t+1}|x_t) = \frac{1}{\sqrt{2\pi\tilde{\nu}_t}} \int \mathbb{Q}_D(y_{t+1}|y_t)\mathbb{Q}_D(y_t|x_t)dy_t \tag{74}$$

as in Koyoma et al. (2010). One other option is to choose a Poisson distribution with intensity $\lambda_i(x_t) \cdot \Delta$, where Δ corresponds to the distribution. In terms of previous time, $x_t = F_{x_{t-1}} + \varepsilon_t$, where F represents some Kalman filter.
Apply Laplace's method again

$$\hat{\mathbb{Q}}(y_{t+1}|x_t) \cong \tilde{q}(y_{t+1}) + \sum_{j=1}^{N} \frac{\tilde{q}^{(2j)}(y_{t+1})}{2^j \Gamma(j+1)} \nu^{-j}, \tag{75}$$

where $\Gamma(j+1)$ is the Gamma function and

$$\tilde{q}_t^{(\ell)} \equiv \left. \frac{\partial^\ell \mathbb{Q}(y_{t+1}|y_t)}{\partial y_t} \right|_{y_t=\tilde{y}_t}. \tag{76}$$

Then using substitution

$$\begin{aligned}
\hat{h}(y_t) &= -\frac{1}{\gamma}\log \mathbb{Q}(x_t|y_t) \times \tilde{\mathbb{Q}}(y_t|x_{t-1}) \\
&\cong -\frac{1}{\gamma}\log \mathbb{Q}(x_t|y_t) \times \left[\mathbb{Q}(y_t|x_{t-1}) + \sum_{j=\nu}^{N} \varepsilon_{t,j}(y_t)\gamma^{-j}\right] \\
&\cong -\frac{1}{\gamma}\log \mathbb{Q}(x_t|y_t) \times \mathbb{Q}(y_t|x_{t-1}) \left[1 + \sum_{j=\nu}^{N} \frac{\varepsilon_{t,j}(y_t)\gamma^{-j}}{\mathbb{Q}(y_t|x_{t-1})}\right] \\
&\cong h(y_t) - \frac{1}{\gamma}\left[1 + \sum_{j=\nu}^{N} \frac{\varepsilon_{t,j}(y_t)\gamma^{-j}}{\mathbb{Q}(y_t|x_{t-1})}\right] \\
&\cong h(y_t) - \frac{1}{\gamma}\sum_{j=\nu}^{N} \frac{\varepsilon_{t,j}(y_t)\gamma^{-j}}{\mathbb{Q}(y_t|x_{t-1})} \\
&\cong h(y_t) - \sum_{j=\nu}^{N} \frac{\varepsilon_{t,j}(y_t)\gamma^{-(j+1)}}{\mathbb{Q}(y_t|x_{t-1})} \\
&\cong h(y_t) - \sum_{j=\nu}^{N} \mathcal{F}_{t,j}(y_t)\gamma^{-(j+1)}, \qquad (77)
\end{aligned}$$

where $\mathcal{F}_{t,j}(y_t) = \frac{\varepsilon_{t,j}(y_t)}{\mathbb{Q}(y_t|x_{t-1})}$ is a collection of terms that depend on $\varepsilon_{t,j}(y_t)$.

Suppose $\hat{y}_t = y_t + \epsilon$ and $\epsilon << 1$, as in Tierney et al. (1989) to guarantee the choice $y_t = \hat{y}_t$. Taking the derivative of both sides of Equation (76) and evaluating at y_t yields

$$\epsilon \cong \sum_{j=\nu}^{N} \frac{\mathcal{F}'_{t,j}}{h''_t}\gamma^{-(j+1)}.$$

Then

$$\hat{y}_t = y_t + \epsilon \cong y_t + \sum_{j=\nu}^{N} \frac{\mathcal{F}'_{t,j}}{h''_t}\gamma^{-(j+1)}. \qquad (78)$$

Now we generalize the ℓ^{th} order derivative of $\hat{h}_t$ by inserting Equation (77) into Equation (76), which gives

$$\begin{aligned}
\hat{h}_t^{(\ell)} &\cong h_t^{(\ell)} - \sum_{j=\nu}^{N} \mathcal{F}_{t,j}^{(\ell)}\gamma^{-(j+1)} \\
&\cong h_t^{(\ell)} - \sum_{j=\nu}^{N} \left[\mathcal{F}_{t,j}^{(\ell)} + \frac{\mathcal{F}'_{t,j}}{h''_t}h^{(\ell+1)}\right]\gamma^{-(j+1)}. \qquad (79)
\end{aligned}$$

Substituting Equations (78) and (79) into Equation (69) leads to

$$\tilde{y}_t \cong y_t + \sum_{j=1}^{\alpha-1} A_j \gamma^{-j} + \sum_{j=\nu}^{N} \frac{\mathcal{F'}_{t,j}}{h''_t} \gamma^{-(j+1)}. \tag{80}$$

Inserting Equation (80) into Equation (76) and expanding with respect to γ^{-1},

$$\begin{aligned} \tilde{q}^{(\ell)}(y_{t+1}) \cong & q^{(\ell)}(y_{t+1}) + \sum_{j=1}^{\alpha-1} A_j q^{(\ell+1)}(y_{t+1}) \gamma^{-j} \\ & + \sum_{j=2}^{\alpha} \left[\sum_{k=2}^{j} \frac{1}{k!} q^{(\ell+k)} \mathcal{C}_{j,k}(A_1, \dots) \right] \gamma^{-j} \\ & + \sum_{j=\nu}^{N} \frac{\mathcal{F'}_{t,j}}{h''_t} q^{(\ell+1)}(y_{t+1}) \gamma^{-(j+1)}. \end{aligned} \tag{81}$$

Substituting Equations (70), (79), and (81) into Equation (75), the final asymptotic expansion of $\hat{\mathbb{Q}}(y_{t+1}|x_t)$ is obtained.

$$\mathbb{Q}(y_{t+1}|x_t) = q(y_{t+1}) + \sum_{j=1}^{\alpha} R_j(y_{t+1}) \gamma^{-j} + \sum_{j=\nu}^{N-1} \frac{\mathcal{F'}_{t,j}}{h''_t} q'(y_{t+1}) \gamma^{-(j+1)}, \tag{82}$$

where

$$R_j(y_{t+1}) = \begin{cases} G_j(y_{t+1}) + A_j q'(y_{t+1}), & 1 \le j \le \alpha - 1 \\ G_j(y_{t+1}), & j = \alpha \end{cases}$$

and

$$\begin{aligned} G_j(y_{t+1}) = & \sum_{s=2}^{j} \frac{1}{s!} \mathcal{C}_{j,s}(A_1, \dots) q^{(s)}(y_{t+1}) + \sum_{s=1}^{j} \frac{\mathcal{C}_{j,s}(B_1, \dots) q^{(2s)}(y_{t+1})}{2^s \Gamma(s+1)} \\ & + \sum_{s=1}^{j-1} \sum_{k=s}^{j-1} \frac{A_{j-k} \mathcal{C}_{k,s}(B_1, \dots) q^{(2s+1)}(y_{t+1})}{2^s \Gamma(s+1)} \\ & + \sum_{s=1}^{j-2} \sum_{k=s}^{j-2} \sum_{n=2}^{j-k} \frac{\mathcal{C}_{j-k}(A_1, \dots) \mathcal{C}_{k,s}(B_1, \dots) q^{(2s+1)}(y_{t+1})}{2^s \Gamma(s+1) n!}, \end{aligned}$$

where $B_j \equiv B_j(\{h_t^{(\ell)}\})$ as in Lemma 3.3.

□

To obtain the asymptotic approximation of the Palm distribution, we identify features on the neighborhood contribution of the distribution support. In that neighborhood, integrals are put into Laplace form to take advantage of Laplace insights. The predictive distribution is stable and the error does not accumulate over time. If an analytic error bound is available, the parameters will have physical significance and guide extracting essential features of the physics underlying the problem, thereby offering identifiable physical quantities into the problem.

CHAPTER 4

INFERENCE AND STATISTICAL PROPERTIES

In disease mapping studies, distributional models based on generalized linear models are not identifiable (see Besag, York, & Molliè, 1991; Gelfand & Sahu, 1999; Goicoa, Adin, Ugarte, & Hodges, 2018). Some of the issues are related to the exchangeable distribution and the conditional autoregressive component where parameters are determined up to an additive constant and sum to zero constraints are considered without clear guidance. Palm distributions offer conditional distributions of a point process given a previous location (see Daley & Vere-Jones, 2008; Kallenberg, 2002). These distributions play an important role in spatial point processes that are followed over time. For maximum likelihood estimates, Bayesian approach based on integrated nested Laplace approximation (INLA) have be used (see Coeurjolly, Møller, & Waagpeterson, 2017; Gómez-Rubio & Palmí-Perales, 2019; Rue & Martino, 2009). Theoretical properties and estimation have been investigated under the assumption that the underlying area is fixed. We propose to extend this theory to a dynamic spatio-temporal model.

In spatial statistics, procedures and techniques to detect clustering recursively or consecutively over time are much needed. We propose dynamic inference of disease process by reducing spatial variance in a data set. The likelihood function is fundamental to statistical inference, as it takes the data set as given, then expresses the likeliness of different parameters for the distribution of the cluster. Bayesian approximation of the posterior means are hard to estimate, especially for nonlinear or non-Gaussian setup. Monte Carlo methods offer a better, more accurate computational framework with spatial and temporal synergy.

Given the location of offspring points, the goal is to estimate area and associated model parameters. Unfortunately, point process models are often intractable, making it impossible to find a solution from the likelihood function or a closed form of the likelihood function (Diggle, 2014). Monte Carlo methods will be utilized to overcome the challenge of intractability and identify temporal tendencies at selected areas. We also propose areas that do not overlap by applying isotropic properties of point processes and by nesting areas at time t within areas from previous Markov chain time with flexible boundaries. The nested Monte Carlo

estimation method relies on mapping and uniqueness properties of the Markov process. In performing estimation, three components must be determined:

1. the family of density/distribution functions;
2. the intensity parameters; and
3. the transition measure.

4.1 SPATIAL THEORETICAL FRAMEWORK

We first present an overview of spatial point processes and time clustering framework based on the likelihood. Poisson and Cox point process can be expressed via their likelihoods; however, for the Markov point process it is necessary to define a pseudo-likelihood.

4.1.1 POISSON POINT PROCESS

For the homogeneous Poisson process, let $\lambda :=$ arrival rate/occurrence of cases per unit time. Set $P_{n(s)}(t) :=$ probability of observing $n(s)$ events at time t in a location $s \in D$; and $P_{n(s)}(t + \Delta t) :=$ probability of observing $n(s)$ events in interval length time Δt. Thus

$$p = P_1(\Delta t) = \lambda \Delta t,$$

$$1 - p = P_0(\Delta t) = 1 - \lambda \Delta t,$$

and the number of cases Y_i given $\lambda \sim Pois(\lambda)$,

$$\begin{aligned} P(Y_t = n) =& \frac{e^{-\lambda t}(\lambda t)^n}{n!} \\ P(Y_t = 1) =& e^{-\lambda t}(\lambda t). \end{aligned}$$

In general,

$$\begin{aligned} P_{n(s)}(t + \Delta t) =& P_{n(s_t)}(t) \cdot P_0(\Delta t) + P_{n(s_{t-1})} \cdot P_1(\Delta t) \\ =& P_{n(s_t)}(t)(1 - \lambda \Delta t) + P_{n(s_{t-1})}(\lambda \Delta t). \end{aligned}$$

Then

$$\frac{P_{n(s)}(t+\Delta t) - P_{n(s)}(t)}{\Delta t} = \lambda P_{n(s_{t-1})} - \lambda P_{n(s)}(t).$$

Letting $\Delta t \to 0$ turns into a differential equation:

$$\frac{dP_{n(s)}(t)}{dt} = \lambda P_{n(s_{t-1})} - \lambda P_{n(s)}(t).$$

When $n(s) = 0$,

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t),$$

with the solution:

$$P_0(t) = ce^{-\lambda t}.$$

The likelihood based on the domain area D is $f(y_i|\lambda_i z_i)$ where

$$z_i = \begin{cases} 1 & \text{if } i \in D, \\ 0 & \text{otherwise,} \end{cases}$$

and the density function can be expressed as

$$f(y_i|\lambda_i z_i) = \frac{\lambda(y_i)}{\mu}^{z_i}.$$

Resnick (2002) proposed a transformation of the homogeneous Poisson point process to obtain to the nonhomogeneous Poisson process with local intensity $\alpha(t)$. For the nonhomogeneous Poisson point process with varying intensity across space and/or time, the model is constructed such that its likelihood function is tractable. It has been applied in Tanaka et al., (2008) and Diggle (2014). Therefore, if $\mathcal{P}$ denotes the Poisson point process at points/locations $X = \{s_1, s_2, \ldots, s_n\}$ on a finite region D it captures the limiting expected count $\lambda(s)$ called intensity. Recalling the conditioning property defined in Section 2.1.2, the factorized product of Poisson distribution will have probability density $\lambda(s)/\mu$, where $\mu = \int_D \lambda(s)ds$, and the probability of observing n points is $\frac{\mu^n}{n!}e^{-\mu}$. Since the observations are independent, the likelihood can be expressed as:

$$L_\lambda = e^{-\mu}\frac{\mu^n}{n!}\prod_{i=1}^{n}\frac{\lambda(s_i)}{\mu} \propto e^{-\int_D \lambda(s)ds} \cdot \prod_{i=1}^{n}\lambda(s_i) \tag{83}$$

Then the log-likelihood is of the form (Diggle, 2014):

$$\ell_\lambda \propto \sum_{i=1}^{n} \log \lambda(s_i) - \int_D \lambda(s) ds, \tag{84}$$

where D is the region. In many cases, $\log \lambda(s_i)$ is expressed as a regression model such as $\log \lambda(s) = \sum_{j=1}^{m} \beta_j z_j(s)$, where $z_j(s)$ represents explanatory variables that are spatially referenced. If $z(s)$ are modeled stochastically, this becomes a Cox process.

4.1.2 COX POINT PROCESS

Recall from Section 2.1.3 that for a Cox process directed by η, any random measure on D is conditionally Poisson given η. Assuming the model for η is indexed by a parameter θ, the Poisson likelihood in Equation (83) becomes (see Diggle, 2014):

$$L(\theta, \eta) \propto \left\{ \int_D \eta(s) ds \right\}^{-n} \prod_{i=1}^{n} \eta(s_i). \tag{85}$$

Then the log-likelihood for the Cox process is the expectation with respect to η of the Poisson likelihood in Equation (85) for a given η,

$$\ell_\theta = \mathbb{E}_\eta[L(\theta, \eta)]. \tag{86}$$

The density at time (t_{r-1}, t_r) is

$$f(m|D_{k(t)}) = \lim_{s \in D_{k(t)}} \frac{\mathbb{E}(N(s) = m)}{\mu(D_{k(t)})}, content... \tag{87}$$

for $m = 0, 1, 2, \ldots$. However, such likelihood is not easily implemented as the coverage density of points increases resulting in high dimensionality of the integration (Adams et al., 2009; Diggle, 2014); in fact, the likelihood is intractable.

4.1.3 MARKOV POINT PROCESS

As discussed in Section 2.1.4, Markov (or Gibbs) processes are defined in terms of joint densities of pair potentials. This construct leads to the likelihood function as a potential tool for inference; however, the normalizing constant c in Equation (13) is not tractable.

To avoid complication of estimation due to lack or tractability, Besag (1975) introduced pseudo-likelihood for lattice process, and later a point process version was derived by Besag, Milne, and Zachary (1982). For a general multivariate distribution $f(x_1, \ldots, x_n)$, the pseudo-likelihood is defined as the product of full conditionals (Diggle, 2014):

$$PL = \sum_{i=1}^{n} \log f(x_i | x_j, j \neq i). \tag{88}$$

Markov point processes can be expressed as ratios of Janossy densities j_i, which alleviates the issues of intractability of the normalizing constant that disappears (Daley & Vere-Jones, 2003):

$$PL = \left(\prod_{i=1}^{n} \frac{j_n(f(x_1, \ldots, x_n))}{j_{n-1}\left(f(x_1, \ldots, x_n) \setminus \{x_i\}\right)} \right) \exp\left(- \int_D \frac{j_{n+1}(f(x_1, \ldots, x_n) \cup \{u\})}{j_n(f(x_1, \ldots, x_n))} \right). \tag{89}$$

4.1.4 PALM FUNCTION

When a sample of points is approximated by a nonhomogeneous Poisson point process, Tanaka et al. (2008) defined the corresponding Palm log-likelihood function as:

$$\ell(\mu, \nu, \tau) = \sum_{\{i,j; i \neq j, r_{ij} < R\}} \log\{N(D_k^t) \cdot \lambda_0(d_{ij})\} - N(D_k^t) \int_{D_k^t} \lambda_0(r) \cdot 2\pi r dr. \tag{90}$$

The sum is taken over all pairs i, j such that the distance d_{ij} between s_i and s_j is smaller than some radius from an original point that is sufficiently greater than the range of correlation of the Neyman-Scott process, i.e. for our construct, the points s_i and s_j are contained in a specific region D_k^t, which can also be described as a neighborhood.

4.2 SPATIO-TEMPORAL THEORETICAL FRAMEWORK

Spatial point patterns do not generally have a natural ordering; however, the addition of time allows for a natural ordering of observations that lends itself to building a statistical approach based on conditioning on the past. Extension of the theoretical framework from the Section 4.1 to include time reveals unique challenges.

Set $(n_1, t_1), (n_2, t_2), \ldots, (n_m, t_m)$ to be the observed pairs of occurrences and times. Then

we define the number of points generated as $n(s,t_r) = \#\{s_i, s_i \in D, t_i < t_r\} = n_{t_r}$. This can also be seen as the realization of the point process X, where the number of occurrences is $n = X(D_k \times [0,T])$. We build the conditional intensity function $\lambda(s;t)$ based on the entire history $\mathcal{H}_t$ of the spatio-temporal process up to time t (Cressie & Wikle, 2011):

$$\lambda(s;t) \equiv \lim_{|ds|,|dt|\to 0} \frac{\mathbb{E}(X(ds;dt)|\mathcal{H}_t)}{|ds||dt|}. \tag{91}$$

Given the finite region $D \times T$ where the spatio-temporal process is observed, we define marginal spatial and temporal intensities as:

$$\lambda_T(s) = \int_T \lambda(s,t)dt \quad \lambda_D(t) = \int_D \lambda(s,t)ds. \tag{92}$$

We can also further restrict the marginal temporal intensity to a subarea D_k. This basic construct will be used to develop inference in the spatio-temporal setting.

4.2.1 SPATIO-TEMPORAL POISSON

For data (s_i,t_i), $i = 1,\ldots,n$, the likelihood function of the spatio-temporal Poisson point process is:

$$L_\lambda \propto e^{-\int_0^T \int_D \lambda(s,t)dsdt} \cdot \prod_{i=1}^n \lambda(s_i,t_i) \tag{93}$$

Then the log-likelihood can be expressed as (Diggle, 2014):

$$\ell(\lambda) \propto \sum_{i=1}^n \log \lambda(s_i,t_i) - \int_0^T \int_D \lambda(s,t)dsdt. \tag{94}$$

4.2.2 SPATIO-TEMPORAL COX

Since the Cox point process is simply an inhomogeneous Poisson process with a random intensity, the resulting likelihood for the spatio-temporal Cox point process is (Cressie & Wikle, 2011):

$$L(\theta,\eta) \propto \int_0^T \int_D \eta(s,t)dsdt \prod_{i=1}^n \eta(s_i,t_i)e^{-\int \eta(s,t)dsdt}, \tag{95}$$

where $\eta(s,t)$ is assumed to be a stochastic process. Then the log-likelihood follows as:

$$\ell(\lambda) = \sum_{i=1}^{n} \log \eta(s_i, t_i) - \int_0^T \int_D \eta(s,t) ds dt. \tag{96}$$

4.2.3 SPATIO-TEMPORAL MARKOV

For the spatio-temporal Markov process, the intensity is of the form:

$$\lambda_c((s,t)|\mathcal{H}) = \lambda(s,t) = \alpha(t) \prod_{i=1}^{n_t} e^{-\|s-s_i\|}, \tag{97}$$

with density function

$$f(m|D_{k(t)}) = \alpha(t) \prod_{s \in D_{k(t)}} e^{-\|s-s_{k(t-1)}\|}. \tag{98}$$

To estimate this intensity, the log-likelihood of (s_i, t_i) is of the form (see Diggle, 2014):

$$\ell(\lambda((\cdot,\cdot)|\mathcal{H})) \propto \sum_{i=1}^{n} \log \lambda_c((s_i,t_i)|\mathcal{H}) - \int_0^T \int_D \lambda_c((s,t)|\mathcal{H}) ds dt. \tag{99}$$

Then

$$\ell(f(\cdot|D_{k(t)})) = \sum_{s \in D_{k(t)}} \log f - \int_0^T \int_D f ds dt, \tag{100}$$

where f is some spatio-temporal pair correlation function.

4.2.4 SPATIO-TEMPORAL PALM

Palm distributions offer the conditional distribution of a point process given a point occurs at a particular location. This conditioning can also be based on a point occurring at a previous time, which is important in spatial point processes that are followed over time. For maximum likelihood estimation, Bayesian approach based on integrated nested Laplace approximation (INLA) may be used (see Choi & Baccelli, 2018; Coeurjolly et al., 2017).

Follwoing Equation (91) by incorporating time into the Palm likelihood and conditioning

on the past, $\mathcal{H}$ yields the conditional intensity function (see Diggle, 2014):

$$\lambda_c((s,t)|\mathcal{H}) = \lim_{|ds|\to 0, dt\to 0} \frac{\mathbb{E}(N(D_k^t)|\mathcal{H})}{|D_k^t|}. \tag{101}$$

The density at time (t_{r-1}, t_r) is

$$f(m|D_{k(t)}) = \lim_{s\in D_{k(t)}} \frac{\mathbb{E}(N(s)=m)}{\mu(D_{k(t)})}, \tag{102}$$

for $m = 0, 1, 2, \ldots$.

For non-overlapping areas nested over time, we follow the idea presented in Kedem and Gagnon (2010), such that at time t the function of the observed point process in q_t subareas can be described as $\tau_t = (\tau_{1_t}, \tau_{2_t}, \ldots, \tau_{q_t})$. Here, we set $n_{1_t}, n_{2_t}, \ldots, n_{q_t}$ to be the number of occurrences in subareas $D_{1_t}, D_{2_t}, \ldots, D_{q_t}$, and with associated errors $\varepsilon_t = (\varepsilon_{1_t}, \varepsilon_{2_t}, \ldots, \varepsilon_{q_t})$. In this construct, τ_t is a function of the disc size, Palm distribution, and past values and covariates plus the associated error. Following the system from time $t = 1, \ldots, m$ as in Wen and Kedem (2009), such that

$$\begin{aligned}
\tau_{1_t} &= f(\text{disc size; Palm \& past values \& covariates}) + \varepsilon_{1_t} \text{ for } D_{1_t} \\
&\vdots \\
\tau_{q_t} &= f(\text{disc size; Palm \& past values \& covariates}) + \varepsilon_{q_t} \text{ for } D_{q_t}.
\end{aligned}$$

The log-concave type of distribution is used to model the error associated with the system. This class of distributions is highly flexible and most notably, the maximum likelihood estimator exists and can be computed by using existing algorithms (Walther, 2009). We let the error of the system follow the exponential distribution such that:

$$\begin{aligned}
f(\varepsilon_{1_t}) &= e^{-\alpha_{1_t} + \beta_{1_t} h(\varepsilon_{1_t})} \\
&\vdots \\
f(\varepsilon_{q_t}) &= e^{-\alpha_{q_t} + \beta_{q_t} h(\varepsilon_{q_t})}
\end{aligned}$$

Each subarea has a Palm pdf defined from:

$$p_{t,s} := p_{t,s^t_{k(t)}} = d\mathbb{Q}_{t,s}(x_{k(t)}) = \text{pdf of reduced Palm}$$

$$\stackrel{or}{=} d\mathbb{Q}_{t,s}(x) = p_{k(t)} \tag{103}$$

The product density can also be thought of as the product of kernels K associated with the process, where

$$\begin{aligned}\rho^{(n(t))}(x_1, \ldots, x_{n(t)}) &= \prod K(x_i, s_{k(t)}) \\ &= \prod_{k(t)=1}^{n(t)} K_o(x_i - s_{k(t)}) \mathbf{1}_{D^t_{k(t)}}(x_i).\end{aligned} \tag{104}$$

The kernel can take the form of any density described in Section 2.1.1 or could be any function of spatial correlation such as Moran's index that will be described in Section 4.3. The $\rho^{(n(t))}$ define the Palm distributions as they are conditioning the spatial point process on the locations or the Papangelou conditional intensity (Coeurjolly et al., 2016).

For the random probability measure ξ_t on $(S, \mathcal{S})$, let η be a random element in the Borel space $(T, \mathcal{T})$. Then let

$$\rho^{(n(t))} = \mathbb{Q}_{t,\underset{\sim}{s}}, \ s = (s_1, s_2, \ldots, s_{n(t)}) \in D_1 \times \cdots \times D_{n(t)} \tag{105}$$

be Palm distributions of η_t with respect to ξ_t and let $x_t = (x_{1(t)}, x_{2(t)}, \ldots, x_{n(t)})$ be random elements in $D_1 \times \cdots \times D_{n(t)}$ with $P(x_t \in \cdot | \xi_t \eta_t) = \xi_t^{n(t)}$ almost surely. Then $P(\eta_t \in \cdot | x_t) = \mathbb{Q}_{t,x_t} = \rho^{(n(t))}$ almost surely (see Kallenberg, 2002).

Baudin (1981) stated that the computation of the joint likelihood is very challenging and the complexity makes it difficult in application. Tanaka et al. (2008) and Tanaka and Ogata (2014) applied a nonhomogeneous Poisson process of Palm type from the nearest neighbor distance function. In the simulations that follow, the Palm intensities in each subarea and time are compared to the values of the Moran statistics. The likelihood of the Palm cluster process is used to capture estimates that are validated under Moran statistics discussed below. The Palm process is chosen because the probabilities of spread are associated with cluster sizes. This is a spatial marking where the parent or reference point could be chosen anywhere in the cluster, and spread incubation (the period infection and manifestation) is space controlled by a nearest neighbor distance distribution with a nested framework. Estimation of parameters will not be enough, which leads us to autocorrelation and consecutive time autocorrelation.

4.3 AUTOCORRELATION MEASURE

In the context of the Palm distribution, a typical point has the distribution associated with nearest neighbor. Moran's autocorrelation index I is used to measure spatial dependence between points (Cliff & Ord, 1981). It is a method for quantifying the degree of spatial clustering or dispersion. The global Moran index I is defined as:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}, \tag{106}$$

where z_i and z_j are the standardized values of spatial characteristics or features s_i and s_j of occurrences i and j, and w_{ij} represents a spatial weight between locations i and j. The range of I is -1 to +1, where -1 indicates dispersion and +1 indicates clustering.

Recognizing that a single Morans index may not give a sufficient summary of the spatial autocorrelation measure, Anselin (1995) defined a local indicator of spatial association (LISA) for a variable y_i observed at location i, as a statistic L_i, such that

$$L_i = f(y_i, y_{D_i}), \tag{107}$$

where f is a function, and the y_{D_i} are the values observed in the neighborhood D_i of i. Neighborhoods were defined by critical distance thresholds, but could also be defined based on the geography of the region, e.g. countries, states, or counties. The LISA for an observation is an indicator of spatial clustering around that observation and the sum of LISAs is proportional to the global indicator of spatial association, e.g. Moran's index I or Geary's c. The local Moran was defined as

$$I_i = z_i \sum_{j \neq i} w_{ij} z_j, \tag{108}$$

where z_i, z_j are standardized values and the summation over j includes only neighboring values, while the global Moran is

$$I = \sum_i I_i. \tag{109}$$

To incorporate time into the Moran index, Vaillant et al. (2011) treated the data as time series for each spatial location with cross correlation between series representing spatial correlation. For a partitioned time interval with pairs of observations $(t_{i-1}, t_i],\ i = 1, \ldots, n-$

1, Moran's index was defined based on a nearest neighbor scheme as follows:

$$M_i = \sum_{(x,y)\in D} w_{(x,y)} \mathbf{1}_{[0,t_{r-1}]}(T_x) \mathbf{1}_{(t_{r-1},t_r]}(T_y), \tag{110}$$

where D denotes a discrete set of plant locations, T_x denotes the date (time variable) of an event at location x with $\mathbf{1}_{[0,t_{i-1}]}(T_x)$ denoting the indicator of whether time T_x falls in the interval $[0, t_{r-1}]$, and the weight $w_{(x,y)}$ is nonzero only when x and y are neighbors. The index M_i is focused on binary event indicators from sequential time intervals.

We propose a space-time Moran-type statistic that incorporates concepts from both Anselin (1995) and Vaillant et al. (2011), Equations 108 and 110. For the simplest form, an indicator function is utilized; however, any function could be used. Since the partitioned areas vary from one time to the next, we define D_k^t as subarea $k \geq 1$ at time t and then we will define a Moran-type autocorrelation statistic on each measurable subset D_k^t as:

$$M_k^t = \sum_{s_i,s_j \in D_k^t} w_{s_i,s_j} \mathbf{1}_{[t_{r-1},t_r)}(T_{s_i}, T_{s_j}), \quad k \geq 1, \quad t \geq 1, \tag{111}$$

where w_{s_i,s_j} denotes the spatial weight between points s_i and s_j of disease/event occurrence ($w_{s_i,s_i} = 0$), T_{s_i} and T_{s_j} denote the time of detection of s_i and s_j, and $\mathbf{1}_{[t_{r-1},t_r)}(T_{s_i}, T_{s_j})$ is an indicator of whether times T_{s_i} and T_{s_j} both fall in the interval $[t_{r-1}, t_r)$.

In the definition of Moran's autocorrelation, w_{s_i,s_j} represents a spatial weight between any two distinct event locations generated within a subarea D_k^t, which could be a function, e.g.,

1. binary matrix based on distance threshold or neighborhood:

$$w_{s_i,s_j} = \begin{cases} 1 & d_{s_i,s_j} \leq d_0 \\ 0 & d_{s_i,s_j} > d_0, \end{cases} \quad \text{or} \quad w_{s_i,s_j} = \begin{cases} 1 & \text{if points } s_i \text{ and } s_j \text{ are neighbors,} \\ 0 & \text{otherwise;} \end{cases}$$

2. the inverse power function between two points, $w_{s_i,s_j} = (d_{s_i,s_j})^{-p}$, for p fixed integer;

3. "geographical" weights defined as $w_{s_i,s_j} = \begin{cases} e^{\left(-d_{s_i,s_j}/\bar{d}\right)} & s_i \neq s_j \\ 0 & \text{otherwise,} \end{cases}$
where $\bar{d}$ is the average of all distances between two points s_i and s_j;

4. an estimate of the autocorrelation/semivariance statistic;

Method 1 described above represents quasi-local correlation and local correlation, whereas Methods 2 and 3 represent global correlation and quasi-global correlation, respectively. However, selection of the weight function objectively is a pending question (Chen, 2012).

The geographically weighted regression approach proposed by Anselin (1988) has been widely adopted for spatial heterogeneity. However, as described in Murakami, Yoshida, Seay, Griffith, and Yamagata (2017), it has limitations in the degree of smoothness and its Bayesian variation needs further exploration. For example, as described in Wheeler and Waller (2008), rivers act as barriers in cases of rabies in raccoon population.

Since the spatial autocorrelation we propose is allowed to be as flexible as possible with local action, we use the step function alongside the weighted distance. And while values of this index are not between -1 and 1, it is essentially in the same spirit as Moran's I in that larger values indicate more autocorrelation or clustering.

4.3.1 EXPECTED VALUE

Cliff and Ord (1981) derived the moments of Moran's index in Equation (106) the assumption that observations are random independent drawings from normal populations. Under these assumptions, the expected value of Moran's index is given as:

$$\mathbb{E}(I) = \frac{n}{S_0} \frac{\mathbb{E}\left(\sum_{i,j} w_{ij} Z_i Z_j\right)}{\mathbb{E}\left(\sum_i Z_i^2\right)} = \frac{-1}{n-1}, \tag{112}$$

where $S_0 = \sum w_{ij}$. Under the same assumptions, Anselin (1995) calculated the expected value of the LISA in Equation (108) as:

$$\mathbb{E}(I_i) = \frac{-w_i}{(n-1)}. \tag{113}$$

Next we derive the expected value of our proposed spatio-temporal Moran autocorrelation statistic. Given the assumption that observations z_i^t, $i = 1, \ldots, n_{k(t)}$, are random independent drawings within the same disc k from a given distribution with pdf $Z_\cdot^t \sim P(Z_\cdot^t | Z_\cdot^{t-1})$, the expected value of $Z_\cdot^t$ and $Z_{t,\cdot}^2$ can be derived as following Cliff and Ord (1981). Randomization will be used to create identical observable outcomes that do not alter the behavior of the

process, but the population will not be assumed normal.

$$\begin{aligned}\mathbb{E}_R(Z_\cdot^t) &= \sum_{i=1}^{n_{k(t)}} P(Z_\cdot^t | Z_\cdot^{t-1}) \cdot z_i^t \\ &= P(Z_\cdot^t | Z_\cdot^{t-1}) \sum_{i=1}^{n_{k(t)}} z_i^t = \mathbf{m_1} \qquad (114)\\ \mathbb{E}_R(Z_{t,\cdot}^2) &= \sum_{i=1}^{n_{k(t)}} P(Z_\cdot^t | Z_\cdot^{t-1}) \cdot z_{t,i}^2 \\ &= P(Z_\cdot^t | Z_\cdot^{t-1}) \sum_{i=1}^{n_{k(t)}} z_{t,i}^2 = \mathbf{m_2} \qquad (115)\end{aligned}$$

There are now two methods to be considered when calculating the expectation under randomization of two realizations: fixed i with random j and random i and j.

(i) Given a point z_i^t with distribution Z_i^t, the correlation with regard to any other neighboring point z_j^t can be expressed as follows.

$$\begin{aligned}\mathbb{E}_R(Z_i^t \cdot Z_j^t | Z_i^t = z_i^t) &= \sum_{j=1, j\neq i}^{n_{k(t)}} P(Z_j^t | Z_\cdot^{t-1}) \cdot z_i^t z_j^t \\ &= P(Z_j^t | Z_\cdot^{t-1}) \mathbb{E}_R \left[Z_i^t (Z_1^t + Z_2^t + \cdots + Z_{i-1}^t + Z_{i+1}^t + \cdots + Z_{n_{k(t)}}^t) \right] \\ &= P(Z_j^t | Z_\cdot^{t-1}) \mathbb{E}_R \left[Z_i^t \left(\sum_{j=1}^{n_{k(t)}} z_j^t - Z_i^t \right) \right] \\ &= P(Z_j^t | Z_\cdot^{t-1}) \left\{ \mathbb{E}_R(Z_i^t) \sum_{j=1}^{n_{k(t)}} z_j - \mathbb{E}_R(Z_{t,i}^2) \right\} \\ &= P(Z_j^t | Z_\cdot^{t-1}) \left\{ \mathbf{m_1} \sum_{j=1}^{n_{k(t)}} z_j - \mathbf{m_2} \right\}, \qquad (116)\end{aligned}$$

where $\mathbf{m_1}$ and $\mathbf{m_2}$ are as defined in Equations (114) and (115), respectively.

(ii) Given any point z_i^t with distribution Z_i^t, the correlation with regard to any other neighboring point z_j^t can be expressed as follows.

$$\mathbb{E}_R(Z_i^t \cdot Z_j^t) = \mathbb{E}_R \left[\mathbb{E}_R(Z_i^t \cdot Z_j^t | Z_i^t = z_i^t) \right]$$

$$
\begin{aligned}
&= \mathbb{E}_R \left[P(Z_j^t | Z_\cdot^{t-1}) \left\{ \mathbf{m_1} \sum_{j=1}^{n_{k(t)}} z_j - \mathbf{m_2} \right\} \right] \\
&= \sum_{i=1}^{n_{k(t)}} P(Z_i^t | Z_\cdot^{t-1}) \cdot z_i \left[P(Z_j^t | Z_\cdot^{t-1}) \left\{ \mathbf{m_1} \sum_{j=1}^{n_{k(t)}} z_j - \mathbf{m_2} \right\} \right] \\
&= P(Z_i^t | Z_\cdot^{t-1}) P(Z_j^t | Z_\cdot^{t-1}) \left[\mathbf{m_1} \sum_{i=1}^{n_{k(t)}} z_i \sum_{j=1}^{n_{k(t)}} z_j - \mathbf{m_2} \right] \qquad (117)
\end{aligned}
$$

where $\mathbf{m_1}$ and $\mathbf{m_2}$ are as defined in Equations (114) and (115), respectively.

$\mathbb{E}_R \left(Z_j \cdot Z_i^t \right)$ can also be explored based on the "birthday problem" $\sum_{i=1}^{n_{k(t)}-1} \sum_{j=i+1}^{n_{k(t)}} z_j^t z_i^t$, randomization $z_j^t z_i^t$, or expectation under randomization $\frac{1}{n_{k(t)!}} z_j^t z_i^t$.

Then the proposed spatio-temporal Moran autocorrelation statistic in Equation (111) for a given disc k can be written as

$$
M_k^t = S_0 \sum_i \sum_j Z_i^t Z_j^t, \qquad (118)
$$

where $S_0 = \sum_i \sum_j w_{ij}$.

Then using Equations (114), (115), and (117), the expected value of Equation (118) can be written as

$$
\begin{aligned}
\mathbb{E}(M_k^t) &= S_0 \mathbb{E}_R (Z_i^t Z_j^t) \\
&= S_0 P(Z_i^t | Z_\cdot^{t-1}) P(Z_j^t | Z_\cdot^{t-1}) \left[\mathbf{m_1} \sum_{i=1}^{n_{k(t)}} z_i \sum_{j=1}^{n_{k(t)}} z_j - \mathbf{m_2} \right]. \qquad (119)
\end{aligned}
$$

All points of $D_k^t \coloneqq (D_k \times T_r)$ will be assumed to be neighbors of s_k within the subarea D_k^t. This idea is also presented in Stoyan et al. (1995) under the spherical contact distribution or nearest neighbor distance. So the counting measure is defined as:

$$
N(D_k^t) = \sum_{s_i \in D_k^t} \mathbf{1}_{(t_{r-1}, t_r]}(s_i) = \int_{D_k^t} \xi(s) ds, \qquad (120)
$$

where ξ is a random measure on D with positive intensity. On each measurable subset D_k^t,

define the intensity function of the process at time t as

$$\lambda_k^t(f) = \exp(\zeta_k^t(f)), \text{ where}$$

$$\mathbb{E}(\zeta_k^t(f)) = \mu_k^t(f) = \mathbb{P}_{(X,\phi)}(f)(s_k) := \text{ Palm distribution} \tag{121}$$

of (X, ϕ) with regard to ϕ for any nonnegative measurable function f on D_k^t, i.e

$$Z_k^t(f) = \frac{1}{\mathbb{E}\phi \ D_k^t} \int_{D_k^t} f(\theta_u(X, \phi))\phi(du)$$

will have probability density $\lambda(s)/\mu$, where $\mu = \int_D \lambda(s)ds$, and the probability of observing n points is $e^{-\mu}\frac{\mu^n}{n!}$

Hence,

$$\mathbb{E}(M_t^k) = \sum_{s \in D_k^t} w_{s_i,s_j} \mathbb{Q}_{(X,\xi)} f(s_i^{t-1}). \tag{122}$$

4.4 PATTERN DETECTION

Tango (1995) proposed cluster analysis of disease allowing a measure of closeness between regions and computed power tests, but ignored autocorrelation and time effects. The cluster representation of a space and time process is elucidated in the discrete setting. In a time evolving heterogeneous population, traditional statistical methods lack accuracy in implementing data patterns. In this new proposed approach, local partitioning of areas over time will be used to build clustering under Palm processes, generated under a Markov chain technique. Using data from simulations of virus propagation over space and consecutive times, we discuss the multi-sites clusters and the similarities among these clusters that result from subareas that are nested over time. The spatial structure of the data is then taken advantage of as we depict prevalence in spread based on prevalence mapping after geostatistical kriging area reference (the model of interpolated values is governed by prior covariances) so that spatio-temporal association is captured.

CHAPTER 5

APPLICATION

As shown in Chapter 4, complexity of spatio-temporal inference of stochastic processes directs research into simulation models or algorithms to describe properties, various parameter estimations, and summary statistics. We provide adaptive simulation of the spread X_t by approximations then with the locations $\underset{\sim}{X}_t$ under a thinning algorithm of a homogeneous Poisson point process which enjoys explicitly solvable likelihood estimation functions. We examine simulated examples to illustrate our proposed method under Palm distributions. We then explore the performance of the proposed model on real data sets. Visualization of spatio-temporal data can be achieved through marginal and conditional plots, e.g. space-time plots, time series plots, and spatial maps.

Moran's index is one of the most widely applied spatial measures of autocorrelation and is applied here. However, Moran's testing is challenged by its distributional assumption of normality. Simulations thru Monte Carlo methods as proposed in Yamada and Okabe (2015) show that the normality criteria is often unrealistic. Tiefelsdorf (1998) proposed representation of Moran's index as conditional distributions of its moments with an underlying Gaussian spatial process. However, the normal distribution assumption was rooted in computational deductions. Hence, the interest in identifying clustering becomes more evident. We will compute the proposed Moran statistics, challenge the distributional validity, and make comparisons in space and time with other statistics.

5.1 SIMULATIONS IN R

Results are shown in three simulated examples. We first generate sample observations initially with a Poisson point process. We show the process growth at later times and evidence possible configuration changes in irregular shaped spaces and extend implementation of stationary time series sequence of observations of the spatial point process. The existence and uniqueness as described in Section 3.3 are tested in the simulations. We conduct the simulations using `R` with the `spatstat` package and represent the weights of the Moran's

index as the inverse distance between two points. We use a modified Geary's c and estimates of the Palm parameters to validate if our Moran-like statistics have similar trends to those other known measures of autocorrelation. Geary's c is a measure of autocorrelation based on the standardized squared distance of points (see Geary, 1954; Sokal, Oden, & Thomson, 1998). Palm parameters are used to estimate the average number of points that are generated within a distance r of a typical point (Tanaka & Ogata, 2014). Simulations are performed in fixed subareas (discs), dynamic subareas (discs) and dynamic subareas (Voronoi cells).

5.1.1 FIXED SUBAREAS [1]

We begin with an observed area of 4 x 4 units, partitioned into unit squares (1 x 1 unit), and generate a Poisson point processes within that area. We include time at indicators $t = 0, 1, 2, \ldots, T$, and add that component to the domain subareas D_k^t, $k = 1, 2, \ldots, m$, where $m = 16$ is the total number of subareas. Next, within these subareas, we create discs with radius of half unit. These discs will allow us to generate sample points within the subareas as well as prevent any overlap of points between the subareas. We choose the Poisson point process to randomly generate points in the 16 subareas. Lastly, our Moran statistic is calculated as the sum of the inverse distances between each point that are generated within each of the discs over time.

The steps are as follows: in `R`, we (i) introduce a perturbation (using a Poisson point process) at each local subarea based on some initial location with constant rate λ at an initial time, (ii) find the Moran statistic within each sub area, (iii) generate new points with different Poisson process rates defined for each time period as $\lambda_i = i\lambda, i = 1 \ldots, I$ where I is the number of time intervals as in Equation (49), and (iv) calculate the Moran statistic. We continue steps (iii) through (iv) until all time periods are covered. The `R` code for simulation of the Poisson point process and calculation of Moran statistics is provided in Appendix A.

Algorithm: Iterative local Moran statistics for fixed subareas

Procedure:

(i) Define the first time and subareas and include their centers.

(ii) Iterate local points from a Poisson point process in that subarea.

[1] Material from this section is based on joint work with Nhan Bui, Norou Diawara, Kumar Das, and Lance Waller. The published work is reference Bui et al., 2018.

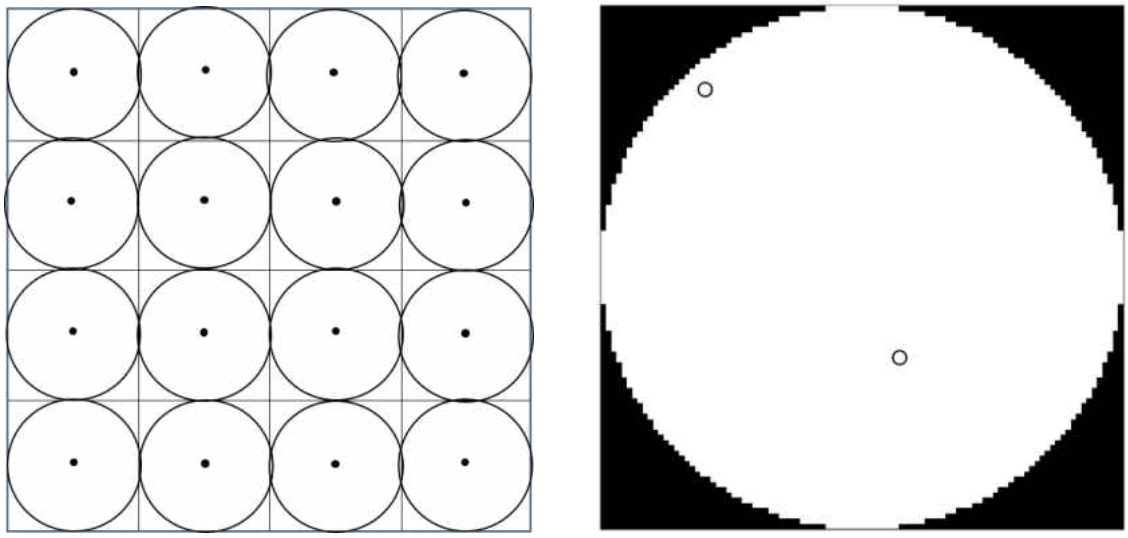
(iii) Compute the Moran statistic within that time and subarea combination.

Repeat the generation of the Poisson point process at the next time period within each subarea. Compute the Moran statistic within each subarea.

Stop when all times and subareas are reached.

End

Figure 5 (a) shows our observed area with 16 subareas and associated discs. To keep it simple, the discs do not overlap and the point process is subsampled and confined to these discs. Another benefit of this model is the tractability of data handling from the output. Figure 5 (b) below shows two points generated within a certain disc. This point generation process is repeated across all 16 discs.



(a) (b)

Figure 5: Fixed subarea simulation (a) 4x4 area plot and (b) points generated in one disc.

The Moran values generated within each disc across the 6 time periods are represented in Table 4 for $\lambda = 2$. The table shows that there is no extra point generated in discs 3 and 14 at the first time period and is denoted with "NA". The algorithm only considers two or more points when computing the Moran statistic, due to the way we defined our weights (inverse distance).

Table 4: Moran values for fixed subarea simulation with $\lambda = 2$

	Time					
disc	1	2	3	4	5	6
1	3.53	7.75	8.90	114.25	116.28	13.92
2	2.29	2.66	50.59	117.16	266.95	43.08
3	NA	5.80	2.19	278.04	65.05	60.36
4	23.04	7.70	14.01	212.57	24.99	63.85
5	19.26	NA	29.60	55.70	245.33	77.13
6	6.28	10.46	192.15	93.76	192.52	54.88
7	2.22	0.00	47.04	24.61	25.14	130.91
8	11.95	20.35	43.11	32.57	185.78	129.88
9	8.41	46.86	46.84	47.26	67.79	83.96
10	11.29	16.86	46.44	91.11	170.05	112.64
11	38.81	7.80	68.91	256.41	125.90	207.29
12	3.30	91.13	25.43	62.48	253.22	237.19
13	29.81	30.73	21.67	71.18	44.23	118.23
14	NA	17.25	26.71	98.46	107.99	173.16
15	2.19	211.91	79.45	44.79	415.01	251.05
16	12.63	20.09	72.53	34.14	61.49	229.93
Global Moran	148.61	450.36	1017.681	1086.424	1887.994	3736.684

The table also displays the global Moran values at each time period in the last line. The values are quite large relative to the measures of the subareas. In addition, for time period 1, discs 3 and 14 did not generate any points; the global Moran does not sufficiently portray a good description of the correlation and leads to exaggerated large values. The local Moran's statistics based on subareas provide a better description of the data according to the number of points generated within the disc.

Another benefit of the proposed approach is the tractability of data handling from the output. The spatial distribution of the generated points over time is displayed in Figure 6. This is not a surprising result since time is a function of the intensity $\lambda|B|$ where B is our observed area. In the simulation we kept intensity fixed and allowed for time interaction, i.e. letting it vary across (sequential) time points. Thus, the density plot shows that for sufficiently large time intervals, the number of points generated will start to spread in the

entire area as expected in the description of the model in the previous section. These results show also that the process is stable and infinitely divisible as suggested in Theorem 2.2.

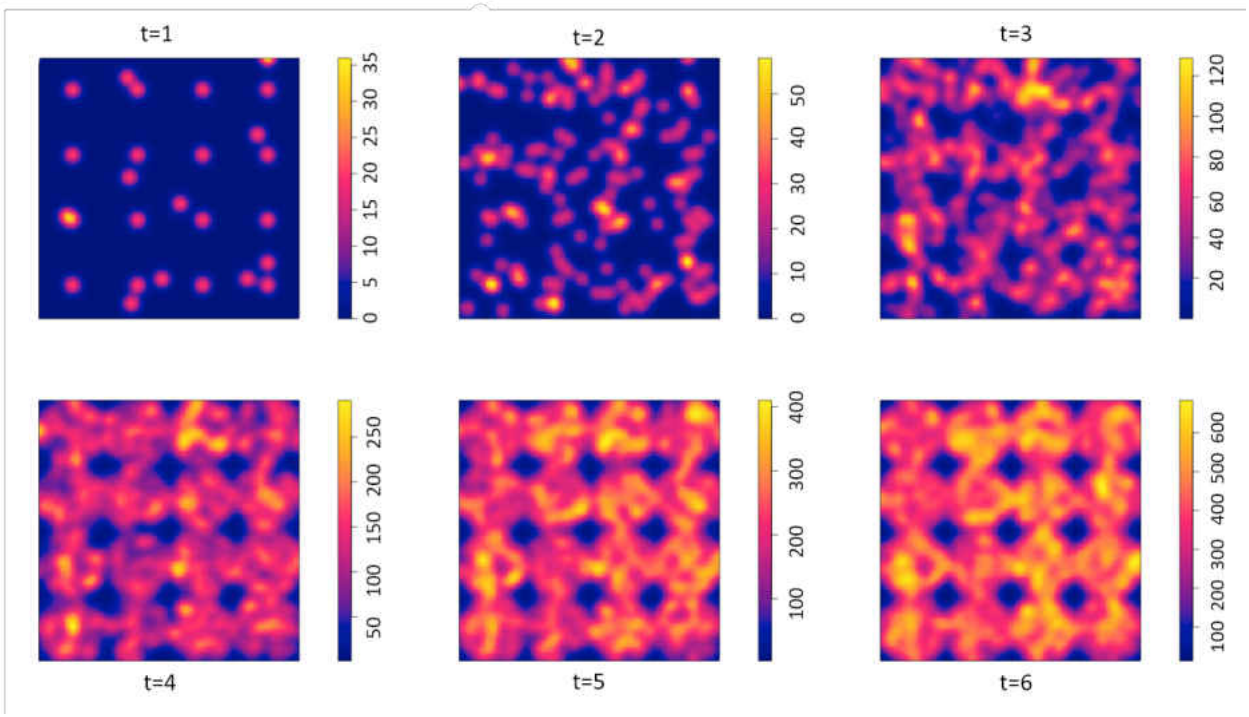


Figure 6: Density over time for fixed subarea simulation.

The local Moran's values provide a better summary of the spread after partitioning the spatial area into subareas and looking at behaviors associated with spread. We further explore the Moran's values, as comparisons of the Moran's statistics will offer more insight about the nature of the autocorrelation. One method used is the mixed linear model. As a generalization of the standard linear model, this method allows us to account for data which exhibit correlation and nonconstant variability as noticed in the previous section. The results show that time is significant. While the model captures the profile of the most important characteristic, time, the data has large variations associated with the Moran's values. Thus, we should be careful with methods that have an underlying normality assumption.

We simulate larger data of Moran's values using Monte Carlo inference technique about the mean of the Moran's statistics hoping to capture the distribution of the Moran's statistics at values of $\lambda = 2$. The algorithm ran at $10,000$ iterations produced the histogram displayed in Figure 7.

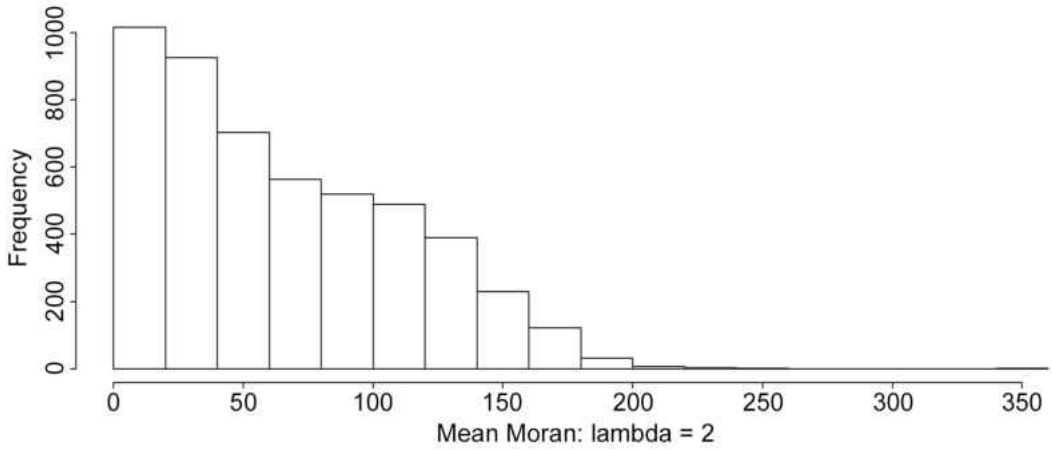


Figure 7: Histogram of Moran's statistics for fixed subareas

Figure 7 shows that the distribution of Moran's values is skewed to the right, showing that estimation techniques of Moran that are based on normality assumptions may be questionable. This is due to the varying nature of the dependency of the intensity. However, if we did not allow for time interaction, then the well established methods associated with Moran's indices are sufficient for the analysis.

Spatial autocorrelation with generalized linear models however is not well understood (see Griffith, 2005, 2009). To further conceptualize the idea of normality violations, we implement the Von Neumann rank test to assess whether or not our sample of points come from an underlying normal population. Using R for the calculations, the Von-Neumann test is constructed as in Taeger and Kuhnt (2014):

Description: Tests if a sample is sampled randomly from an underlying normal population

Assumptions: Data are at least measured on an ordinal scale. And let $X_1, \ldots, X_n$ be a sequence of random variables with observations $x_1, \ldots, x_n$.

Hypothesis: H_0 : Sequence $X_1, \ldots, X_n$ is randomly generated vs H_1 : Sequence is not randomly generated.

Test Statistic: $Z = \left(1 - \frac{V}{2}\right) \sqrt{(N-2)/(N^2-1)}$, where $V = \frac{\sum_{i-1}^{N-1}(X_{i+1}-X_i)^2}{\sum_{i-1}^{N}\left(X_i-\bar{X}\right)^2}$.

We modify the test to account for the fact that our data is two dimensional. Using the

standard Euclidean distance formula as "observations" of X_i's within disc k, we set

$$X_i = \sqrt{(x_1 - d_1)^2 + (x_2 - d_2)^2},$$

where (d_1, d_2) denotes the center of the disc and (x_1, x_2) a randomly generated point. That is, use the distance from center of the discs as a ranking mechanism. If our sample was indeed normally distributed, it would then be reasonable to see more points generated near the center of the disc or clustering at some area of the disc.

The results in Table 5 show non-significant p-values. In fact there is significant evidence that our points are randomly generated within each disc. This verifies that the normality assumption should not be used, randomness is present in some form, and we should look towards alternative models to analyze our data. These constants led to the use of extreme value distribution.

Table 5: Fixed subarea simulation p-values, H_0 : sequence is randomly generated

Disc	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
p-value	.65	0.98	.23	.25	.24	.90	.76	.11	.68	.64	.22	.30	.89	.14	.94	.11

One main concern is the nature of spread over time. The Moran values may grow larger and it might be of interest to investigate the extreme values as discussed in de Jong et al. (1984) or in Tiefelsdorf and Boots (1997). For instance, understanding the areal spread of a rare disease is crucial to quarantine and protection. In this section, we focus on values in our output from disc 1, across all 5 time points (as an illustrative example), that are considered "extreme."

The simulation output has some very large values (outliers) that may actually be from the way we defined our area and Moran's index as suggested in Maruyama (2015). Points generated extremely close ($d << 1$) will result in extremely large Moran values. It is of interest to understand such behavior. Thus, we shift our focus to the first disc at time 1. Using the interquartile range (IQR) as a measure of spread, since it is resistant to outliers, we can see from Table 6 below that $IQR = Q_3 - Q_1 = 5.17$ hence values $3 * (IQR)$ above Q_3 can be considered extreme value distributions.

Table 6: Quantiles for fixed subarea simulation disc 1 at time 1

min	Q_1	median	Q_3	max
0	0	2.158214	5.169383	163.457831

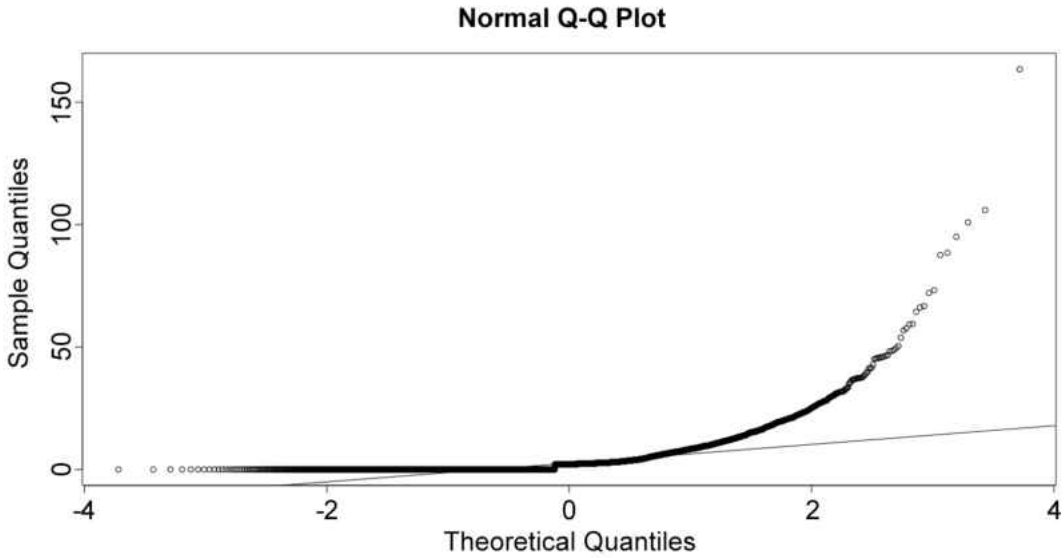


Figure 8: Normal QQ Plot of Moran values from fixed subarea simulation of disc 1 at $t = 1$

The curve shape of the QQ plot increases from left to right indicating the distribution is right skewed. This is further evidence of extreme value distributions; however, there is no clear indication of the distribution of Moran values.

The generalized Pareto Distribution (GPD) is the classical asymptotically motivated model for excesses above a high threshold. If our data points (Moran statistics for disc 1) are *iid* above a threshold u, then the limiting distribution will be a GPD. In applications, the GPD is used as a tail approximation to the population distribution from which a sample excesses $x - u$ above some threshold u are observed.

$$G(x|u,\sigma_u,\xi) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right]_+^{-\frac{1}{\xi}}, & \xi \neq 0, \quad x \geq 0 \\ \\ 1 - \exp\left[-\left(\frac{x-u}{\sigma_u}\right)\right]_+ & \xi = 0, \quad x \geq 0 \end{cases} \tag{123}$$

The GPD is parameterized by the shape and scale parameters ξ and σ_u. In particular, the GPD as expressed in Equation (123)) is expressed as exceedances $x > u$ where σ_u, u, ξ describe scale, location, and shape parameters, respectively as in Scarrott and MacDonald (2012). In this representation the mean is equivalent to the threshold.

Parameters of the model are estimated again with Monte Carlo techniques and compared (under maximum likelihood estimation, method of moments, and probability weighted moments estimation) using the data from the 10,000 runs of the algorithm. We then fit the data to a GPD for the first disk for all time periods. For the first time period, we can see that the mean residual life plot (Figure 9) is linear almost everywhere but in particular, becomes slightly erratic above 50. This plot suggests that threshold $u = 12$ is an appropriate choice as sample size of $n_u = 512$ above excess provides a good balance between bias and variance of parameter estimation for the GPD.

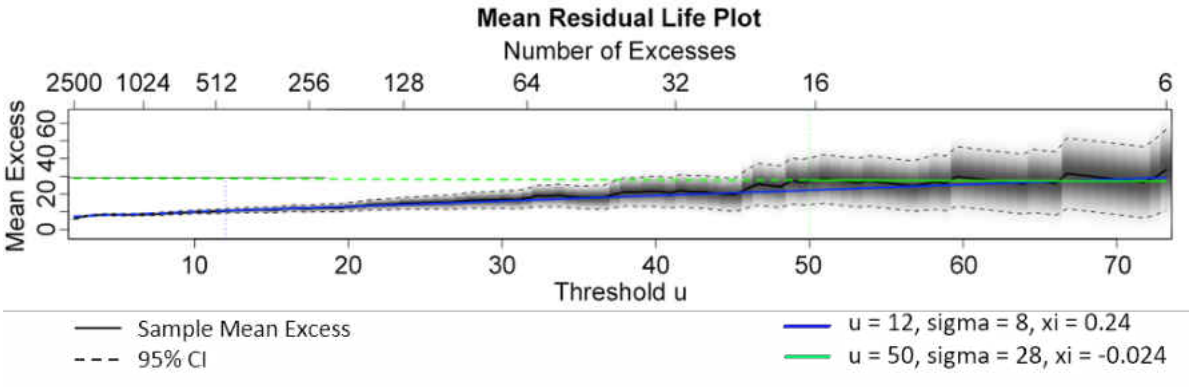


Figure 9: Mean Residual Life Plot for fixed subarea simulation.

Using `R` along with the `POT` package, we fit a GPD using the threshold value $u = 12$. We then select threshold values below and above 12 to observe any trend and find one that fits best. Table 7 shows the parameter estimates using maximum likelihood estimation (MLE), method of moments (MoM), and probability weighted moments (PWM). While the MLE has consistently low standard errors, the PWM may be a better choice for higher threshold values. The MoM on the other hand, has larger standard errors than the other two methods.

Table 7: Estimation comparison

Threshold u	No. Excess n_u	Shape ξ MLE (SE)	Shape ξ MoM (SE)	Shape ξ PWM (SE)	Scale σ_u MLE (SE)	Scale σ_u MoM (SE)	Scale σ_u PWM (SE)
5	1272	0.209	0.228	0.202	6.543	6.410	6.629
		(0.032)	(0.101)	(0.034)	(0.276)	(0.817)	(0.291)
10	611	0.240	0.236	0.240	7.433	7.489	7.441
		(0.049)	(0.181)	(0.051)	(0.468)	(1.743)	(0.477)
12	468	0.236	0.233	0.234	8.001	8.047	8.032
		(0.055)	(0.189)	(0.058)	(0.570)	(1.935)	(0.587)
15	339	0.288	0.248	0.287	7.865	7.865	007.834
		(0.071)	(0.687)	(0.072)	(0.690)	(7.525)	(0.687)
20	193	0.321	0.246	0.317	8.733	9.565	8.665
		(0.099)	(0.599)	(0.101)	(1.052)	(7.537)	(1.024)
25	114	0.276	0.220	0.280	10.964	11.720	10.815
		(0.121)	(0.279)	(0.124)	(1.658)	(4.067)	(1.630)
30	74	0.288	0.208	0.296	12.181	13.361	11.874
		(0.159)	(0.286)	(0.157)	(2.362)	(4.698)	(2.238)

Moreover, in extreme value theory, there are three main domains of attraction: Gumbel, Fréchet, and Weibull. The distributions in the Gumbel domain have the exponential, including normal and Gamma as the limiting distribution of their tails. The Fréchet domain contains distributions with infinite yet heavier tails. And lastly, the Weibull domain contains distributions with lighter tails than the exponential distribution. Thus we want to test for the domain of attraction which is determined by the shape parameter ξ, as in Beisel, Rokyta, Wichman, and Joyce (2007). That is, the test of hypothesis $H_0 : \xi = 0$ vs $H_a : \xi > 0$ is equivalent to testing a Gumbel versus a Weibull domain of attraction. Using a 95% confidence interval for all estimated values of ξ under the maximum likelihood estimation, Table 8 indicates that ξ is likely to be positive. So we reject $H_0 : \xi = 0$ for supporting evidence that our data follows a Weibull domain of attraction.

Table 8: 95% CI for shape parameter ξ

Threshold u	MLE	CI
5	0.209	$(0.147, 0.271)$
10	0.240	$(0.144, 0.336)$
12	0.236	$(0.129, 0..343)$
15	0.288	$(0.150, 0.426)$
20	0.321	$(0.128, 0.514)$
25	0.276	$(0.040, 0.512)$
30	0.288	$(-0.022, 0.598)$

5.1.2 DYNAMIC SUBAREAS[2]

Fixed subareas are not always reasonable, so we will explore dynamic subareas that change over time. First we will conduct the simulation with subareas generated as discs, then move to the more general case using Voronoi cells.

Discs

We begin with an observed unit area (1 x 1 unit) and generate a Poisson point processes within that area. This is used as the first step in a clustering process. Each location site is then used as the center point to generate a disc with diameter equal to the minimum of the distance to the Voronoi edge and the distance to the observed area edge. This effectively means that an infected site can only subsequently infect sites that are closer to it than any other infected site. These Voronoi discs define the subareas $D^t_{k(t)}$ for the point process generated at the next time interval (step). The nested subareas defines a sequence of bounded sets whose limiting properties match results in Theorem 3.1. The Moran statistic is calculated at each step as the sum of the inverse distances between an offspring site and another offspring site. This format will continue through time $t = 4$ adhering to the Markov property ($t = 4$ chosen to follow results from a forestry example in Meddens & Hicke, 2014 or in Vaillant et

[2]Material from this section is based on joint work with Norou Diawara and Lance Waller. The published work is reference Lorio, Diawara, and Waller 2018.

al., 2011). Each subarea will be magnified by a scale parameter, $\alpha > 0$, such that the new radius is α times the previous radius with intensity function $\lambda * t$. Covariates that are time sensitive could have been included following the idea presented in Theorem 2.5.

Algorithm: Iterative local Moran statistics for discs

Procedure:

(i) Define the window and generate a Poisson point process in the area with intensity λ.

(ii) Calculate the Moran statistic.

(iii) Define new subareas with point from previous process as center and radius equal to the minimum of distance to the edge of the window or distance to the edge of the Voronoi cell.

(iv) Rescale subarea and iterate local points from a Poisson point process with intensity $\lambda * t$.

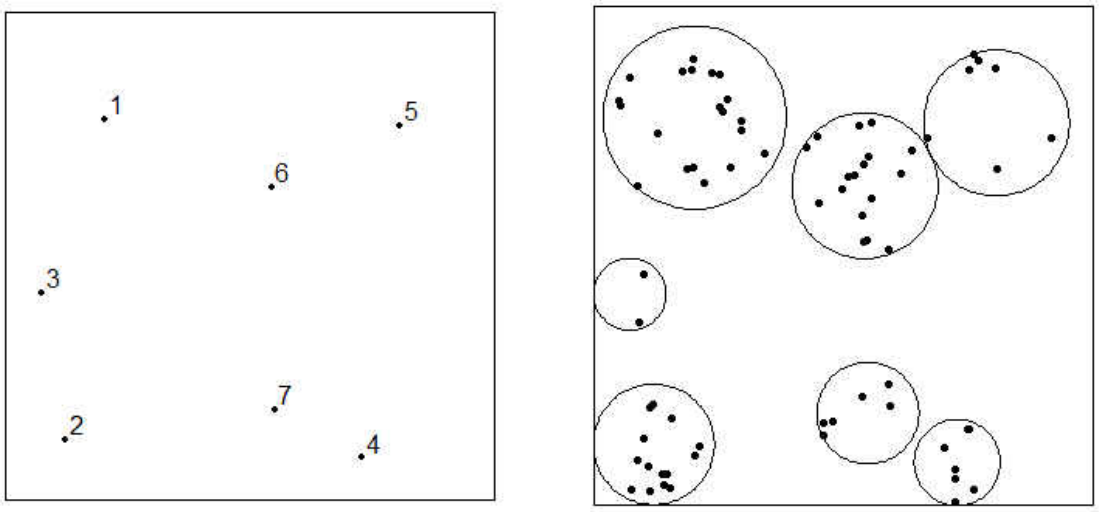
(v) Compute the local Moran statistic for each subarea at that time step.

Repeat steps (iii) through (v).

Stop when all times are reached.

End

Figure 10 shows the original observed area with the seven points generated at $t = 1$ alongside the discs generated from these points and the points generated at $t = 2$.



(a) (b)

Figure 10: Dynamic subarea simulation, $\lambda = 4$ at (a) $t = 1$ and (b) $t = 2$.

The Moran value for $t = 1$ equals 42.57. The number of points generated in each disc contained within the Voronoi subarea, the rescaled area, and Moran values generated at time $t = 2$ are displayed in Table 9 with $\lambda = 4$ and $\alpha = 5$. For example, at point 1 in Figure 10(a), a Voronoi cell is created, then 20 points are generated as shown in Figure 10(b). The number of points generated in other Voronoi cells is displayed in Table 9. Identifiability presented in Subsection 3.3.2 is verified and hence the unique values of spatio-temporal dependence are captured.

Table 9: Moran and Geary's c values for dynamic subareas, $\lambda = 4$ and $\alpha = 5$ at $t = 2$

disc	Points	Area	Moran	Geary's c
1	20	2.650	409.00	6.773
2	14	1.148	315.07	5.217
3	2	0.411	2.09	0.035
4	7	0.585	120.95	2.003
5	7	1.676	55.61	0.921
6	17	1.676	350.45	5.803
7	6	0.817	51.21	0.848

The process continues at time $t = 3$ with disc 1 having 20 subareas (see Figure 11), disc 2 having 14 subareas, etc. The non-zero Moran values for disc 1 at time $t = 3$ are listed in Table 10.

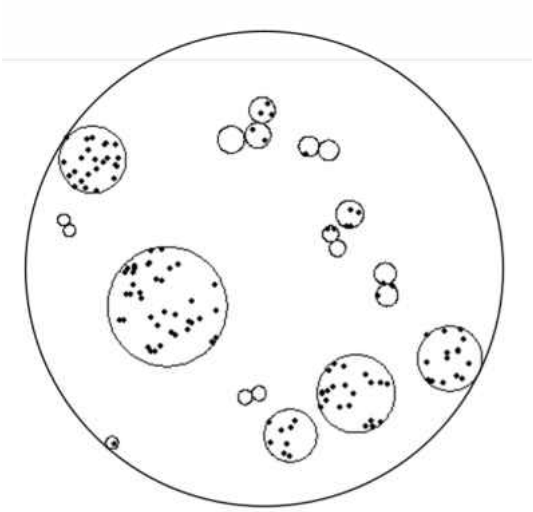


Figure 11: Dynamic subarea simulation disc 1 at $t = 3$.

Table 10: Non-zero Moran and Geary's c values for dynamic subarea simulation disc 1 at $t = 3$

disc	Points	Area	Moran	Geary's c
1	24	1.269	676.94	8.941
4	2	0.135	2.93	0.039
5	39	4.037	1176.21	15.536
6	22	1.723	628.83	8.306
7	8	0.785	84.38	1.115
8	2	0.072	7.90	0.104
9	15	1.183	262.56	3.468
12	4	0.212	29.42	0.389
14	3	0.185	13.75	0.182
19	2	0.185	3.45	0.046

After four time intervals, the clustering of points generated is displayed in Figure 12, with red shades representing the most dense areas and white shades representing areas with no instances of disease.

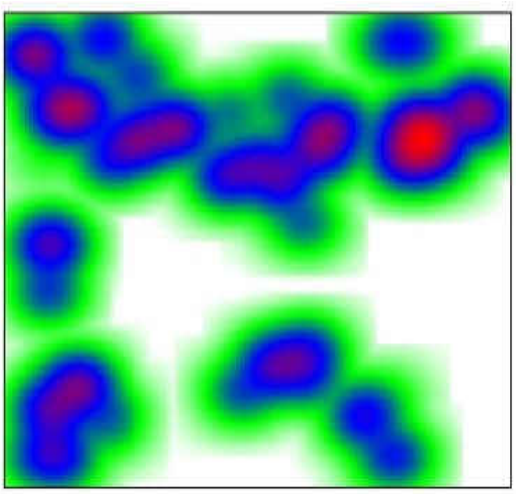


Figure 12: Apparent clustering for dynamic subarea simulation through $t = 4$.

Such a representation would not have been clearly observed if the area was not partitioned and if the rate of spread was not tabulated in space and time. This framework will allow us to detect clusters and density estimation.

To analyze the model, we estimate the parameter λ for the Poisson distribution given the data generated in the simulation is proposed using the maximum likelihood method as presented in Baddeley et al. (2016). The scale parameter is incorporated into the Poisson distribution through the use of the area, $|D_k^t|$, of the Voronoi subarea D_k^t. Using Equation (49), if we assume that the number of points generated in subarea D_k^t has density given as:

$$P\left(X^t = y_i | \underset{\sim}{X}^{t-1} = i\right) = f(y_i(t)|\lambda(t)) = \frac{e^{-\lambda(t)|D_k^t|}(\lambda(t)|D_k^t|)^{y_i(t)}}{y_i(t)!}, \quad y_i \in \mathbb{N},$$

whereas in Equation (2), $Q^t = \dfrac{(\lambda|D_k^t|)^{y_i(t)}}{y_i(t)!}$.

The parameter estimate becomes $\hat{\lambda}(t) = \dfrac{\sum y_i}{\sum |D_k^t|}$ with a standard error of $s.e. = \sqrt{\dfrac{\hat{\lambda}}{n}}$, where $n = n(t) =$ number of subareas D_k^t.

We tabulate the estimates of λ at $t = 2, 3$, and 4 with different scale parameters, α. Table 11 shows these parameter estimates of λ with standard error in parenthesis. For $\alpha = 5$ at

time $t = 2$, $\hat{\lambda} = 8.14$ with associated SE=1.07 and the true parameter is $\lambda = 8$. The data appear to follow a Poisson distribution with intensity $\lambda * t$; however, the estimated parameter values for $\alpha = 2.5$ seem less reliable due to the sparse number of points generated at each time step.

Table 11: $\hat{\lambda}$ (SE) values at times 2, 3, and 4 for dynamic subarea simulation.

	α			
time	2.5	5	7.5	10
2	10.26 (1.21)	8.14 (1.08)	8.08 (1.07)	8.23 (1.08)
3	10.76 (0.68)	12.45 (0.41)	11.69 (0.27)	12.08 (0.20)
4	10.63 (0.67)	16.24 (0.18)	16.01 (0.08)	16.04 (0.04)

The parameters of the Palm intensity function $\lambda_o(\cdot)$ can also be estimated as in Equations (97) and (101) using the `palm` package in `R` (Tanaka et al., 2008). This estimation is based on a difference process $Z_{D_k^t}$ that contains all differences $s_i - s_j$ for points s_i and s_j, $i \neq j$, of the original point process $N_{D_k^t}$, defined as

$$Z_{D_k^t} = \{s_i - s_j : s_i \text{ and } s_j \in N_{D_k^t} \text{ with } i \neq j\}.$$

If this difference process is assumed to be a nonhomogeneous Poisson point process as discussed in Subsection 4.1.4, we can take advantage of the second-order properties (stationary and isotropic) of the original point process to estimate the parameters of the Palm intensity function that may otherwise be intractable. We estimate these parameters choosing a point generated at a time $t = 1$ as the disc to follow through the next time step. At each subsequent time step we will need to choose a subarea to follow. We will focus on disc one and its subareas for initial analysis to determine if there is any trend in the estimated parameters of the Palm intensity function.

It is clear from the Table 12 that the choice of subarea greatly impacts both the estimated Palm intensity and the Moran statistic. Palm parameter estimates were not computed for disc 1 at $t = 4$ since there was only one point generated.

Table 12: Palm parameter estimates (standard deviation) and Moran values for dynamic subareas

time	disc	Points	Area	Palm intensity (sd)	Moran
2	1	20	2.650	0.538 (0.043)	409.00
3	1	24	1.269	24.000 (0.5144)	676.94
4	1	1	0.025	NA	0
2	1	20	2.650	0.538 (0.043)	409.00
3	9	15	1.183	14.071 (0.257)	262.558
4	11	44	2.37	44.000 (0.627)	1888.88

We also investigated increasing the scale parameter α as function of time, $\alpha(t) = \alpha^{t-1}$. Using an initial value of $\alpha = 2.5$, the parameter estimates at $t = 2, 3, 4$ were 10.26 (1.21), 13.44 (0.76), and 15.98 (0.31), respectively. By making the scale parameter a function of time, the parameter estimates appear more precise for small starting values, but under this construct, using $\alpha(t) = 10^{t-1}$, generates up to 30,000 points per subarea at $t = 4$. As a result, we will use $\alpha = 5$ for the scale parameter in the following analysis.

Next we investigate any trend in Moran values based on number of points generated and area of the disc. Figure 13 shows Moran values plotted versus area in orange and versus number of points in blue at $t = 2$ and Figure 14 shows a similar plot for the subareas of disc 1 at $t = 3$. There appears to be an approximate linear trend between Moran and number of points generated, even though there is no apparent relationship between area and Moran.

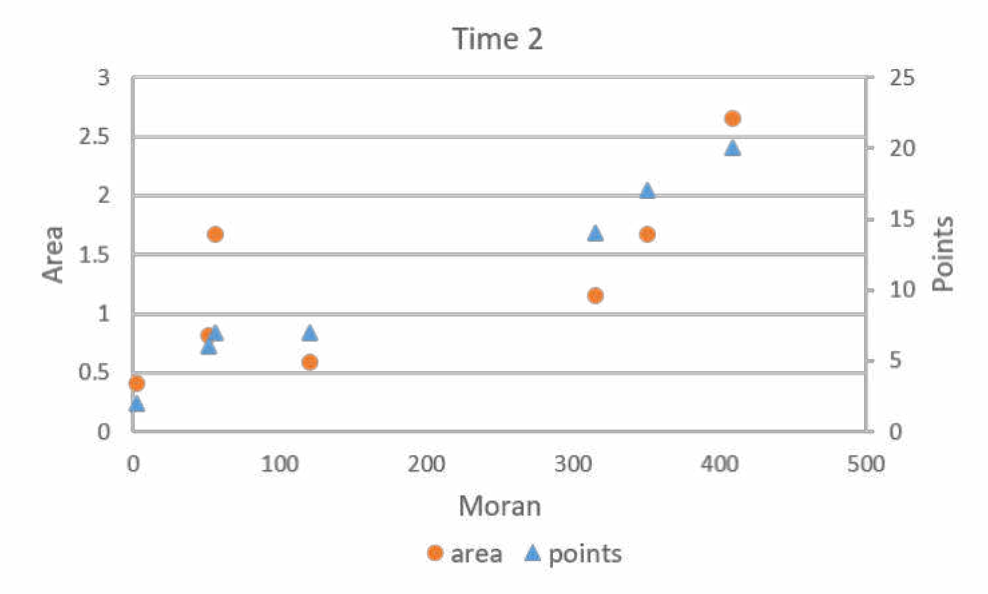


Figure 13: Trend analysis for dynamic subareas at $t = 2$.

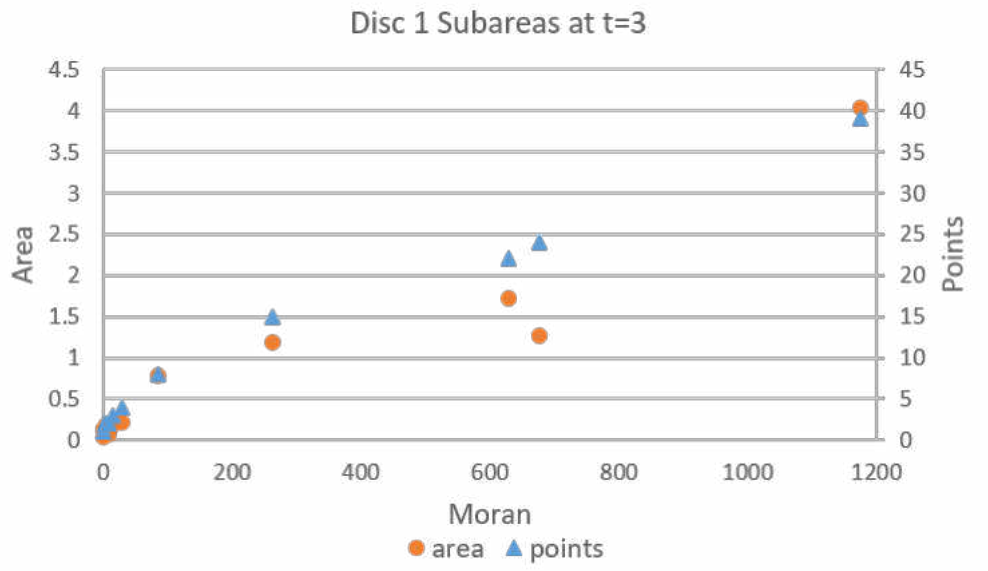


Figure 14: Trend analysis for dynamic subareas of disc 1 at $t = 3$.

The 3-D plot in Figure 15 shows that at each successive time point the number of points generated versus the Moran gets progressively larger.

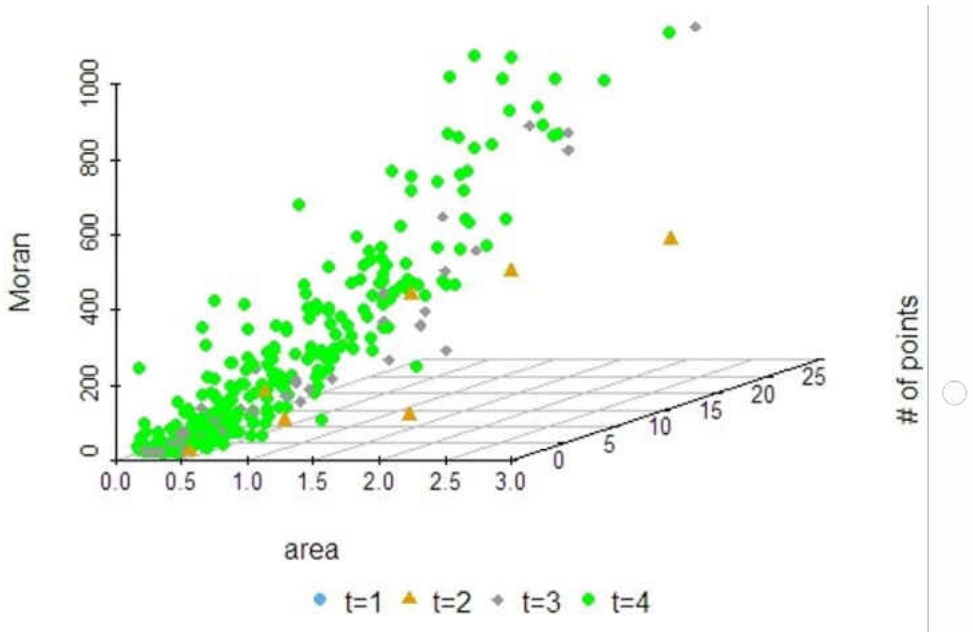


Figure 15: All Moran values for dynamic subarea simulation.

Focusing on the seven points generated at $t = 1$ and the associated discs at $t = 2$, we investigate the Markov property of the simulation results. Table 13 shows the number of points generated within each of the original seven discs at times $t = 2, 3, 4$ and also includes the area of each disc.

Table 13: Dynamic subarea simulation counts at $t = 2, 3, 4$

		time		
disc	area	2	3	4
1	2.650	20	124	806
2	1.148	14	66	391
3	0.411	2	16	138
4	0.585	7	25	143
5	1.676	7	113	745
6	1.676	17	89	565
7	0.817	6	58	226

The transition matrix for each disc Q_k^t is developed using the ratio of points generated at each time to the total number of points generated over all times. This matrix is upper

triangular since we cannot transition backward in time or stay in the same time.

$$Q^t = \begin{bmatrix} 0 & \frac{n_{12}}{\sum} & \frac{n_{13}}{\sum} & \frac{n_{14}}{\sum} \\ 0 & 0 & \frac{n_{23}}{\sum} & \frac{n_{24}}{\sum} \\ 0 & 0 & 0 & \frac{n_{34}}{\sum} \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad Q_1 = \begin{bmatrix} 0 & 0.0211 & 0.131 & 0.848 \\ 0 & 0 & 0.133 & 0.867 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We then conduct cluster analysis to determine similarities between discs. The Canberra distance measure, $\sum \frac{|X_k - X_{k'}|}{|X_k + X_{k'}|}$, is a measure of similarity and dissimilarity between groups and is used to create a distance matrix from the counts in Table 13. Figure 16 shows the dendrogram that was generated. It can be inferred that disc association is related to area since those with similar areas are clustered first.

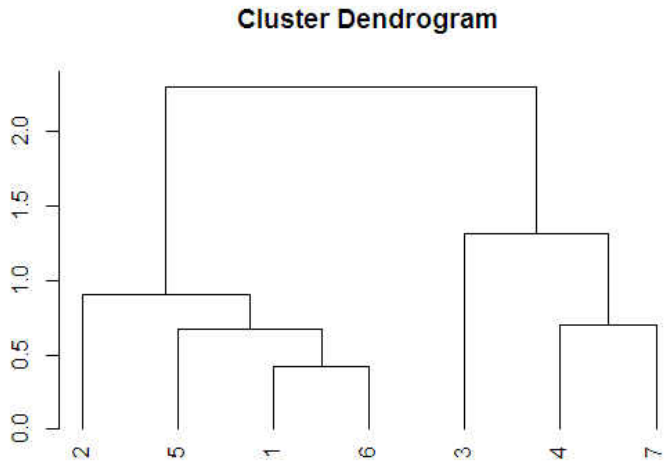


Figure 16: Cluster analysis for dynamic subarea simulation.

Voronoi cells

The dynamic subarea simulation can be generalized to Voronoi cells by relaxing the restriction of the point process to a disc. Again, we begin with an observed unit area (1 x 1 unit) and generate a Poisson point processes within that area. The observed points are then used to generate Voronoi cells that cover the entire region. This effectively means that an infected site can subsequently infect any sites that are closer to it than any other infected

site. These Voronoi cells define the subareas $D^t_{k(t)}$ for the point process generated at the next time interval (step). The Moran statistic is calculated at each step as the sum of the inverse distances between an offspring site and another offspring site. This format will continue through time $t = 4$ adhering to the Markov property where subarea will be magnified by a scale parameter $\alpha = 5$ and the intensity is a function of time $\lambda * t$. We used the same seed as in the previous simulation with discs, so results at $t = 1$ are the same, but the subareas generated for $t = 2$ from those original points is not restricted to a disc.

Algorithm: Iterative local Moran statistics for Voronoi cells

Procedure:

(i) Define the window and generate a Poisson point process in the area with intensity λ.

(ii) Calculate the Moran statistic.

(iii) Define new subareas as Voronoi cells generated from points.

(iv) Rescale subarea and iterate local points from a Poisson point process with intensity $\lambda * t$.

(v) Compute the local Moran statistic for each subarea at that time step.

Repeat steps (iii) through (v).

Stop when all times are reached.

End

Figure 17 shows the Voronoi regions generated from the seven points at $t = 1$ and also shows the points generated within each of those subareas at $t = 2$. The original point from $t = 1$ is overlayed for reference as a heavy filled black circle and a different marker is used for each subarea at $t = 2$.

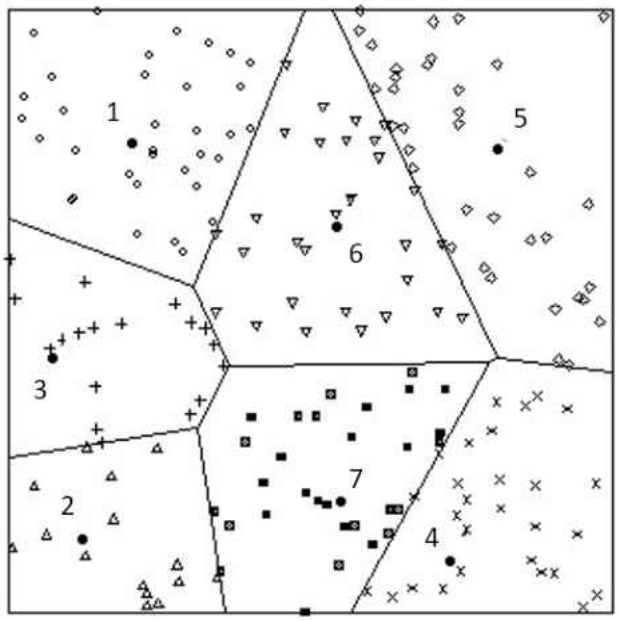


Figure 17: Voronoi cell dynamic subarea simulation, $\lambda = 4 * t$ through $t = 2$.

The number of points generated in each Voronoi subarea, the rescaled area, and Moran values calculated at time $t = 2$ are displayed in Table 14 with $\lambda = 4$ and $\alpha = 5$. It should be noted that the entire area is represented in Table 14, whereas when we restricted the subareas to largest disc contained within the Voronoi cell, the spread was restricted to less than 36% of the original region (see Table 9).

Table 14: Moran values for dynamic subareas, $\lambda = 8$ and $\alpha = 5$ at $t = 2$

subarea	Points	Area	Moran
1	34	4.148	823.02
2	15	2.384	199.53
3	18	2.569	250.65
4	28	3.314	548.67
5	36	4.789	828.48
6	30	4.259	707.62
7	30	3.536	642.59

The process continues at time $t = 3$ with subarea 1 having points generated in 32 subareas (see Figure 18), subarea 2 having points generated in 16 subareas, etc. The Moran values

for subarea 1 at time $t = 3$ are listed in Table 15. It should also be noted that there are at least two points generated within each subarea, therefore there all subareas have non-zero Moran values.

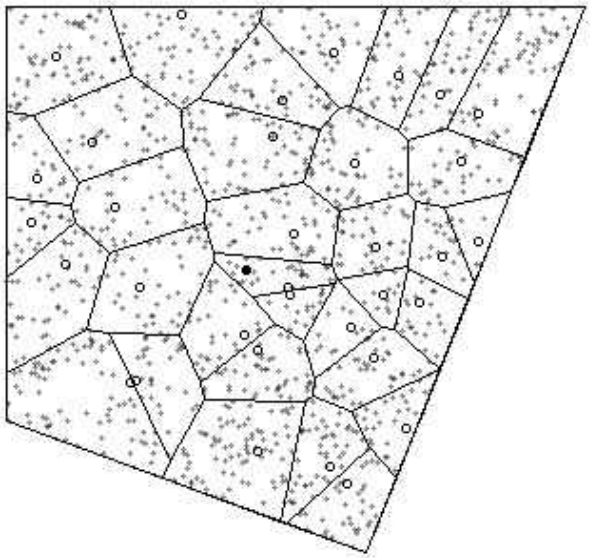


Figure 18: Voronoi cell dynamic subarea simulation, $\lambda = 4 * t$ through $t = 3$.

Table 15: Moran values for dynamic (Voronoi) subarea 1, $\lambda = 4 * t$ and $\alpha = 5$ at $t = 3$

subarea	Points	Area	Moran	subarea	Points	Area	Moran
1	57	4.217	2436.71	18	17	1.506	301.48
2	48	4.342	1655.88	19	77	6.175	3582.01
3	59	4.437	2263.11	20	15	1.062	261.71
4	56	4.429	2152.50	21	8	1.457	48.99
5	33	1.894	1044.22	22	23	1.251	618.57
6	28	2.383	997.88	23	19	1.859	405.67
7	19	2.077	325.88	24	45	3.352	1769.17
8	43	2.998	1413.20	25	38	2.845	1185.61
9	30	2.560	698.72	26	33	2.632	889.17
10	32	2.928	770.86	27	24	2.571	493.73
11	37	3.116	1065.18	28	51	4.374	2020.12
12	50	4.774	1730.29	29	35	2.489	1051.16
13	31	3.790	711.87	30	26	1.714	764.15
14	24	1.661	565.57	31	34	3.184	1145.31
15	13	1.088	206.00	32	41	3.554	1456.87
16	56	5.201	1875.26	33	24	2.594	498.36
17	43	4.157	1311.38	34	64	5.026	2684.04

We fill focus on the top right subarea in Figure 18 at time $t = 4$ and the points generated are displayed in Figure 19. While there are still some parts of the subarea that have few points (events), clustering is not as obvious as in the previous dynamic subarea simulation. The region approaches full coverage much more quickly. As described in Equation (123), the Moran values can become quite large and exceed a specified threshold, making an estimate of the standard deviation unrealistic/unusable.

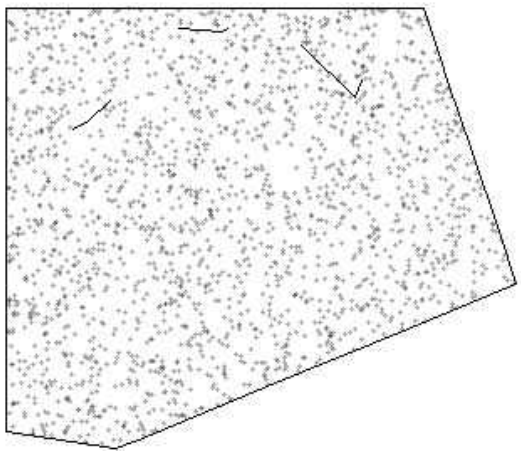


Figure 19: Voronoi cell dynamic subarea simulation, $\lambda = 4 * t$ through $t = 4$.

5.2 REAL DATA EXAMPLES FOR FIXED SUBAREAS

To exactly match the realization of our simulations, data collected would need to include a number of events that are observed at a specific time and a unique location for each of these events. Often, however, data is aggregated for an area and counts are recorded. To use this data, it is necessary to utilize a function of the count data. We provide two real data examples using this type of aggregated data.

5.2.1 VIRGINIA SEAT BELT DATA

Data on seat belt use were collected in regions of Virginia over multiple years. We will focus on vehicle miles traveled (VMT) in thousands of miles, which is a proxy for other variables such as population, regional location, etc. (i.e., high VMT tend to be on the I-95 corridor and urban crescent where the money is in Virginia). VMT is broken into three categories based on millions of miles traveled on average each year: high (≥ 1001), medium (501 to < 1001), and low (< 501). This data is from the data collected and analyzed in the annual reports prepared for the Virginia Department of Motor Vehicles Highway Safety Office by Porter, Diawara, and Jenkins (2016). A proportion of seat belt use for each region was calculated. The data for the 15 regions of interest over five years (2012-2016) is shown in Table 16. A graphical representation of the data is provided in Figure 20.

Table 16: Virginia seat belt data

Region	2012	2013	2014	2015	2016	VMT
Alleghany	0.7285	0.7717	0.7127	0.6563	0.5768	Low
Carroll	0.7882	0.8421	0.7315	0.7858	0.7758	Medium
Fairfax	0.8974	0.8929	0.8807	0.9112	0.8764	High
Halifax	0.8681	0.8052	0.7451	0.8139	0.7680	Low
Henry	0.8706	0.7945	0.6891	0.8122	0.7823	Medium
Loudoun	0.7692	0.8469	0.8427	0.8462	0.8396	High
Mecklenburg	0.7890	0.7685	0.7298	0.7164	0.6574	Low
Prince George	0.8340	0.7793	0.7154	0.7234	0.7057	High
Rockbridge	0.7494	0.7540	0.6667	0.7607	0.8095	Medium
Shenandoah	0.7447	0.7720	0.6796	0.7612	0.8381	Medium
Southampton	0.8880	0.8460	0.8109	0.8448	0.7968	Low
Southeast	0.8670	0.8373	0.8074	0.8242	0.8227	High
Stafford	0.8315	0.8408	0.8537	0.9035	0.8758	High
Tazewell	0.7493	0.7425	0.6025	0.8183	0.7300	Low
Washington	0.8430	0.7605	0.7842	0.7986	0.8225	Medium

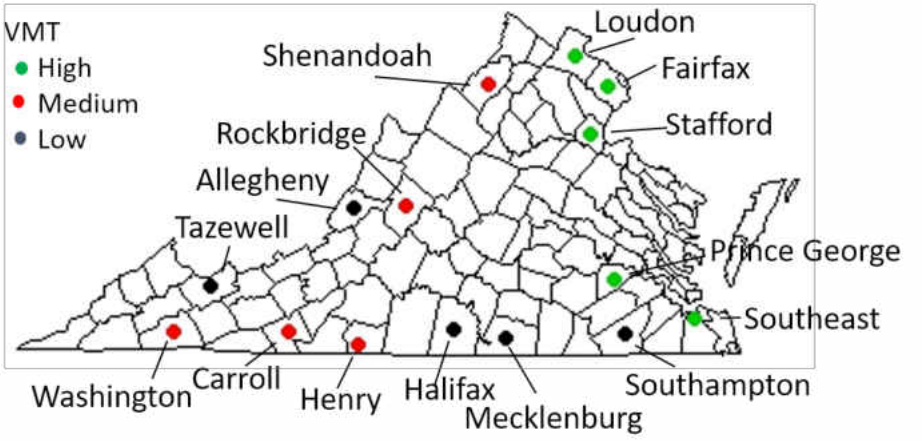


Figure 20: Map of Virginia regions of data collection.

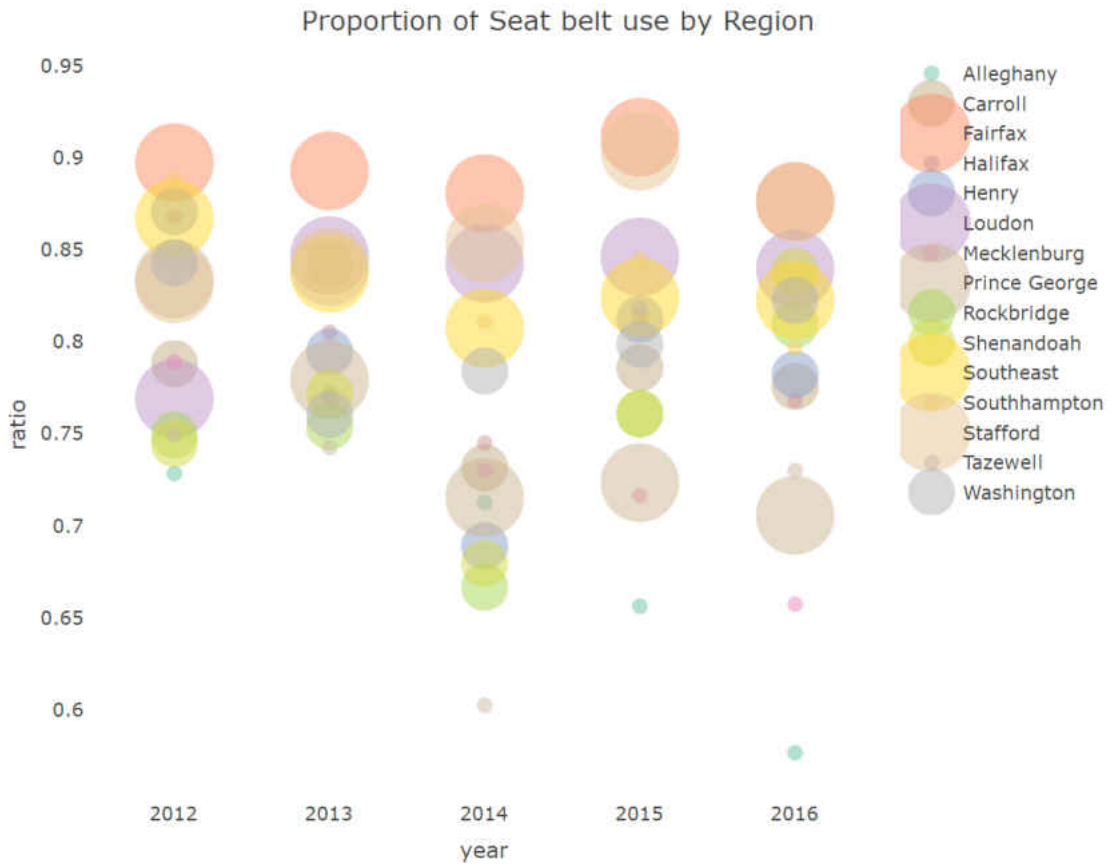


Figure 21: Proportion of seat belt use with bubble size determined by VMT.

To use this data with our fixed subareas model, we utilize a function of the proportion of seat belt use within each county, and calculate the Moran's statistics as follows:

$$M_{i,j}^{t} = \sum_{i,j} w_{i,j} f(p_i, p_j),$$

where $w_{i,j} = e^{d_{ij}/\bar{d}}$ is the geographic weight between two regions i and j, d_{ij} is the great circle distance (in km) between two region centroids, and $f(p_i, p_j)$ is some function of the proportion of observed seat belt use at locations i and j for time t. For the analysis, we used $f(p_i, p_j) = \frac{p_i + p_j}{2}$, the average proportion of seat belt use in regions i and j.. The overall Moran was calculated for each year along with the "localized" Moran for each VMT category, which is presented in Table 17. The local Moran, which can be thought of as a subordinate of the global Moran, measures the local statistical quantities of shared characteristics based on the High, Medium or Low VMT.

Table 17: Global and local Moran values for VA seat belt data

	2012	2013	2014	2015	2016
Global	35.0785	34.5749	32.3207	34.4378	33.5927
High	3.2882	3.2698	3.1775	3.2519	3.1907
Medium	3.8682	3.8234	3.4321	3.8098	3.9338
Low	3.3491	3.2925	3.0056	3.2047	2.9262

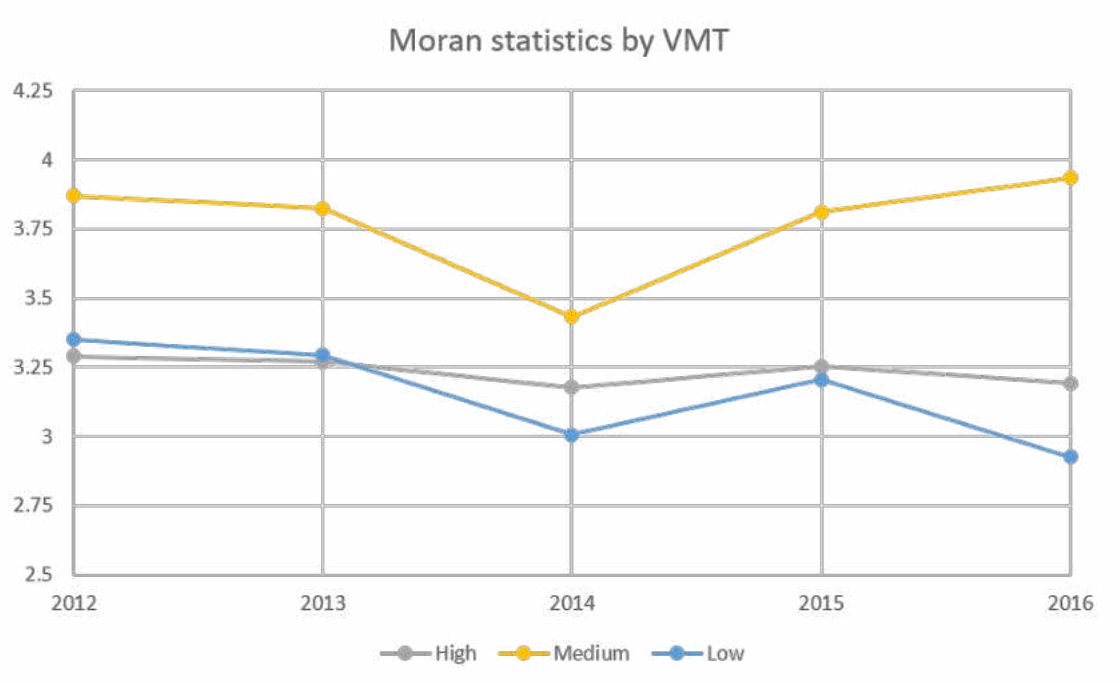


Figure 22: Moran statistics by VMT.

From the Figure 22, we can see that the regions of High and Low VMT are more predictable in their average proportion of yearly seat belt use (i.e., there is more consistency) than in Medium VMT based on geographic spread/locations from year to year. The range coverage of the global and local Moran values is quite different. To make the comparison a little more uniform, the values are divided by their respective row sums, and the standardized results are displayed in Table 18, along with their graph in Figure 23.

Table 18: Standardized global and local Moran values for VA seat belt data

	2012	2013	2014	2015	2016
Global	0.2063	0.2034	0.1901	0.2022	0.1976
High	0.2033	0.2021	0.1964	0.2010	0.1972
Medium	0.2050	0.2026	0.1819	0.2019	0.2085
Low	0.2123	0.2087	0.1905	0.2031	0.1855

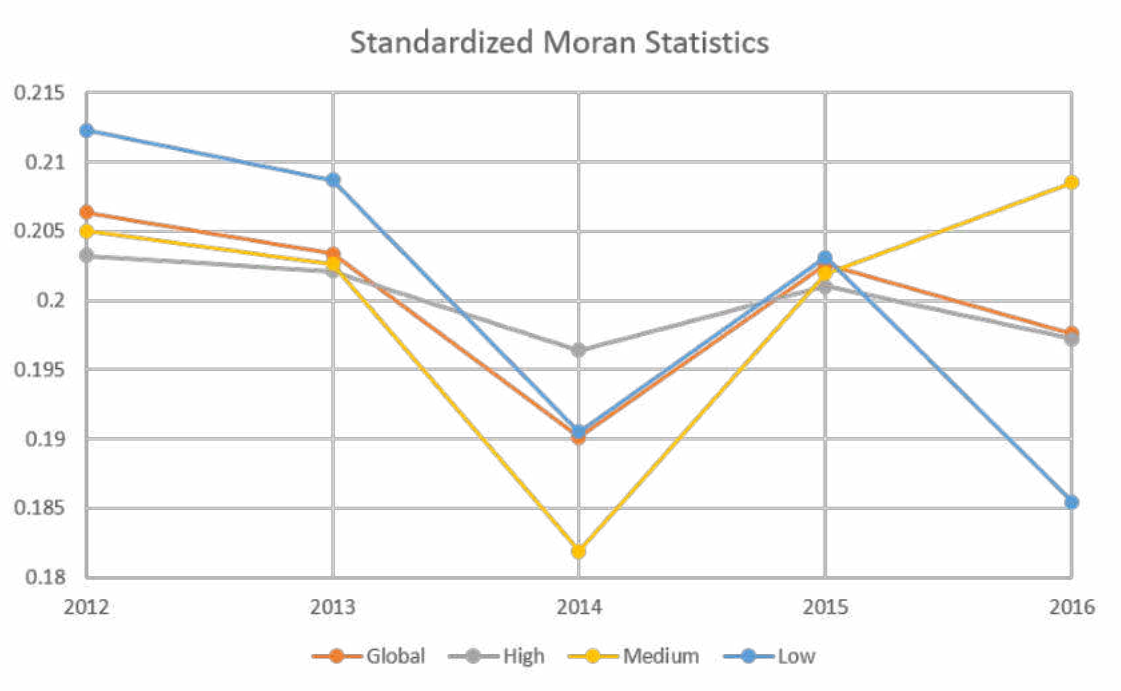


Figure 23: Moran statistics by VMT.

The global Moran considering all regions seems to be generally consistent from year to year; however, a further analysis based on the VMT categories alludes to some differences among the regions. Within the High VMT category, the correlation appears consistent and also appears to deviate less from the global Moran, than the Low and Medium VMT. In 2015, all VMT categories were similar to the global Moran. In general, the characteristics of High VMT are less heterogeneous than Low VMT. There is extremeness of Moran values in the Low and Medium VMT. It also appears that over time the proportion of seat belt use in High and VMT will stagnate, which may indicate that an intervention may be needed to improve seat belt use, especially for Low VMT.

5.2.2 FNIRS DATA[3]

Functional near-infrared spectroscopy (fNIRS) data was recorded for two participants, a 68 year-old male (P1), 91 months post onset of a left hemisphere stroke and a 61 year-old male (P2), 86 months post onset of a left hemisphere stroke, during each of 18 treatment days during an intervention period. A complete description of participants, stimuli, and behavior outcome measures can be found in Johnson, Lott, and Prebor (2018). Changes in hemodynamic activity of oxygenated (HbO) and deoxygenated (HbR) hemoglobin levels of the anterior prefrontal cortex were recorded using a 16-optode continuous wave fNIRS system, and total hemoglobin (HbT) concentrations (HbO + HbR) were calculated. The locations of interest in the prefrontal cortex included left and right hemisphere and isolated Brodmann Areas (BA) 9 and 46 (optodes 1, 2, and 15), and BA 10 (optodes 3-14 and 16) as shown in Figure 24, which were determined based on a recent class of Montreal Neurological Institute (MNI) virtual spatial registration for older adults (Chen, Blumen, Izzetoglu, & Holtzer, 2017). Indeed, models of biological processes must also account for interconnected individual neurons (see Riedler & Thieullen, 2015).

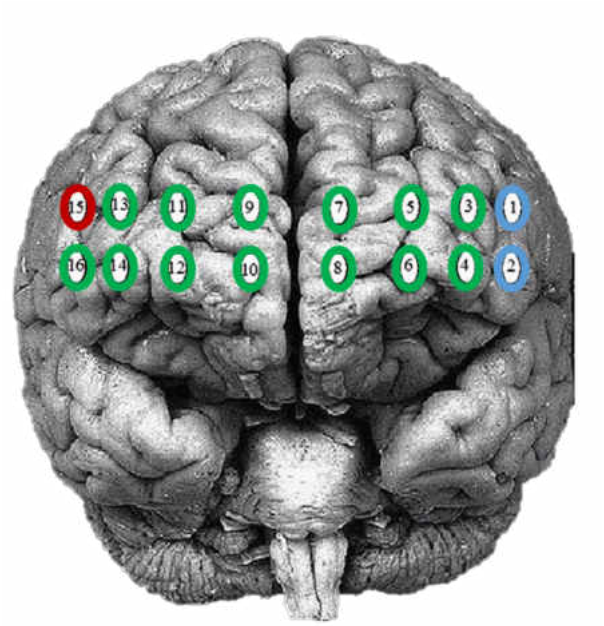


Figure 24: Optode grouping for regions of interest. Red=Brodmann area 9; Green=Brodmann area 10; Blue=Brodmann area 46.

[3]Material from this section is based on joint work with Norou Diawara and Rachel Johnson. The work has been submitted for publishing and is referenced Lorio, Diawara, and Johnson.

A general framework for the spatial model is that the regions of the brain interact after adjustment for the distance between them. The goal is to identify any regions of the brain that have strong correlation(s). To use this data with our fixed subareas model, we utilize a function of the observed fNIRS intensity at each region, and calculate the Moran's statistics as follows:

$$M_{i,j}^{t} = \sum_{i,j} w_{i,j} f(o_i, o_j),$$

where $w_{i,j} = e^{d_{ij}/\bar{d}}$ is the geographic weight between two regions i and j, and $f(o_i, o_j)$ is some function of the observed fNIRS intensity in regions i and j. We will take $f(o_i, o_j)$ as the absolute value of the difference in observed fNIRS intensities at locations i and j for the t^th day and for each stage, thus adding the temporal aspect of the model.

We generate a distance matrix d_{ij}, where distance is measured as one step (up, down, left, right) along the shortest distance path between regions and when the regions coincide, (i.e. $i = j$), the distance is taken from a central node (along the brain stem). We first focus on three regions of optodes 1, 2, and 15 (BA 9 and BA 46). The resulting distance matrix utilizing this distance scheme is:

$$d_{ij} = \begin{bmatrix} 4 & 1 & 7 \\ 1 & 4 & 8 \\ 7 & 8 & 4 \end{bmatrix}.$$

The geographical weights w_{ij} are then calculated using the distance matrix above:

$$w_{ij} = \begin{bmatrix} 0.2765 & 0.7251 & 0.1054 \\ 0.7251 & 0.2765 & 0.0764 \\ 0.1054 & 0.0764 & 0.2765 \end{bmatrix}.$$

Then we calculate the Moran statistics for each participant at each stage utilizing the weight matrix above and the observed HbO intensities. Figure 25 shows the graphs of the Moran statistics; a Loess smoothing curve has been added to easily detect trends between participants.

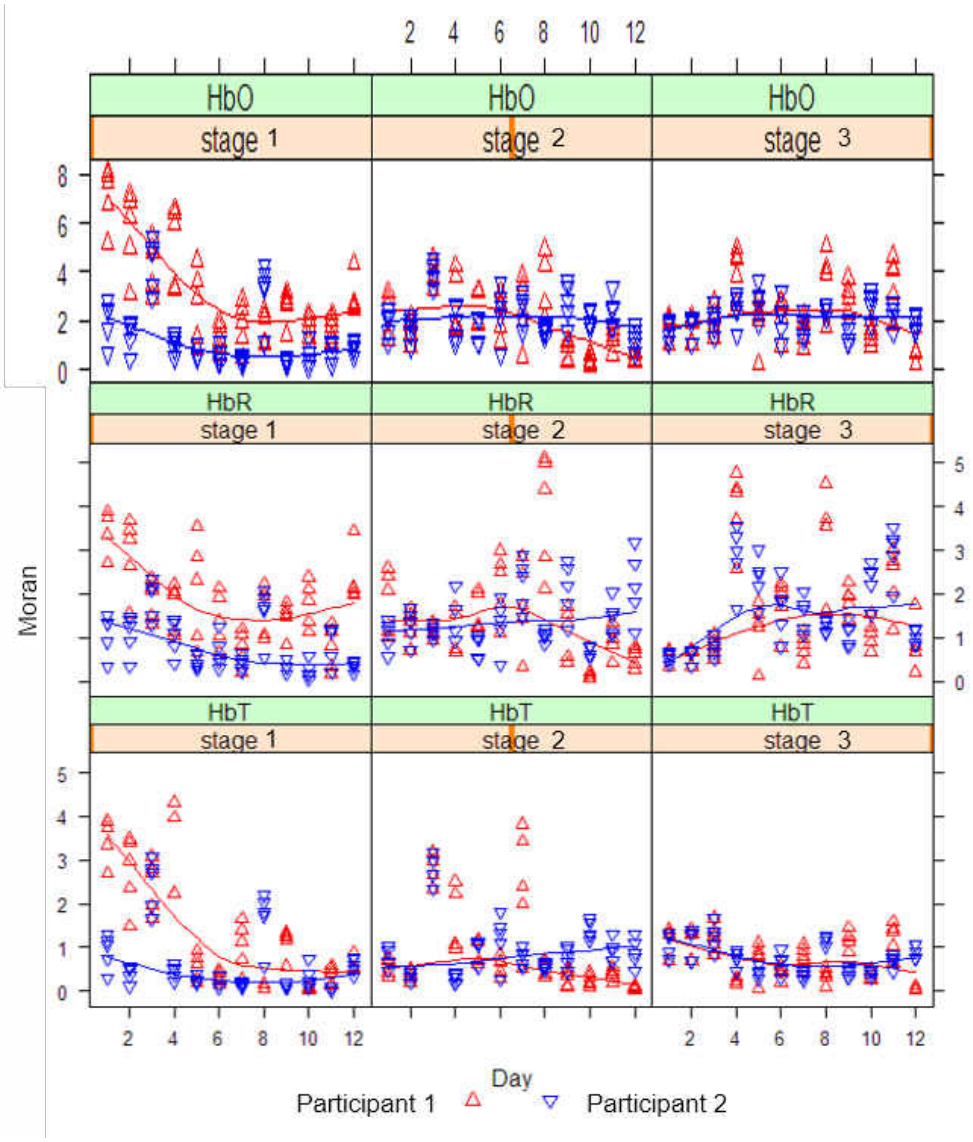


Figure 25: Moran statistic for Grouping 1 for HbO, HbR, and HbT by stage for days 1-12.

The degree of correlation measured here by the Moran statistics between the participants indicates that the correlation between optodes 1 and 2 (BA 46) vs optode 15 (BA 9) is always higher for P1 (red) in stage 1, whereas in stages 2 and 3, the correlations cross. This last remark is especially visible in stage 2 after day 6. This shows that there is a magnitude of PFC activity that are represented in stage 1 of HbO, HbR, and HbT that lessen in stage 2 and 3 over time. This indicates for P1 the cognitive control processes to maintain the rules to respond and the need to inhibit previously learned behavior decrease with practice (Glascher et al., 2012). There is a decrease in the cognitive effort similar to the tasks

completed following a model as in stage 1. Indeed, there are signs that BA 46 (optodes 1 and 2) activity changed based on the stage of practice across days. Under HbO, oxygenation increases mainly in stage 1, suggesting that P1 had a higher level of activity than P2. It appears that over time the activity will stabilize with practice.

CHAPTER 6

SUMMARY

While spatio-temporal data have been collected for many years, novel analysis techniques for use with these data are still developing. Moran's index is a widely used measure of spatial dependence. In this dissertation, we have proposed a spatio-temporal Palm distribution, a spatio-temporal Palm likelihood, and a spatio-temporal Moran-type autocorrelation statistic. We utilized stationarity of the Markov process to obtain characterization of the process. We have also derived the expected value of the proposed Moran statistics. To illustrate the setup, we have simulated data under the Markov process model for different lattice formats and presented Moran spatio-temporal estimates. To evaluate the performance of the Moran values, we compared them with adjusted Geary's c statistics. Additionally, estimates of Palm parameters were calculated and analyzed over time to study potential dependence and compare trends with our Moran statistic. The local estimates are also compared with the global estimates to show that the local Moran statistics provide a better description of the data. We controlled for targeted subareas, and such an approach provides a framework for understanding disorganized or disordered evolution of some natural phenomena in the form of cluster analysis. The distribution of subareas can be found based on our density estimation or concentration under the modified Moran statistics in space and time. Lastly, we presented applications of methods to real data in human behavior (seat belt use) and in neuroscience. In practice, this method can be applied to growth data with the goal of assisting with when and where to implement an intervention.

From a statistical and research simulation viewpoint, combination of information from diverse sources consolidate the finding. We have focused on a spatial temporal model with measures of dependency and have addressed a localized version of Moran's index. The process is simple, efficient and easy to implement;. The random field that we have proposed matches the process that is being observed over events in the case of plant disease and spread, and captures flexibility in space partitioning. We constructed a multilevel Poisson process with dependent marginals, and the dependence among the levels is captured by the rate of increase of the disease spread over time, steered by a common factor in the scale. The main consequence of our results is that the Moran statistics considered as a measure

of correlation are compared from explicit algorithms (Monte Carlo simulation) of time and disc areas. The proposed fitted Moran statistics of time and space are much more locally consistent and capture the variation in much more interactive utility stages than a simple global Morans index; however, the Moran-type statistic is very sensitive to the choice of weight function. The simulation results also show that long range, spread features, and variability yielded spatial clustering of occurrence events. Although the global Moran's index test statistic is quite popular, chosen weights may influence the internal fluctuations. The cluster representation of a space and time process is elucidated in the discrete setting.

However, in the continuous time setting, the process is not easily described. Extensions to different dynamics other than nonhomogeneous discrete Markov chains will be considered in future studies and will widen the scope of the model over other correlation structures. In the seat belt data example, expanding the model to include other covariates will allow us to build local interventions based on demographics and other behaviors. In the brain image data example, it is worth noting that only a finite number of measures are accessible, so it is natural to build a model where accessible neurons are allowed to vary over space and time based on activation. Additionally, the spatio-temporal Palm likelihood presented in Chapter 4 can be further extended to the multivariate case.

REFERENCES

[1] Adams, R. P., Murray, I., & MacKay, D. (2009), "Tractable Nonparametric Bayesian Inference in Poisson Processes with Gaussian Process Intensities," *Appearing in Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada.

[2] Applebaum, D. (2004), *Lévy Processes and Stochastic Calculus*, Cambridge University Press.

[3] Anderson, I. T., & Hahn, U. (2015), "Matérn thinned Cox Processes," Aarhus, Denmark: Aarhus University for Centre for Stochastic Geometry and Advanced Bioimaging.

[4] Anselin, L. (1995), "Local Indicators of Spatial Association - LISA," *Geographical Analysis*, 27(2), 93–115.

[5] Baddeley, A. (2007), "Spatial Point Processes and their Applications," *Stochastic geometry, Lecture Notes in Mathematics*, 1892, Berlin: Springer.

[6] Baddeley, A., Rubak, E., & Turner, R. (2016), *Spatial Point Patterns: Methodology and Applications with R*, Boca Raton, FL: CRC Press.

[7] Baudin, M. (1981), "Likelihood and Nearest-Neighbor Distance Properties of Multidimensional Poisson Cluster Processes," *Journal of Applied Probability*, 18(4), 879–888. 10.2307/3213062.

[8] Beisel, C., Rokyta, D., Wichman, H., & Joyce, P. (2007), "Testing the Extreme Value Domain of Attraction for Distributions of Beneficial Fitness Effects," *Genetics*, 176, 2441–2449.

[9] Besag, J. (1975), "Statistical Analysis of Non-Lattice Data," *Journal of the Royal Statistical Society, Series D*, 24(3), 179–195.

[10] Besag, J., Milne, R., & Zachary, S. (1982), "Point Process Limits of Lattice Processes," *Journal of Applied Probability*, 19(1), 210–216.

[11] Besag, J., York, J., & Mollièe, A. (1991), "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43, 1–21.

[12] Bradley, J. R., Cressie, N., & Shi, T. (2016), "A Comparison of Spatial Predictors When Datasets Could Be Very Large," *Statistics Surveys*, 10, 100-131.

[13] Brix, A., & Diggle, P. J. (2001), "Spatiotemporal Prediction for Log-Gaussian Cox Process," *Series B Statistical Methodology*, 63(4), 823–841.

[14] Bui, N., Lorio, J., Diawara, N., Das, K., and Waller, L. (2018), "New Approaches to Model Simulated Spatio-Temporal Moran's Index," *Journal of Probability and Statistical Science*, 16(1), 11–24.

[15] Casella, G., & Berger R. L. (2002), *Statistical Inference* (Second Edition), Pacific Grove, CA: Duxbury Press/Thomson Learning.

[16] Cheilaris, P., Khramtcova, E., Langerman, S., & Papadopoulou, E. (2014), "A Randomized Incremental Approach for the Hausdorff Voronoi Diagram of Non-crossing Clusters*," In: *Theoretical Informatics: 11th Latin American Symposium*, eds. A. Pardo and A. Viola, pp. 96-107, Berlin: Springer.

[17] Chen, Y. (2012), "On the Four Types of Weight Functions for Spatial Contiguity Matrix," *Letters Spat Resources Sci*, 5, 65–72.

[18] Chen, M., Blumen, H. M., Izzetoglu, M., & Holtzer, R. (2017), "Spatial Coregistration of Functional Near-Infrared Spectroscopy to Brain MRI," *Journal of Neuroimaging*, 27(5), 453–460.

[19] Chiu, S. N., Stoyan, D., & Kendall, W. S., and Mecke, J. (2013), *Stochastic Geometry and Its Applications,* Wiley.

[20] Choi, C., & Baccelli, F. (2018), "Poisson Cox Point Processes for Vehicular Networks," *I*EEE Transactions on Vehicular Technology, 67(10), 10160–10165.

[21] Cliff, A. D., & Ord, J. K. (1981), *Spatial Processes: Models and Applications,* London: Pion.

[22] Coeurjolly, J., Møller, J., & Waagapeterson, R. (2016), "A Tutorial on Palm Distributions for Spatial Point Processes," *International Statistical Review*, 0, 0, 1-17 doi:10.1111/insr.12205

[23] Coeurjolly, J. F., Møller, J., & Waagepetersen, R. (2017), "Palm Distributions for Log Gaussian Cox Processes," *Scandinavian Journal of Statistics*, 44, 192-203. doi: 10.1111/sjos.12248.

[24] Cressie, N. (1993), *Statistics for Spatial data*, Hoboken, NJ: John Wiley & Sons, Inc..

[25] Cressie, N., & Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, Hoboken, NJ: John Wiley & Sons, Inc.

[26] Cronie, O., & van Lieshout, M. N. M. (2016), "Summary statistics for inhomogeneous marked point processes," *Annals of the Institute of Statistical Mathematics*, 68(4), 905–928.

[27] Daley, D. J., & Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes, Vol I: Elementary Theory and Methods* (Second Edition), New York: Springer.

[28] Daley, D. J., & Vere-Jones, D. (2008), *An Introduction to the Theory of Point Processes, Vol II: General Theory and Structure* (Second Edition), New York: Springer.

[29] Daniels, M. J., Zhou, Z., & Zou, H. (2006), "Conditionally Specified Space-time Models for Multivariate Processes," Journal of Computational and Graphical Statistics, 15, 157–177.

[30] de Jong, P., Sprenger, C., & van Veen, F. (1984), "On extreme Values of Moran's I and Geary's C," *Geographical Analysis*, 16, 17–24.

[31] Diawara, N., Waller, L., King, R., & Lorio, J. (2018) "Simulations of Local Morans Index in a Spatio-temporal Setting," *Communications in Statistics - Simulation and Computation*, DOI: 10.1080/03610918.2018.1425441

[32] Diggle, P. (1985), "A Kernel Method for Smoothing Point Process Data," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34(2), 138–147

[33] Diggle, P. J., Moraga, P., Rowlingson, B., & Taylor, B.M. (2013), "Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm," *Statistical Science*, 28(4), Special Issue on Mathematics of Planet Earth, 542–563.

[34] Diggle, P. J. (2014), *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns* (Third Edition), Boca Raton, FL: CRC Press.

[35] Dixon, P. M. (2002), *Ripley's K Function,* In: *Encyclopedia of Environmetrics*, 3, ed. A. H. El-Shaarawi and W. W. Piegorsch, Chichester: John Wiley & Sons, 1796–1803.

[36] Edgar, G. A. (2000), "Packing Measure in General Metric Space," *Real Analysis Exchange*, 26(2), 831–852.

[37] Elith, J., Leathwick, J. R., & Hastie, T. (2008), "A Working Guide to Boosted Regression Trees," *Journal of Animal Ecology*, 77, 802–813.

[38] Geary, R. C. (1954), "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, 5(3), 115–145.

[39] Gelfand, A. E., & Sahu, S. K. (1999), "Identifiability, Improper Priors and Gibbs Sampling for Generalized Linear Models," *Journal of the American Statistical Association*, 94, 247–253.

[40] Griffith, D. (2005), *Spatial Autocorrelation,* In: *Encyclopedia of Social Measurement*, 3, ed. K. Kempf-Leonard, pp. 581-590, Amsterdam: Elsevier.

[41] Griffith, D. (2009), *Methods: Spatial Autocorrelation,* In: *International Encyclopedia of Human Geography*, eds. R. Kitchin and N. Thrift, pp. 396-402, New York: Elsevier.

[42] Goicoa, T., Adin, A., Ugarte, M. D., & Hodges, J. S. (2018), "In Spatio-temporal Disease Mapping Models, Identifiability Constraints Affect PQL and INLA Results," *Stochastic Environmental Research and Risk Assessment*, 32, 749–770.

[43] Gómez-Rubio, V., & Palmí-Perales, F. (2019), "Multivariate Posterior Inference for Spatial Models with the Integrated Nested Laplace Approximation," *J. R. Stat. Soc. C*, 68, 199–215.

[44] Guidoum, A. C. (2015), "kedd: Kernel Estimator and Bandwidth Selection for Density and its Derivatives," R package version 1.0.3, http://CRAN.R-project. org/package=kedd

[45] Hastie, T. J., & Tibshirani, R. J. (1990), *Generalized Additive Models,* London: Chapman & Hall.

[46] He, X. (2013), "Lebesgue Approximation of Superprocesses," *Doctoral dissertation*, retrieved from https://etd.auburn.edu

[47] Hess K., & Gentleman R. (2010), "muhaz: Hazard Function Estimation in Survival Analysis," R package version 1.2.5, http://CRAN.R-project.org/package=muhaz.

[48] Johnson, R. K., Lott, A., & Prebor, J. (2018), "A Comparison of Outcome Measures for Speech Motor Learning in Acquired Apraxia of Speech Using Motor Learning Guided Treatment," *Clinical Archives of Communication Disorders*, 3(1), 1–13.

[49] Kallenberg, O. (1983), *Random Measures,* (Third Edition), Berlin/New York: Akademie Verlag/Academic Press.

[50] Kallenberg, O. (2002), *Foundations of Modern Probability,* (Second Edition), New York: Springer.

[51] Kallenberg, O. (2005), *Probabilistic Symmetries and Invariance Principles,* New York: Springer.

[52] Kedem, B., & Gagnon, R.E. (2010), "Semiparametric Distribution Forecasting," *Journal of Statistical Planning and Inference*, 140, 3734–3741.

[53] Kiêu, K., Adamczyk-Chauvata, K., Monoda, H., & Stoica, R. S. (2013), "A Completely Random T-tessellation Model and Gibbsian Extensions," *Spatial Statistics*, 6, 118–138.

[54] Kingman, J. F. C. (1993), *Poisson Processes,* Oxford University Press.

[55] Koyama, S., Prez-Bolde, L. C., Shalizi, C. R., & Kass, R. E. (2010), "Approximate Methods for State-Space Models," *Journal of the American Statistical Association*, 105(489), 170–180.

[56] Last, G., & Penrose, M. (2017), *Lectures on the Poisson Process*, Cambridge University Press.

[57] Last, G., Penrose, M. D., Schulte, M., & Thaele, C. (2014): "Moments and central limit theorems for some multivariate Poisson functionals," *A*dvancements in Applied Probability, 46, 348–364.

[58] Langrené, N., & Warin, X. (2018), "Fast and Stable Multivariate Kernel Density Estimation by Fast Sum Updating,"

[59] Lindgren, F., & Rue, H. (2015), "Bayesian Spatial Modelling with R-INLA," *J*ournal of Statistical Software, 63(19), 1–25.

[60] Li, H., Calder, C. A., & Cressie, N. (2007), " Beyond Moran's *I*: Testing for Spatial Dependec=nce Based on the SAR Model," *G*eographical Analysis, 39, 357–375.

[61] Lorio, J., Diawara, N., & Johnson, R. (2018) "Consideration of Spatial Varying Coefficient Model of Prefrontal Cortex Activity Changes during Speech Motor Learning in Apraxia of Speech," Manuscript submitted for publication.

[62] Lorio, J., Diawara, N., & Waller, L. (2018), "Density Estimation of Spatio-temporal Point Patterns Using Moran's Statistics," *International Journal of Statistics and Probability*, 7(2), 80–90. doi: 10.5539/ijsp.v7n2p80

[63] Martin, R. L., & Oeppen, J. E. (1975). "The Identification of Regional Forecasting Models Using Space:Time Correlation Functions," *Transactions of the Institute of British Geographers*, 66, 95-118. https://doi.org/10.2307/621623

[64] Maruyama, Y. (2015), *An Alternative to Moran's I for Spatial Autocorrelation.* Retrieved from http://arxiv.org/abs/1501.06260.

[65] McCullagh, P., & Nelder, J.A. (1989), *Generalized Linear Models* (Second Edition), New York: Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

[66] Meddens, A. J. H., & Hicke, J. A. (2014), "Spatial and Temporal Patterns of Landsat-based Detection of Tree Mortality Caused by a Mountain Pine Beetle Outbreak in Colorado, USA," *Forest Ecology and Management*, 322-15, 78–88.

[67] Møller, J., & Waagepetersen, R.P. (2007)," Modern Statistics for Spatial Point Processes*," em Scandinavian Journal of Statistics, 34, 643–684.

[68] Møller, J., & Díaz-Avalos, C. (2010), "Structured Spatio-Temporal Shot-Noise Cox Point Process Models, with a View to Modelling Forest Fires," *Scandinavian Journal of Statistics*, 37(1), 2–25, doi: 10.1111/j.1467-9469.2009.00670.x

[69] Møller, J., Syversveen, A. R., & Waagepetersen, R. P. (1998), "Log Gaussian Cox Processes," *Scandinavian Journal of Statistics*, 25(3), 451–482.

[70] Moran, P. A. P. (1950), "Notes on Continuous Stochastic Phenomena," *Biometrika*, 37(1-2), 17–23.

[71] Murakami, D., Yoshida, T., Seay, H., Griffith, D.A., & Yamagata, Y. (2017), "A Moran Coefficient-based Mixed Effects Approach to Investigate Spatially Varying Relationships," *Spatial Statistics*, 19, 68–69.

[72] Norris, J. (2018), Advanced Probability, Lecture notes, University of Cambridge, dated 04 October 2018.

[73] Odland, J. (1988), *Spatial Autocorrelation,* Newbury Park, CA: Sage Publications.

[74] Osada, H., & Shirai, T. (2016), "Absolute Continuity and Singularity of Palm Measures of the Ginibre Point Process," *Probability Theory Related Fields*, 165, 725-770.

[75] Palm, C. (1943), "Intensitätsschwankungen im Fernsprechverk," *Ericssons Techniks*, 44, 1-189.

[76] Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, 33, 1065–1076.

[77] Peng, M. (2008). "Palm Measure Invariance and Exchangeability for Marked Point Processes," *Doctoral dissertation*, retrieved from https://etd.auburn.edu

[78] Porter, B. E., Diawara, N., & Jenkins, J. (2016), "2016 Seat Belt Use in Virginia," Norfolk, Virginia: Old Dominion University for the Virginia Highway Safety Office.

[79] Reich, B. J., Hodges, J. S., & Zadnik, V. (2006), "Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models," *Biometrics*, 62, 1197–1206.

[80] Resnick, S. I. (2002), *Adventures in Stochastic Processes,* Boston: Birkhäuser.

[81] Riedler, M. G. & Thieullen, M. (2015), "Spatio-temporal Hybrid (PDMP) Models: Central Limit Theorem and Langevin Approximation for Global Fluctuations. Application to Electrophysiology," *Bernoulli*, 21(2), 647–696.

[82] Ripley, B. D. (1976), "The Second-Order Analysis of Stationary Point Processes," *J*ournal of Applied Probabilit, 13(2), 255-266.

[83] Rizzo, M. L. (2008), *Statistical Computing with R,* London: Chapman & Hall/CRC.

[84] Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 832–837.

[85] Rue, H., & Martino, S. (2009), "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations," *Journal of the Royal Statistical Society B*, 71, Part 2, 319–392.

[86] Saule, E., Panchananam, D., Hohl, A., Tang, W., and Delmelle, E. (2017), "Parallel Space-Time Kernel Density Estimation," International Conference on Parallel Processing, 483–492.

[87] Scarrott, C., & MacDonald A. (2012), "A Review of Extreme Value Threshold Estimation and Uncertain Quantification," *Revstat - Statistical Journal*, 10(1), 33–60.

[88] Sokal, R. R., Oden, N. L., & Thomson, B. A. (1998), "Local spatial autocorrelation in a biological model," *Geographical Analysis*, 30(4), 331-354.

[89] Stoyan, D., Kendall, W. S., & Mecke, J. (1995), *Stochastic Geometry and Its Applications* (Second Edition), Wiley.

[90] Stoyan, D., & Stoyan, H. (1998), "Non-Homogeneous Gibbs Process Models for Forestry - A Case Study." *Biometrical Journal*, 40(5), 521–531.

[91] Taeger, D., and Kuhnt, S. (2014), *Statistical Hypothesis Testing With SAS and R*, Chichester, UK: Wiley.

[92] Tanaka, U., & Ogata, Y. (2014), "Identification and Estimation of Superimposed Neyman-Scott Spatial Cluster Processes," *Annals of the Institute of Statistical Mathematics*, 66, 687–702.

[93] Tanaka, U., Ogata, Y., & Stoyan, D. (2008), "Parameter Estimation and Model Selection for Neyman-Scott Point Processes," *Biometrical Journal*, 50, 43–57.

[94] Tiefelsdorf, M (1998), "Some Practical Applications of Moran's I's Exact Conditional Distribution," *Papers in Regional Science*, 77(2), 101-129.

[95] Tiefelsdorf, M., & Boots, B. (1997), "A Note on the Extremities of Local Moran's $I_i s$ and Their Impact on Global Moran's I," *Geographical Analysis*, 29(3), 248–257.

[96] Tierney, L., Kass, R., & Kadane, J. (1989), "Fully exponential Laplace approximations to expectations and variances of nonpositive functions," *Journal of the American Statistics Association*, 84, 710-716.

[97] Vaillant, J., Puggioni, G., Waller, L. A., & Daugrois, J. (2011), "A Spatio-temporal Analysis of the Spread of Sugarcane Yellow Leaf Virus," *Journal of Time Series Analysis*, 32, 392–406.

[98] van Lieshout, M. N. M. (2000), *Markov Point Process and their Applications,* London: Imperial College Press.

[99] van Lieshout, M. N. M. (2006), "Markovianity in Space and Time," *IMS Lecture Notes Monograph Series in Dynamics & Stochastics*, 48, 154-168.

[100] Walther, G. (2009), "Inference and Modeling with Log-concave Distributions," *Statistical Science*, 24(3), 319–327.

[101] Wen, S., & Kedem, B. (2009), "A Semiparametric Cluster Detection Method - a Comprehensive Power Comparison with Kulldorffs Method," *International Journal of Health Geographics*, 8, 73–89.

[102] Wheeler, D. C., & Waller, L. A. (2008), "Mountains, Valleys, and Rivers: the Transmission of Raccoon Rabies over a Heterogeneous Landscape," *Journal of Agricultural, Biological, and Environmental Statistics*, 13(4), 388–406.

[103] Wikle, C. K., & Hooten, M. B. (2010), "A general science-based framework for dynamical spatio-temporal models," *TEST*, 19(3), 417–451.

[104] Wojdylo, J. (2006), "On the Coefficients that Arise from Laplaces Method," *Journal of Computational and Applied Mathematics*, 196, 241–266.

APPENDIX A

FIXED SUBAREA SIMULATION R CODE

```
library(spatstat)

pois.points <- function(lambda,time){
sim.results <- data.frame(run=NA,time=1,mean.Moran=NA,
sd.Moran=NA,Moran.total=NA)
start <- ppp(x=rep((0:3)+.5,1,each=4),y=rep((0:3)+.5,4),
xrange=c(0,4),yrange=c(0,4))
length.of.points <- 16 # number of discs
row=1
#time=1
for (run in 1:1) {
radius <- rep(.5,length.of.points)
lower.x <- rep(0:3,1,each=4)
lower.y <- rep(0:3,4)
list.of.discs <- list()
list.of.points <- list()
n.i <- c()
data.frame.of.points <- data.frame(discID=(1:16),
        x=start$x,y=start$y)

# generate Poisson point process for each disc
for (i in 1:length.of.points){
u <- owin(xrange=c(lower.x[i],lower.x[i]+1),
        yrange = c(lower.y[i],lower.y[i]+1))
mask <- as.mask(u,eps=0.01)
bigX <- raster.x(mask)
bigY <- raster.y(mask)
disc <- owin(xrange=c(lower.x[i],lower.x[i]+1),
        yrange = c(lower.y[i],lower.y[i]+1),
        mask = ((bigX-start$x[i])^2 + (bigY-start$y[i])^2
```

```
        <= radius[i]^2 ))
list.of.discs[[i]] <- disc
list.of.points[[i]] <- rpoispp(lambda=lambda*time,
        win=list.of.discs[[i]])
n.i[i] <- 1+length(list.of.points[[i]]$x)
if (n.i[i]>1) {
# work out how many points to add
insert.rows <- (dim(data.frame.of.points)[1]+1):
        (dim(data.frame.of.points)[1]+(n.i[i]-1))
data.frame.of.points[insert.rows,] <-
        data.frame(discID=rep(i,(n.i[i]-1)),
        x=list.of.points[[i]]$x,
        y=list.of.points[[i]]$y)
}
}

# calculate global Moran statistic
Moran.total <- sum(1/dist(data.frame.of.points[,2:3]))
split.points <- split(data.frame.of.points,
data.frame.of.points$discID)
Moran <- c()
# calculate Moran statistic for each disc
for (i in 1:length.of.points){
if (n.i[i] == 1){
Moran[i] = NA
} else {
Moran[i] <- sum(1/dist(split.points[[i]][,2:3]))
}
}

mm = mean(Moran,na.rm=TRUE)
ms = sd(Moran,na.rm=TRUE)
sim.results[row,] <- c(run,time,mm,ms,Moran.total)
row = row+1
return(data.frame.of.points)
}
}
```

VITA

Jennifer L. Matthews
Commander, United States Navy
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D	Old Dominion University, Norfolk, VA (May 2019). Major: Computational and Applied Mathematics (Biostatistics).
MS	Naval Postgraduate School, Monterey, CA (June 2006). Major: Operations Research.
MS	National University of Singapore, Singapore (March 2006). Major: Defence Technology and Systems.
BS	United States Naval Academy, Annapolis, MD (May 2000). Major: Mathematics.

Publications

Lorio, J., Diawara, N., and Waller, L. (2018), "Density Estimation of Spatio-temporal Point Patterns Using Moran's Statistics," *International Journal of Statistics and Probability*, 7 (2), 80–90. DOI: 10.5539/ijsp.v7n2p80

Bui, N., Lorio, J., Diawara, N., Das, K., and Waller, L. (2018), "New Approaches to Model Simulated Spatio-Temporal Moran's Index," *Journal of Probability and Statistical Science*, 16(1), 11–24.

Diawara, N., Waller, L., King, R., and Lorio, J. (2018), "Simulations of Local Moran's Index in a Spatio-temporal Setting," *Communications in Statistics - Simulation and Computation*.

Matthews, J. L., Diawara, N., & Waller, L. A. (2018) "Quantifying spatio-temporal characteristics via Moran's statistics," Manuscript accepted for publication In *Modern Statistical Methods for Spatial and Multivariate Data*, New York: Springer.

Typeset using LaTeX.