Title
Weighted Cross J-Function and Its Application to African Avian Flu Data

Author
Zanontian, Linda Ania

Publication Date
2016

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Weighted Cross J-Function and Its Application to African Avian Flu Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Statistics

by

Linda Ania Zanontian

2016

Abstract of the Dissertation

Weighted Cross J-Function and Its Application to African Avian Flu Data

by

Linda Ania Zanontian

Doctor of Philosophy in Statistics

University of California, Los Angeles, 2016

Professor Frederic R. Paik Schoenberg, Chair

It is common to use geostatistical techniques to analyze epidemiological data. However, we might gain further insight by viewing these types of data as a point pattern due to the spatial nature of the dataset which would allow us to use the spatial information of each point and apply point process techniques. Point process techniques are applied to the avian influenza virus data, and a summary statistic called the weighted cross J-function is proposed. The ordinary cross J-function is extended to a weighted version by incorporating weights to account for inhomogeneity because this dataset appears to exhibit non-constant intensity. Unlike the ordinary cross J-function, the weighted cross J-function takes into account the varying background rate of the point process by incorporating weights for each point in the point patterns. The advantage of the weighted cross J-function is that it is used to measure the interaction between events in two point processes and to detect clustering or inhibition between them, in order to recognize where spatial interaction appears most prevalent. We introduce the weighted cross J-function, discuss its properties, demonstrate it with simulations and show its application to the African Avian Flu dataset.

The dissertation of Linda Ania Zanontian is approved.

Michael Edward Shin

Yingnian Wu

Hongquan Xu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2016

In loving memory of Hakob Medsbaba and Rosa Medsmama ...

my guardian angels

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to express my gratitude to my advisor, Rick Schoenberg, who has been instrumental in the completion of my PhD. He gave me a chance when most people would have looked away and I cannot thank him enough for that. He has truly been an exceptional advisor and has provided me constant support, encouragement and guidance. Additionally, I would like to thank the members of my committee, Hongquan Xu, Yingnian Wu and Michael Shin, who provided me with important feedback and suggestions. Thank you to Kevin Njabo for providing the avian flu dataset used in this research.

I would also like to acknowledge the amazing group of people of the UCLA Statistics Department who have become my second family. I would especially like to thank Nicolas Christou who has given me endless support from the moment I met him, whether it was helping me with my teaching skills or providing me feedback on my research. From being his student to being his TA, he has made this entire experience so much more enjoyable. Rob Gould who has always been so kind and has widened my horizons by giving me opportunities to work with Mobilize and DataFest which have been absolute pleasures. Amy Braverman who introduced me to the fascinating world of JPL and helped refine my writing and research skills. Mahtash Esfandiari who encouraged me to pursue my graduate school studies in Statistics after taking an introductory statistics class with her. Tom Ferguson who always gave me a smile of encouragement. Jan de Leeuw who gave me the opportunity of the lifetime to sit in on his meetings with graduate students to learn about Statistics when I was a shy undergrad with no direction. The knowledge I took away from those years was truly unparalleled. Glenda Jones who is a truly exceptional human being. She has provided me with limitless support, great laughs and a lifelong friendship that I will always cherish. Thank you to all my wonderful friends and colleagues, especially, Yuliya Marchetti and Irina Kukuyeva who have been amazing listeners, even more amazing role models and great lunch buddies. Medha Uppala who has always been there for me with a helping hand or funny videos to cheer me up. James Molyneux who has made this last year so enjoyable. LeeAnn Trusela who has been so caring and kind and sent me positive and encouraging messages

when I needed them most.

I am grateful for my family for supporting me through this journey. My grandparents for being my biggest fans and always being so proud of me. My cousin, Ani, for her endless support, positive motivation and encouragement. Thank you for standing by my side through it all. My brother, Armen, for always being my partner in crime. Thank you for always being there to talk and give me advice on my work. It has been special experience knowing you were on the same journey as me! My sister-in-law, Sarah, for always taking my side no matter what. My wonderful husband, Anto, for being my rock. Thank you for all your patience and support. I definitely could not have done this without you! My loving mother-in-law and father-in-law who have loved me and supported me from the moment I met them. Thank you to all my aunts, uncles, cousins and close friends for the constant encouragement.

Finally, I would like to express my immense appreciate to my amazing parents for being so incredibly caring and supportive of me throughout my life. They have raised me to be a strong, independent woman and have taught me to dream big and achieve whatever I set my mind to. I have truly been blessed with such loving, dedicated and accomplished parents. This one is for you, Mom and Bob!

Vita

2008	B.S. (Mathematics/Applied Science), UCLA, Los Angeles, California
2010	M.S. (Statistics), UCLA, Los Angeles, California
2008-2016	Teaching Assistant, Reader and Graduate Student Researcher UCLA Department of Statistics
2014	Teaching Assistant of the Year UCLA Department of Statistics
2015	C.Phil., Statistics, UCLA, Los Angeles, California

Publications and Presentations

Gharibans, L.; Braverman, A.; Mattmann, C.; Garcia, J.; Crichton, D. (2010). "Networks for Analysis of Distributed Data". Poster Presentation: *SAMSI Program on Complex Networks*.

Njabo, K.Y.; **Zanontian, L.**; Sheta, B.N.; Samy A.; Galal, S.; Schoenberg, F.P. (2016). "Living with avian flu - persistence of the H5N1 highly pathogenic avian influenza virus in Egypt". *Veterinary Microbiology*, 187, 82-92.

CHAPTER 1

Introduction

Infectious diseases, which are caused by pathogenic microorganisms and transferred from human to human, continue to proliferate and are becoming a serious problem (Jones et al., 2008). With modern medicine and improved technology the general expectation is that this would not be an issue, but overuse of antibiotics is creating resistant infectious diseases, or superbugs, that are projected to be one of the biggest threats to humanity in the coming decades (Nordmann et al., 2007; Calderone, 2015a). Figure 1.1 illustrates the decrease in deaths due to infectious diseases since the introduction of antibiotics, such as penicillin, in the 1940s. Unfortunately, due to overuse, antibiotics which were once able to treat bacterial infections are now slowly becoming ineffective. According to the Centers for Disease Control and Prevention (CDC) an estimated 2 million people are annually affected by drug-resistant bacteria annually, resulting in 23,000 deaths and more than $50 billion in excess health-case costs and lost productivity (CDC, 2013). The World Health Organization has recognized that the growth of infection diseases is a serious issue and if we do not respond vigorously as a global community we risk entering post-antibiotic era which might hinder modern medicine (Nizet, 2015). Calderone (2015b) describes and illustrates that by the year 2050 officials say infectious diseases will be the main cause of death, more than cancer, potentially killing 10 million people annually (Figure 1.2).

Understanding and predicting the behavior of infectious diseases is an important first step to overcome spread of drug-resistant bacteria. Additionally, with the advent of data accessibility and availability, real-time tracking of infectious diseases is now possible and may prove to be more efficient and effective in preventatively halting the spread of it (Carneiro and Mylonakis, 2009). This is especially true for regions where it is not feasible to widely

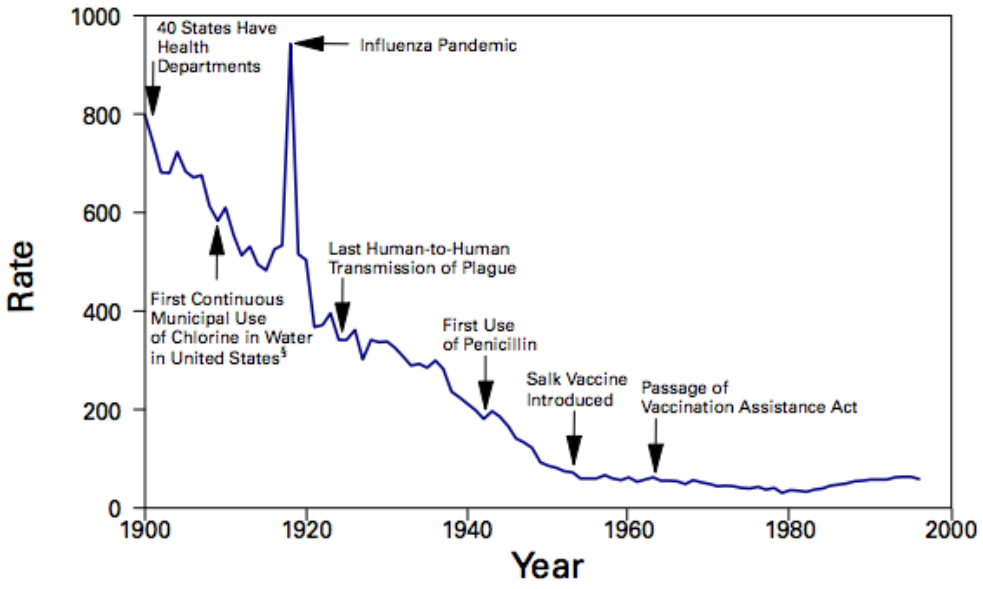

Figure 1.1: Crude death rate per 100,000 population annually for infectious diseases in the United States between 1900 and 1996.

source by: http://www.cdc.gov/mmwr/PDF/wk/mm4829.pdf

vaccinate due to financial or geographical limitations (Oshitani et al., 2008). An intuitive approach to this problem from a statistical viewpoint is the application of spatial statistical methods, specifically point process techniques, which can be applied to events that randomly occur in nature, such as outbreaks of an infectious disease. These techniques allow us to gain further insight on the characteristics of the event which might be useful for prediction and forecasting of the event. There has been extensive research done in the application of point process techniques to wildfires (Xu and Schoenberg, 2011; Nichols et al., 2011; Peng et al., 2011) and earthquakes (Ogata, 1999, 1988; Gordon et al., 2015; Clements et al., 2012), to name a few, but more can be achieved in the application to epidemiological data, such as infectious diseases. Although an intuitive approach, point process techniques have not been effectively utilized as an approach to stop the spread of infectious diseases, hence we decided to explore this idea in this dissertation.

For this dissertation, we had the opportunity to work with the highly pathogenic avian influenza virus (HPAIV), H5N1, which continues to threaten Egypt despite serious efforts

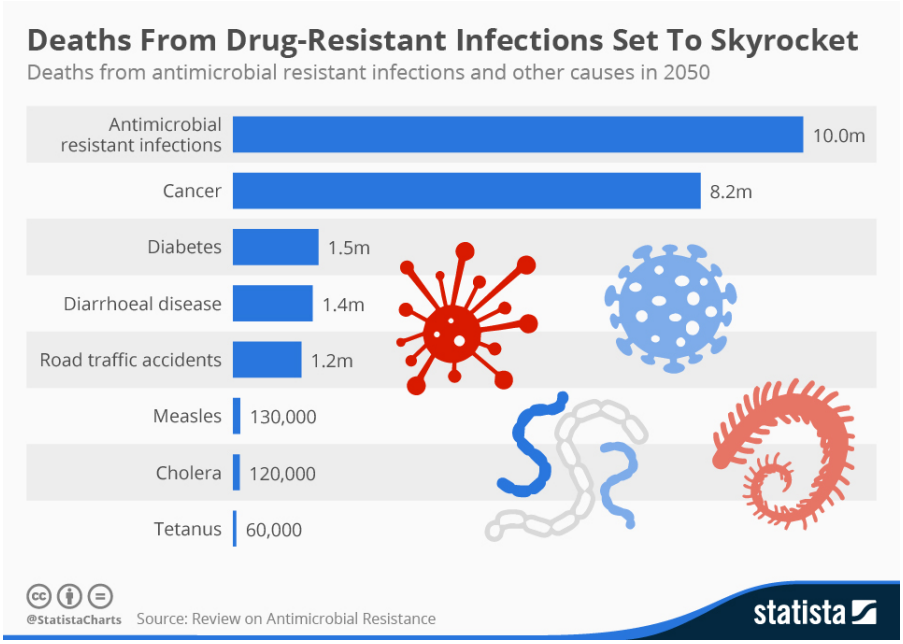

Figure 1.2: Projection of deaths from drug-resistant infections in the year 2050. Infectious diseases will be the cause of a projected 10 million deaths in humans.

source by: https://www.statista.com/chart/3095/drug-resistant-infections/

of prevention. Egypt has become an epicenter for the virus and it is a country where the virus is endemic. The avian flu was first reported amongst domestic poultry at commercial farms and in the backyard flocks in three northern Egyptian governorates in February 2006 (Kandeel et al., 2010). Within one month it was already being detected in twelve other governorates. Control measures, such as culling, disinfection, vaccination and controlled poultry movement, have been put into effect however virus outbreaks continues to occur.

Ever since the first avian flu outbreak in Asia in 1996, H5N1 has been a serious health concern. Even though the avian influenza virus predominantly affects birds, it can also affect mammals, including humans, and can cause serious illness and mortality. The amount of human influenza occurrences continues to grow. Between 2003 and 2015, 844 laboratory-confirmed human cases of infection from 16 countries were reported to the World Health Organization (WHO) (Njabo et al., 2016). Of those 844 recorded cases, 55% or 449 people have died. In Egypt, 336 confirmed human influenza cases have been reported and 99 of human deaths due to the influenza until March 2015 (El Masry et al., 2015; Report, 2015). As of January 2015, there have been 143 new laboratory-confirmed human cases of H5N1, of which 42 have died, according to the WHO. Of the 143 cases 136 are from Egypt and only 1 from China, where the avian flu first began. The demographic that is highly affected by the flu is young people and women (Report, 2015). The ages range from 1 year old to 75 years old with a median of 26 years old. 23% of the affected cases are under the age of 10 (Report, 2015).

Egypt continues to be the most affected country by the avian flu H5N1 outside of Asia. There are a number of theories that have tried to explain the reason for this. One theory is it could be that Egypt lies in the middle of major intercontinental avian flyways which link Africa, Europe and Asia (Tian et al., 2015). Due to its location, Egypt can be viewed as the hub between these nations and therefore the most likely pathway between the regions. Additionally, the Nile River can be the reasoning behind the persistence of the avian flu, as the Nile River Delta is the location where the flu is most concentrated. If we look closer at the Nile River Detla , three major risk factors that have been identified as the reason for continued proliferation of the flu according to Abdelwhab and Hafez (2011). One is the high

density of domestic waterfowl located at the river delta. The second is high density of rural human population since it is desirable to be near the water. Lastly, the river delta has an abundance of water and irrigation networks.

Even though the avian flu mainly affects poultry, humans can get the flu through contact with infected birds. Since chickens are very much a part of the Egyptian society and agricultural economy, it is understandable how a large number of humans are introduced to the flu in Egypt. Even though, the infected poultry are mainly believed to be found amongst backyard flocks, which also are believed to provide the most infection to humans, they can also be found amongst live bird markets and confined animal feeding operations. These areas have played a major role in the surge of zoonotic infection, diseases that are transferred from animals to humans. Egypt produces about 750 million to one billion poultry annually (El Masry et al., 2015). An estimated 4 to 9.5 million poultry are found in households or backyards specifically in confined spaces in houses, on rooftops or free range in the backyard with little or no biosecurity measures (Fasina et al., 2012). According to the World Health Organization, the avian flu is not highly transferrable from human to human even though a few cases have been reported (Report, 2015).

The avian flu is a very serious endemic in Egypt and it will most likely require a large, coordinated effort to eliminate it from a country so densely populated. Therefore it is crucial for multidisciplinary approaches to study and understand the avian flu and its properties, where it exists, where it is highly concentrated and how it is transferred. One approach would be to apply spatial statistical tools, such as point process techniques, to understand the behavior of the avian flu. With the collaborative efforts of Egyptian Animal Health Research Institute/National Laboratory of Quality Control of Poultry Production (AHRI/NLQP) and the General Organization of Veterinary Services (GOVS), data was collected from four Egyptian governorates: Damietta, El Gharbia, Fayoum and Menofia. Six types of birds, chickens, ducks, geese, turkeys, pigeons and wild birds, were sampled and tested for avian flu between the years 2009 and 2012. This dataset is analyzed in Chapter 4 of this dissertation.

Typically epidemiological data is studied by applying geostatistical techniques, as seen in Nuvolone et al. (2008) and Seng et al. (2005). The data can be organized into counts of

observations within predefined gridded cells which would allow the application of standard grid-based spatial statistical methods. However, we might be able to gain further information by applying point process techniques as well. The spatial distribution of epidemiological or disease data can be seen as point processes as it is a random collection of events occurring in a bounded region of space, where each influenza outbreak is represented by a point on the surface of Earth. In these types of datasets, spatial and/or temporal information is recorded of each observation allowing us to use the point process approach.

Point process techniques are applied to the avian influenza virus data, and a summary statistic called the weighted cross J-function is proposed. The ordinary cross J-function is extended to a weighted version by incorporating weights to account for inhomogeneity because this dataset appears to exhibit non-constant intensity. By applying this new summary statistic to the avian flu dataset we hope to characterize and understand the spatial-temporal aspect of the events and to pinpoint exact moments of interaction between two groups along with the type and range of the interaction.

The application of spatial statistics techniques to epidemiological data may help us to understand the characteristics of the disease which may then allow us to create and implement a preventive solution. The idea of applying spatial statistics in this particular way can potentially open new doors of research and help us understand other diseases across the world in an entirely new way. Currently, the Zika virus has become a very important topic in current events. The virus has spread quickly since its first large outbreak in humans on the Pacific island of Yap in 2007 (Kindhauser et al., 2016). It has made its way Africa to Asia to the Pacific Islands to Central and South Americas. It is now a serious endemic in Brazil where it has been estimated to have harmed 1.5 million people and has been linked to pregnancy microcephaly and Guillain-Barrè syndrome. The Zika virus is similar to the outbreak of the avian flu in Egypt where we can apply point process techniques, specifically the weighted cross J-function, to understand the nature of the disease, the interactions that are occurring and in the larger sense suggest some preventive solutions.

The remainder of this dissertation is organized as follows. Chapter 2 provides a survey of point processes along with the second order properties and nearest neighbor techniques.

Chapter 3 introduces the aforementioned weighted cross J-function and demonstrates the performance of the estimation of the weighted cross J-function with simulations. Chapter 4 details the application of the weighted cross J-function to avian influenza virus data, and then Chapter 5 concludes the dissertation and suggests important topics for future research in this area.

CHAPTER 2

Point Processes Properties and Techniques

2.1 Overview of Point Processes and Their Properties

A spatial point process is a random collection of data points or "events" occurring in a region of two or more dimensional space (Diggle, 2013). For example, if we take an area and mark all the wildfires that have occurred in that given space for a set interval of time, this map will consist of a random pattern of points in a two dimensional space, because wildfires are random occurrences in nature. Hence, there will be a random number of points in the given area and their locations will be random as well (Baddeley et al., 2007). This example is illustrated in Figure 2.1. Other examples of spatial data where point process techniques can be used are trees in forest, earthquakes and lightning strikes.

Mathematically, a point process is defined as a random measure N on a space $S \subseteq \mathbb{R} \times \mathbb{R}^3$ of space-time, taking values in the non-negative integers Z^+ or infinity, where Z represents the observed attribute. The measure N(A) represents the number of points in the subset A of S. Often the spatial region of interest is rectangular but it can also have an irregular shape due to geographical constraints, such as city limits, rivers or other geographical features. For example, if you want to analyze the flu outbreaks in the state of California it might look like Figure 2.2, where the area of the point process is irregular because it takes the shape of the state borders.

When using point process techniques, the event's location and/or time is the variable to be analyzed (Cressie, 1993). If the point process contains both spatial and temporal information then it is a spatial-temporal point process. For example, in the case of wildfires, the time the wildfires occurred can be recorded along with the location of the wildfire occurrences.

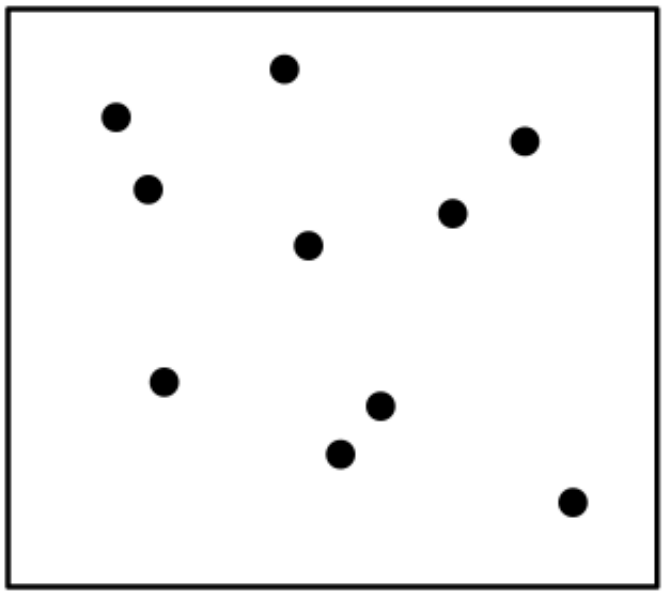

Figure 2.1: An example of a point process in two dimensions, space and time.

Typically the spatial locations are recorded as longitude, latitude and/or depth. When space and time are both present, we have a point process in a three dimensional space (space x time). A point process can also have one dimension, where only the time of a sequence of randomly occurring events is recorded. For example, the random instances of times that rainfall occurs in the city of Los Angeles in an interval of time can be modeled as a point process with one dimension (time). The number of times rainfall occurs will be random along with the times it occurs. Figure 2.3 represents a point process in time such as the rainfall example.

Most of the time, the observation window W is contained within S but sometimes there are missing data problems which are referred to as edge effects (Møller and Waagepetersen, 2007). Edge effects occur when an observed point inside the observed window is close to the border. When a circle of radius r is drawn around each observed point to determine its nearest neighbors, the circles of the points closest to the boundaries of the window will extend outside the boundary. The points might have nearest neighbors outside the window which would not be counted. This problem is illustrated in Figure 2.4. The red point sits near the boundary of W hence its nearest neighboring points fall outside window W. Therefore,

Figure 2.2: An example of a point process with an irregular shape.

Figure 2.3: An example of a point process with one dimension, time.

bias can be introduced due to edge effects. One approach to account for this problem is the border method where observation points lying within r units from the boundary of W are not taken into consideration (Baddeley et al., 2006). Additional edge effects correction methods have been discussed in Baddeley (1999). Edge effects are critical in the analysis of spatial point processes, especially when using the K-function, further described in Section 2.2.

The intensity, or background rate, can be used to characterize the behavior of the point process. Lambda (λ), or intensity, is the mean number of events per unit area or the rate of occurrence of the events. The mean number of points in any set A of N is equal to λ multiplied by the area of A (Illian et al., 2008).

$$E\left(N\left(A\right)\right) = \lambda \cdot v\left(A\right)$$

where A is any subset of $\mathbb{R}^d$.

The homogeneous Poisson process is the most basic point process because there no interaction exists between the events, meaning the number of points in the space follows a Poisson distribution and the number of points in each event are independent. The Poisson distribution is a discrete probability distribution where the probability of a set number of independent events occurring in a fixed interval of space and/or time with a known average rate, λ. Homogeneity implies that the intensity, or background rate, is constant. Hence, this could be used as a baseline to distinguish between events with positive interactions or events with negative interactions since the Poisson process has complete spatial randomness, meaning no type of interaction between events (Gelfand et al., 2010). A complete spatial randomness (CSR) point process refers to a stationary, or spatially homogeneous, Poisson

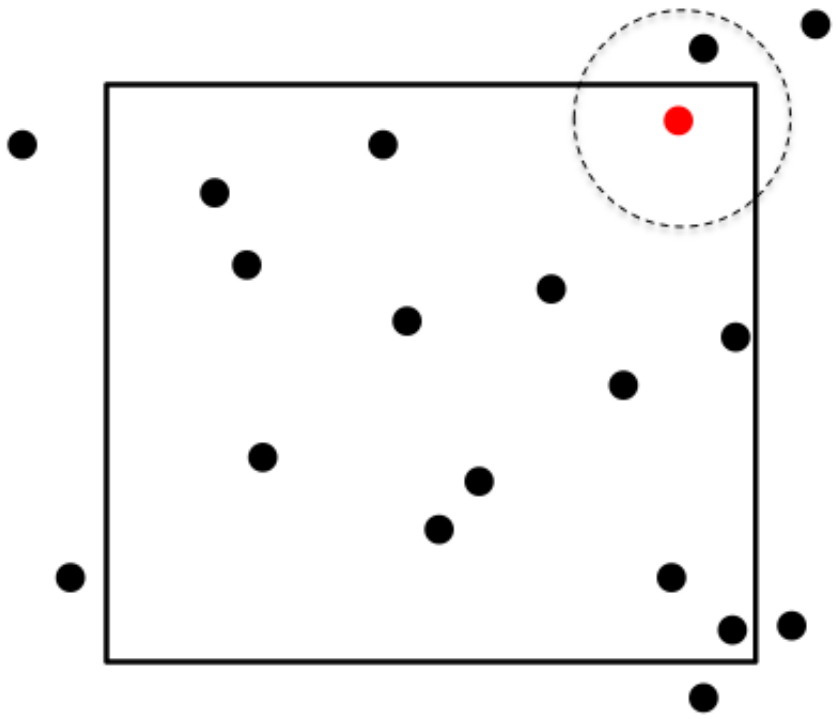

Figure 2.4: An example of a point process with the edge effects problem. The red point is the reference point to find its nearest neighbor points. The open circle with radius r is used to find the nearest neighbors to the red point.

process. A point process N in $\mathbb{R}^d$ is stationary if N and the translated point process N_x have the same distribution for all points x, meaning

$$N \equiv N_x$$

where $N = \{x_1, x_2, ...\}$ and $N_x = \{x_1 + x, x_2 + x, ...\}$. (Illian et al., 2008)

The complete spatial randomness (CSR) pattern can be used as a basis to distinguish between homogeneous patterns with interaction. If the pattern deviates from a CSR pattern then we conclude that there is an interaction, either attraction or repulsion, between the events. If there is clustering, the average distance between an event and its nearest neighbor is smaller than the average distance between an event and its nearest neighbor in a CSR

pattern. If there is inhibition, the average distance between an event and its nearest neighbor is larger than the average distance between an event and its nearest neighbor in a CSR pattern (Schabenberger and Gotway, 2004).

There is also a non-stationary Poisson process, or inhomogeneous point process, where there is still no interaction between the events but the intensity, or background rate, does vary and is not constant. For example, in a Poisson cluster process, a set of 'parent' events are generated, then 'offspring' events are generated around the parent events, meaning there is clustering around each 'parent' event. This is similar to a Poisson process with a non-constant intensity hence it would be considered to be an inhomogeneous point process. An example of a clustered Poisson process is the Neyman-Scott process where the number of offspring events are independent and identical for each parents. An extension of the Poisson process for a clustered point pattern, typically used in biological processes such as locations of bacteria, is the Cox process, also known as the doubly stochastic Poisson process or mixed Poisson process (Møller and Waagepetersen, 2003b). Here the events are dependent and the intensity function is a realization of a stochastic process (Schabenberger and Gotway, 2004).

Point processes can have clustering, inhibition (regularity) or independence. Clustering occurs when there is interaction between the events in the point process and inhibition occurs when there is repulsion between the events in the point process. It is possible for a point process to exhibit clustering at some scales and inhibition at others, e.g. Schoenberg and Bolt (2000). Independence occurs when there is no interaction between the events meaning that each event occurs in the space with equal probability. In the case of independence, i.e. where the number of points N(A1) in one area A1 is independent of the number of points N(A2) in any other area A2, for any two disjoint areas, the process is called a Poisson process. Therefore N(A) is a Poisson random variable for any measurable region A and the mean is simply lambda times the area of A. For independent point processes, λ remains constant meaning the intensity does not vary. For non-stationary point processes the intensity or the background rate of events varies.

A point process can be unmarked or marked. A marked point process is a random collection of data points with additional information, or marks, besides location and/or

time. For example, along with the location of the trees in the forest, information is collected about the type of trees. Marks can be continuous or discrete (Schabenberger and Gotway, 2004). For example, the height of trees in a forest is continuous whereas the types of trees in a forest is discrete. A mark can also be binary, meaning there are only two choices in the category. For example, if the trees in the forest are categorized as either oak or spruce then these marks would be binary. An unmarked point process contains no additional information besides location and time.

A multivariate point process is referred to a point process consisting of events of different types. This can also be referred to as a multitype point pattern (Møller and Waagepetersen, 2003a). The types of trees in a forest is an example of a multivariate or multitype point process. Hence, the terms multivariate and marked patterns can be used interchangeably depending on how the pattern is defined. We can look at the types of trees as different events, for example, if there are two types, oak and spruce, then we have a bivariate point process. But we can also combine all the trees into one event with two different marks which indicate the pattern type (Schabenberger and Gotway, 2004).

2.2 Second Order Properties

Second order properties are significant in the analysis of spatial point processes. Much previous study has focused on methods to summarize and estimate second order properties.

2.2.1 *K*-Function

The K-function, $K(r)$, is the second order reduced moment function ("Ripley's K Function") (Dixon, 2002). The K-function is used to analyze a single point process with the assumption of homogeneity and to detect clustering or inhibition. The K-function is the observed count of all pairs of points (N), within distance r, in the point process.

$$K(r) = \frac{1}{\lambda} E(N)$$

If the K-function from the data is different than the K-function for CSR then there is some sort of interaction, either clustering or inhibition. If the observed is less than the theoretical it implies inhibition and if the observed is greater than theoretical it implies clustering. The K-function is not unique to a point process; two point processes can have the same K-function (Dixon, 2002).

The K-function takes into account edge effects. When the observed region of study, A, is part of a larger area of region we typically have edge effects because some of the events might be close to the border of region A. Since the nearest neighbor is the closest neighboring event and the nearest neighbors of these near border events might be outside the region of study, the distance between the pair of events will be unknown. Therefore there might be some bias when using nearest neighbor techniques. There are three common methods to correct for edge effects: using barrier areas inside the border of A, regarding a rectangular observed region A as a torus or sphere where there are no edges, and making adjustments to regard for the unobserved events by making a guard area outside of the bounded region (Cressie, 1993; Diggle, 2013).

2.2.2 Weighted K-Function

The weighted K-function, or inhomogeneous K-function, is a modified K-function to allow for non-constant intensity or background rate to account for inhomogeneity (Baddeley et al., 2000b). The weighted K-function uses weights to balance the inhomogeneity of the point process. For example, Veen and Schoenberg (2006) apply the weighted K-function to Southern California earthquakes.

As defined in Veen and Schoenberg (2006), to assess the model $\lambda_0(x, y)$, the weighted K-function can be defined as

$$K_w(h) = \frac{1}{\lambda_2^* A} \sum_r w_r \sum_{s \neq r} w_s 1(|p_r - p_s| \leqslant h)$$

where $\lambda_* := inf\{\lambda_0(x, y); (x, y) \in A\}$ is the infimum of the conditional intensity over the observed region, $w_r = \lambda_* / \lambda_0(p_r)$ where $\lambda_0(p_r)$ is the modeled conditional intensity at point (p_r) and h is the distance. Methods for edge effects correction can also be applied to the

inhomogeneous K-function (Baddeley et al., 2000b).

2.2.3 Cross *K*-Function

The K-function takes into account location but no other information about the events, such as heights or types of trees, which are called marks. However further insight can be gleaned by marks hence multivariate or marked point processes can account for this extra information (Schabenberger and Gotway, 2004). For example, knowing the type of trees along with the recording of the locations might provide insight of whether certain types of trees have a positive or negative interaction, or they simply co-exist with no interaction.

The cross K-function, $K_{ij}(r)$, involves two or more point processes within the same space. It is used to analyze how the two events relate to each other, whether there is clustering, inhibition or independence. A bivariate spatial point process, where there are two events, is an example of when the cross K-function is utilized.

$$K_{ij}(r) = \lambda^{-1} E[N_{ij}]$$

where N_{ij} is number of type j events with distance r of a randomly chosen type i event.

2.2.4 *L*-Function

The L-function stabilizes the variance of the K-function. Generally, it is easier to use the L-function in practice because the variance is approximately constant under CSR, meaning $L(r) = r$. Similar to the K-function, if the observed is less than the theoretical it implies inhibition and if the observed is greater than theoretical it implies clustering.

$$L(r) = \sqrt{K(r)/\pi}$$

where $K(r)$ is the K-function and r is the radius of distance. When plotting $\widehat{L}(r) - r$ against r, the horizontal line at 0 represents complete spatial randomness. Clustering is indicated by $\widehat{L}(r) - r > 0$ whereas repulsion or inhibition is indicated by $\widehat{L}(r) - r < 0$.

2.3 Nearest Neighbor Techniques

Summary statistics, such as nearest neighbor techniques, have been developed for point processes. The underlying idea of nearest neighbor techniques is to look at the distance between a point and its nearest neighboring point to gather information regarding interactions in the point process.

2.3.1 *F*-Function

The F-function, $F(r)$, is the empty space function, is the cumulative distribution function of the distance from a fixed point to the nearest point of point process X less than or equal to r. Let X be a stationary point process on $\mathbb{R}^d$ where the intensity is λ.

$$F_j(r) = \mathbb{P}\left(X \bigcap B(0,r) \neq \varnothing\right)$$

where $B(0,r)$ is the closed ball of radius $r \geq 0$ centered at the origin 0 (Baddeley et al., 2000b).

2.3.2 *G*-Function

The G-function, $G(r)$, is the nearest neighbor distance distribution function, is the cumulative distribution function of the distance from a point in point process X to the nearest other point of point process X less than or equal to r.

$$G(r) = \mathbb{P}^{0!}\left(X \bigcap B(0,r) \neq \varnothing\right)$$

where $B(0,r)$ is the closed ball of radius $r \geq 0$ and $\mathbb{P}^{0!}$ denotes the reduced Palm distribution at the origin (Baddeley et al., 2000b). If we have a stationary Poisson process with intensity λ then the reduced Palm distribution $\mathbb{P}^{0!}$ is the same as $\mathbb{P}$ which means G $\equiv$ F (Van Lieshout and Baddeley, 1996).

2.3.3 *J*-Function

The J-function, $J(r)$, is the relationship between G-function and F-function for all $r \geq 0$ such that $F(r) < 1$. (Van Lieshout and Baddeley, 1996).

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

It is used to measure the interaction between events in a point process and to detect clustering or inhibition. If $J(r)$ is less than 1 it implies clustering and if it is greater than 1 it implies inhibition. $J(r)$ will equal exactly 1 for a Poisson process because $F(r) \equiv G(r)$ which implies complete spatial randomness. For distances, r, greater than the range of spatial interaction $J(r)$ will be constant. Therefore the J-function can be used to not only quantify spatial interaction but also the range of the spatial interaction, which will help to provide better insight for the point process. For example, if we have the locations of random trees in a forest, by using the J-function we can determine if a positive or negative interaction exists within the trees and at which radius it occurs. The J-function has been extended to inhomogeneous point processes (Van Lieshout, 2010), inhomogeneous spatio-temporal point processes (Cronie and Van Lieshout, 2015) and marked point patterns (Van Lieshout, 2006).

2.3.4 Inhomogeneous *J*-Function

The inhomogeneous J-Function, or weighted J-Function, is the J-Function for inhomogeneous point processes where the background rate or λ is non-constant. The inhomogeneous J-Function is the relationship between the inhomogeneous, or weighted, G-Function and the inhomogeneous F-Function.

$$J_{inhom}(r) = \frac{1 - G_{inhom(r)}}{1 - J_{inhom(r)}}$$

where $G_{inhom(r)}$ is the inhomogeneous G-function, $F_{inhom(r)}$ is the inhomogeneous F-function and $r \geq 0$. (Van Lieshout, 2010)

2.3.5 *J*-function for marked point processes

Similar to the K-function, the J-function takes into account only the location and/or time of the point process. However, some point patterns contain additional information, called marks, which are attached to each point (Stoyan and Stoyan, 1994). For example, the number of leaves on each tree in a forest is a discrete mark whereas the width of the trunk of the trees in the forest is a continuous mark. If there are two distinct types of discrete marks, say blue and red or yes and no, then we can consider this a point process with binomial marks. Methods used for these types of point processes with discrete marks will be discussed in the next section. A method has been introduced for point processes with continuous marks in Van Lieshout (2006).

2.3.6 Cross *J*-Function

The cross J-function is the J-function for bivariate point patterns (Van Lieshout and Baddeley, 1999). A multivariate point pattern is a spatial point pattern where the points belong to one of two or more specific "types" such as type i or type j. The most common use of multivariate point patterns is the bivariate point pattern or two-type pattern where there are two distinct types of points, i and j. The cross J-function can also be used to gather insight on point processes with discrete binomial marks.

$$J_{ij}(r) = \frac{1 - G_{ij}(r)}{1 - F_j(r)}$$

for all $r \geq 0$ and $F_j(r) < 1$. $F_j(r)$ is the empty space function for the process X_j of points of type j within a distance, r.

$$F_j(r) = \mathbb{P}\left(X_j \bigcap B(0,r) \neq \varnothing\right)$$

where $B(0,r)$ is the closed ball of radius $r \geq 0$ centered at the origin 0. $G_{ij}(r)$ is the cumulative distribution function of the distance from a point of type i, X_i, to a point of type j, X_j within a distance, r.

$$G_{ij}(r) = \mathbb{P}^{!(0,i)}\left(X_j \bigcap B(0,r) \neq \varnothing\right)$$

where $B(0,r)$ is the closed ball of radius $r \geq 0$ and $\mathbb{P}^{!(0,i)}$ denotes the reduced Palm distribution at the origin.

Therefore, $J_{ij}(r)$ is the probability that there is no X_j within a distance, r, of X_i divided by the probability that there is no X_j within a distance, r, of a fixed point. If X_i and X_j are independent, then $F_j(r) \equiv G_{ij}(r)$ and $J_{ij}(r) \equiv 1$ (Gelfand et al., 2010). When $J_{ij}(r) > 1$, $F_j(r) < G_{ij}(r)$, hence there is inhibition between X_i and X_j meaning the presence of a point of X_i decreases the probability of having point of X_j nearby. When $J_{ij}(r) < 1$, $F_j(r) > G_{ij}(r)$, hence there is clustering between X_i and X_j meaning the presence of a point of X_i decreases the probability of having point of X_j nearby.

The cross J-function does not account for the inhomogeneous point processes, which are point processes with a non-constant or varying background rate. Therefore, further insight might be gained by taking inhomogeneity into consideration.

CHAPTER 3

Weighted Cross J-Function

3.1 Motivation and Definition

The weighted cross J-function, which is an extension of the cross J-function (Van Lieshout and Baddeley, 1999) to the inhomogeneous case, which can be used to describe the degree of interaction between two different types of birds whose occurrence rates are not constant, as it appears to be the case in the data analyzed in this dissertation. The ordinary cross J-function is the quotient of one minus the nearest neighbor distance distribution function (G-function), and one minus the empty space function (F-function), and is useful for examining clustering or inhibition, relative to the overall rates, of two point patterns N_i and N_j and assumes that the point patterns are spatially homogeneous. To account for inhomogeneity, we incorporated weights for each point in the point patterns, with each weight corresponding to the inverse of the estimated intensity at its location. For each bird, the intensity estimates were obtained by kernel smoothing the occurrences of detected infections within the species, with a Gaussian kernel and default plug-in bandwidth, using R.

To provide a meaningful measure of clustering between point processes N_i and N_j, we propose the following formula for the weighted cross J-function:

$$\tilde{J}_{ij}(r) = 1 - \frac{G_{ij}(r)}{F_j(r)}$$

where G_{ij} is the cross G-function, F_j is the F-function, and thus $\tilde{J}_{ij}$ is estimated using the estimated weighted cross G-function and the estimated weighted F-function for all $r \geq 0$.

The weighted cross J-function is used to measure the interaction between events in two point processes and to detect clustering or inhibition between them, in order to recognize

where spatial interaction appears most prevalent. If $\tilde{J_{ij}} < 0$ then there is clustering between points of type i and type j within distance r. $\tilde{J_{ij}} > 0$ implies that there is inhibition between points of type i and type j within distance r. $\tilde{J_{ij}}$ will equal 0 when there is independence between both point processes.

Since we are incorporating weights to account for the inhomogeneity of the data, we are now taking the sum of all the weights of the points within a distance, r, of a chosen point instead of taking the proportion of points with a distance, r, of the chosen point, as with the ordinary cross J-function. This means that F can equal 1 whereas in the ordinary cross J-function F was defined to be less than 1, hence it is beneficial to change the formula to not have a problem of division by zero. We demonstrated by simulations that the new formula does in fact account for this problem and also provides more accurate results for the moment of interaction between two events. To demonstrate a clustered simulation, process N_i is simulated from an inhomogeneous Poisson process with a triangular intensity on the square [0, 1] x [0, 1]. N_j is simulated such that for each point tau of N_i, 5 points were simulated independently with a uniform spatial distribution on a circle centered at tau with radius 0.05. We demonstrated the original formula, the ratio of one minus weighted cross G-function and one minus weighted cross F-function, in Figure 3.1 and the updated formula, one minus the ratio of weighted cross G-function and one minus weighted cross F-function, in Figure 3.2. It can be seen that the estimate of the updated formula in Figure 3.2 performs better and detects the clustering within radius 0.05.

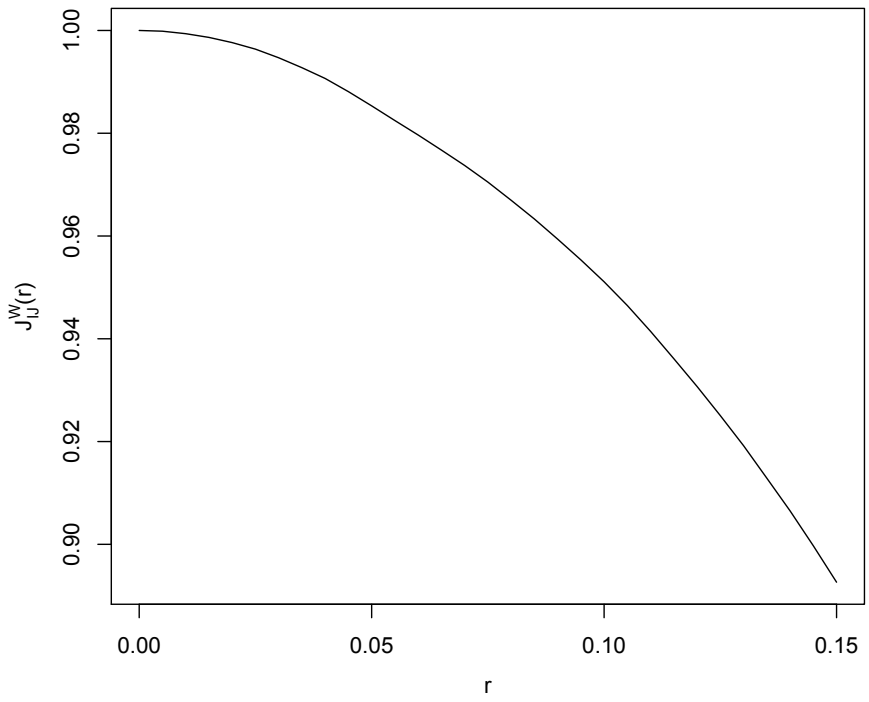

Figure 3.1: Estimates of the weighted cross J-function, $\tilde{J}_{ij}$, between simulated N_i and simulated N_j, as a function of distance, r, for various choices of pairs of point patterns N_i and N_j using the original formula.

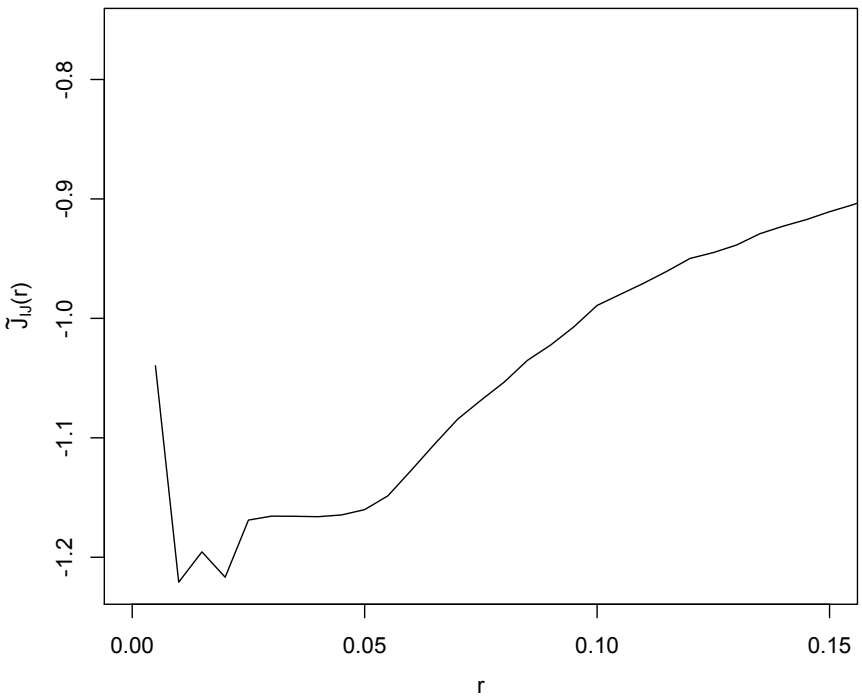

Figure 3.2: Estimates of the weighted cross J-function, $\tilde{J}_{ij}$, between simulated N_i and simulated N_j, as a function of distance, r, for various choices of pairs of point patterns N_i and N_j using the updated formula.

3.2 Estimation

Van Lieshout (2010) uses generating functionals to prove that the inhomogeneous J-function for temporal point processes is unbiased. Let $W \subset \mathbb{R}^d$ be a compact set with non-empty interior and assume point process X is observed in window W. Also assume the intensity, λ, is known. The generating function for the inhomogeneous F-function is as follows,

$$F_{inhom}(t) = 1 - G\left(1 - u_t^0\right)$$

for all for all $t \geq 0$. Suppose $L \subseteq W$ is a finite point grid. The following shows the estimation of the inhomogeneous F-function,

$$1 - \widehat{F_{inhom}}(t) = \frac{\sum_{l_k \in L \bigcap W_{\Theta t}} \prod_{x \in X \bigcap B(l_k,t)} \left[1 - \frac{\overline{\lambda}}{\lambda(x)}\right]}{\#L \bigcap W_{\Theta t}}$$

where $W_{\Theta t}$ is the eroded set $\{x \in W : B(x,t) \subseteq W\}$ and $\overline{\lambda} = inf_{x \in W} \lambda(t) > 0$. Van Lieshout (2010) continues to prove that this is unbiased.

Similarly, the generating function for the inhomogeneous G-function is as follows,

$$G_{inhom}(t) = 1 - G^{!a}\left(1 - u_t^a\right)$$

for all for all $t \geq 0$. The following shows the estimation of the inhomogeneous G-function,

$$1 - \widehat{G_{inhom}}(t) = \frac{\sum_{x_k \in X \bigcap W_{\Theta t}} \prod_{x \in X \setminus \{x_k\} \bigcap B(x_k,t)} \left[1 - \frac{\overline{\lambda}}{\lambda(x)}\right]}{\#X \bigcap W_{\Theta t}}$$

where $W_{\Theta t}$ is the eroded set $\{x \in W : B(x,t) \subseteq W\}$ and $\overline{\lambda} = inf_{x \in W} \lambda(t) > 0$. Van Lieshout (2010) continues to prove that this is ratio-unbiased. This implies that $J_{inhom}(r)$ is ratio-unbiased as well. Typically in spatial statistics, it is difficult to obtain unbiasedness for many estimators, which are ratio estimators, so often these estimators are proven to be ratio-unbiased by showing the numerator and denominator are both unbiased. For example, the following would be considered to be ratio-unbiased, $\widehat{\theta} = Y/Z$ where $\theta = EY/EZ$ (Møller and Waagepetersen, 2003b).

Cronie and Van Lieshout (2015) have shown for the inhomogeneous J-function for spatial-temporal point processes, where they choose to define $J_{inhom}(r,t)$ as the quotient of one

minus $G_{inhom}(r,t)$ and one minus $F_{inhom}(r,t)$ is ratio-unbiased. In Cronie and Van Lieshout (2014), a new summary statistic, $J^{ij}_{inhom}(r)$, for marked point processes is proposed, where it is defined as the ratio of $1 - D^{ij}_{inhom}$, the nearest neighbor function, and $1 - F^{j}_{inhom}$, the empty space function. Cronie and Van Lieshout (2014) prove that $J^{ij}_{inhom}(r)$ is ratio-unbiased. Unlike Cronie and Van Lieshout (2014), we have chosen to update our weighted cross J-function formula for bivariate point process, which could behave similar to marked point processes with discrete marks, to achieve more meaningful results when accounting for inhomogeneity. A similar approach can be taken to prove that the weighted cross J-function for spatial point processes is also ratio-unbiased. The mean and variance of $\tilde{J}_{ij}(r)$ will be addressed in future work. We would like to note that much of the work here was performed independently of (Cronie and Van Lieshout, 2014) and (Cronie and Van Lieshout, 2015).

3.3 Simulations

We used simulations to demonstrate the successful estimation of the weighted cross J-function. To demonstrate a clustered simulation, process N_i is simulated from an inhomogeneous Poisson process with a triangular intensity on the square [0, 1] x [0, 1]. N_j is simulated such that for each point tau of N_i, 10 points were simulated independently with a uniform spatial distribution on a circle centered at tau with radius 0.05. As expected the estimate of weighted cross J-function is less than 0 and shows that most of the clustering occurs within radius of 0.05, as shown in Figure 3.3. However, in comparison, the estimate of the unweighted cross J-function detects clustering as well but does not clearly indicate that the clustering is within radius of 0.05 (Figure 3.4). Please note that the formula for the unweighted cross J-function was adjusted for ease of comparison with the weighted cross J-function, so $J_{ij} < 0$ represents clustering.

For the simulation demonstrating inhibition between events, process N_i is simulated from an inhomogeneous Poisson process with a triangular intensity on the square [0, 1] x [0, 1]. N_i and N_j are simulated by first generating independent Poisson processes with triangular intensity, and then deleting each point of N_j independently with probability 80% if it is within

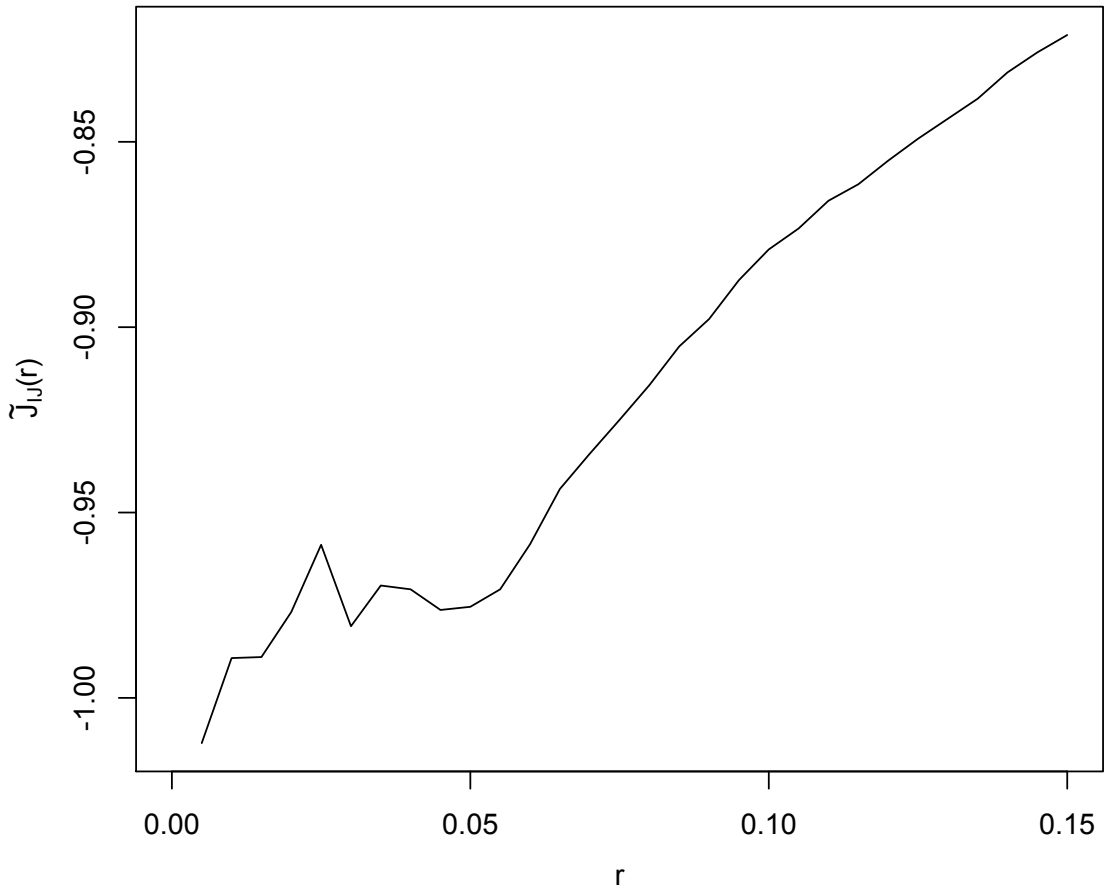

Figure 3.3: Estimates of the weighted cross J-function, $\tilde{J}_{ij}$, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Clustering between simulated N_i and simulated N_j.

a radius of 0.05 of any point of N_i. The estimate of the weighted cross J-function produces positive results which confirms inhibition within radius of 0.05 (Figure 3.5). However, the results of the estimate of the unweighted cross J-function does not clearly exhibit inhibition, as shown in Figure 3.6.

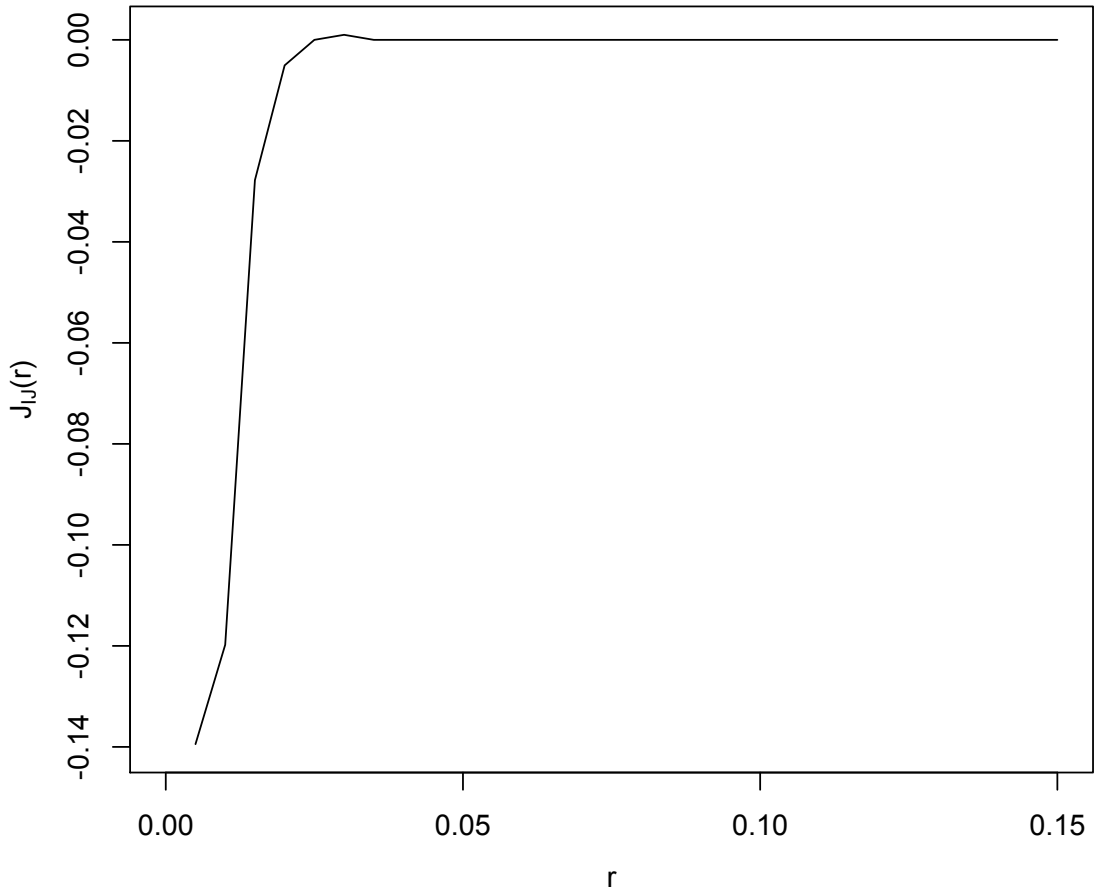

Figure 3.4: Estimates of the unweighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Clustering between simulated N_i and simulated N_j.

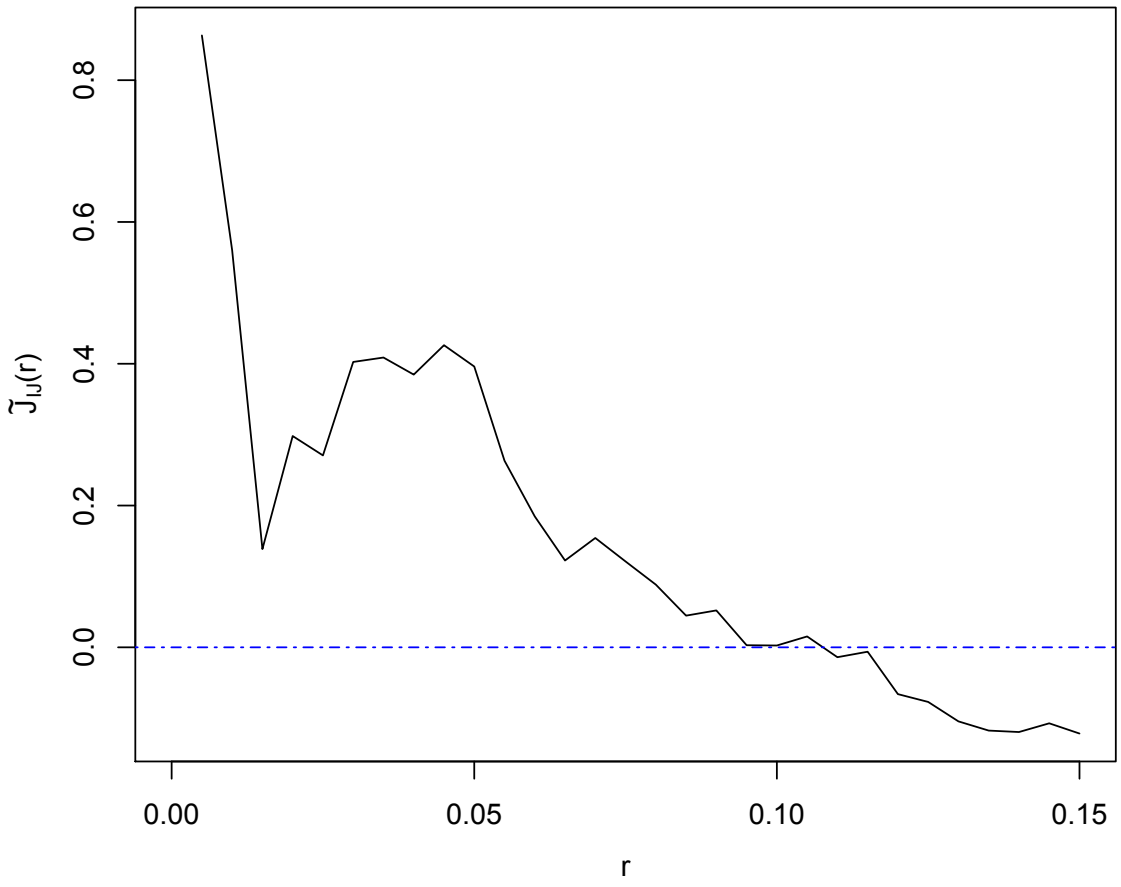

Figure 3.5: Estimates of the weighted cross J-function, $\tilde{J}_{ij}$, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Inhibition between simulated N_i and simulated N_j.

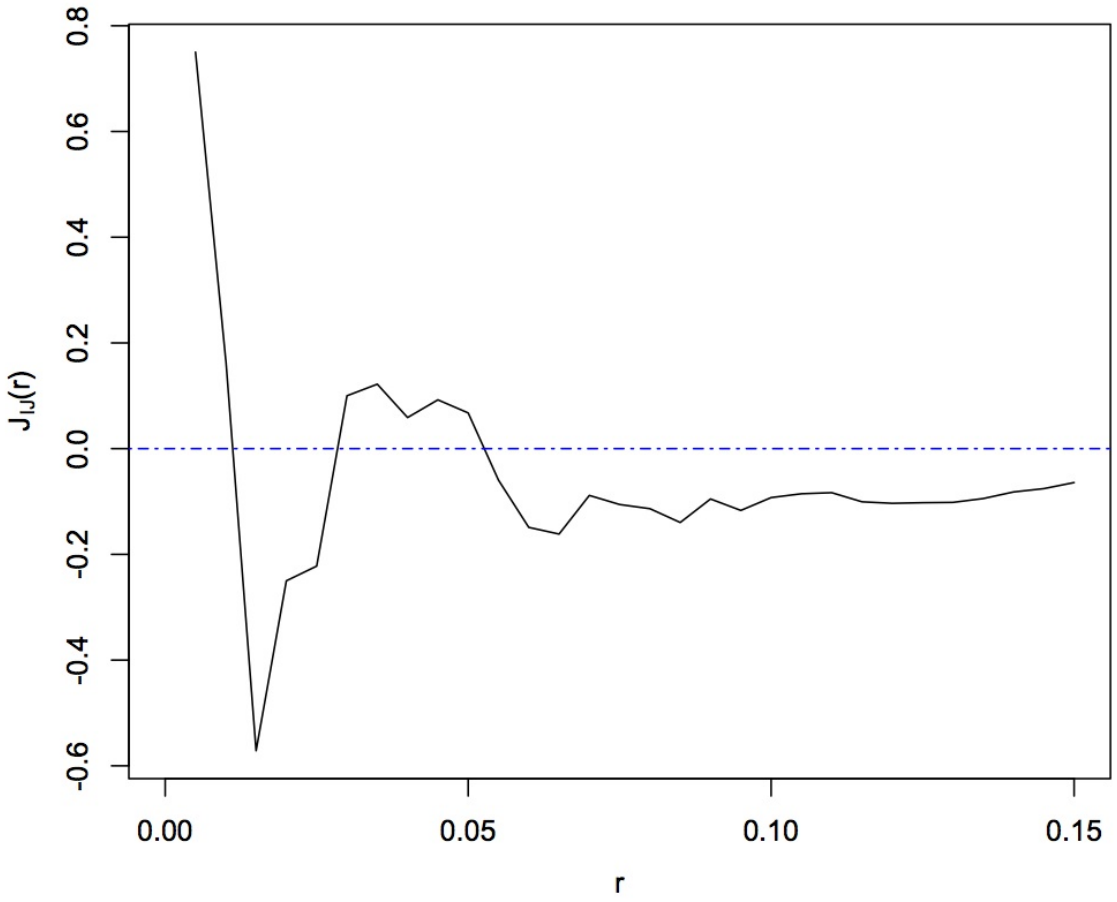

Figure 3.6: Estimates of the unweighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Inhibition between simulated N_i and simulated N_j.

For the simulation demonstrating independence between events, process N_i and N_j are each simulated from an inhomogeneous Poisson process with a triangular intensity on the square [0, 1] x [0, 1]. Hence the plots should show independence. We expect $\tilde{J}_{ij}(r)$ to equal 0. The plots for unweighted cross J-function shown in Figure 3.7 and weighted cross J-function shown in Figure 3.8 should look the same because there is no interaction between the two groups. Therefore the weights should not matter. The plots show $\tilde{J}_{ij}(r)$ to approximately equal 0, the discrepancies can be due to noise or edge effects.

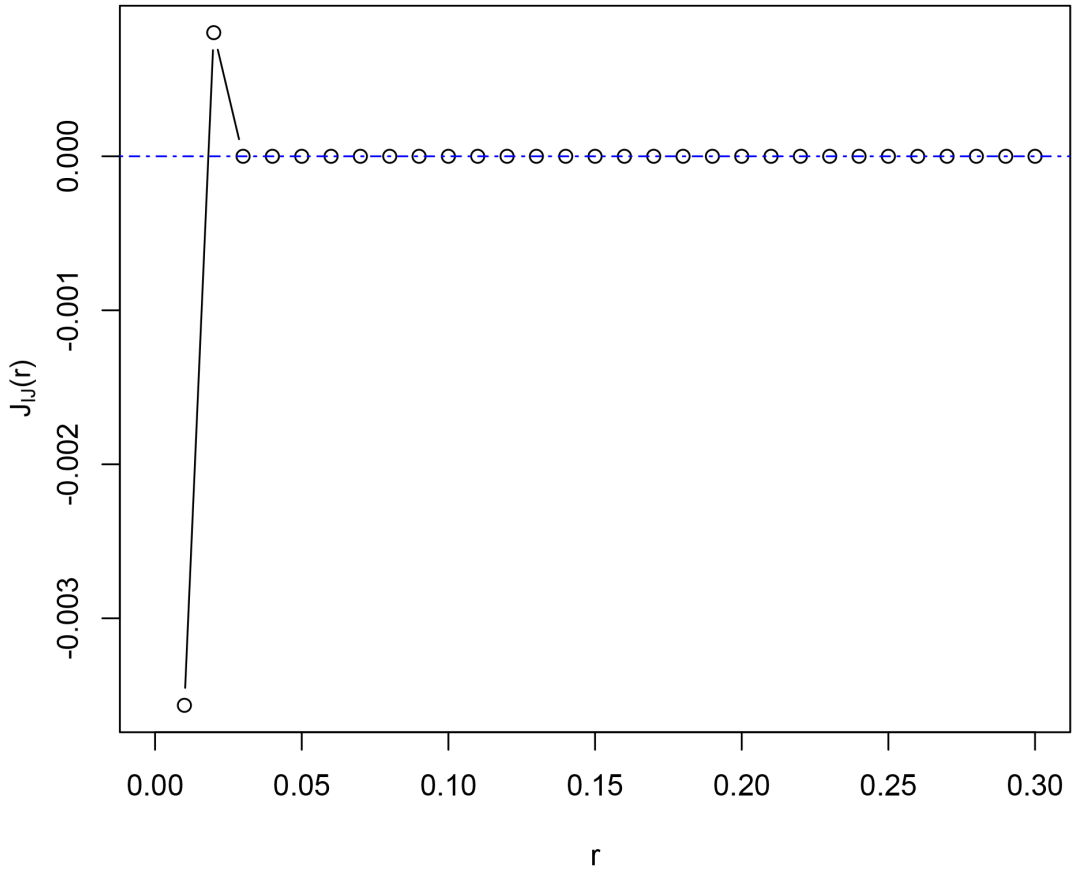

Figure 3.7: Estimates of the unweighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Independence between simulated N_i and simulated N_j.

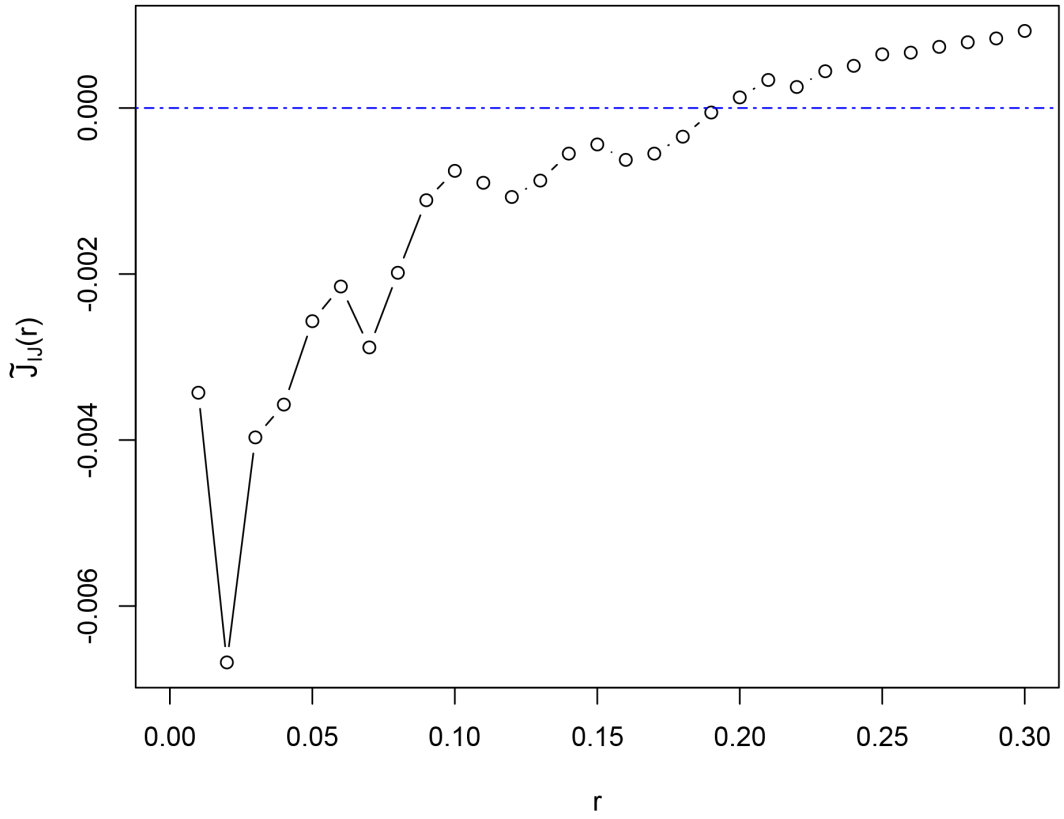

Figure 3.8: Estimates of the weighted cross J-function, $\tilde{J}_{ij}$, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Independence between simulated N_i and simulated N_j.

Additionally, we have provided simulations of different scales and sizes to demonstrate the successful estimation of the weighted cross J-function, as seen in Figures 3.9 through 3.12. In Figure 3.9, process N_i contains 1000 points for all the plots whereas process N_j varies between 2000 points, which means 2 points of process N_j are placed near each point of process N_i within radius r, and 10000 points, which means 10 points of process N_j are placed near each point of process N_i within radius r, where the radius varies between 0.03 and 0.10. In Figure 3.10, process N_i consists of 2000 points for all the plots whereas process N_j varies between 2000 points, which means 2 points of process N_j are placed near each point of process N_i within radius r, and 5000 points, which means 5 points of process N_j are placed near each point of process N_i within radius r, where the radius varies between 0.05 and 0.10. Figure 3.11 shows the estimate of the weighted cross J-function where process N_i contains 3000 points and process N_j contains either 2000 points, which means 2 points of process N_j are placed near each point of process N_i within radius 0.08, or 3000 points, which means 3 points of process N_j are placed near each point of process N_i within radius 0.08. Figure 3.12 shows the estimate of the weighted cross J-function where process N_i contains 5000 points and 5 points of process N_j are placed near each point of process N_i within radius 0.05.

These simulations demonstrate how the estimate of the weighted cross J-function behaves under different scenarios. The sharpness of the dip of the estimate of the weighted cross J-function seems to be connected with the level of clustering, i.e. the number of points of process N_j around each point of process N_i. When there is more intense clustering, meaning the number of points of process N_j increases around each point of process N_i, the dip is generally sharper as seen in the following plots (1) $N_i = 1000$, $N_j = 5000$ and r = 0.03, (2) $N_i = 2000$, $N_j = 5000$ and r = 0.05, (3) $N_i = 3000$, $N_j = 3000$ and r = 0.08, and (4) $N_i = 5000$, $N_j = 25000$ and r = 0.05. Similarly, as the radius of interaction decreases, the dip of the estimate becomes sharper, as seen in Figure 3.9 for $N_j = 2000$ and $N_j = 5000$ at r = 0.03. When the clustering is less intense, e.g. only 2 points of process N_j around each point of process N_i, then the dip is much less pronounced and sometimes not even noticeable. However, estimating the exact number of N_j per N_i from the weighted cross

J-function seems difficult because of the signal to noise ratio.

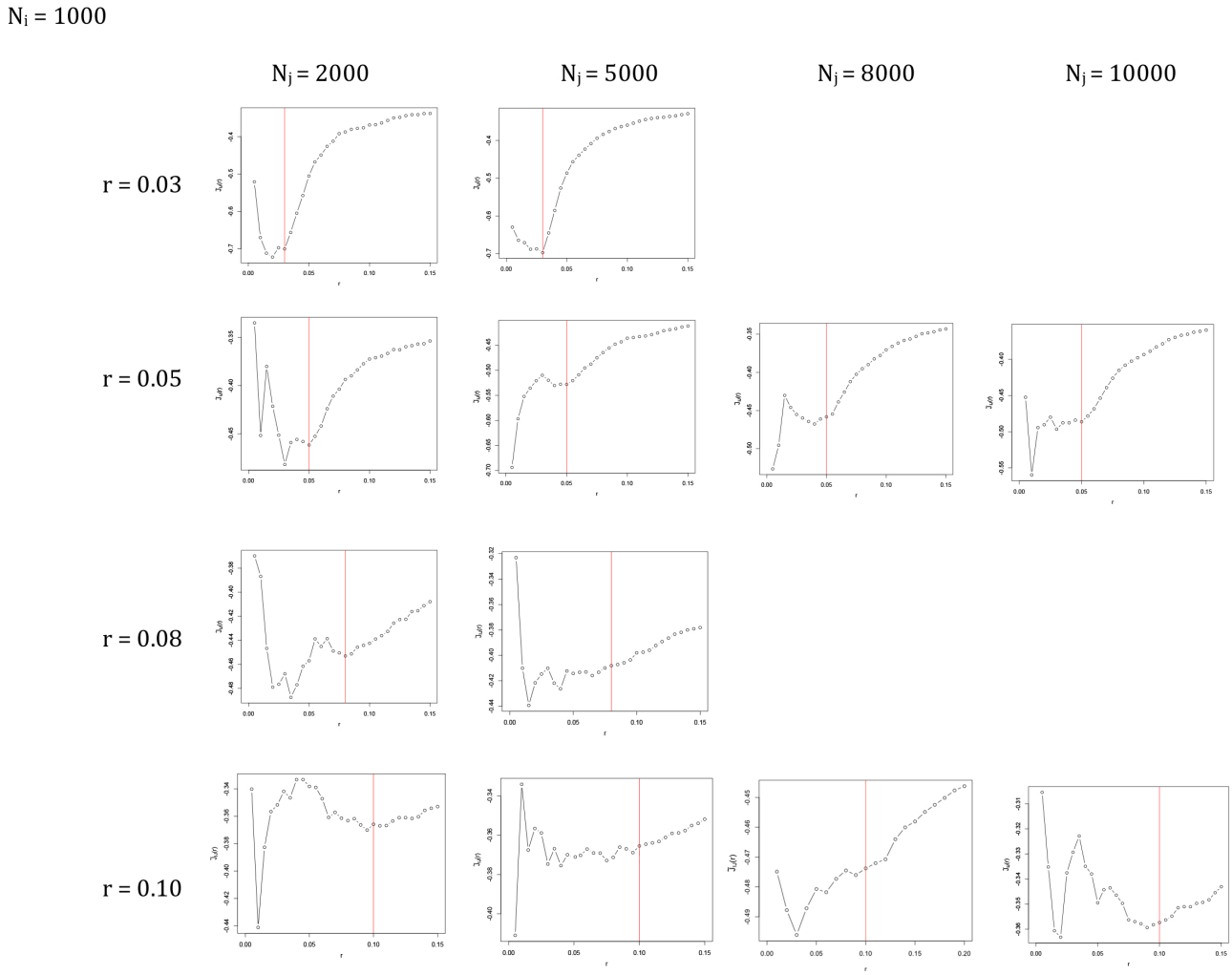

Figure 3.9: Estimates of the weighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Clustering between simulated N_i and simulated N_j with 1000 points of N_i and n points of N_j placed near each N_i within radius r.

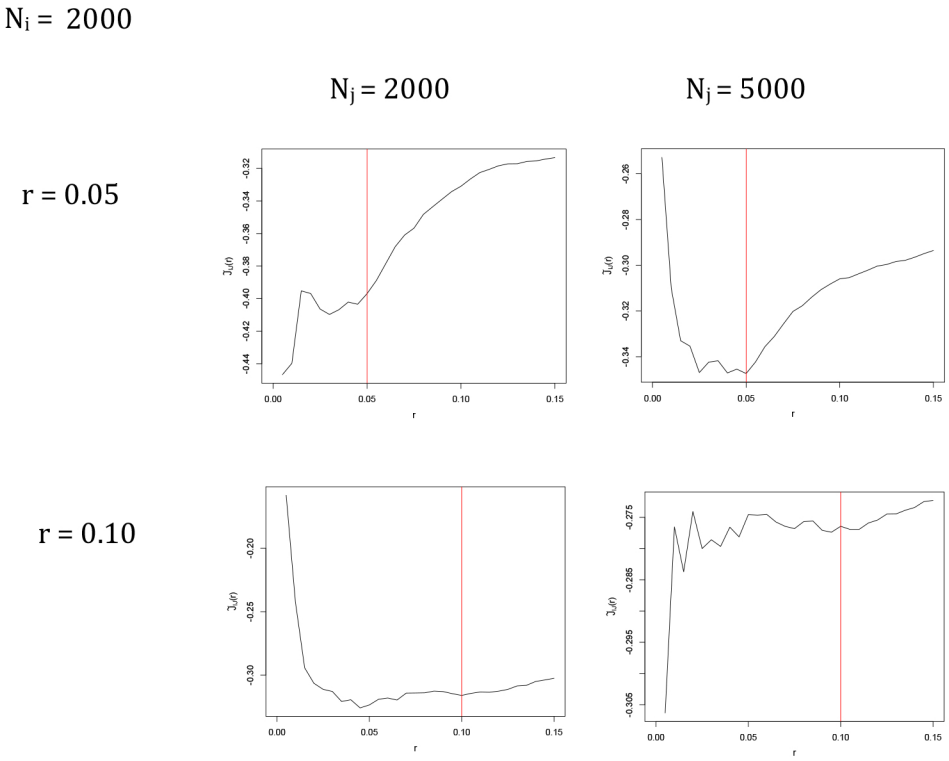

Figure 3.10: Estimates of the weighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Clustering between simulated N_i and simulated N_j with 2000 points of N_i and n points of N_j placed near each N_i within radius r.

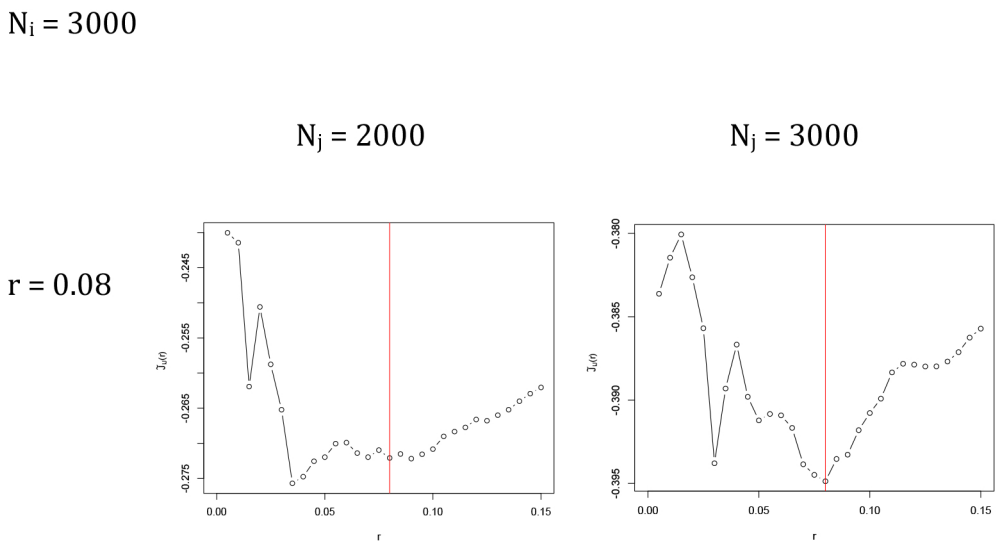

Figure 3.11: Estimates of the weighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Clustering between simulated N_i and simulated N_j with 3000 points of N_i and n points of N_j placed near each N_i within radius 0.08.

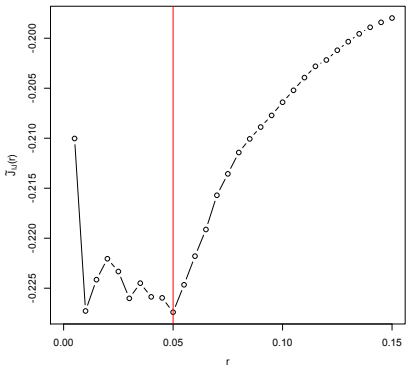

Figure 3.12: Estimates of the weighted cross J-function, J_{ij}, between simulated N_i and simulated N_j, as a function of radius r, for various choices of pairs of point patterns N_i and N_j. Clustering between simulated N_i and simulated N_j with 5000 points of N_i and 5 points of N_j placed near each N_i within radius 0.05.

CHAPTER 4

Application of Weighted Cross J-Function

4.1 Introduction

The main focus of this dissertation was inspired from a collaboration with Dr. Kevin Njabo, UCLA Institute of the Environment and Sustainability Center for Tropical Research Africa Director and Assistant Adjunct Professor, to address why the Avian Flu disease is still present in Egypt, by statistically analyzing and modeling how it is spreading, which might suggest a preventive solution to it spreading further. We use spatial statistics techniques to analyze a dataset collected by Dr. Njabo which contains information regarding infected and not infected animals in Egypt villages. Often these types of data are analyzed using gridded, geostatistical techniques but perhaps additional information can be gleaned from a point process approach which does not require the analyst to count totals over grids but instead use detailed spatial information on each observation. We are able to view this dataset as a point process since it contains both location and time of events in a space. Hence we can apply techniques, such as nearest neighbor methods which show interactions between infected and non-infected animals, and spatial time plots to show how the events change over time. Through applying point process techniques we developed a new method, weighted cross J-function, which applies weights to the cross J-function to help us pinpoint exact moments of interaction between two groups besides the type and range of interaction.

The application of spatial statistics techniques to these types of data may help us to understand the characteristics of the disease which may then allow us to come up with a preventive solution. The idea of applying spatial statistics in this particular way to epidemiological data can potentially open new doors of research and help us understand other

diseases across the world in a new way.

4.2 African Avian Flu Data

With the collaborative efforts of Egyptian Animal Health Research Institute/National Laboratory of Quality Control of Poultry Production (AHRI/NLQP) and the General Organization of Veterinary Services (GOVS), the data were collected from four Egyptian governorates: Damietta, El Gharbia, Fayoum and Menofia. Six types of birds, chickens, ducks, geese, turkeys, pigeons and wild birds, were sampled and tested for the highly pathogenic avian influenza virus (HPAIV), H5N1, between the years 2009 and 2012. The observation window of the data is between longitudes 30.4333 and 31.8739 and latitudes 28.3333 and 31.5250. Figure 4.1 shows the map of Egypt with points representing the villages which were tested for the avian flu. In the dataset, if a bird was tested and found to be non-infected with the disease a 0 was recorded. If a bird is found to be infected, a 1 was recorded. If multiple birds in the same village were found to be infected then a number pertaining to the amount of infected birds was recorded. For example, if 3 infected birds were found then a 3 was recorded. Table 4.1 shows the counts of each infected and non-infected bird type. The highest counts of infection were found in chickens and ducks. Out of all the pigeons and wild birds that were tested, only 1 pigeon was tested with infection and no wild birds were found with infection. The villages that were studied were not chosen at random, but were chosen based on the researchers' path of travel. Nearby villages might share the same coordinates in the data. The coordinates of the tested birds were recorded based on which village they were tested in.

	Chickens	Ducks	Geese	Turkeys	Pigeons	Wild Birds
Infected	365	317	48	21	1	0
Non-infected	194	238	511	538	558	559

Table 4.1: The counts of infected and non-infected of each of the six bird types in the dataset.

Table 4.2 displays the counts of the infected birds by year. It appears that the highest

numbers of infected chickens were found in 2010 and 2011, 140 recordings and 162 recordings respectively, whereas the number drops to less than half in 2012. As for infected ducks, the lowest recorded amount was 17 recordings in 2009 and the count grew very quickly to 110 recordings in 2010 and it continued at 95 recordings for 2011 and 2012. The counts of infected geese tend to be low, 5 recordings in 2009 with an increase to 21 recordings 2010, then 13 recordings in 2011 and 9 recordings in 2012. The counts of infected turkeys stay stable at 49 recordings from 2009 and 2011 and then drops drastically to 3 recordings. There was only 1 recording of an infected pigeon during the entire research period, in year 2010, and no infected wild birds were recorded.

	Chickens	Ducks	Geese	Turkeys	Pigeon	Wild Birds
2009	14	17	5	49	0	0
2010	140	110	21	49	1	0
2011	162	95	13	49	0	0
2012	49	95	9	3	0	0

Table 4.2: The counts of infected birds by year.

Tables 4.3 and 4.4 show the counts of each infected and non-infected bird type subsetted by each of the four governorates in Egypt. Overall, the largest amount of birds were sampled is in the governorate of Menofia, which is in the center of the tested governorates near El Gharbia, whereas the least amount were sampled in the governorate of Damietta, which is the northern most tested governorate. In general, the most infection was found in Menofia and the least in Damietta. Specifically, most of the chickens were tested in Menofia where 154 were recorded to be infected. There were 86 and 81 recordings of infected chickens in El Gharbia and Fayoum, respectively and the lowest recording of 44 infected chickens was in Damietta. Similarly, most of the ducks were tested in Menofia where 169 infected ducks were recorded. Approximately 55 infected ducks were recorded in El Gharbia and Fayoum and 38 infected ducks were recorded in Damietta. The most recorded infected geese were found in Fayoum and Menofia whereas less than 5 geese were recorded with infection in Damietta and El Gharbia. Interestingly, the most infection amongst turkeys was recorded in Damietta,

where 10 infected turkeys were tested. There were less than 5 infected turkeys found in each of the other three governorates that were tested.

	Chickens		Ducks		Geese	
Governorate	Infected	Non-Infected	Infected	Non-Infected	Infected	Non-Infected
Damietta	44	14	38	17	3	55
El Gharbia	86	24	54	56	4	106
Fayoum	81	15	56	40	20	76
Menofia	154	141	169	125	21	274

Table 4.3: The counts of infected and non-infected of chickens, ducks and geese by governorate.

	Turkeys		Pigeons		Wild Birds	
Governorate	Infected	Non-Infected	Infected	Non-Infected	Infected	Non-Infected
Damietta	10	48	0	58	0	58
El Gharbia	2	108	0	110	0	110
Fayoum	4	92	1	95	0	96
Menofia	5	290	0	295	0	295

Table 4.4: The counts of infected and non-infected of turkeys, pigeons and wild birds by governorate.

Figure 4.1 shows the map of Egypt with the capital, Cairo, and the Nile River. The green villages represent all the villages that were tested for birds infected with the avian flu. The four governorates have also been marked on the map to highlight the regions where the data were collected.

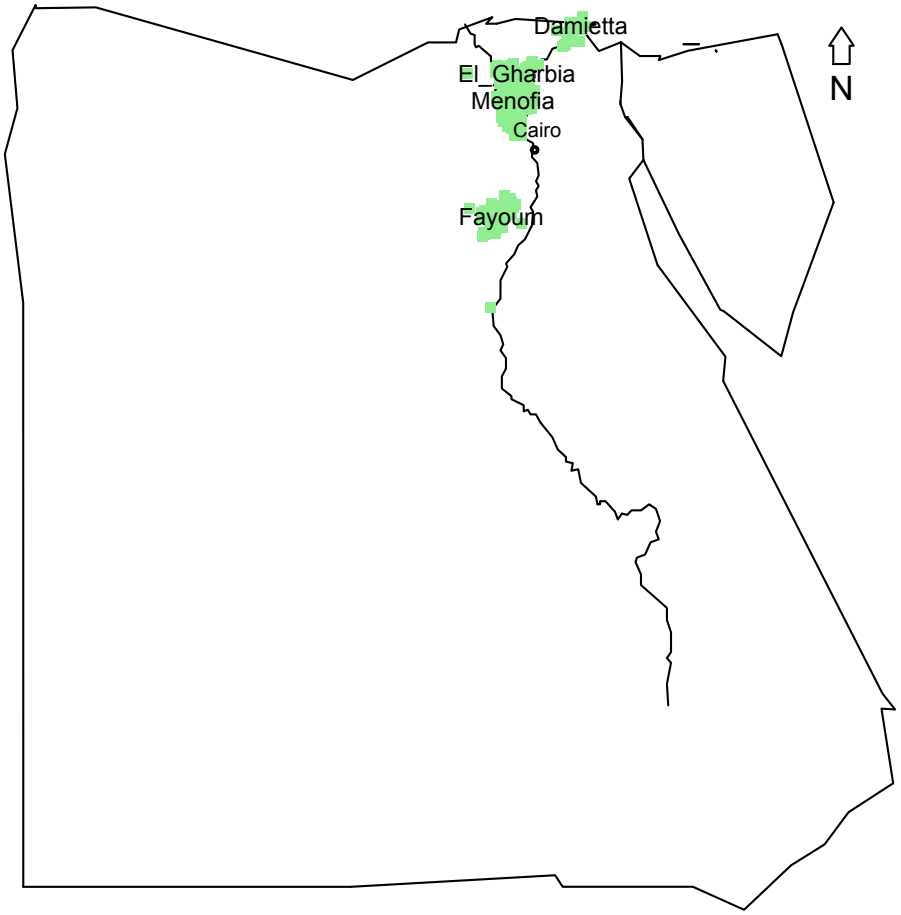

Figure 4.1: Map of Egypt. The green points represent the villages that were tested for Highly Pathogenic Avian Influenza Virus (HPAIV), H5N1, during the years 2009-2012. Also, the 4 governorates, Damietta, El Gharbia, Fayoum and Menofia, are labeled.

Infected and non-infected birds were plotted within the observed area across four years (2009-2012) to visualize the infection, as shown in Figures 4.2 and 4.3. Based on our results, most of the infection was found in chickens and the least amount of infection was found in wild birds. Within the infected birds, multiple cases of infection were found in chickens and ducks in the southern villages, which might suggest the disease was more dominant in those areas.

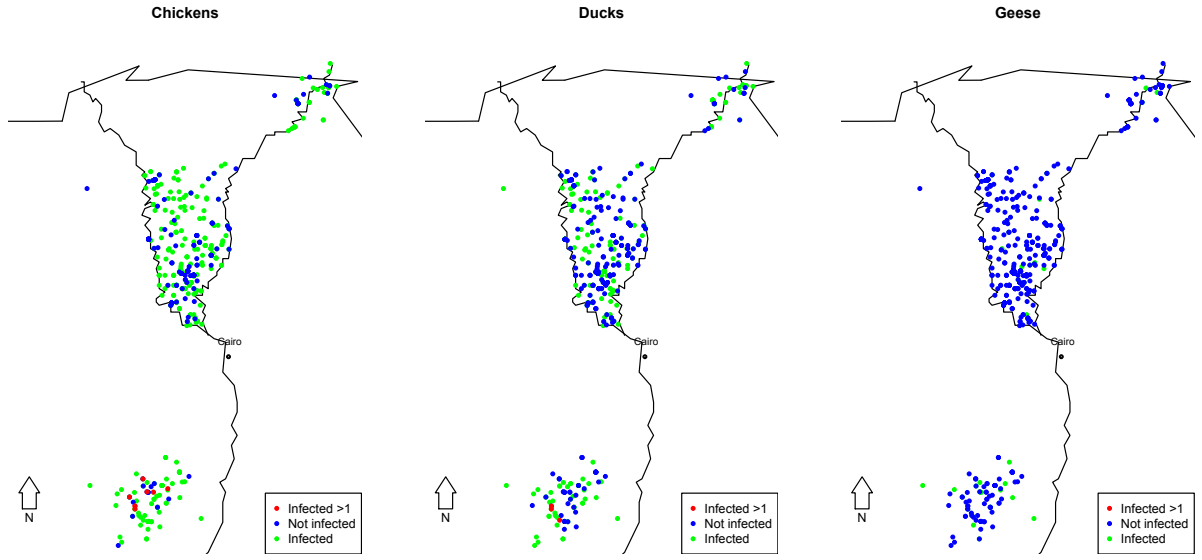

Figure 4.2: Maps of infected and non-infected animals (chickens, ducks, geese) years 2009-2012. The green points represent the locations where infected birds were found. The red points represent the locations where more than one infected bird was found, and the blue points represent the locations where no infected birds were found.

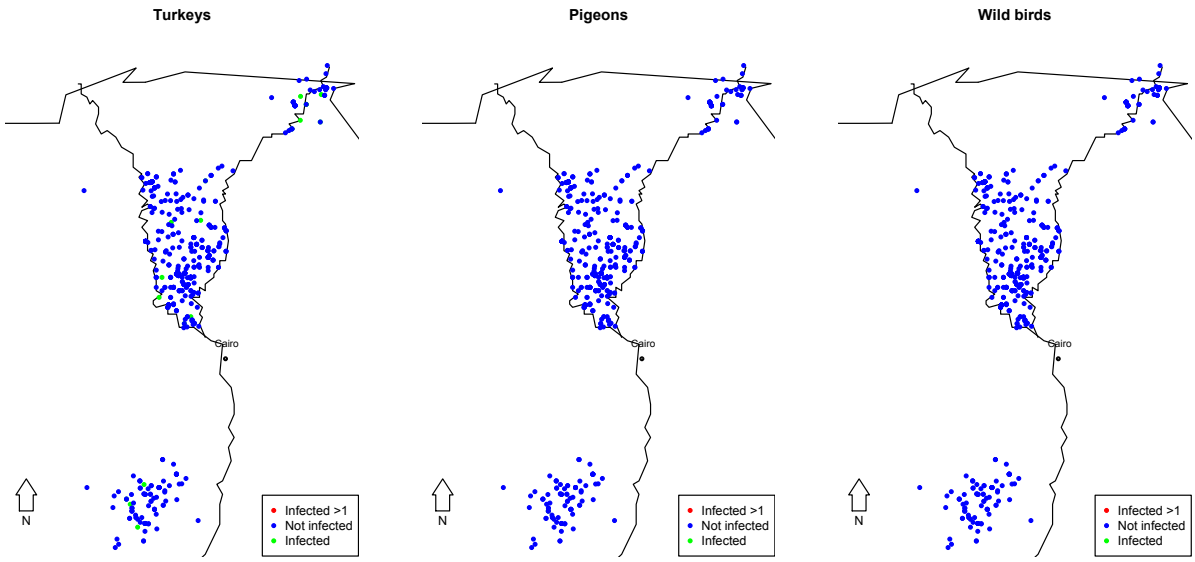

Figure 4.3: Maps of infected and non-infected animals (turkeys, pigeons, wild birds) years 2009-2012. The green points represent the locations where infected birds were found. The red points represent the locations where more than one infected bird was found, and the blue points represent the locations where no infected birds were found.

We can see the amount of infection amongst each group of birds in Figures 4.4 and 4.5 which only show where the infection amongst the birds is present. Each map represents one bird type. Again, it can be seen that multiple cases of infection amongst birds, specifically chickens and ducks, are present in the southern villages on the map. Very little infection was recorded in pigeons and wild birds.

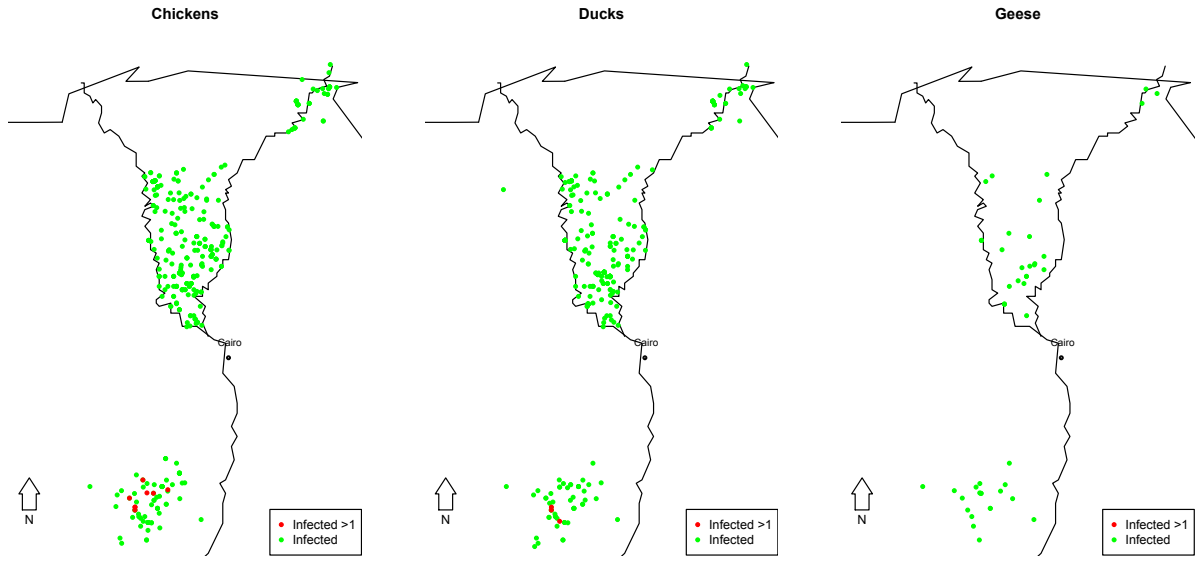

Figure 4.4: Maps of only infected animals (chickens, ducks, geese) years 2009-2012. The green points represent the locations where infected birds were found. The red points represent the locations where multiple infected birds were found.

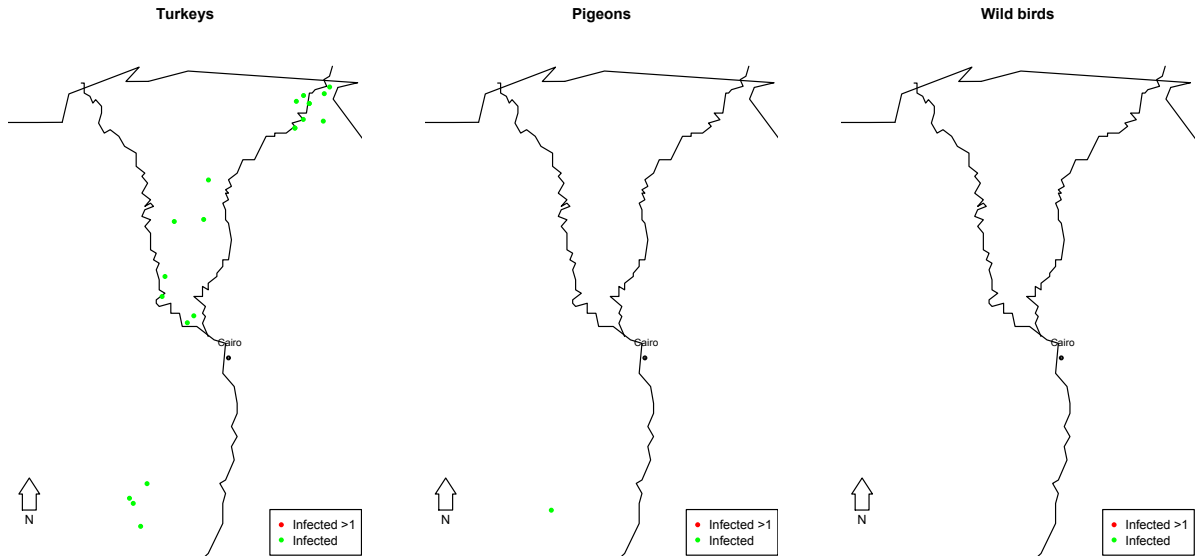

Figure 4.5: Maps of only infected animals (turkey, pigeon, wild birds) years 2009-2012. The green points represent the locations where infected birds were found. The red points represent the locations where multiple infected birds were found.

To view the change in disease over time, we plotted all the infected birds by year, as shown in Figures 4.6 and 4.7. Each map represents one type of bird. Our results show that the disease was more present in the years 2010 and 2011. Also, the disease was more apparent in the southern villages in 2009 and slowly migrates to the northern villages over the following years. These maps help us to understand how the disease spread to different regions over time.

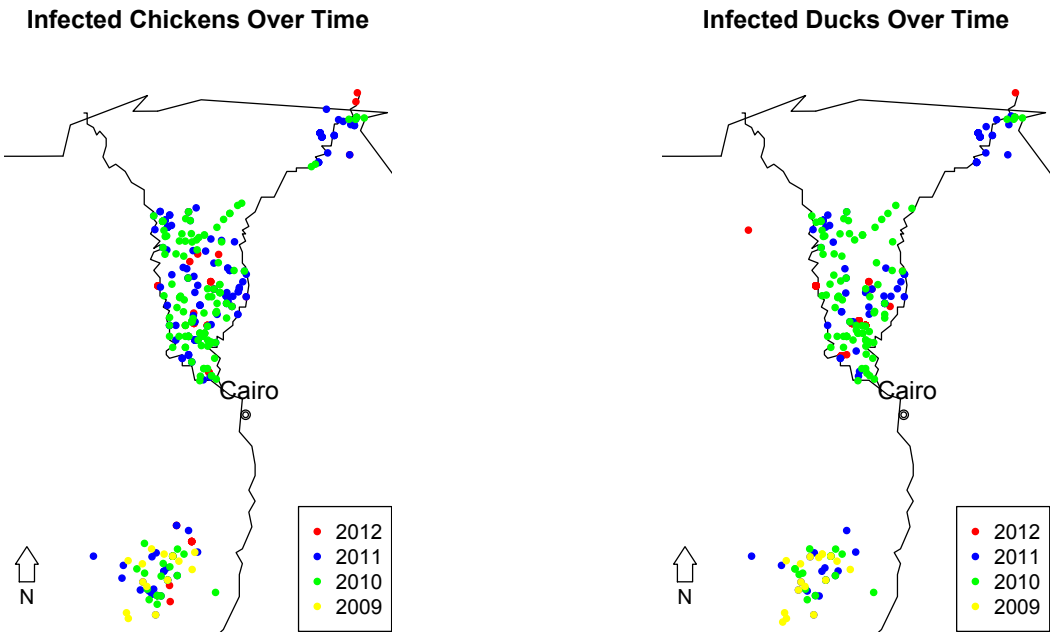

Figure 4.6: Maps of infected birds over time. Avian flu (H5N1) was more present in the years 2010 and 2011, and was more apparent in the southern villages in 2009 and appears to migrate to the northern villages in 2010 and 2011.

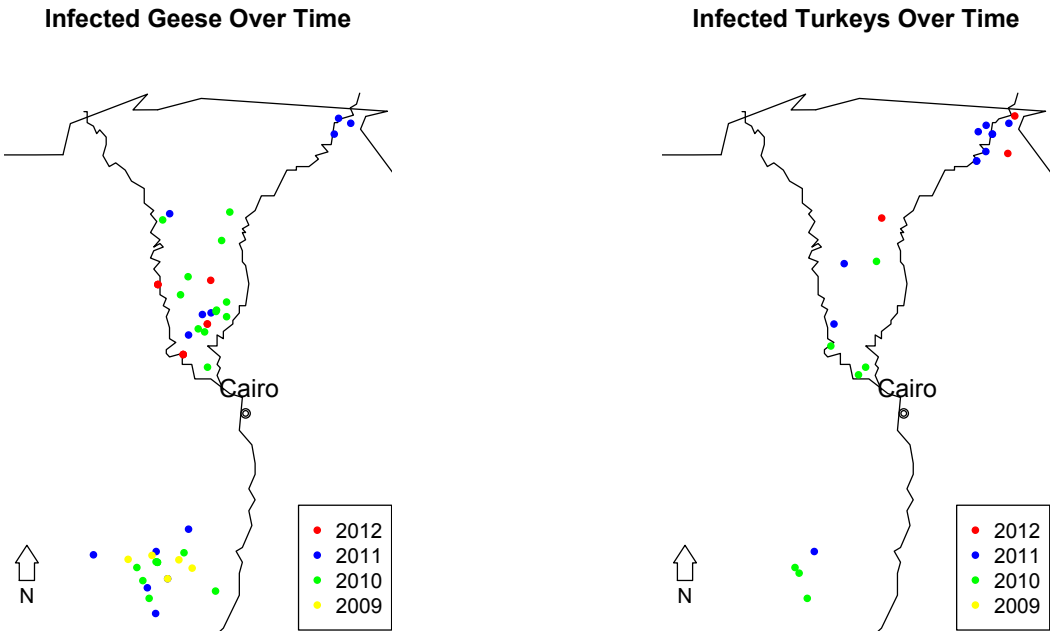

Figure 4.7: Maps of infected birds over time. Avian flu (H5N1) was more present in the years 2010 and 2011, and was more apparent in the southern villages in 2009 and appears to migrate to the northern villages in 2010 and 2011.

4.3 Application to African Avian Flu Data

As described earlier, a new summary statistic, the weighted cross J-function, was proposed while analyzing this dataset because the point patterns appeared to exhibit non-constant intensity rates. We extended the cross J-function to point patterns where the intensity is not constant, by incorporating weights for each point in the point pattern, with each weight corresponding to the inverse of the estimated intensity at its location, which resulted in more sensible and interpretable results in the case of inhomogeneous point processes. The weighted cross J-function was applied to each pair of birds in the dataset to inspect for spatial interactions.

Figures 4.8 through 4.13 display the weighted cross J-function estimates of each pair of birds, shown by the solid line, along with the upper and lower 95% bounds, shown by the dotted lines, of 100 simulations of the same intensity and observed in the same window as the data. As the radius, r, increases, the bounds seem to become narrower. When the radius, r, is small there is more interaction and when radius, r, increases less interaction occurs as expected.

Our results detected clustering, that is, a positive interaction between infected chickens, 365 recorded, and infected ducks, 317 recorded, within radius of 10.5 km, after inhomogeneity in both the chickens and ducks has been accounted for; there is a lack of apparent interaction at larger distances as seen in Figure 4.8. The results also suggest clustering present between all the other infected pairs of birds. There seems to be the most positive interactions between the following pairs, infected chickens, 365 recorded, and infected geese, 48 recorded, (Figure 4.9), infected ducks, 317 recorded, and infected turkeys, 21 recorded, (Figure 4.12), happening within radius of 21 km, especially around radius of 10.5 km, and it levels off after 42 km. When comparing infected chickens, 365 recorded, to infected turkeys, 21 recorded, in Figure 4.11, we see the most positive interaction occurring within radius of 31.5 km, especially up to radius of 21 km, and levels off after 42 km, whereas with infected geese, 48 recorded, and infected turkeys, 21 recorded (Figure 4.13) the most positive interaction also occurs within radius of 31.5 km, especially around 10.5 km to 21 km, but levels

off at a longer distance of 50 km. Similar to infected chickens and infected geese, infected ducks, 317 recorded, and infected geese, 48 recorded, (Figure 4.10) have the most interaction within radius of 21 km, with most of the interaction occurring at 10.5 km and it levels off after 31.5 km. The clustering seems to be mostly due to the inhomogeneity, so although the locations are correlated, the weighted cross J-function is not statistically significant.

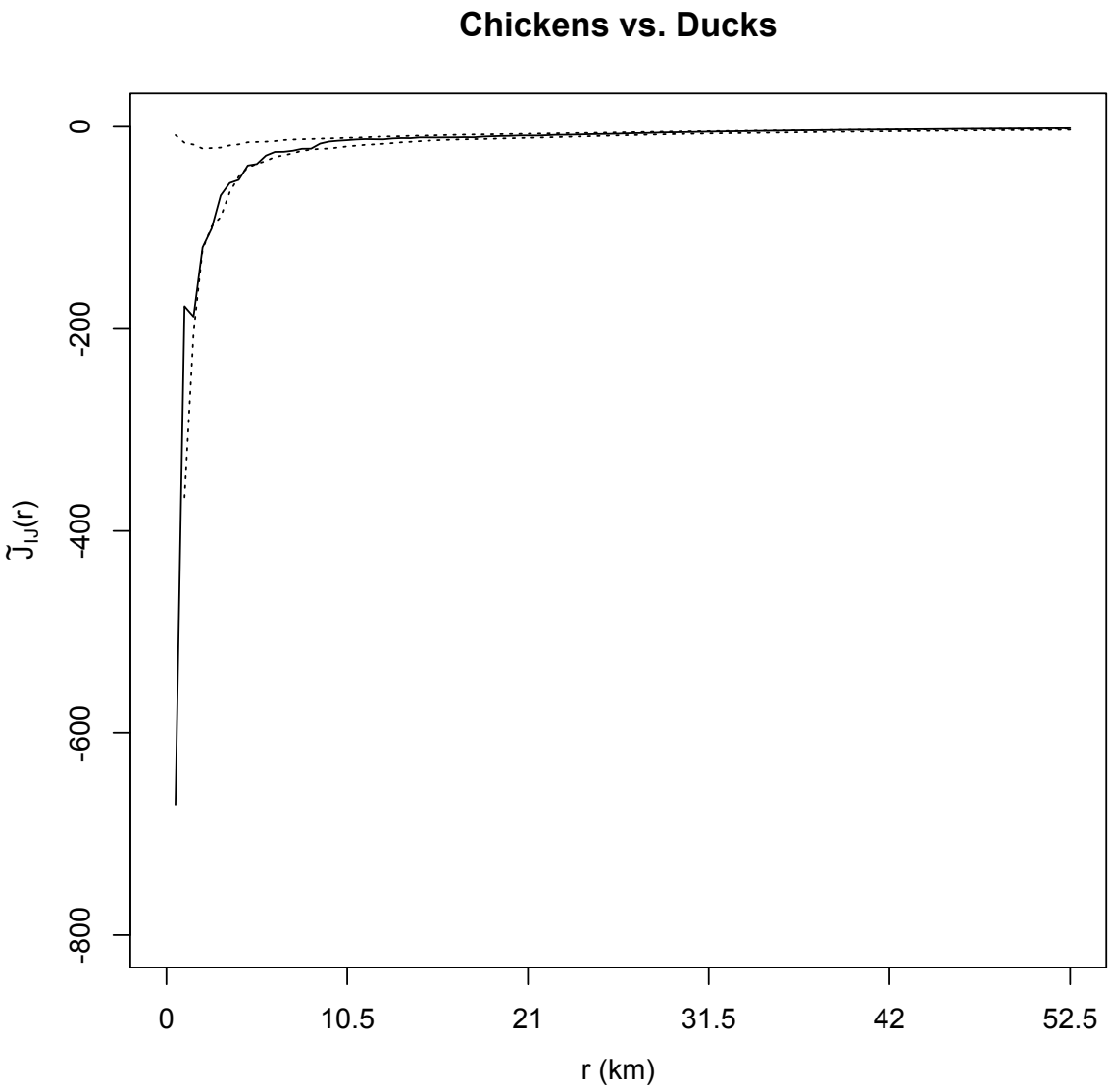

Figure 4.8: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected ducks, as a function of radius r, shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

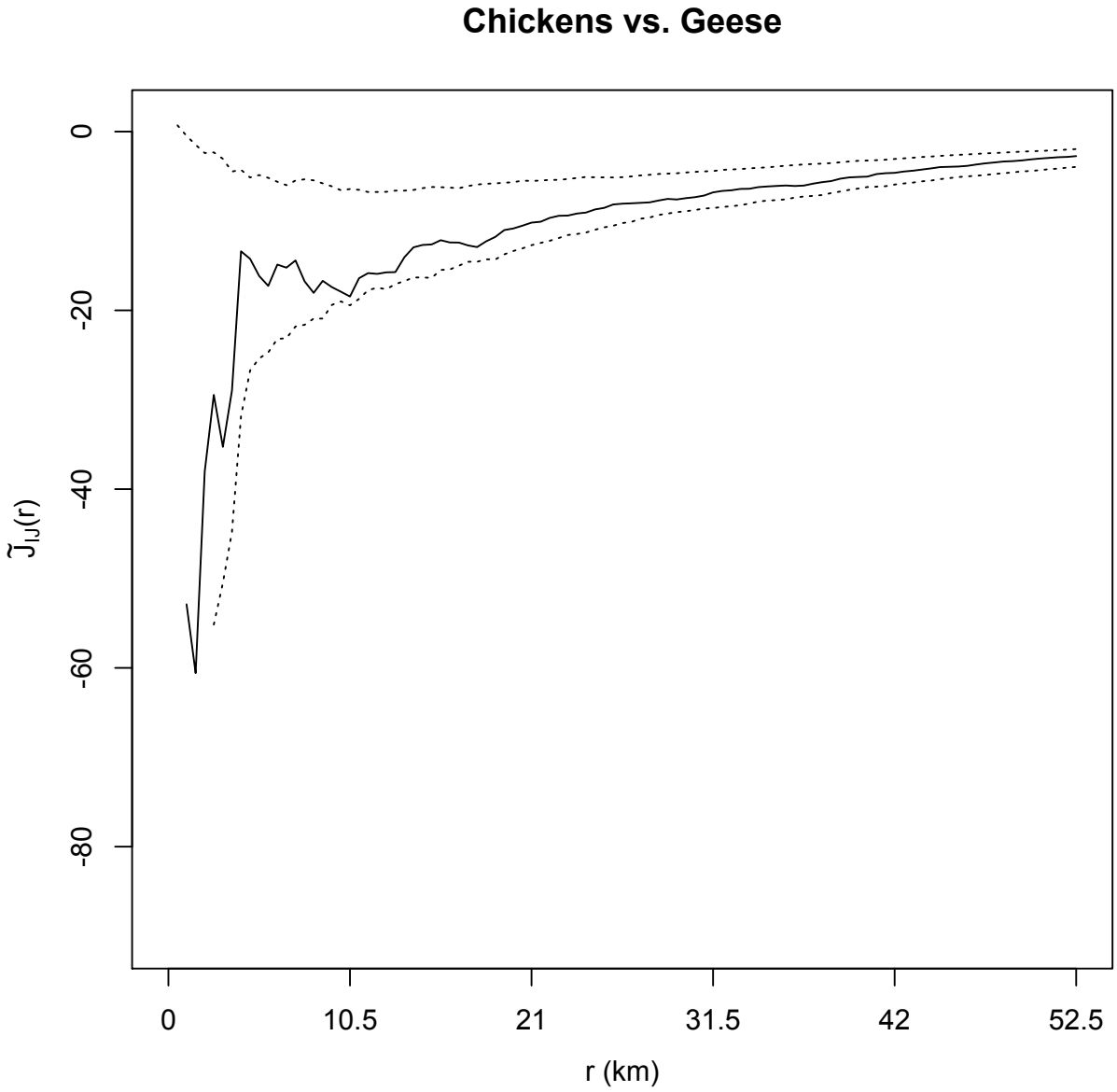

Figure 4.9: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected geese, as a function of radius r, shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

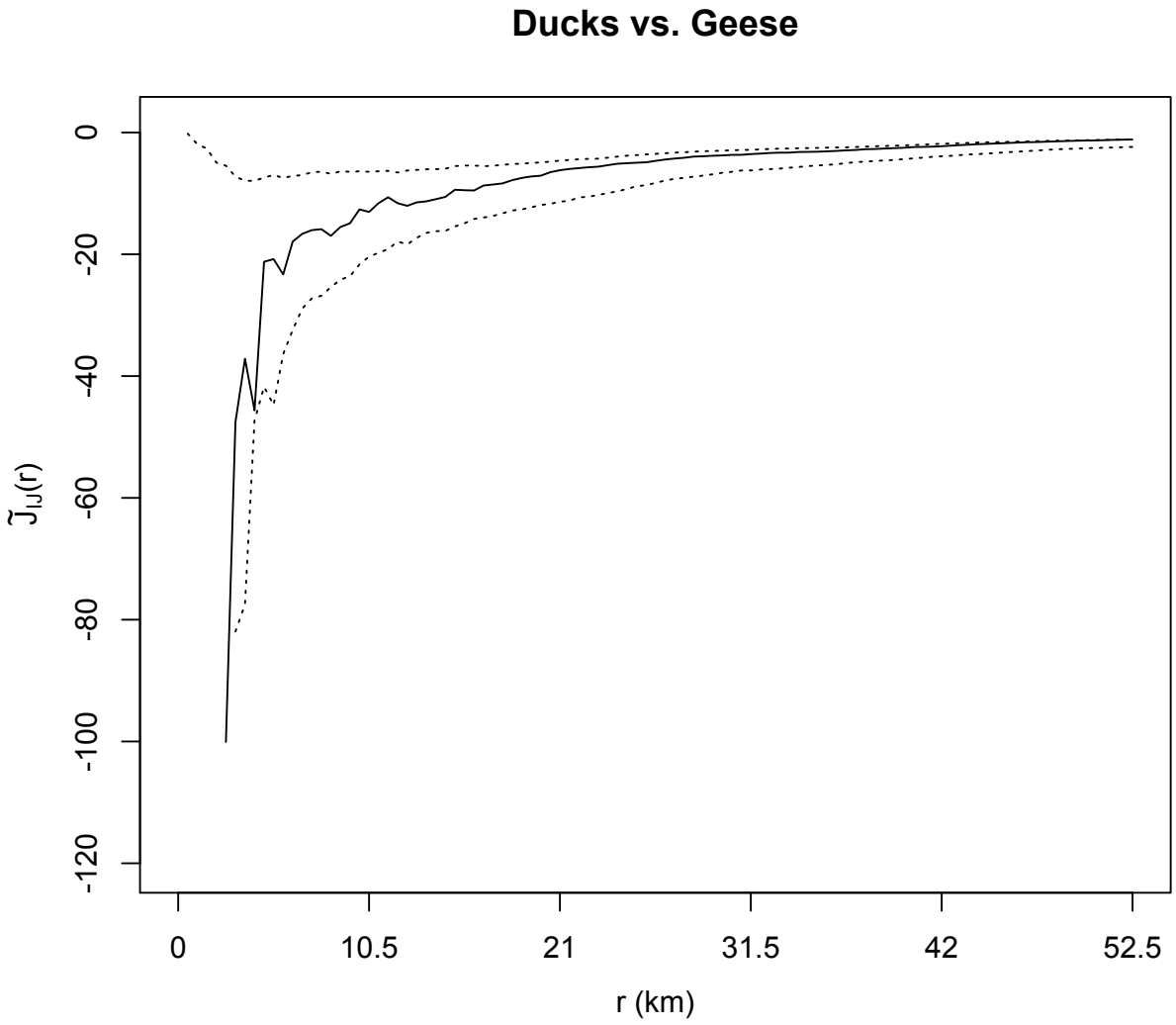

Figure 4.10: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected ducks, and N_j, infected geese, as a function of radius r, shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

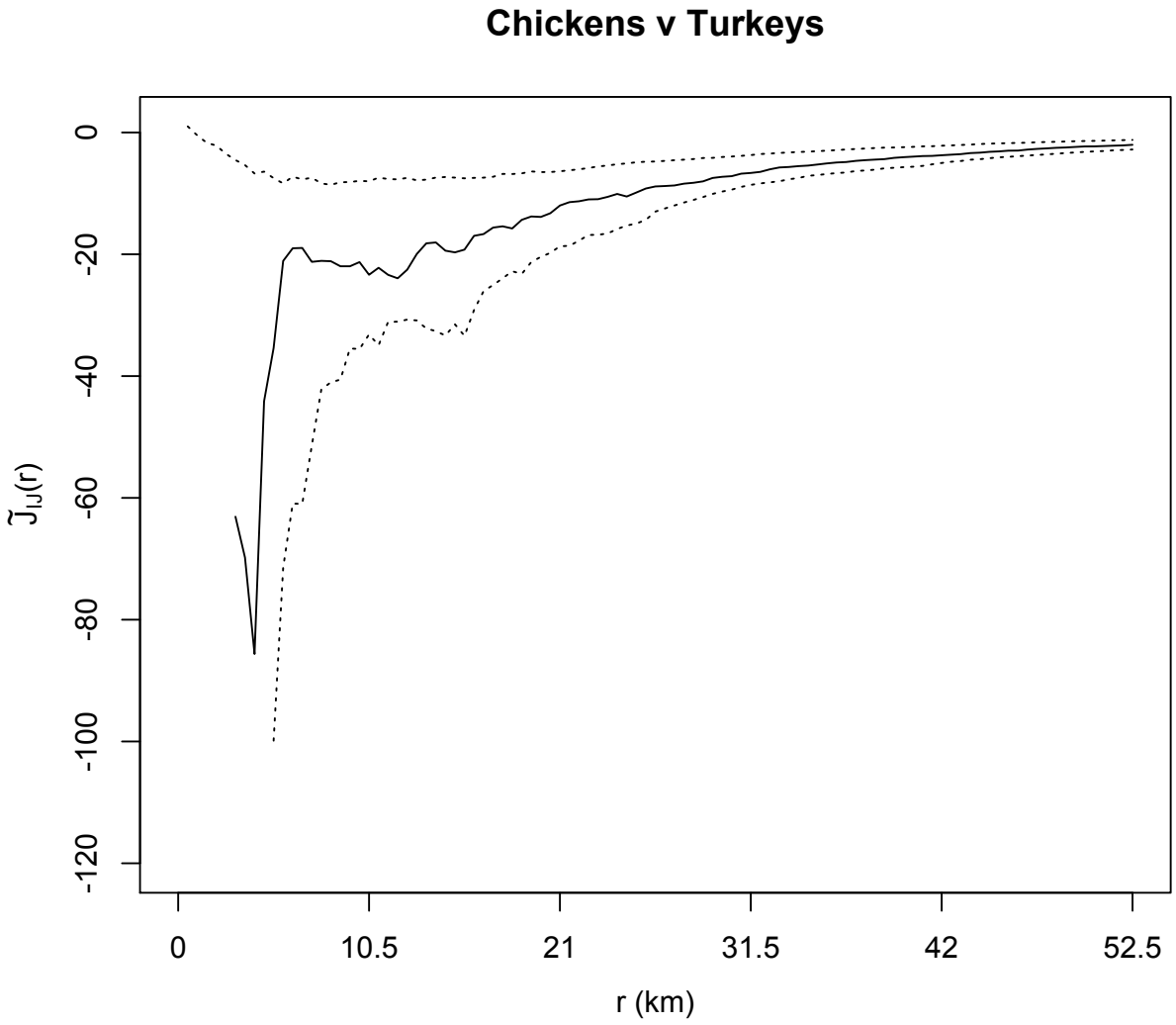

Figure 4.11: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected turkeys, as a function of radius r, shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

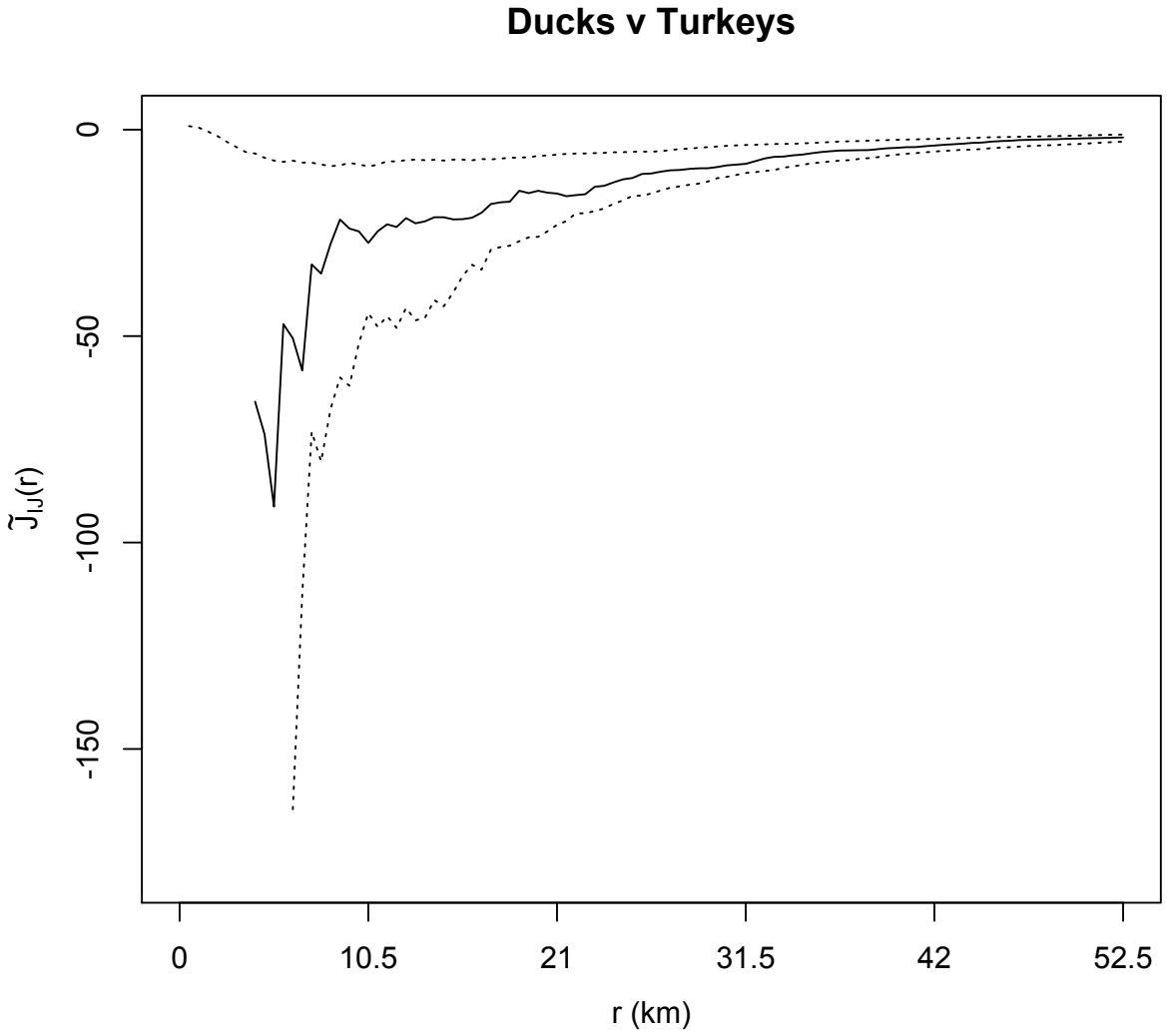

Figure 4.12: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$), between N_i, infected ducks, and N_j, infected turkeys, as a function of radius r, shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

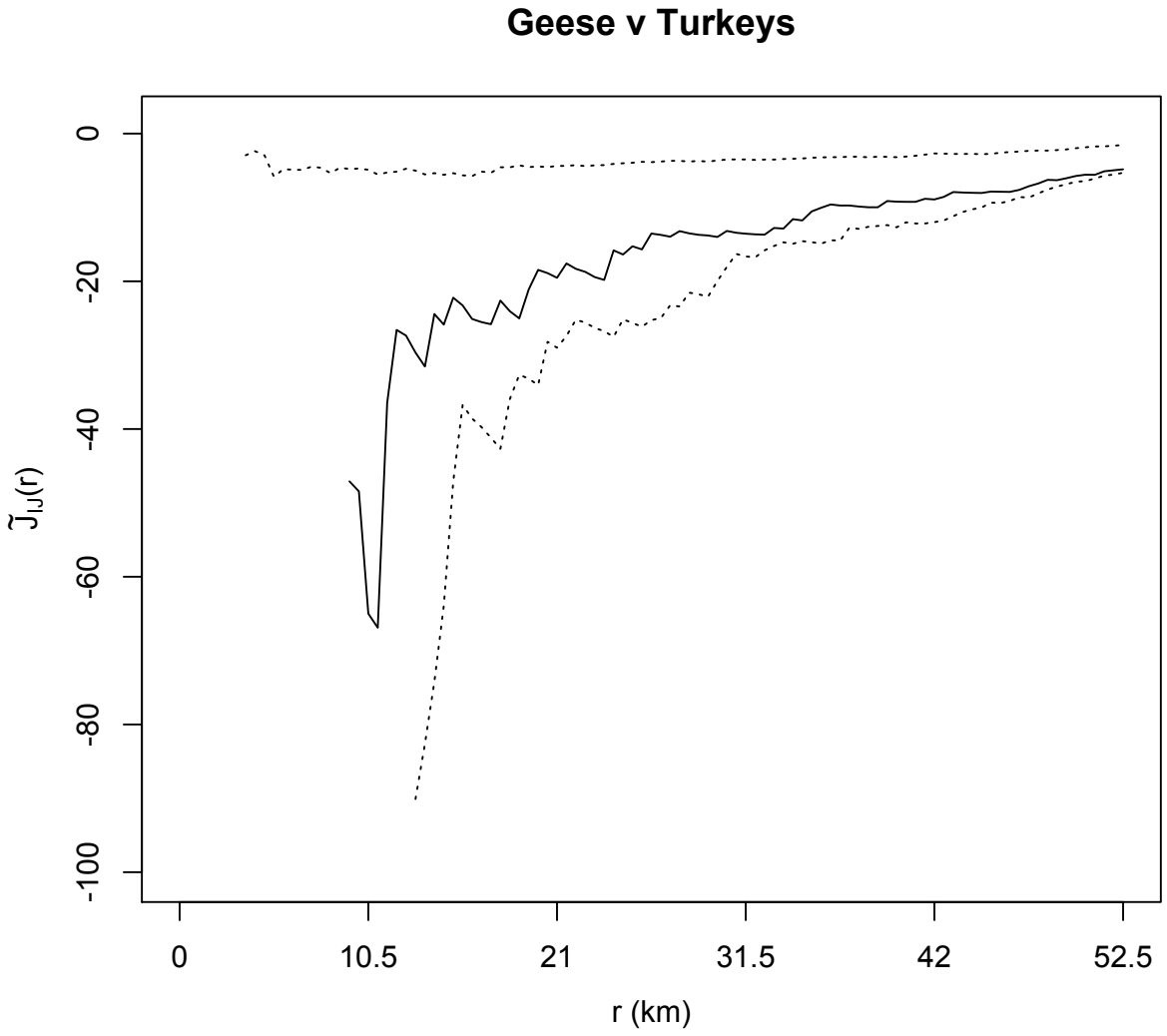

Figure 4.13: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected geese, and N_j, infected turkeys, as a function of radius r, shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

We wanted to explore the weighted cross J-function a bit further so we decided to look at the specific regions, or governorates, in the dataset. Since the governorate of Menofia had the most amount of data points collected we decided to start there. Figures 4.14 through 4.17 present the estimates of the weighted cross J-function for each set of birds, shown by the solid line, along with the upper and lower 95% bounds, shown by the dotted line, for 100 simulations of the same intensity and observed in the same window as the data. There were 154 recorded infected chickens and 169 infected ducks in Menofia. When looking at the interaction of infected chickens and infected ducks, (Figure 4.14), we saw that the estimate of the weighted cross J-function was similar to that of infected chickens and infected ducks of the whole recorded area in the dataset, where little positive interaction was occurring within radius 10.5 km. However the bounds tend to be much wider. Similar to the infected chickens and infected ducks, the infected ducks, 169 recorded, and infected geese, 21 recorded, had an estimate of the weighted cross J-function that was similar to that of the whole area, where most of the positive interaction or clustering occurred within radius of 21 km and leveled off afterwards (Figure 4.16). However the bounds tend to be much wider than that of the whole area. The estimate of the weighted cross J-function was slightly different for the infected chickens, 154 recorded, and infected geese, 21 recorded, of Menofia (Figure 4.15). Here we see clustering or positive interaction between infected chickens and infected geese within radius of 63 km, mainly before radius of 42km, whereas for the whole area we saw most of the activity within radius 21 km. As for the infected geese, 21 recorded, and infected turkeys, 5 recorded, of Menofia (Figure 4.17), we see clustering or positive interaction within radius of approximately 55 km similar to that of the whole area.

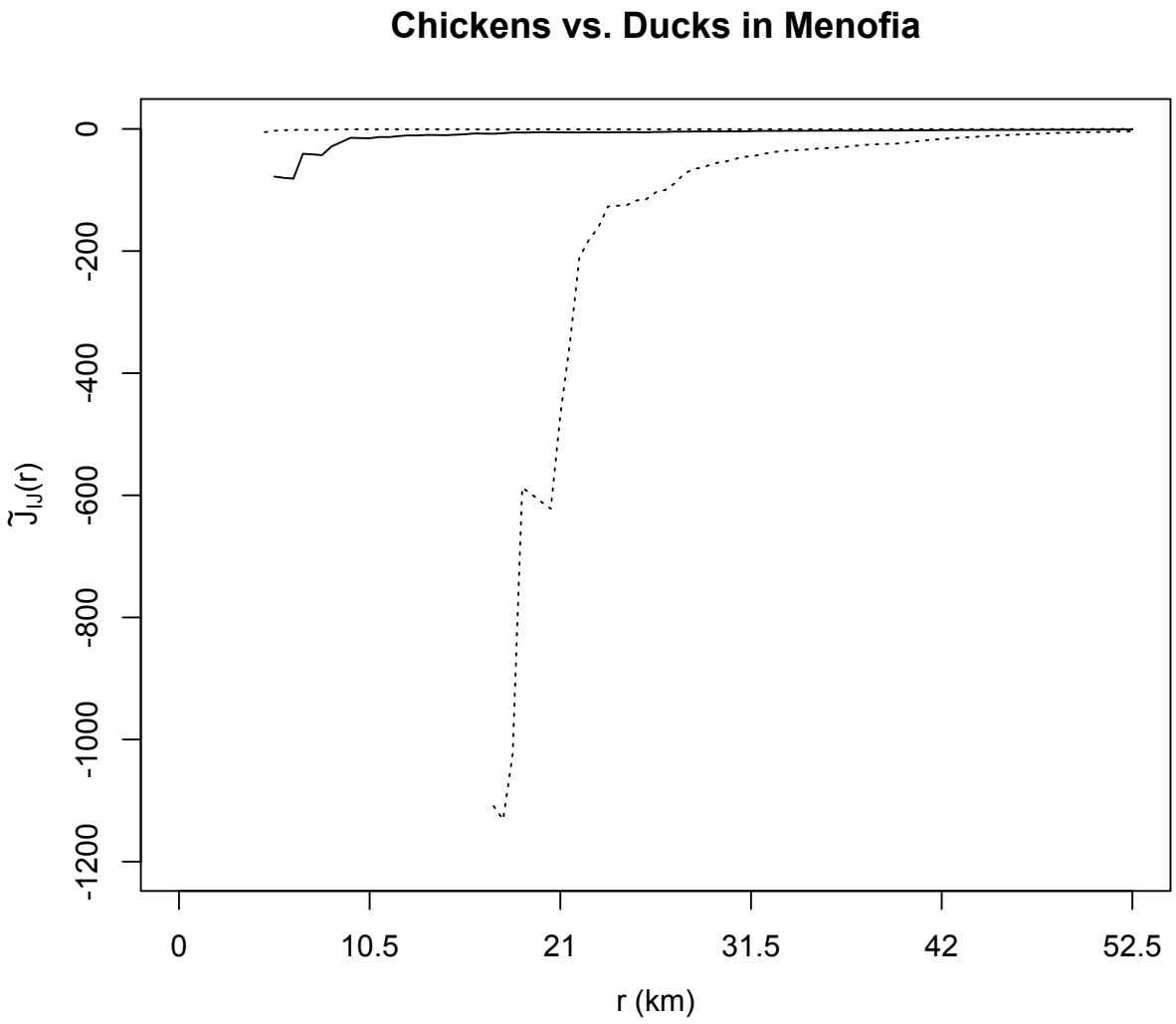

Figure 4.14: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected ducks, as a function of radius r, in the governorate of Menofia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

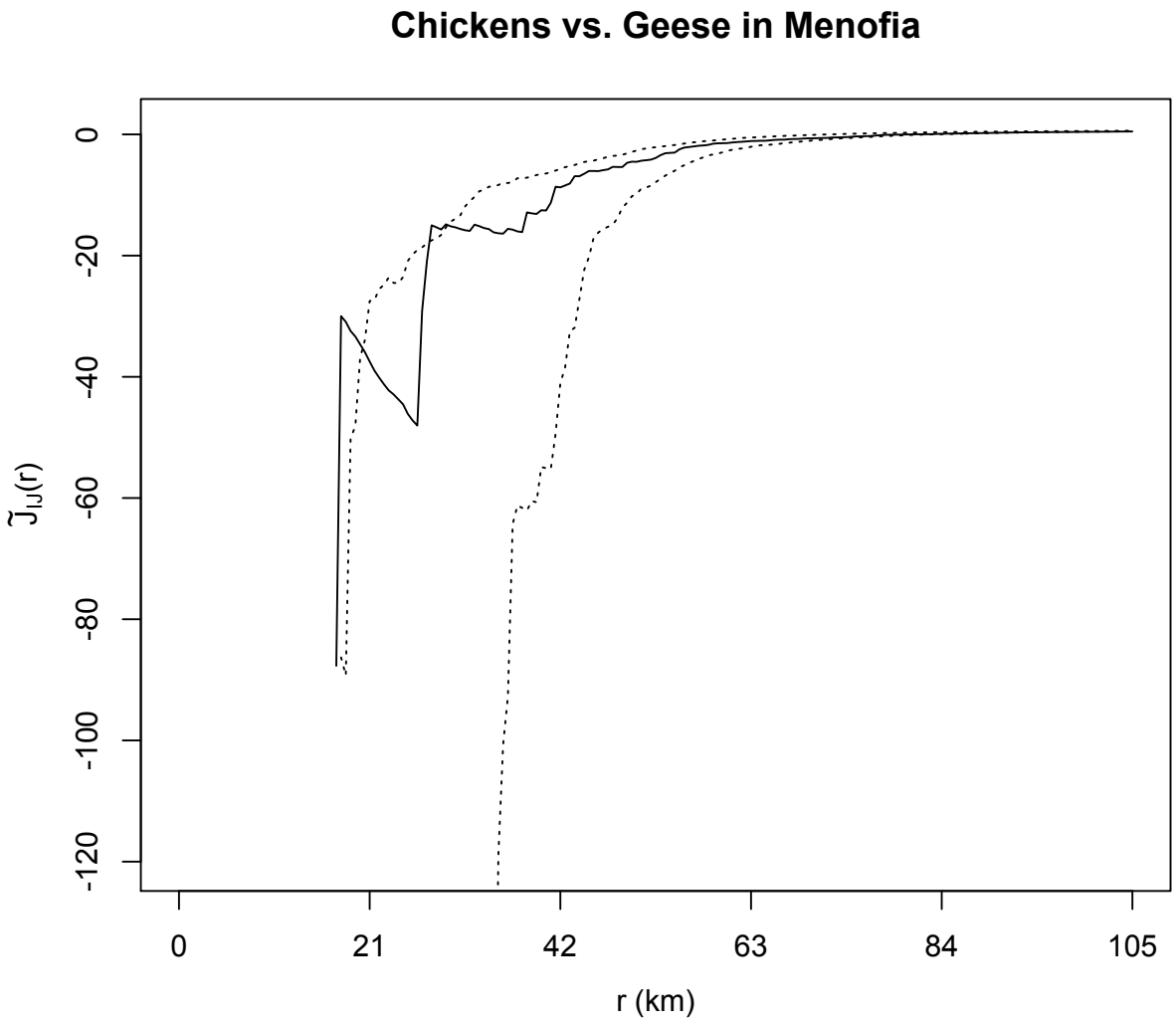

Figure 4.15: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected geese, as a function of radius r, in the governorate of Menofia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

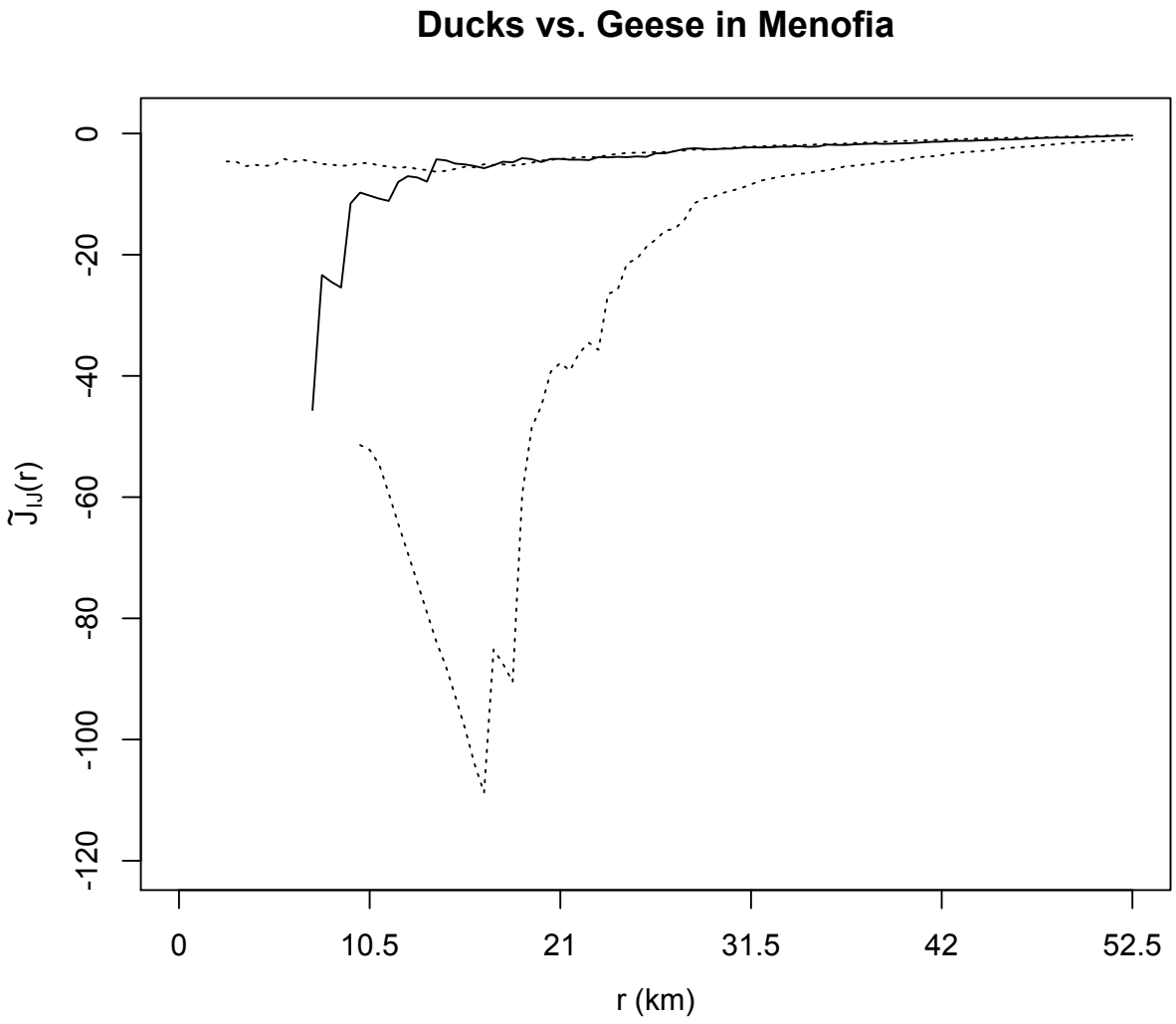

Figure 4.16: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected ducks, and N_j, infected geese, as a function of radius r, in the governorate of Menofia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

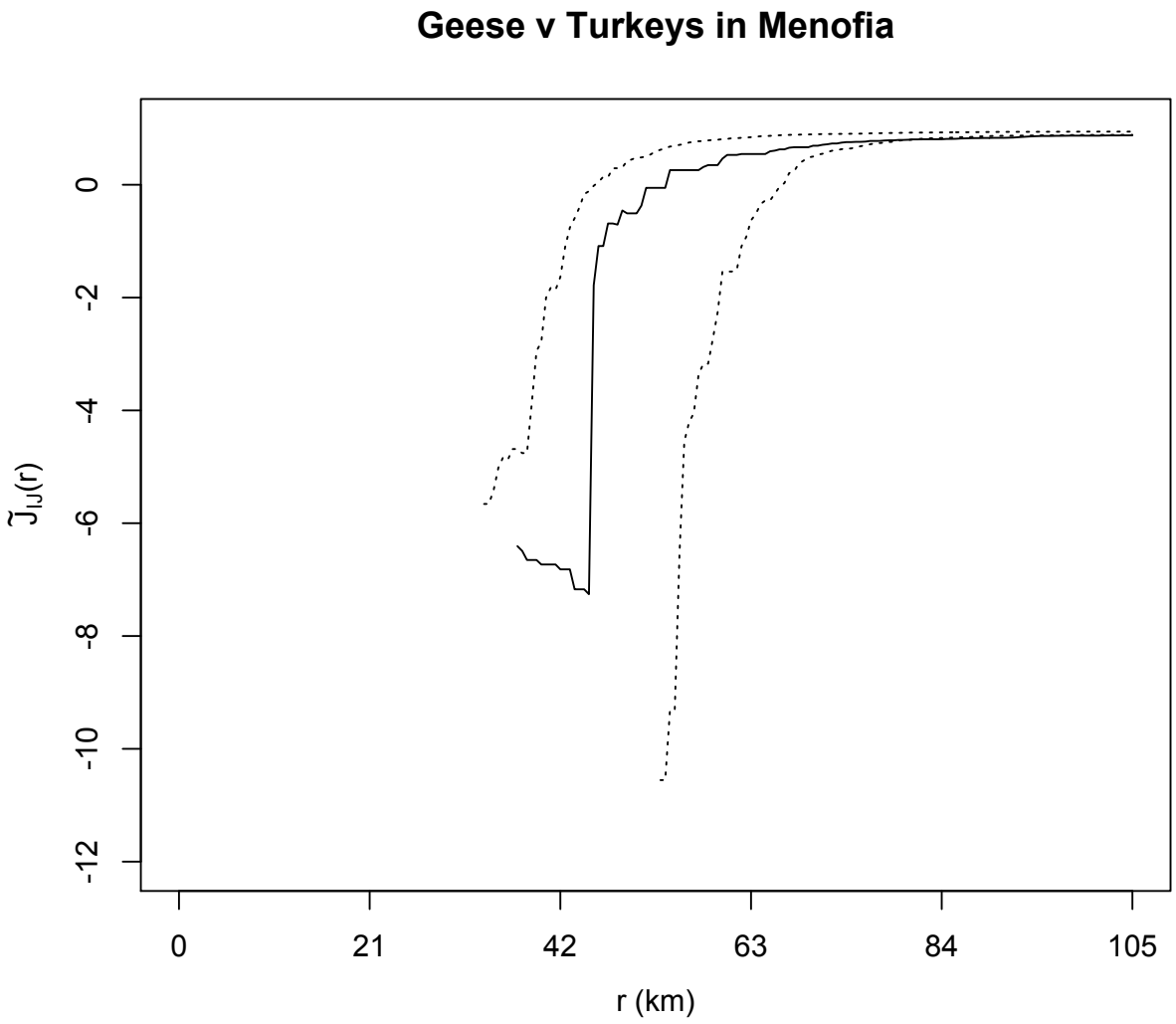

Figure 4.17: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected geese, and N_j, infected turkeys, as a function of radius r, in the governorate of Menofia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

Also, we explored the weighted cross J-function for the governorate, El Gharbia. Figures 4.18 through 4.20 present the estimates of the weighted cross J-function for each set of birds, shown by the solid line, along with the upper and lower 95% bounds, shown by the dotted line, for 100 simulations of the same intensity and observed in the same window as the data. Here we have interesting results, there seems to be inhibition amongst all three pairs of birds. This could be due to small amount of data points, the inhomogeneity of the data, or some bias due to edge effects. The estimates of the weighted cross J-function show inhibition or negative interaction between infected chickens, 86 recorded, and infected ducks, 54 recorded, with the most interaction occurring at small radii (Figure 4.18). Similarly, the interaction between infected chickens, 86 recorded, and infected geese, only 4 recorded, is also a negative interaction which the most inhibition occurring at small radii (Figure 4.19). Again, this could be due to the small amount of recordings. As for infected ducks, 54 recorded, and infected geese, 4 recorded, in the governorate of El Gharbia, there is inhibition amongst the pair of birds especially within radius of 10.5 km (Figure 4.20). Inferences made on presence only data like this really have to be taken very cautiously because of missing data. With a more complete dataset, the weighted cross J-function can successfully detect the type, range and moment of interactions between two point processes.

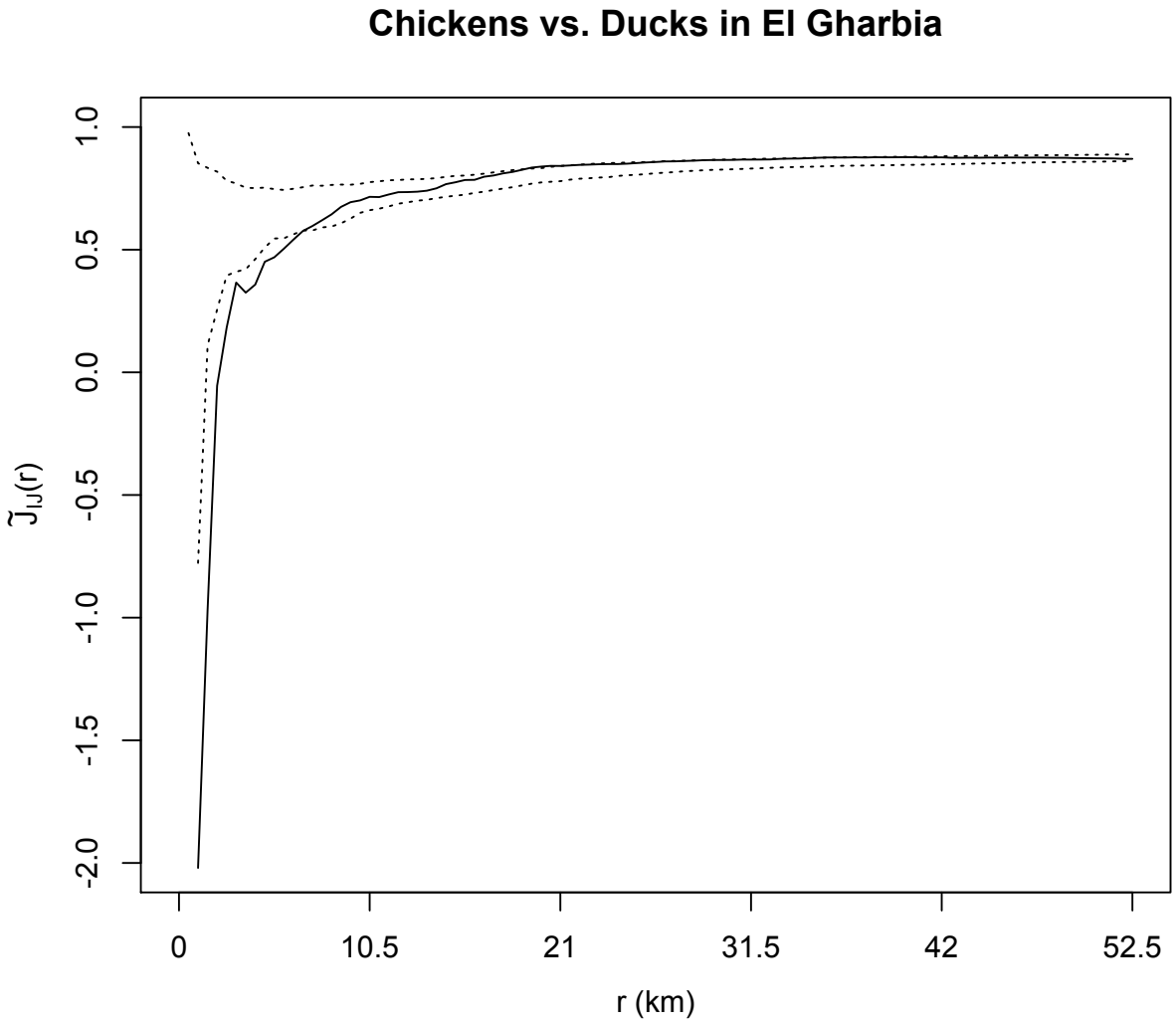

Figure 4.18: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected ducks, as a function of radius r, in the governorate of El Gharbia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

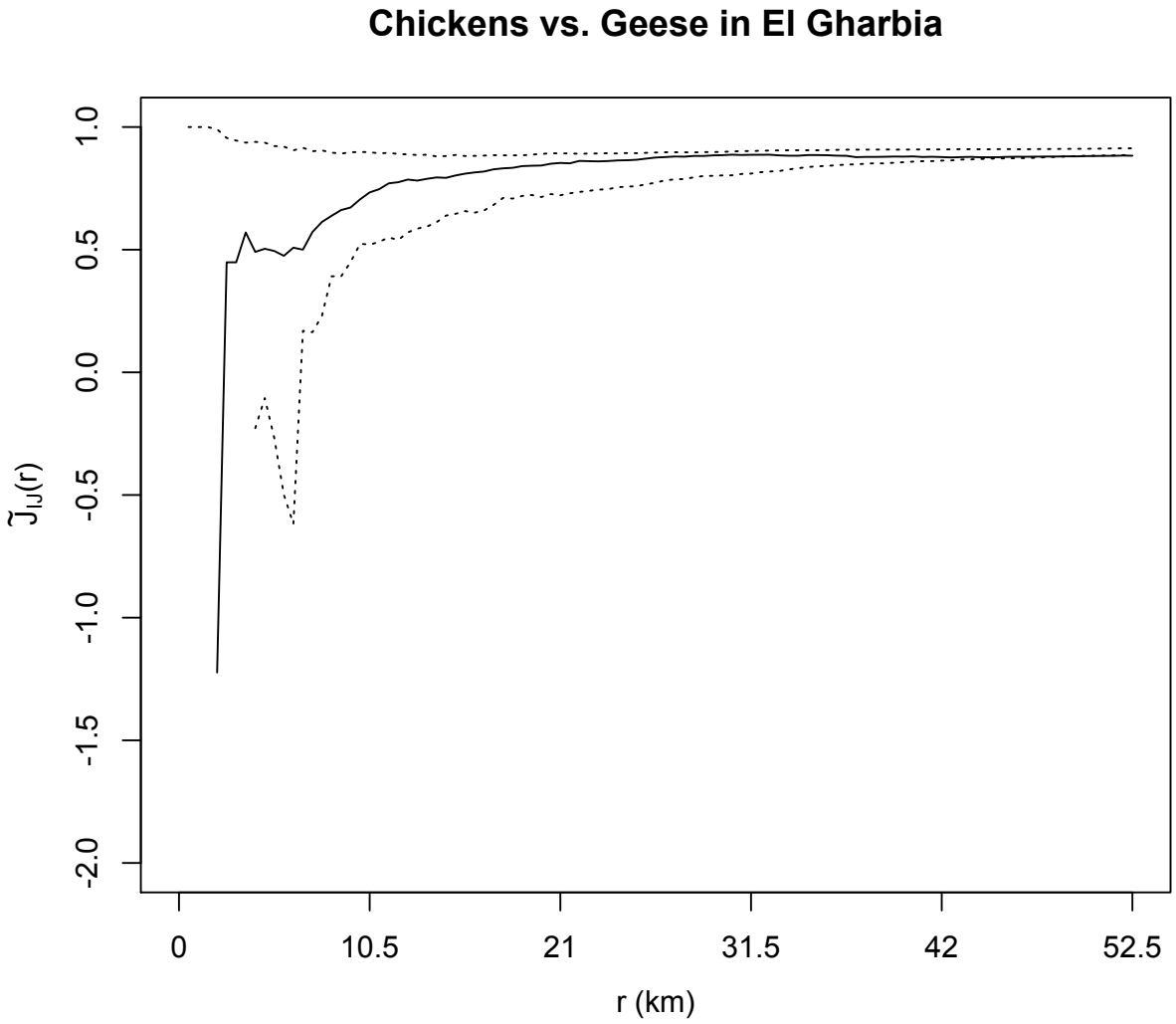

Figure 4.19: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected chickens, and N_j, infected geese, as a function of radius r, in the governorate of El Gharbia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

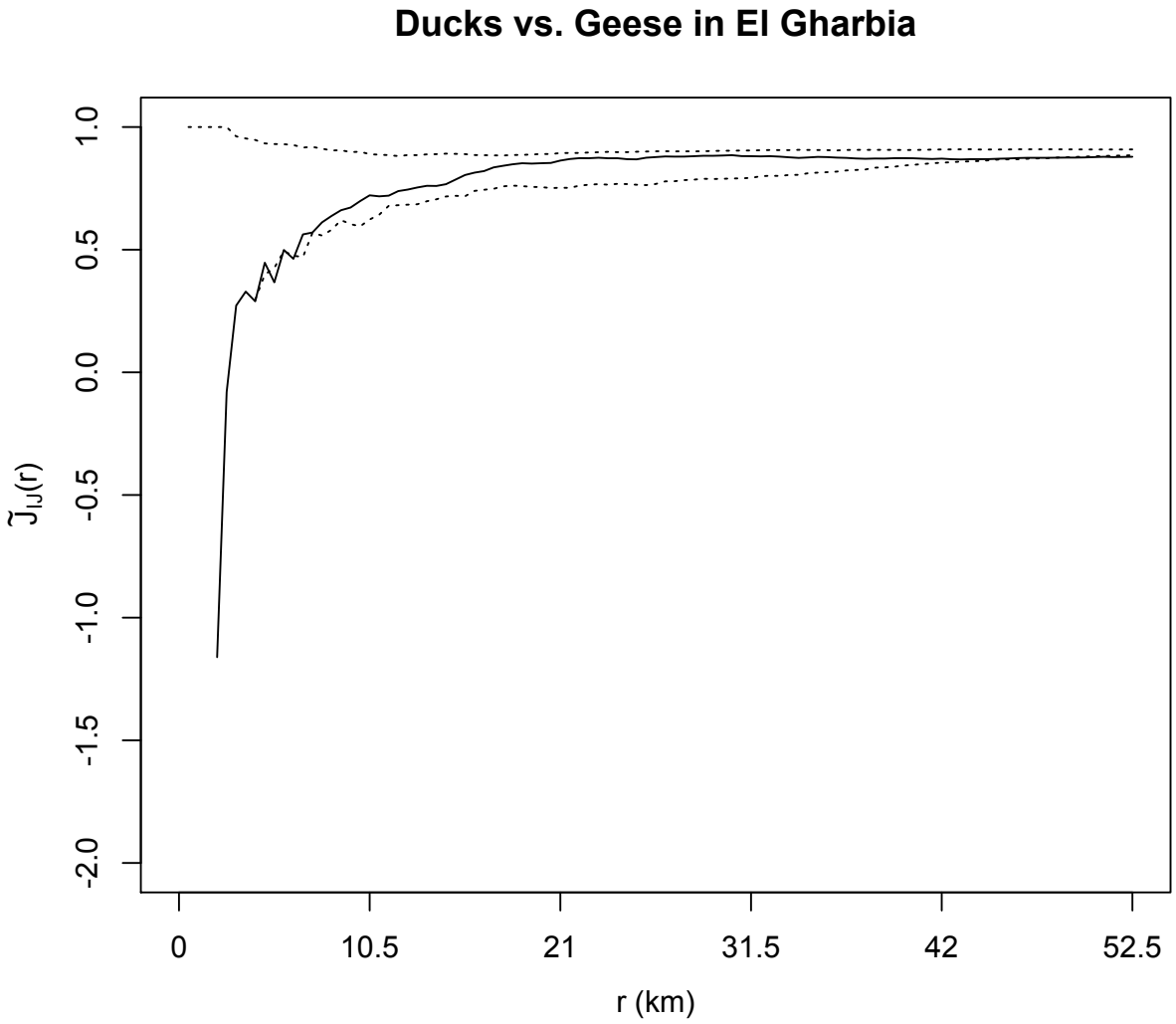

Figure 4.20: Estimates of the weighted cross J-function, $\tilde{J}_{ij}(r)$, between N_i, infected ducks, and N_j, infected geese, as a function of radius r, in the governorate of El Gharbia shown by solid line. Upper and lower 95% bounds of 100 simulations with the same intensity are shown by dotted lines.

CHAPTER 5

Discussion

Point process techniques can play an important role in understanding the spread of infectious diseases, such as the highly pathogenic avian influenza virus. By analyzing the information stored in each individual data point, we can gain further information on the characteristics of the disease outbreaks allowing us to come one step closer to finding ways to halt the spread of the disease.

In this dissertation, point process techniques are applied to the avian influenza virus data, and a summary statistic called the weighted cross J-function is proposed, which is used to indicate the type and range of interaction between two point processes and additionally to indicate the moment of interaction between two events. The advantage of the weighted cross J-function is that it accounts for inhomogeneity in point processes. To adjust for the varying intensity or non-constant background rate we incorporated weights and we demonstrated the effectiveness of the weighted cross J-function with simulations.

In the application of the weighted cross J-function, after accounting for the inhomogeneity in the data, we found that clustering, or positive interaction, amongst all pairs of birds. However, the clustering seems to be mostly due to the inhomogeneity, so although the locations are correlated, the weighted cross J-function is not statistically significant. Inferences made on presence only data like this really have to be taken very cautiously because of missing data. With a more complete dataset, the weighted cross J-function can successfully detect the type, range and moment of interactions between two point processes.

An important direction for future work is the investigation of the statistical properties, such as variance, of the weighted cross J-function. Veen and Schoenberg (2006) investigated the statistical properties of the weighted J-function, noting that the estimator essentially

counts pairs of events and thus can be well approximated by a binomial distribution, and it seems that a similar approximation can be made for the weighted cross J-function. Standard errors can be computed for the estimate. Additionally, the weighted cross J-function could be extended spatial-temporal point process, containing both space and time information. Also, the summary statistic proposed in this dissertation can extend to other datasets to detect interactions between two events such as oil spills and beached whales, the stock market and current events, or disease outbreaks and hospital malpractices or accidents.

The weighted cross J-function can be used to understand the behaviors and characteristics of an infectious disease, such as the highly infectious Zika virus, which will allow us to explore preventive solutions for halting the spread of the disease. In general, this research is a beneficial contribution to epidemiological research and opens new doors for research of infectious diseases.

Bibliography

E.M. Abdelwhab and H.M. Hafez. An overview of the epidemic of highly pathogenic H5N1 avian influenza virus in Egypt: epidemiology and control challenges. *Epidemiology and Infection*, 139, 5 2011.

A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan. *Case Studies in Spatial Point Process Modeling*. Springer, 2006.

A. Baddeley, I. Bárány, and R. Schneider. Spatial Point Processes and their Applications. *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004*, pages 1–75, 2007.

A.J. Baddeley. Spatial sampling and censoring. *Stochastic geometry: likelihood and computation*, 2:37–78, 1999.

A.J. Baddeley, M. Kerscher, K. Schladitz, and B.T. Scott. Estimating the J function without edge correction. *Statistica Neerlandica*, 54(3):315–328, 2000a.

A.J. Baddeley, J. Møller, and R. Waagepetersen. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000b.

J. Calderone. Superbugs will soon be a bigger problem than cancer, 2015a. URL http://www.businessinsider.com/how-common-will-antibiotic-resistant-infections-be-in-the-future-2015-6?_ga=1.107395531.819411270.1463183224.

J. Calderone. This chart shows how incredibly far we've come in the past 100 years of human health, 2015b. URL http://www.techinsider.io/rates-of-infectious-disease-deaths-in-the-past-100-years-2015-7.

H.A. Carneiro and E. Mylonakis. Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks. *Clinical infectious diseases*, 49(10):1557–1564, 2009.

US CDC. *Antibiotic Resistance Threats in the United States, 2013*. Centers for Disease Control and Prevention, US Department of Health and Human Services, 2013. URL http://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf.

R.A. Clements, F.P. Schoenberg, and A. Veen. Evaluation of space-time point process models using super-thinning. *Environmetrics*, 23(7):606–616, 2012.

N. Cressie. *Statistics for Spatial Data*, volume 900. Wiley New York, 1993.

O. Cronie and M.N.M. Van Lieshout. Summary statistics for inhomogeneous marked point processes. *Annals of the Institute of Statistical Mathematics*, pages 1–24, 2014.

O. Cronie and M.N.M. Van Lieshout. A J-function for Inhomogeneous Spatio-temporal Point Processes. *Scandinavian Journal of Statistics*, 42(2):562–579, 2015.

P. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC Press, 2013.

P.M. Dixon. Ripley's K function. *Encyclopedia of environmetrics*, 2002.

I. El Masry, H. Elshiekh, A. Abdlenabi, A. Saad, A. Arafa, F.O. Fasina, J. Lubroth, and Y.M. Jobre. Avian Influenza H5N1 Surveillance and its Dynamics in Poultry in Live Bird Markets, Egypt. *Transboundary and emerging diseases*, 2015.

F.O. Fasina, A.M. Ali, J.M. Yilma, O. Thieme, and P. Ankers. The cost–benefit of biosecurity measures on infectious diseases in the Egyptian household poultry. *Preventive veterinary medicine*, 103(2):178–191, 2012.

A.E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. *Handbook of Spatial Statistics*. CRC Press, 2010.

J.S. Gordon, R.A. Clements, F.P. Schoenberg, and D. Schorlemmer. Voronoi residuals and other residual analyses applied to CSEP earthquake forecasts. *Spatial Statistics*, 14:133–150, 2015.

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 70. John Wiley & Sons, 2008.

K.E. Jones, N.G. Patel, M.A. Levy, A. Storeygard, D. Balk, J.L. Gittleman, and P. Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993, 2008.

A. Kandeel, S. Manoncourt, E. Abd el Kareem, A.N. Mohamed Ahmed, S. El-Refaie, H. Essmat, J. Tjaden, C.C. De Mattos, K.C. Earhart, A.A. Marfin, and N. El-Sayed. Zoonotic Transmission of Avian Influenza Virus (H5N1), Egypt, 2006-2009. *Emerg Infect Dis*, 16 (7):1101–7, 2010.

M.K. Kindhauser, T. Allen, V. Frank, R.S. Santhana, and C. Dye. Zika: the origin and spread of a mosquito-borne virus. *World Health Organization*, 2016. URL http://www.who.int/bulletin/online_first/16-171082/en/.

J. Møller and R.P. Waagepetersen. An Introduction to Simulation-Based Inference for Spatial Point Processes. In *Spatial statistics and computational methods*, pages 143–198. Springer, 2003a.

J. Møller and R.P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. CRC Press, 2003b.

J. Møller and R.P. Waagepetersen. Modern statistics for spatial point processes*. *Scandinavian Journal of Statistics*, 34(4):643–684, 2007.

K. Nichols, F.P. Schoenberg, J.E. Keeley, A. Bray, and D. Diez. The application of prototype point processes for the summary and description of California wildfires. *Journal of Time Series Analysis*, 32(4):420–429, 2011.

V. Nizet. Stopping superbugs, maintaining the microbiota. *Science translational medicine*, 7(295):295ed8–295ed8, 2015.

K.Y. Njabo, L. Zanontian, B.N. Sheta, A. Samy, S. Galal, F.P. Schoenberg, and T.B. Smith. Living with avian FLU - Persistence of the H5N1 highly pathogenic avian influenza virus in Egypt. *Veterinary Microbiology*, 187:82–92, 2016.

P. Nordmann, T. Naas, N. Fortineau, and L. Poirel. Superbugs in the coming new decade; multidrug resistance and prospects for treatment of Staphylococcus aureus, Enterococcus spp. and Pseudomonas aeruginosa in 2010. *Current opinion in microbiology*, 10(5):436–440, 2007.

D. Nuvolone, R. Fresco, S. Maio, S. Baldacci, A. Angino, F. Martini, M. Borbotti, G. Viegi, and R. della Maggiore. Application of Geostatistical Methods for Public Health Risk Mapping. 2008.

Y. Ogata. Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

Y. Ogata. Seismicity Analysis through Point-process Modeling: A Review. *Pure and applied geophysics*, 155(2-4):471–507, 1999.

H. Oshitani, T. Kamigaki, and A. Suzuki. Major issues and Challenges of Influenza Pandemic Preparedness in Developing Countries. *Emerg Infect Dis*, 14(6):875–80, 2008.

R.D. Peng, F.P. Schoenberg, and J.A. Woods. A Space-Time Conditional Intensity Model for Evaluating a Wildfire Hazard Index. *Journal of the American Statistical Association*, 2011.

WHO Report. Egypt: upsurge in H5N1 human and poultry cases but no change in transmission pattern of infection. *World Health Organization*, 2015. URL http://www.emro.who.int/egy/egypt-news/upsurge-h5n1-human-poultry-cases-may-2015.html.

O. Schabenberger and C.A. Gotway. *Statistical Methods for Spatial Data Analysis*. CRC press, 2004.

F. Schoenberg and B. Bolt. Short-term Exciting, Long-term Correcting Models for Earthquake Catalogs. *Bulletin of the Seismological Society of America*, 90(4):849–858, 2000.

S.B. Seng, A.K. Chong, and A. Moore. Geostatistical Modelling, Analysis and Mapping of Epidemiology of Dengue Fever in Johor State, Malaysia. In *The 17th Annual Colloquium*

of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand, pages 24–25. Citeseer, 2005.

D. Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. Wiley, 1994.

H. Tian, S. Zhou, L. Dong, P Thomas, V. Boeckel, Y. Cui, S. H. Newman, J.Y. Takekawa, D.J. Prosser, X. Xiao, Y. Wu, B. Cazelles, S. Huang, R. Yang, B. Grenfell, and B. Xu. Avian influenza H5N1 viral and bird migration networks in asia. *Proceedings of the National Academy of Sciences*, 112(1):172–177, 2015.

M.N.M. Van Lieshout. A J-function for marked point patterns. *Annals of the Institute of Statistical Mathematics*, 58(2):235–259, 2006.

M.N.M. Van Lieshout. A J-function for inhomogeneous point processes. *arXiv preprint arXiv:1008.4504*, 2010.

M.N.M. Van Lieshout and A.J. Baddeley. A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica*, 50(3):344–361, 1996.

M.N.M. Van Lieshout and A.J. Baddeley. Indices of Dependence between Types in Multivariate Point Patterns. *Scandinavian Journal of Statistics*, 26(4):511–532, 1999.

A. Veen and F.P. Schoenberg. Assessing Spatial Point Process Models for California Earthquakes Using Weighted K-functions. In *Case Studies in Spatial Point Process Modeling*, pages 293–306. Springer, 2006.

H. Xu and F.P. Schoenberg. Point process modeling of wildfire hazard in Los Angeles County, California. *The Annals of Applied Statistics*, 5(2A):684–704, 2011.