Spring 5-31-2021

Co-localization Analysis of Bivariate Spatial Point Pattern

Stephan C. Komladzei
University of New Orleans, sckomlad@uno.edu

Co-localization Analysis of Bivariate Spatial Point Pattern

A Thesis

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Master of Science
in
Mathematics

by

Stephan Cobby Komladzei

B.Sc. Kwame Nkrumah University of Science and Technology, 2016
M.S. University of New Orleans, 2021

May, 2021

Dedicated to Mum, Dad and Junior

Acknowledgements

First of all, I would like to give praise to the Almighty God for giving me grace and strength throughout my study.

I would also like to thank the University of New Orleans especially the the faculty members of the Department of Mathematics for giving me this great opportunity to complete my masters' degree and from whom I have learnt a lot, as far as Statistics is concerned.

I would like to express the deepest appreciation to Dr. Xueyan Liu, whom for without her patience and guidance, I would not have made it this far with regards to my thesis. I will forever remain grateful for her tremendous supervision and help to make this thesis a success.

My sincere gratitude goes to my thesis committee members, Dr. Tumulesh Solanky and Dr. Linxiong Li, for their time, guidance and input which guided me throughout the entire research process. I thank them for their support.

Last but not the least, I would also want to thank my family and friends who supported me with much love and prayers. Without them, I am nothing.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CBC	**C**oordinate-**B**ased **C**olocalization
CSR	**C**omplete **S**patial **R**andomness
KCBC	**K(r)** function **C**oordinate-**B**ased **C**olocalization
UCL	**U**pper **C**onfidence **L**imit
LCL	**L**ower **C**onfidence **L**imit

Abstract

Spatial point pattern analysis investigates the localizations of random events in a defined spatial space usually conveyed in the form of images. Spatial distribution of two types of events observed in these images reflects their underlying interactions, which is the focus of co-localization analysis in spatial statistics. Malkusch et al. (Malkusch, et al., 2012) recently proposed the Coordinate-based Colocalization (CBC) method for co-localization analysis. However, the method did not incorporate edge corrections for point proportions and ignored their correlations over nested incremental observational regions. Hence, it yields false positive results for even complete spatial random distributions. In this research, we propose the new K(r) function Coordinate-based Colocalization (KCBC) method to quantify co-localization of two species by utilizing local bivariate Ripley's K and Pearson's Correlation Coefficient. Simulation studies are conducted to demonstrate the unbiasedness of the new method. An application to real life data was provided to illustrate its applicability.

Keywords: Colocalization, Spatial point pattern, Spatial statistics, Pearson's correlation, Ripley's K(r) function.

1. Introduction

1.1.Spatial Point Patterns

Spatial statistics studies statistical methods that are particularly for analyzing data that has spatial characteristics associated with it where the locational information plays one of the most important roles (Unwin, 2009). Spatial point pattern data therefore generally refers to the collection of points randomly located on some underlying mathematical space, i.e., the data on the locations of events on some predefined mathematical space (Unwin, 2009). A classic example is the data on the locations of the centers of 42 biological cells observed under optical microscopy in a histological section (Ripley, 1977), where the locations show a regular distribution, see Figure 1 (a). Whilst the descriptive analysis uses visualization and numerical methods in analyzing the structure of the spatial point pattern, inferential statistical methods analyze the pattern as a stated hypothesis, where complete spatial randomness (CSR) is used as the reference to determine how the other spatial point patterns deviate from it. Different from cluster point processes (see Figure 1 (b) for an example), CSR is formed from a homogenous Poisson process (Figure 1 (c)) and it describes a point process whereby points occur within a given study area in a completely random fashion so that each point in the point pattern is independent of another and they all have equal probability of occurring at any location in the given study region. CSR patterns therefore do not exhibit co-localization or dispersion if there are two or more CSR processes marked as different species in the same study region.

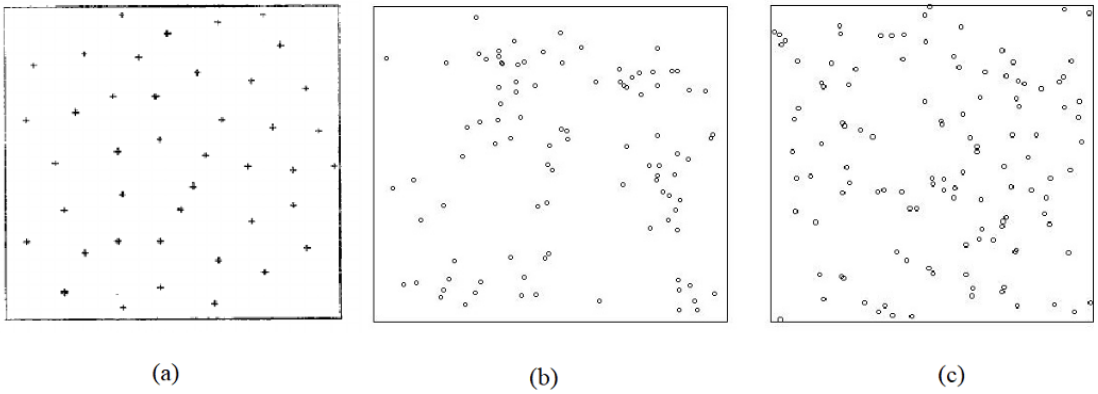


Figure 1: Three different types of point patterns. (a) Regular pattern: a sample square containing localizations of 42 biological cells (Graph adopted from (Ripley, 1977)); (b) A simulated image derived from a Matern Cluster process; (c) A simulated complete spatial random point pattern.

1.2.Complete Spatial Randomness (CSR)

Complete spatial randomness (CSR) is foundational to quantitative and inferential description of spatial point patterns. A CSR is realization of a homogeneous Poisson process (Diggle P. , 1983) in a bounded region of a 2-dimensional or a 3-dimensional space which absorbs a single parameter λ, as the intensity (mean number of events per unit area) of the spatial point pattern, with the actual number of events observed in the region being *N*. Therefore, *N* is an observation of the Poisson process with mean $\lambda|A|$, where $|A|$ is the area of the observational study region in the spatial point pattern. Conditional on *N*, CSR has the following characteristics (Diggle P. J., (1986)):

i. Each event is equally likely to occur at any location of the study region.

ii. All the events are located independently of each other.

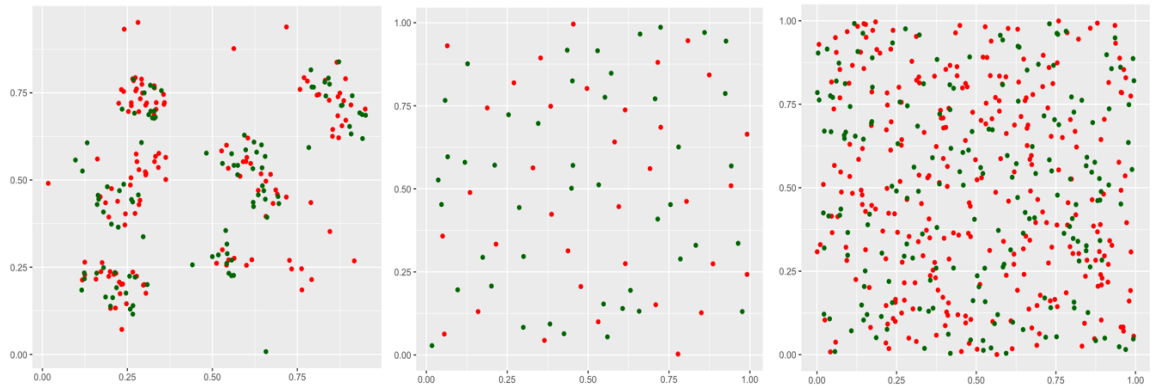
The CSR pattern therefore usually serves as a basis for hypothesis testing in analyzing spatial distributions. One of the most popular methods, the Ripley's K functions (Dixon, 2002), uses the CSR as the benchmark accompanied by Monte Carlo simulations.

1.3.Co-localization Analysis

Analyzing the interactions between different species in spatial point patterns exist largely in many disciplines such as biology, epidemiology, ecology, natural resource management and criminology to give different ideas in making decisions. Co-localization analysis therefore can be seen as the technique used to investigate the interactions in a completely mapped spatial point patterns. It involves the measurement of the spatial overlap between spatial events to see if the events are located in the same area or close to one another (Fig. 2(a)). The opposite case is that the events of different channels tend to locate in different areas or far from one another (Fig. 2(b)), showing a dispersive interaction. These spatial patterns have marked information such as the types of species/channels in the point pattern. When such information is known, analyzing co-localization in the spatial point patterns with multiple channels can be thought to reflect two components: the co-occurrence and co-distribution along or around the different channels in the spatial point pattern.

Over the years, few methods have been developed for co-localization analysis. The most common ones include the mean nearest distance methods, the Ripley's K function, cross correlation, among others. (Dixon, 2002) summarized how the multivariate Ripley's K(r) functions describe the spatial point processes at many distance scales. He showed that under independent CSR (Fig 2(c)), the cross-channel *K(r)* function is $K_{12}(r) = \pi r^2$. The results are unbiased due to edge corrections in the calculation of the K functions.

Recently, a new method, named as Coordinate-Based Co-localization (CBC) method, was proposed by Malkusch in 2012 (Malkusch, et al., 2012) and later employed in their software LAMA (Malkusch & Heilemann, (2016)). The method was developed for analyzing interactions of different types of biomolecules using their localizational information in super-resolution microscopic imaging data.



(a) (b) (c)

Figure 2: Three different types of bivariate point patterns (red and green species).; (a) A simulated image derived from a Matern Cluster process; b) A simulated regular pattern (c) A simulated complete spatial random point pattern

The CBC method selects one channel as the base channel and around each base-channel signal the CBC calculates neighboring point proportions of both channels, respectively, within a series of equally spaced radii of the center between r = 0 and r = Rmax. The method uses the particular Rmax = 250nm as the largest distance Rmax to accommodate the diffraction limit of light (Malkusch & Heilemann, (2016)). The Spearman correlation of the point proportions of the two types of molecules across the series of concentric circles is then calculated as a quantitative measure of co-localization index for that particular signal at the center. The mode of correlation values around all base-channel signals is utilized as the co-localization index for the whole image. However, ignoring the association of point proportions within different sizes of nested circles around the center and failing to incorporate edge corrections yield false positive results even for the independent CSR point processes.

1.4.Problem Statement and Objective of the Thesis

The limitation in the CBC algorithm therefore serves the motivation for this thesis. The main idea for this research is to incorporate bivariate local *K* functions to correct edge effects and then the K functions of both channels are calculated over a sequence of neighboring rings of

equal widths around each base-channel signal. The rings instead of concentric solid circles are used in order to remove the inherent correlation of K function values across different sizes of nested circles due to the natural overlapping. In the last, the Pearson Correlation Coefficient of the local bivariate K functions between the two channels across the series of rings is used to quantify co-localization at the base-channel signal. The average of correlations at all base-channel signals serves as the measure of co-localization of cross-channel towards the base-channel for the whole image. We name the novel method the KCBC method. Basically, if the correlation is toward 1, the two channels are considered to be positively co-localized; if the correlation is closer to -1, the two channels show a dispersive interaction; if the correlation is at around 0, then we can conclude the two types of events have no interactions. The objective is to develop the new method in order to correct the biasedness in the CBC method and produce unbiased results for two independent CSR point patterns. Hence later, the method can be used for hypothesis testing using CSR as reference.

1.5.Organization of the Thesis

In this thesis, Chapter 2 reviews the K function, Ripley's K function, bivariate Ripley's K function and local bivariate Ripley's K function. In Chapter 3, we briefly go over the CBC method. Chapter 4 is devoted to demonstrating the methodology of our KCBC method. Simulation studies and an empirical application are provided in Chapter 5 and Chapter 6 respectively to illustrate the validity and applicability of our new method. Conclusion and future work are discussed in Chapter 7. The R software and the package *spatstat* are used for our simulation and empirical studies.

2. The K(r) Function

2.1.Introduction

In this chapter, we would review the K function, the edge corrected K function known as the Ripley's K function, the bivariate Ripley's K function for spatial patterns with two types of events and finally the local bivariate Ripley's K function.

2.2.Theoretical K(r) Function

The K function is a second-order statistics used to analyze data on the location of events (Dixon, 2002). It analyzes spatial point patterns by measuring the expected number of events found within some distance *r* from a randomly selected event which is then normalized by the intensity (mean number of events per unit area). It has the ability to describe a spatial point pattern at different distance scales, however, a spatial point pattern is not uniquely defined by the K function, i.e., different spatial point patterns can have the same K(r) function (Baddeley & Silverman, 1984). Mathematically, the K(r) is defined as

$$K(r) = \lambda^{-1} E[N_0(r)] \tag{2.1}$$

where, λ is the intensity (number of points per unit area) of the image and $E[N_0(r)]$ counts the expected number of events found within the radius *r* of a randomly selected event in the spatial point pattern. $E[N_0(r)]$ can however be estimated as,

$$\hat{E}[N_0(r)] = N^{-1} \sum \sum_{j \neq i} I(d_{ij} < r) \tag{2.2}$$

where, d_{ij} represents the distance between the randomly selected i^{th} point to the j^{th} point, N is the observed number of points in the image and I is the indicator function which has a value of 1 if d_{ij} is within distance *r* and 0 otherwise.

Therefore, the K(r) function can be estimated as,

$$\widehat{K}(r) = \hat{\lambda}^{-1} \frac{\sum\sum_{j \neq i} I(d_{ij} < r)}{N}, \tag{2.3}$$

where $\hat{\lambda} = N/|A|$, and $|A|$ is the area of the observation window.

2.3.The Ripley's K(r) function

Since observation window is bounded, it is intuitive and necessary to consider edge effects especially when r is large. Ignoring edge effects is likely to yield biased results for the events close to the boundary of the study region and/or as r gets large. This is because for large radii and for events close to the boundaries of the study region, neighboring points outside the boundaries are not observed and hence not counted even if they fall inside the neighborhood of distance r of the selected events. In view of this, edge effects corrections must be considered in order to avoid biased results. The most popular edge-corrected estimator used in the literature is known as the Ripley's K function proposed by Ripley (Ripley B. D., 1981) which adds a weight *w* to the indicator function in *(2.3)*,

$$\widehat{K}(r) = \hat{\lambda}^{-1} \sum_i \sum_{j \neq i} w(x_i, x_j)^{-1} \frac{I(d_{ij} < r)}{N}, \tag{2.4}$$

The function $w(x_i, x_j)$ has a value equal to the proportion of the circumference of the circle with center x_i and passing through the point x_j found within the study region. The Ripley K function can be estimated for all r; however, it is more efficient and common practice to use an *r*, that is less than one quarter of the smallest dimension of the enclosing rectangle of the study region.

2.4.Bivariate Ripley's K function

The bivariate Ripley's K function is a Ripley's K function that is in particular for localizations of two channels. For example, for an image having the channels A and B, it measures the expected number of events in channel B found within some distance r from a randomly selected

event in channel A which is then normalized by the intensity of the signals in channel B and vice versa. Considering edge effect corrections, it is estimated as,

$$\widehat{K}_{AB}(r) = \left(\hat{\lambda}_B\right)^{-1} \sum_{x_i \in A} \sum_{x_k \in B} w(x_i, x_k)^{-1} \frac{I(d_{ik} < r)}{N_A}, \quad (2.5)$$

where, $\hat{\lambda}_B$ is the intensity of the channel B (number of points of channel B per unit area).

2.5.Local Bivariate Ripley's K(r) Function

The local bivariate Ripley's K is of the bivariate Ripley K function at a particular base-channel signal. For an image having channels A and B, it calculates the number of extra events in channel B found within distance r of a selected signal point x_A in channel A, which is then normalized by the intensity of the signals in channel B. Considering edge effect corrections, the local bivariate Ripley's K estimator is given as below,

$$\widehat{K}_{AB}(r, x_i) = (\hat{\lambda}_B)^{-1} \sum_{x_k \in B} w(x_i, x_k)^{-1} I(d_{ik} < r). \quad (2.6)$$

Essentially, bivariate Ripley's K function is the average of local bivariate Ripley's K functions across all base-channel signals to provide overall information of the whole image. Local bivariate Ripley's K function can instead be used to investigate the distribution of cross-channel events around each particular base-channel event if people are more interested in interactions of two channels for a particular subregion or for a particular location/event.

Theoretically, under CSR, Ripley's K functions and local bivariate Ripley's K functions are expected to be equal to the area πr^2.

3. Coordinate-Based Colocalization (CBC) Method

3.1.Introduction

In this chapter we review the CBC method. It serves as one of the main motivations to build our proposed KCBC method in Chapter 4.

3.2.The CBC Method

As we have introduced previously, the CBC method (Malkusch, et al., 2012) calculates the correlation of point proportions of two channels (A, B) across a series of distances around each base-channel signal. To find the co-localization value at a base-channel signal point A_i in channel A, it uses the scaled ratio of the total count of signal points on a sequence of circular neighborhoods ranging from a minimum r=0 value to a maximum value R_{max} of each base-channel signal point to get two ratios for the two channels.

$$D_{A_{i,}A}(r) = \frac{N_{A_i,A}(r)}{N_{A_i,A}(R_{max})} \cdot \frac{R_{max}{}^2}{r^2}, \tag{3.1}$$

$$D_{A_{i,}B}(r) = \frac{N_{A_i,B}(r)}{N_{A_i,B}(R_{max})} \cdot \frac{R_{max}{}^2}{r^2}, \tag{3.2}$$

where, $N_{A_i,A}(r)$ is the number of points in species A found within the r neighborhood of the selected base-channel signal point A_i, $N_{A_i,B}(r)$ is the number of points in species B found within the r neighborhood of the selected base-channel signal point A_i.

A rank correlation coefficient, the Spearman's correlation, is then calculated at each base signal point for the ratios in *(3.1)* and *(3.2)*:

$$S_{A_i} = \frac{\sum_{r_j=0}^{Rmax}\left(O_{D_{A_i,A}}(r)-\bar{O}_{D_{A_i,A}}\right)\left(O_{D_{A_i,B}}(r)-\bar{O}_{D_{A_i,B}}\right)}{\sqrt{\sum_{r_j=0}^{Rmax}(O_{D_{A_i,A}}(r)-\bar{O}_{D_{A_i,A}})^2}\sqrt{\sum_{r_j=0}^{Rmax}(O_{D_{A_i,B}}(r)-\bar{O}_{D_{A_i,B}})^2}}, \tag{3.3}$$

where, $O_{D_{A_i,A}}(r)$ is the rank of $D_{A_{i,}A}(r)$ and $\bar{O}_{D_{A_i,A}}$ is the average of $O_{D_{A_i,A}}(r)$ (Malkusch, et al., 2012).

Finally, each of the Spearman rank correlation is calibrated into a scaled ratio rank correlation and then serves as the colocalization value for each signal point A_i. The calibrated rank correlation is calculated as:

$$C_{A_i} = S_{A_i} . e^{\left(-\frac{E_{A_i,B}}{R_{max}}\right)}, \quad (3.4)$$

where, $E_{A_i,B}$ is the distance from A_i to its nearest neighbor in the cross species B. This calibration is to reduce false positives if the nearest cross-channel neighbor is too far away. The overall co-localization index at each r is given by the mode of the co-localization values of all the base-channel signal points. Under CSR the value for $D(r)$ is expected to be 1 for all r, hence the scaled rank correlation coefficient is expected to be 0. However, because the CBC algorithm does not consider edge correction and ignores the correlation of point proportions over different sizes of nested circles it is prone to produce false positive errors, resulting in upward biased colocalization values even under CSR.

4. K(r) Coordinate-Based Colocalization (KCBC) Method

4.1. Introduction

In this chapter, we describe our novel method for co-localization analysis. The local Ripley's K function and the CBC method serves as the foundation for the construction of the KCBC method.

4.2.Construction of the KCBC Method

To construct the KCBC method, we incorporate local Ripley's K function to modify the CBC method in order to correct the false positive errors. We denote our study region by Υ, which contains n signal points X from two species A and B, i.e., $\Upsilon: X = X_A \cup X_B$. To calculate the co-localization index for each base signal point A_i, the local bivariate Ripley K(r) function is first used to estimate the expected number of signal points from both channels $A\ and\ B$ found within the r neighborhood of the particular point A_i which is then scaled by the intensity of image. The estimates are calculated as:

$$\widehat{K}_{AA}(r, A_i) = (\hat{\lambda}_A)^{-1} \sum_{x_k \in A} w(x_i, x_k)^{-1}\ I(d_{ik} < r), \tag{4.1}$$

$$\widehat{K}_{AB}(r, A_i) = (\hat{\lambda}_B)^{-1} \sum_{x_k \in B} w(x_i, x_k)^{-1}\ I(d_{ik} < r). \tag{4.2}$$

The local bivariate Ripley's K functions are estimated on a sequence of r values ranging from some minimum r to a maximum value, i.e., $0 = r_0, r_1, \ldots, r_J = R_{max}$. In our KCBC method, the selected base signal point A_i is added to the local bivariate K(r) values

$$\widehat{K}_{AA}(r_j, A_i) = \widehat{K}_{AA}(r_j, A_i) + (N_A - 1)^{-1}\ , \tag{4.3}$$

Hence the local K functions are now interpreted as expected number of base-channel signals around an observed base-channel event normalized by the image intensity. Under CSR, the local K functions are equal to the area of the corresponding circle

$$K_{AA}(r_j, A_i) = \pi r_j^2, \tag{4.4}$$

$$K_{AB}(r_j, A_i) = \pi r_j^2 \, . \tag{4.5}$$

To eliminate the correlation of these K functions over different sizes of solid circles due to overlapping, we subtract the neighboring K functions and get the K functions over disjoint rings:

$$\widehat{K}_{AA}(r_j, A_i) \backslash \widehat{K}_{AA}(r_{j-1}, A_i) = \widehat{K}_{AA}(r_j, A_i) - \widehat{K}_{AA}(r_{j-1}, A_i) \, , \tag{4.6}$$

$$\widehat{K}_{AB}(r_j, A_i) \backslash \widehat{K}_{AB}(r_{j-1}, A_i) = \widehat{K}_{AB}(r_j, A_i) - \widehat{K}_{AB}(r_{j-1}, A_i) \, , \tag{4.7}$$

for $j = 1, \dots, J$. Normalization is then applied using the areas of the rings to get our new local K values:

$$N_{\widehat{K}_{AA}}(r_j, A_i) = \frac{\widehat{K}_{AA}(r_j, A_i) - \widehat{K}_{AA}(r_{j-1}, A_i)}{\pi r_j^2 - \pi r_{j-1}^2} \, , \tag{4.8}$$

$$N_{\widehat{K}_{AB}}(r_j, A_i) = \frac{\widehat{K}_{AB}(r_j, A_i) - \widehat{K}_{AB}(r_{j-1}, A_i)}{\pi r_j^2 - \pi r_{j-1}^2} \, . \tag{4.9}$$

This process is repeated for each base-channel signal point A_i, and at each A_i, the Pearson's correlation coefficient P_{A_i} for the normalized pairs $(N_{\widehat{K}_{AA}}(r_j, A_i) \, , N_{\widehat{K}_{AB}}(r_j, A_i))$ is computed, where $i = 1, \dots, N_A$ and $j = 1, \dots, J$. The Pearson's correlation coefficient P_{A_i} serves as the co-localization index for the point A_i.

$$P_{A_i} = \frac{\sum_{j=1}^{J} \left[N_{\widehat{K}_{AA}}(r_j, A_i) - \overline{N_{\widehat{K}_{AA}}(A_\iota)} \right] \left[N_{\widehat{K}_{AB}}(r_j, A_i) - \overline{N_{\widehat{K}_{AB}}(A_\iota)} \right]}{\sqrt{\sum_{j=1}^{J} \left[N_{\widehat{K}_{AA}}(r_j, A_i) - \overline{N_{\widehat{K}_{AA}}(A_\iota)} \right]^2 \sum_{j=1}^{J} \left[N_{\widehat{K}_{AB}}(r_j, A_i) - \overline{N_{\widehat{K}_{AB}}(A_\iota)} \right]^2}}, \tag{4.10}$$

where $\overline{N_{\widehat{K}_{AA}}(A_\iota)} = \frac{1}{J} \sum_{j=1}^{J} N_{\widehat{K}_{AA}}(r_j, A_i)$ and $\overline{N_{\widehat{K}_{AB}}(A_\iota)} = \frac{1}{J} \sum_{j=1}^{J} N_{\widehat{K}_{AB}}(r_j, A_i)$. P_{A_i} takes on the values between -1 and 1. The KCBC algorithm sets P_{A_i} to -1 in cases where there is extreme exclusion, 0 for lack of interaction and 1 when the two channels are completely colocalized. At the R_{max}, the mean of correlations across all base-channel signals is computed as the overall measure of co-localization of the cross-channel events towards the base channel signal points.

5. Simulation Studies

5.1. Introduction

In this chapter, we conduct simulation studies to investigate the performance of the KCBC algorithm proposed in Chapter 4.

5.2. Simulation Studies on CSR Images.

In the study, we set the unit square $[0,1] \times [0,1]$ in R^2 as our observation window. Independent homogenous Poisson processes were simulated for red species with intensity of 300 and green species of intensity of 175. A total of 500 images were obtained. One example image is given in Figure 3. In order to examine the effect of selection of R_{max} to the index, we applied KCBC to each image at different choices of R_{max}. Also, for purposes of comparison, the CBC method was applied to the same images for each R_{max}, too. Means of co-localization indices by KCBC and modes of the measures from CBC method are collected, respectively, for all simulated images. Results of the two methods are summarized in Table 1 and displayed in Figure 4.

The box plot in Figure 4 shows clearly that, results of CBC method for very small R_{max} are almost not biased, however, as R_{max} gets larger, the CBC yields upward biased co-localization index values. This upward bias is corrected in the KCBC method by using the local bivariate Ripley's K functions over disjoint rings. Results from the KCBC method show that the upward biasedness of co-localization values in the CBC has been significantly reduced for all R_{max}. This reflects no interactions between the two species. The results are extremely well for relatively larger r's. See Table 1 for the information of 95% confidence intervals for each R_{max}.

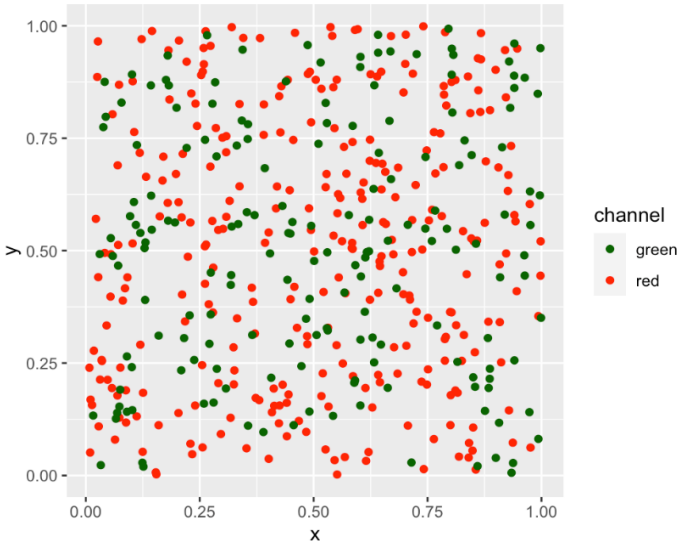


Figure 3: A simulated independent CSR Patterns of 290 red events and 189 green events.

Rmax	CBC Index	LCL	UCL	KCBC Index	LCL	UCL
0.05	0.00	0.0011	0.0015	0.00	0.002	0.004
0.07	0.02	0.0196	0.0296	0.01	0.007	0.009
0.09	0.20	0.1909	0.2148	0.01	0.007	0.010
0.11	0.24	0.2281	0.2486	0.01	0.006	0.009
0.13	0.23	0.2157	0.2345	0.01	0.005	0.008
0.15	0.20	0.1903	0.2082	0.01	0.004	0.007
0.17	0.17	0.1652	0.1833	0.00	0.002	0.006
0.19	0.16	0.1499	0.1676	0.00	0.001	0.005
0.21	0.16	0.1477	0.1664	0.00	0.000	0.004
0.23	0.15	0.1391	0.1584	0.00	0.000	0.004
0.25	0.15	0.1407	0.1596	0.00	0.000	0.004

Table 1: KCBC and CBC results of 500 simulated CSR images. The table shows the summary of the modes and the 95% confidence intervals, lower confidence limit (LCL) and upper confidence limit (UCL) of the CBC values and mean and 95% confidence interval of the KCBC indexes at each $\boldsymbol{R_{max}}$.

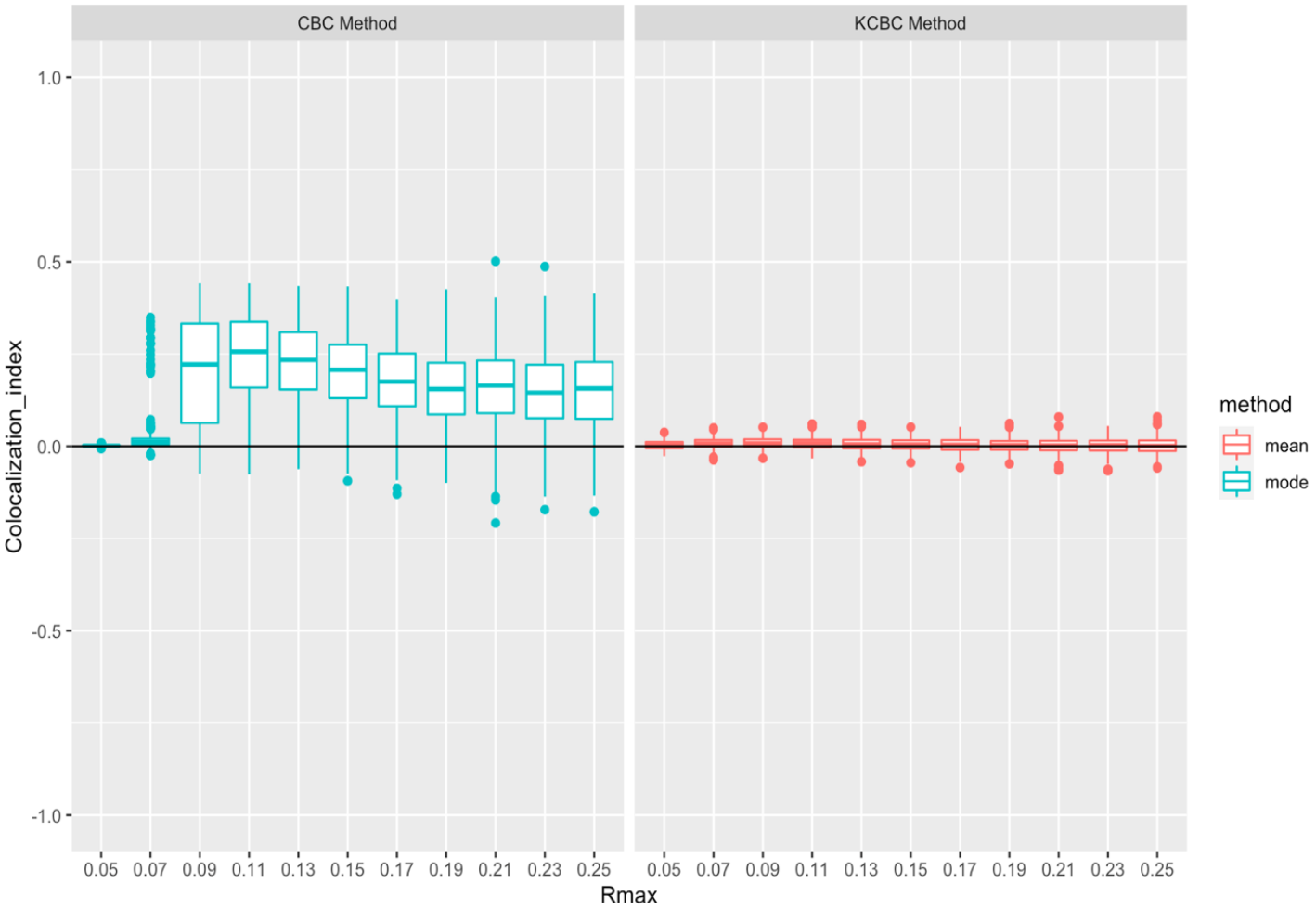


Figure 4: Boxplots showing the summary of co-localization indexes for the 500 simulated CSR images at the each R_{max} obtained by CBC and KCBC, respectively.

6. An Empirical Study

6.1.Introduction

In this chapter, we apply the KCBC method to a real-life data which contains locational information of two type ants. The *ants* (Harkness & Isham, 1983) dataset can be found in the *spatstat.data* package accessible from the CRAN repository (https://cran.r-project.org/web/packages/spatstat.data/index.html).

6.2.Background of the Ants Data

The *ants* dataset shown in *Figure 5* demonstrates the spatial distribution of locations of 97 nests of two types of ant species (68 Cataglyphis and 29 Messor). The nests for all the ant species are beneath the ground and most often have only one opening at the surface of the ground which serves as the passage for the ants. Earlier studies reveals that each species has its unique way of gathering food. The Messor species collects seeds as food by embarking the search in "trunk tails" before the individual Messor ants disperse. The trunk tails from each nest however do not intersect. The Cataglyphis species on the other hand leaves their nests in ones and in every direction in search of food. They collect dead insects which mostly are the dead Messor ants and usually travel at least 50ft before they return to their nests. The Messor ants are however most often hunted and killed by *Zodarium frenatum*, a hunter spider at in and around the opening of their nests and then they are carried away by the spider or other Messor ants. There is therefore an obvious biological relationship between the types of ant species with dead Messor ants serving as food for Cataglyphis ants. It is therefore of much interest to determine whether the locations of the nests of the Cataglyphis ants are as a result of the biological relationship that exist between the two species.

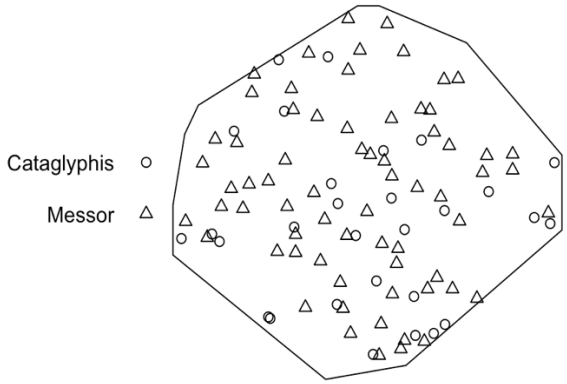


Figure 5:Map of the 97 locations of nests of two species of ants, 68 Messor wasmanni (Δ) and 29 Cataglyphis bicolor (o) in an irregular region 425 feet in diameter.

The map covers a total area of about 1 hectare in which there is an open scrub in the north-western part of the map and field lands at the south-western part of the entire region. The ants however are not able to build in the parts of the scrubs lands where there are bushes are growing. The map also shows that, whilst the nests of the Messor ants are located all over the region, the Cataglyphis ants has few nests in the scrub, and they also are located at some particular sections of the entire region. Our map can therefore be divided into two subregions (A and B) as shown in *Figure 6*, where A covers both the scrub land and field land, and B covers only the field land.

6.3.Application of the KCBC Method on the Empirical Data

We apply the KCBC method to the empirical data to answer our question of interest. In investigating whether the nests of the Cataglyphis ants are as a result of the biological relationship that exist between the two species, we use both the subregions A and B as the study regions and the results are compared.

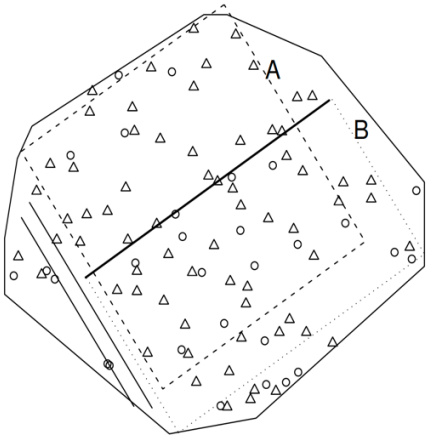
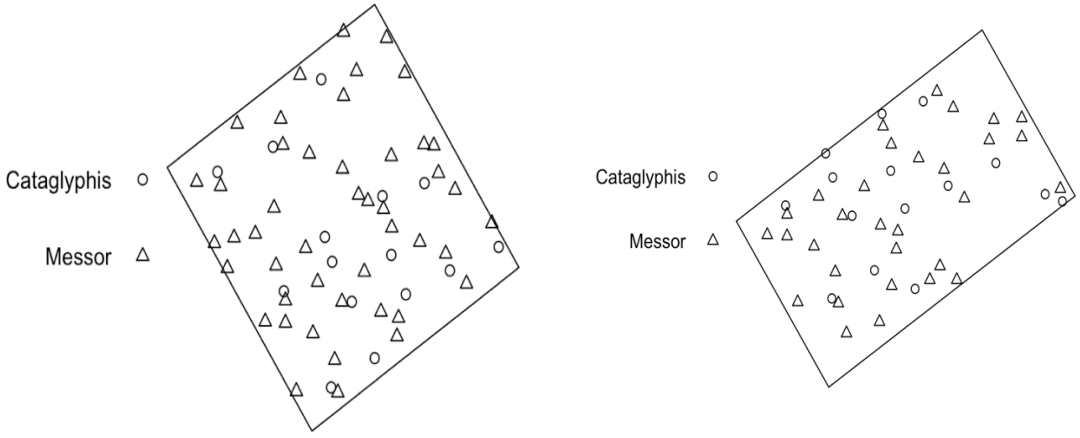


Figure 6: Map of the 97 locations of nests of two species of ants divided into subregion A and subregion B, Messor (∆) and Cataglyphis (o) (Graph adopted from (Baddeley & Turner, 2005)).

We first perform our analysis using subregion A *(Figure 7(a))*. This subregion consists of the location of nests found in both the scrub and the field lands. The observation window for the subregion is a in R^2 rectangular region defined on [8,686] × [31,717]. Messor is used as the base species and Cataglyphis as the second species and we aim to investigate whether the nests of the two species interact with each other by either through co-clustering together or, there is any form of repulsion between the nests of the two species such that they tend to be more regularly spaced than in the case of CSR.

On a sequence of r values ranging from $0 = r_0, r_1, \ldots, r_J = R_{max}$, where R_{max} = (50, 62, 86, 98, 110, 122, 134, 146, 158, 170), the local bivariate Ripley's K function is first used to estimate the expected number of signal points from both species found within the r neighborhood of the each base signal point which is then scaled by the intensity of image.

The bivariate local K values are then transformed into disjoint rings as described in the methodology and then each pair is normalized, and the Pearson's correlation coefficient for the normalized pairs is computed for each base signal point.



(a) (b)

Figure 7: Bounded study regions for the ant data. (a) Subregion A, (b) Subregion B

The Pearson's correlation coefficient at the selected base signal point is the co-localization index for that signal point. The colocalization index at all the R_{max} is displayed in *Figure 8.* The results show that the data does not display a clear pattern of interactions. The KCBC values at all the R_{max}'s are centered around 0. Hence there is no strong evidence to prove that there is any interaction between the nest of the two species. The CBC yields a dispersion for the smallest $R_{max} = 50$, does not support either colocalization or dispersion for $R_{max} \in [62,86]$, but shows a positive co-localization of Cataglyphis nests toward Messor nests for mostly $R_{max} > 98$ except for $R_{max} = 158$; however, KCBC supports a dispersion of Cataglyphis nests from Messor nests for all R_{max} except for $R_{max} = 74$.

Another test is carried out using the subregion B, which consists of nests found only in the field land. The subregion B is defined on a R^2 rectangular region measuring [136,803] × [-47,549]. On a sequence of r values ranging from $0 = r_0, r_1, \ldots, r_J = R_{max}$, where R_{max} = (45,55.5,66,76.5,87.97.5,108,118.5,129,139.5,150), the same procedure used in finding the co-localization index at each base signal point (Messor) in region A using the KCBC method is applied to region B. The results are shown in *Figure 9* and *table 3.* We could see the results are similar as those for Region A. CBC shows a dispersion for $R_{max} = 55.5$, no evidence of any

relation for $R_{max} = 45$, but significantly positive co-localization for $R_{max} > 66$; while KCBC shows a dispersion for all R_{max} except $R_{max} = 66$. Obviously the two methods yield different results.

Rmax	CBC Index	LCL	UCL	KCBC Index	LCL	UCL
50	-0.0011	0.0001	0.0003	-0.0255	-0.0052	-0.0021
62	0.0046	0.0002	0.0004	-0.0277	0.0026	0.0074
74	0.0142	0.0023	0.0030	0.0154	0.0066	0.0120
86	0.0180	0.0043	0.0058	0.0011	0.0117	0.0175
98	0.3382	0.0089	0.0113	-0.0287	0.0172	0.0231
110	0.3340	0.0167	0.0221	-0.0104	0.0182	0.0241
122	0.2945	0.0219	0.0509	-0.0311	0.0219	0.0280
134	0.3498	0.0609	0.1014	-0.0011	0.0265	0.0331
146	0.2684	0.1724	0.2133	-0.0382	0.0250	0.0320
158	0.0144	0.2037	0.2395	-0.0466	0.0238	0.0305
170	0.2149	0.2199	0.2552	-0.0126	0.0275	0.0342

Table 2. Co-localization indices obtained from KCBC and CBC algorithms for Region A. The modes of the CBC values and the means of the KCBC indices at various $\boldsymbol{R_{max}}$ *ranging from 50 to 170. The table also shows the summary of the modes and the 95% confidence intervals, lower confidence limit (LCL) and upper confidence limit (UCL) of the CBC values and mean and 95% confidence interval of the KCBC indexes at each* $\boldsymbol{R_{max}}$.

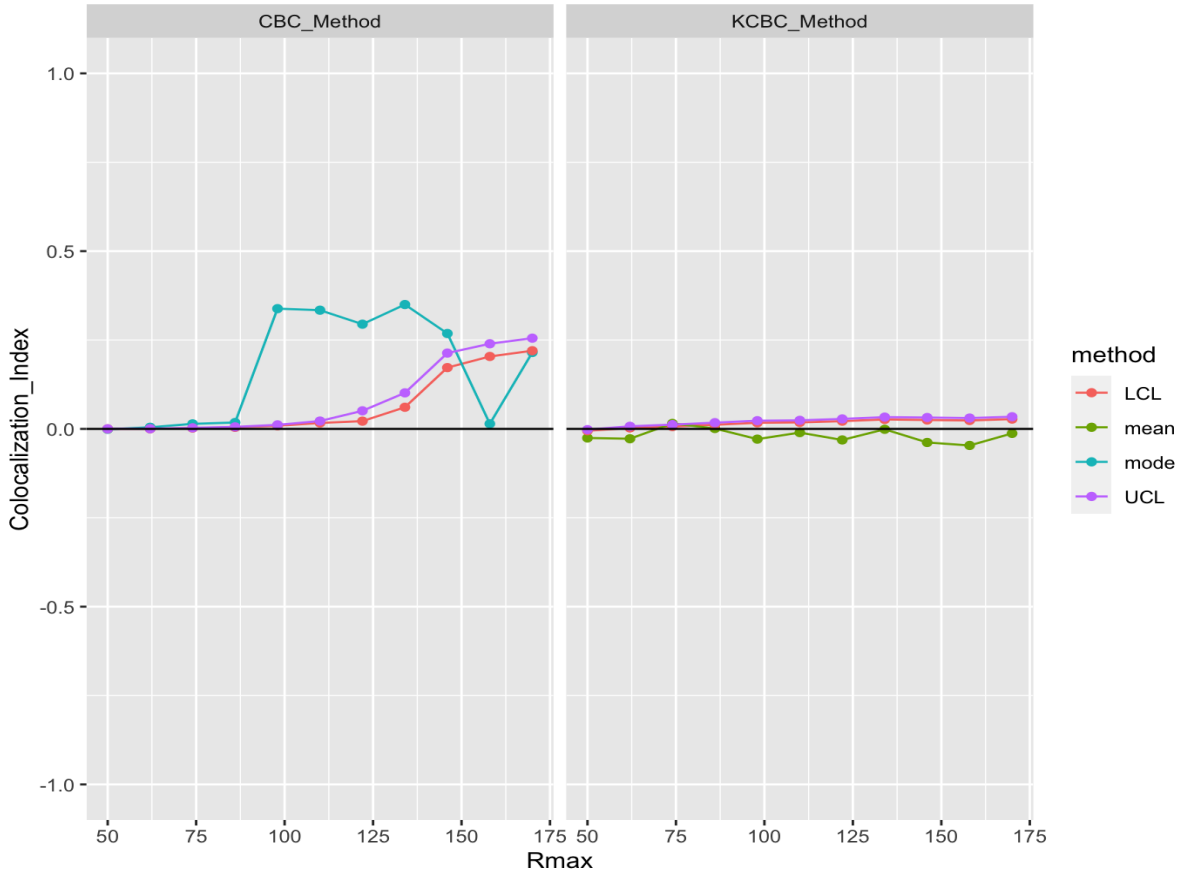


Figure 8:Line plot showing the co-localization indexes for the CBC and KCBC methods at each R_{max} *for subregion A.*

Rmax	CBC Index	LCL	UCL	KCBC Index	LCL	UCL
45	-0.0001	-0.0003	0.0000	-0.0270	-0.0057	-0.0019
55.5	-0.0001	0.0005	0.0009	-0.0364	-0.0045	0.0002
66	0.0328	0.0022	0.0031	0.0146	0.0010	0.0069
76.5	0.1145	0.0090	0.0162	-0.0458	0.0049	0.0111
87	0.1733	0.0230	0.0346	0.0017	0.0080	0.0148
97.5	0.2059	0.0495	0.0666	-0.0020	0.0093	0.0164
108	0.3427	0.0773	0.0974	-0.0178	0.0111	0.0186
118.5	0.3304	0.1022	0.1248	0.0053	0.0151	0.0227
129	0.3110	0.1235	0.1474	-0.0509	0.0138	0.0221
139.5	0.2959	0.1388	0.1634	0.0396	0.0150	0.0232
150	0.2366	0.1535	0.1788	-0.0090	0.0172	0.0255

Table 3: KCBC and CBC indices of Region B. The modes of the CBC values and the means of the KCBC indices at various $\boldsymbol{R_{max}}$ *ranging from 45 to 150 are given. The table also shows the summary of the modes and the 95% confidence intervals, lower confidence limit (LCL) and upper confidence limit (UCL) of the CBC values and mean and 95% confidence interval of the KCBC indexes at each* $\boldsymbol{R_{max}}$.

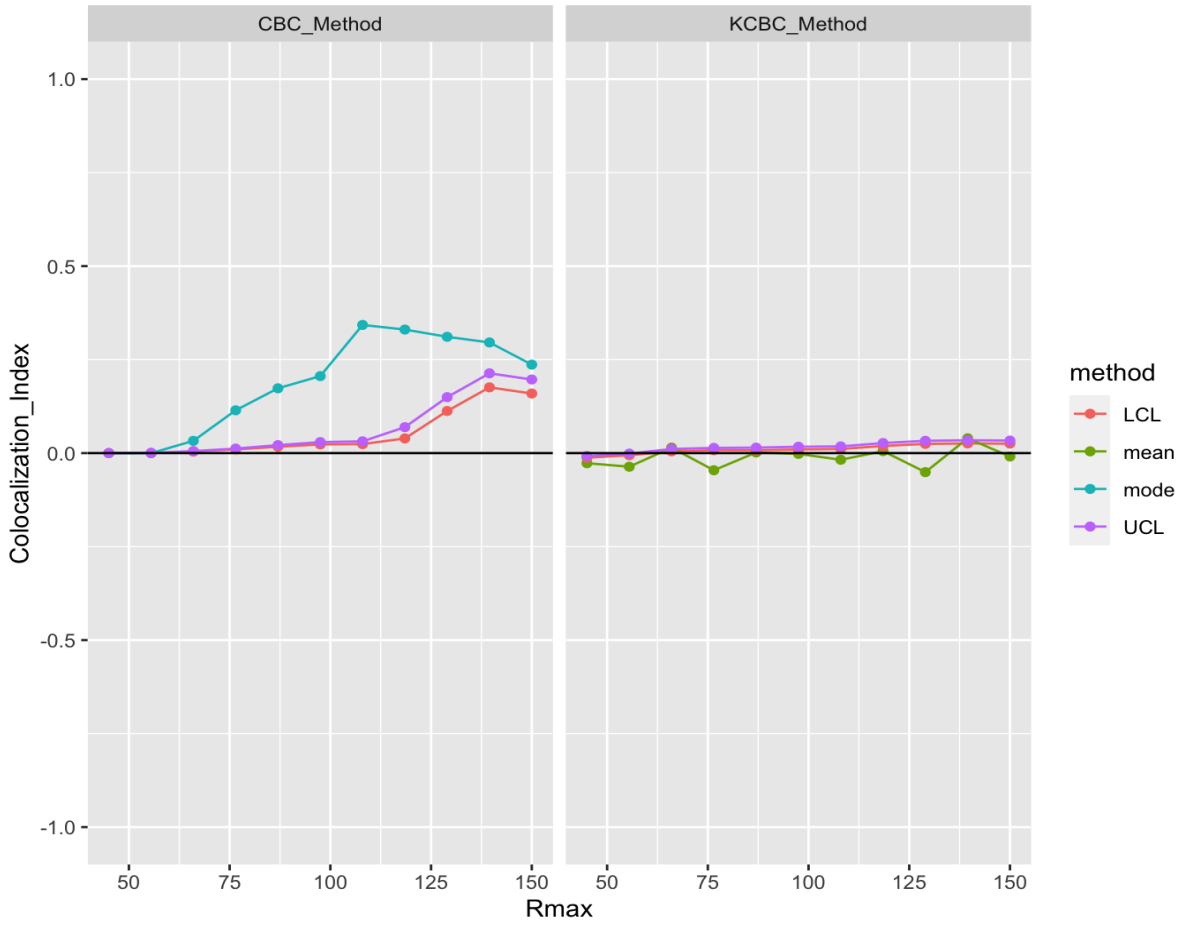


Figure 9: Line plot showing the co-localization indices for the CBC and KCBC methods at each R_{max} *for subregion B.*

7. Conclusion and Future Work

7.1. Introduction

In our final chapter, we summarize the research conducted in the thesis and discuss directions of future work.

7.2. The KCBC Method

We have discussed co-localization analysis and how the ways they are analysed. We saw that the co-localization indices seem to be upward biased when edge effect corrections are ignored especially when r is too large, and when the correlation of point proportions in nested circles is not taken care of in the CBC method. To correct the false positives, we develop the new KCBC method.

The main idea in KCBC is to employ the Ripley's K function over disjoint rings to correct the errors as a result of edge effects and ignorance of correlation of point proportions over nested circles. We conducted a simulation study using 500 CSR images to investigate the performance of the KCBC method. Results for the simulation study justifies the unbiasedness of the KCBC model for the CSR processes, especially for large r's. Finally, we applied the KCBC method on a real-life dataset to investigate interactions between two types of ants' nests. The two methods give quite different results.

7.3. Future work

In building our model, it was expected that the KCBC values approach +1 when there is co-localization and -1 when there is repulsion in the bivariate spatial point pattern. In future research, we will investigate more types of point patterns, such as cluster images, regular

images, one-dimensional images, etc. Also, in our CSR simulations, there is a very slight bias for medium valued distances though there is no bias for either small or large distances. We will investigate the reasons in future studies.

Bibliography

ants. (n.d.). Retrieved from CRAN repository: https://cran.r-project.org/web/packages/spatstat.data/index.html

Baddeley, A. J., & Silverman, B. W. (1984). A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics, 40*, 1089–1093.

Baddeley, A., & Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software, Volume 12*(Issue).

Diggle, P. (1983). *Statistical Analysis of Spatial Point Patterns.*

Diggle, P. J. ((1986)). Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *Journal of Neuroscience Methods, 18*, 115-125.

Dixon, P. M. (2002). Ripley's K function. *Encyclopedia of Environmetrics*, Volume 3, pp 1796 – 1803.

Dunn, K. W., Kamocka, M. M., & McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology, Vol. 300, No. 4*, C723–C742.

Harkness, R., & Isham, V. (1983). A Bivariate Spatial Point Pattern of Ants' Nests. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 32(3)*(doi:10.2307/2347952), 293-303.

Liu, X., Xu, J., Guy, C. S., Romero, E. B., Green, D., Cheng, C., & Zhang, H. (n.d.). Unbiased and Robust Analysis of Co-localization in Super-resolution Images.

Malkusch, S., & Heilemann, M. ((2016)). Extracting quantitative information from single-molecule superresolution imaging data with LAMA – LocAlization Microscopy Analyzer. *Sci. Rep., 6*(34486; doi: 10.1038/srep34486).

Malkusch, S., Endesfelder, U., Mondry, J., Gelléri, M., Verveer, P. J., & Heilemann, M. (2012). *Coordinate-based colocalization analysis of single-molecule localization microscopy data.* Springer-Verlag.

Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z., & Hagen, G. M. (2014, Aug 15). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics. , 30*((16)), 2389–2390.

Ripley. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society*(Series B 39), 172–212.

Ripley, B. D. (1981). *Spatial Statistics.* John Wiley & Sons.

Unwin, D. J. (2009). Spatial Statistics. *International Encyclopedia of Human Geography* , Pages 452-457.

Vita

The author, Stephan Cobby Komladzei, was born in Koforidua, Ghana. He obtained his undergraduate degree from the Kwame Nkrumah University of Science and Technology in Ghana, where he majored in BSc Actuarial Science. In January 2019, he enrolled as a graduate student at the Department of Mathematics, University of New Orleans (UNO) to pursue a masters' degree in Mathematics. At the Department of Mathematics at UNO, his concentration was mostly in Statistics related courses. He also worked as a graduate teaching assistant at the Department of Mathematics in his duration of study at UNO. His research interests lie in spatial statistics, clinical trials and statistical learning methods.