Describing heterogeneity in spatial point processes

by

Christine Kraamwinkel

Submitted in partial fulfillment of the requirements for the degree

Magister Scientiae

In the Department of Statistics
In the Faculty of Natural and Agricultural Sciences
University of Pretoria

July 2017

I, *Christine Kraamwinkel,* declare that this mini-dissertation (100 credits), which I hereby submit in partial fulfillment of the requirements for the degree Magister Scientiae in Applied Statistics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature:

Date:

Summary

In this mini-dissertation we briefly describe the context and development of spatial statistics, spatial sampling and point patterns. Thereafter spatial homogeneity is considered in detail.

Before selecting an appropriate sampling design in the spatial context, it is important to know whether the data is first- and second-order homogeneous. Currently the method of kernel smoothing is used to construct density plots which can be used to visually and subjectively infer on first-order homogeneity. We propose the use of hypothesis tests, developed for the comparison of K Poisson intensities from independent samples, in the spatial setting as a more rigorous statistical approach to testing for first-order homogeneity. We also discuss the data assumptions required for these hypothesis tests and provide suggestions for the users.

Acknowledgements

I would like to thank Dr Inger Fabris-Rotelli for her patience, motivation, insight, inspiration and the coffee she provided. When at times the challenges where overwhelming, she provided guidance and helped to put things in perspective. I could not have completed this work without her input and mentoring.

In addition, the following grants have made this research possible:

Lastly, thank you to my family and friends for all their love and support, for accepting that I could not always be there and for patiently listening to me talking about my work. I could not have done this without them.

Contents

Chapter 1

Introduction

In statistics, the term 'spatial' most commonly refers to the fact that data contains a geographical reference so that we can pinpoint on a map where each observation was taken [49], but can also refer to points in the brain or pixel locations in an image. Tobler's law of geography states that observations that are close together are more likely to be similar than observations that are further apart [98]. The field of spatial statistics deals with data in this setting and developed in conjunction with geographical information systems (GIS) through the field of regional science, an emergent field of enquiry in the 1950s [47]. Pioneering work in the field of mathematical statistics proved fundamental to the development of spatial data analysis. These works include the paper by Whittle [104] who extended autoregressive models for analysing variation in time series to formally represent spatial variation. The spatially autoregressive models introduced therein were amongst the first to appear in statistical literature and made it possible, when working with regular lattice data, to specify a formal representation of certain types of spatial structure for hypothesis testing, to test for model significance and to assess the goodness of fit to the data. Prior to this, it was only possible to test the null hypothesis of no spatial autocorrelation (in other words, no spatial structure) against a non-specific alternative hypothesis [45, 55, 70, 71].

In 1951, Krige [59, 60] proposed the use of statistical analysis of the spatial behaviour of gold grades, both collectively and individually, in order to improve valuation methods for the determination of the probable tonnage and grade of remaining ore. In this method, the interpolated values are modeled by a Gaussian process governed by prior covariances. The theoretical basis for this method was developed by the French mathematician and geologist, Matheron [67], by extending the Wiener-Kolmogorov stochastic prediction theory to the case of spatial processes defined on a continuous geographical space. This theory laid the groundwork for the field of geostatistics. Although the fields of geostatistics and spatial

data analysis initially developed as separate research fields, they have supported and added value to each other over time [47].

Many different types of spatial data exist along with many forms of spatial data analysis. These data types include polygons, lines, points, point processes and lattice/grid data [7, 10, 12, 49, 101]. Data that is recorded from a set of fixed locations on a continuous surface is known as 'data from a surface'. This data on a continuous surface might be measurements on land use, soil characteristics, air pollution, snow depth or precipitation levels [3, 19, 56, 106]. The attributes of interest exist everywhere, but are measured only at a few locations. Point observations for the attribute being measured could be sufficient for a continuous surface such a snow depth, whilst a domain with volume may be necessary for a surface such as air pollution. The attributes being measured are usually of a continuous nature, for instance diameters or heights of trees, but may also be categorical, for instance on/off, case/control, colour, etc. [39, 82]. Data recorded from point or area objects located in a geographical space is referred to as data from objects [62]. An example of point objects would be a set of towns scattered across a region. In this case, the attributes being measured may be quantitative or qualitative in nature. For both data from surfaces and objects, once the data is collected, the location of each observation is treated as fixed. [49]

Spatial data sets contain two aspects of variation, namely variation in data values disregarding the information provided by a locational index, and spatial variation which accounts for variation across the map, i.e. in locatia. The description of these calls for different strategies and terminology. A model which explains the variation in an attribute may also explain spatial variation, but this is not always the case [49]. Spatial analysis aims to describe the spatial variation in attribute values across a study area, or the spatial pattern of variation in terms of other attributes [49]. Descriptions may involve the detection of clusters or concentrations of high or low values. In areas such as geostatistics, spatial analysis can be used to provide estimates or predictions of attribute values at unsampled locations, or can be used for the mapping of attributes on the basis of a sample.

Spatial point patterns are datasets that give the locations of objects, also referred to as events, which are distributed within a study region [40]. Examples of points that may be studied includes any map-type data such as earthquake epicentres, crimes, bird nests, trees or even new cases of a disease/virus. Points may be located on a two-dimensional plane, on the surface of the earth, or in a three-dimensional volume or in space-time, referred to as a spatio-temporal point pattern. Points may contain additional attribute information, called marks, which may be categorical or continuous in nature. Datasets

could also contain explanatory data, referred to as covariates, which do not form part of the response and can be any type of data. Examples of covariates include spatial functions and spatial patterns. Spatial functions, for instance terrain altitude, are defined at all spatial locations and can be displayed as a pixel image or a contour plot. Spatial patterns, such as a map of geological faults, can be displayed as point patterns or as line segment patterns [7, 40]. The motivation for investigating point patterns is often to determine whether points exhibit some form of interpoint dependence [10].

The location of points or areas can often be seen as outcomes of some process [49]. The description of these location patterns is referred to as point pattern data with data commonly obtained from objects, not surfaces, as previously described. In these types of problems, the location of observations are not seen as fixed. A point pattern is defined as a set of locations that are irregularly distributed within a designated region and is presumably generated by some form of stochastic mechanism [40]. In the univariate case, a point process generates points, known as events, which consist of two variables (x, y) referring to a position in $\mathbb{R}^2$. An example of a univariate case would be the positions of trees in a forest. For ease of notation, we will refer to the location of an event as $\mathbf{x} = (x, y)$. In the case of a multivariate point process, additional information is collected for the events under consideration. In such a case, a point will consist of the variables $(\mathbf{x}, f(\mathbf{x}))$ where $\mathbf{x}$ refers to the position of the point and $f(\mathbf{x})$ is a discrete function. This function, known as a covariate or mark, can be of any form and no fixed information is included in the function [7]. A covariate refers to any variable that can be recruited as an explanatory variable and can potentially be observed at any spatial location within the observation window, not only at the observed point. Covariates can also be obtained by overlaying additional maps or layers such as geological polygons and doing a spatial join [2]. On the other hand, marks are attributes associated with each of the observed points. As an example of a multivariate point process we can expand on the univariate example of trees in a forest and collect additional information on attributes such as the heights, circumferences or species of trees.

Sampling within a spatial context is also important. Sampling is used to estimate characteristics of the population when we are unable to investigate the population as a whole. In an ideal world a sample would be a perfectly scaled-down version of the original population in the sense that every characteristic of the population would be matched in the sample [64]. Although this ideal is almost impossible to meet, researchers aim to get as close to this as possible. When certain subsets of the population are under- or over-represented in the sample, we run the risk that the estimates of the population characteristics become biased. Probability sampling methods, also known as random sampling

methods, can be used to obtain representative samples. The properties of these methods are well documented and sampling estimation errors can be calculated [64].

When sampling in the spatial context, the aim is to collect samples which are typically used to estimate the mean or total for some parameter in an area, for the optimization of parameter estimates for unsampled populations, or to predict the location of a movable object [103]. Two theories are commonly applied in these settings, namely design- and model-based sampling. The difference in these two approaches involves the sources of randomness in the technique. In design-based sampling, the population of values are regarded as unknown but fixed and randomness primarily arises from the random selection of sampling sites through traditional sampling methods [22]. On the contrary, model-based sampling regards the unknown population values as variable with observed values representing a single realization of a stochastic process, also known as a superpopulation. In these designs, sampling sights are fixed and inferences are based on the validity of the stochastic model. Unlike design-based schemes, a model-based sampling design is constructed by minimizing a predefined objective function leading to a single, optimal sampling design [103]. Design-based methods are therefore a combination of probability sampling and design-based inference, where model-based sampling is a combination of purposive sampling and interpolation [34].

Consider wildlife data and the process of obtaining samples thereof. Even though wildlife researchers are aware of the advantages of random sampling, these methods are usually not implemented due to practical complications. In practice, most samples are convenience samples [4, 91] so the selection probabilities of the elements cannot be described, making it impossible to derive statistically valid estimators and their errors. This implies that commonly used statistical methods may not be directly applicable or may not yield reliable results [4, 79, 91]. Typically, it is assumed that the convenience samples obtained in wildlife research approximate random samples so that inferences can be made about the population, however, these assumptions remain mostly unfounded and untested. In wildlife research, probability sampling methods such as simple random sampling (SRS) are not practical since all elements in the population may not be known, available or accessible. Instead, prior knowledge is often used to select elements, or in some cases, any available element is included. For example, when studying an endangered animal in the wild, a researcher would typically include any available animal in the study. Another common reason for convenience sampling in wildlife research is that many sampling opportunities could occur as a result of opportunity, rather than planning. Convenience sampling is not only a problem in wildlife research but also in general studies of non-controlled environments such as internet surveys, or other self-selected samples.

Nusser et al. [79] performed a landscape-based simulation study in order to explore the properties of estimators from convenience samples in relation to probability samples. The simulations included spatial information and considered the habitat preferences of the wildlife population, disease distribution and potential biases of a convenience sampling approach. A study such as this can help researchers understand the impact of taking a convenience and probability samples, but is limited by the assumptions being made. Krumm et al. [61] compared the prevalence of chronic wasting disease (CWD) among vehicle-killed mule deer to the disease prevalence among mule deer of the same sex sampled in the vicinity of collision sites. This study is useful in showing how a study based on convenience sampling of deer killed in vehicle collisions would overestimate the disease prevalence in natural deer populations.

These challenges also extend to research on domesticated animals. Rabies is a serious yet neglected public health threat in rural, underserved communities in Africa [57]. The effective control of this infectious disease relies on vaccinating a sufficient threshold proportion of the host population to effect herd immunity [41]. This threshold proportion is a function of the basic reproductive number, the number of secondary cases of infection generated by a typical infectious individual in a otherwise fully susceptible population. Hampson et al. [50] predicted that rabies outbreaks in dogs can be controlled if at least 40% of the population is immune at any time, however the rapid turnover in free-roaming dog populations, along with a lack of affordable and accessible veterinary services, makes achieving this goal challenging. It is therefore recommended that mass vaccinations are implemented annually by the state or other agencies in order to obtain a coverage percentage of 60-70% [50, 72]. Using census data obtained for villages in the Mara province, Northern Tanzania, Kraamwinkel et al. [58] investigated the effectiveness of traditional sampling techniques compared to a variation of a simple spatial-type sampling technique, namely the EPI cluster survey method.

These examples illustrate a need for more effective and appropriate sampling methods, also valid within a spatial domain. The principle behind Tobler's law of geography violates the critical assumption made in conventional sampling, namely that observations are independent and identically distributed within the population (or within subpopulations) and that selection probabilities of elements are known [64]. In reality, and specifically in the setting of wildlife and animal research, the data to be collected is usually spatially autocorrelated and heterogeneous with selection probabilities seldomly known. When using conventional sampling designs, this leads to inefficient and non-representative samples with questionable estimation value [87]. Spatial sampling techniques are aimed at getting results of a higher quality at a lower cost [103]. Cost as well as quality constraints are

commonly applied in order to ensure that sampling costs do not exceed a given budget whilst the quality of the results meet a given minimum requirement.

An important first step in spatial sampling involves the clarification of the sampling objectives. These objectives could include the estimation of population parameters, the prediction of values at unsampled sites, the mapping of an area or the identification of the position of a target [103]. It is crucial to know whether the objective is about the population or the superpopulation. In this context, a population refers to the 'here-and-now' whilst a superpopulation refers to a single realization of a stochastic process and deals with 'where' type questions [102]. For a population, we are often interested in estimating an attribute, making design-based methods more applicable, whilst for a superpopulation, we are interested in estimating the parameters of the underlying process, implying the need for model-based methods [103]. The appropriateness and considerations needed for each of these methods are discussed by Haining [49], Wang et al. [102] and de Gruijter et al. [34]. A useful summary and examples are supplied in Wang et al. [103].

Model-based sampling allows for spatial information to be included in the sampling design and aims to ensure representative sampling by making use of information regarding the underlying spatial structure of the population [49, 66]. However, before we are able to choose and fit an appropriate model, it is crucial that we understand this underlying data structure by considering first- and second-order homogeneity [102]. A surface which is first-order homogeneous has the same intensity (or mean) at different locations on the map. This implies that the expected number of observations is proportional to the size of the observation area. On the other hand, a surface which is second-order heterogeneous has different spatial autocorrelation structures in different sections of the map. This relates not to the expected number of points in a region but rather to the expected number of pairs of points in that region. This concept is discussed in more detail in Chapter 2. The focus of this mini-dissertation will be on methods that can be used to describe heterogeneity in point patterns.

Many tests have been developed for second-order heterogeneity in the spatial point process context and the theory of these is discussed in Section 2.2 with an application in Section 3.2. To our knowledge, tests for first-order homogeneity in the spatial context are not yet well-developed, with researchers mostly relying on the visual inspection of density plots or assumptions about the data to decide whether the data is homogeneous or not. These density plots are obtained through kernel estimation and are affected by the choice of whether or not to apply edge correction. An example of this is shown if Figure 1.1 where kernel estimation was applied to a simulated homogeneous point pattern with and without various edge corrections. As an example, if a judgement call had to be made regarding the

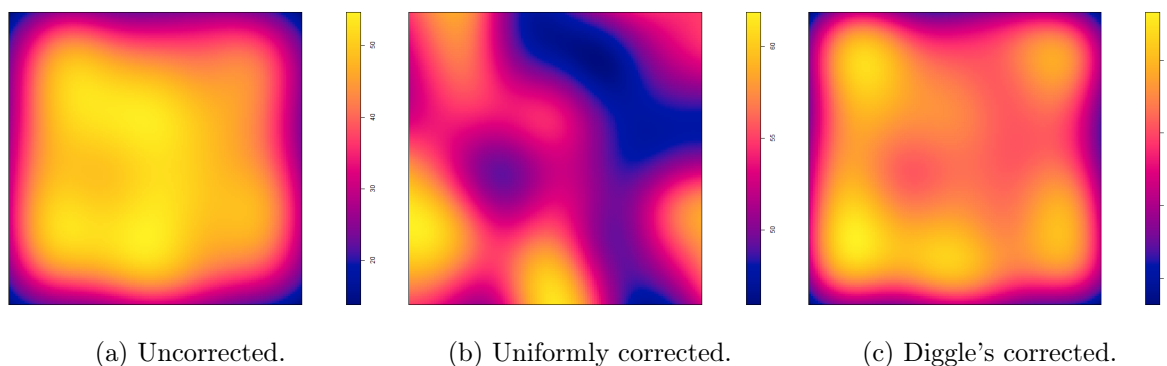(a) Uncorrected.   (b) Uniformly corrected.   (c) Diggle's corrected.

Figure 1.1: Kernel density estimates of a simulated homogeneous point pattern using various edge corrections and a Gaussian kernel.

homogeneity of this pattern using Figure 1.1b, a researcher might mistakenly classify the pattern as first-order heterogeneous. This example is discussed in more detail in Chapter 2 of this mini-dissertation. Tests for first-order homogeneity in point patterns are however merely the comparison of intensities at different locations on the map, leading us to apply tests developed in the context of time intervals to this setting. The theory of these tests is discussed in Section 2.3 with an application to real-world data provided in Section 3.3.

In Chapter 2 we will give a description of point patterns, their applications and the theory of the most common current methods used for assessing first- and second-order homogeneity. Our focus is on assessing first-order homogeneity and these are illustrated using current an proposed methods through a simulation study. In Chapter 3 we apply the first- and second order homogeneity tests and methods discussed in Chapter 2 to real-life point patterns and discuss the possible limitations of these.

In summary, in this mini-dissertation we aim to

- discuss briefly second-order homogeneity in order to indicate to the reader the impact of not evaluating first-order homogeneity correctly,

- discuss methods available for testing first-order homogeneity, and

- pinpoint any shortcomings, limitations and advantages of these tests.

Chapter 2

Tests for heterogeneity in spatial point patterns

2.1 Introduction

Before we can model a spatial point process we need to investigate the first-order and second-order homogeneity of the observed pattern. Our choice of model will depend on whether the intensity of the points vary across the pattern, namely first-order heterogeneity, and whether inter-point interactions vary from location to location, namely second-order heterogeneity.

Complete spatial randomness (CSR) is used to describe points in a point pattern that are randomly distributed [40, 44]. In considering CSR, we are in fact investigating second-order homogeneity as we are looking at the interaction between points. Three possibilities can occur when testing for CSR, namely that data will be regular, clustered or random [7, 10, 40]. Regular implies that points tend to avoid each other, leading to a pattern within the data with points distributed in a regular fashion (see Figure 2.2a). A clustered pattern, also referred to by some authors as an 'aggregated' pattern, implies that points may be grouped or clustered together. An example of a clustered pattern can be seen in Figure 2.1b and shows the locations of redwood seedlings. In this specific case, the clustering can be explained biologically as seedlings were known to be clustered around redwood stumps. The locations of these stumps were however not recorded, limiting the conclusions that can be made through statistical analysis.

A random pattern implies that the there is no pattern to the points, neither regular nor clustered. Figure 2.1a shows the locations of 65 Japanese black pine saplings. It is clear that no obvious pattern can be observed and points appear to be completely random. The theory and tests for complete spatial randomness are well developed and we will consider

these in Section 2.2. Lastly, we could also consider spatio-temporal point patterns (see Figure 2.2b) in which information is provided on both the location and the times where events occurred within a specified spatial region and time interval, but this is beyond the scope of this mini-dissertation. The topic is however addressed in the books by Cox and Lewis [25], Cressie [28] and Diggle [40], as well as the recent article by González et al. [46].
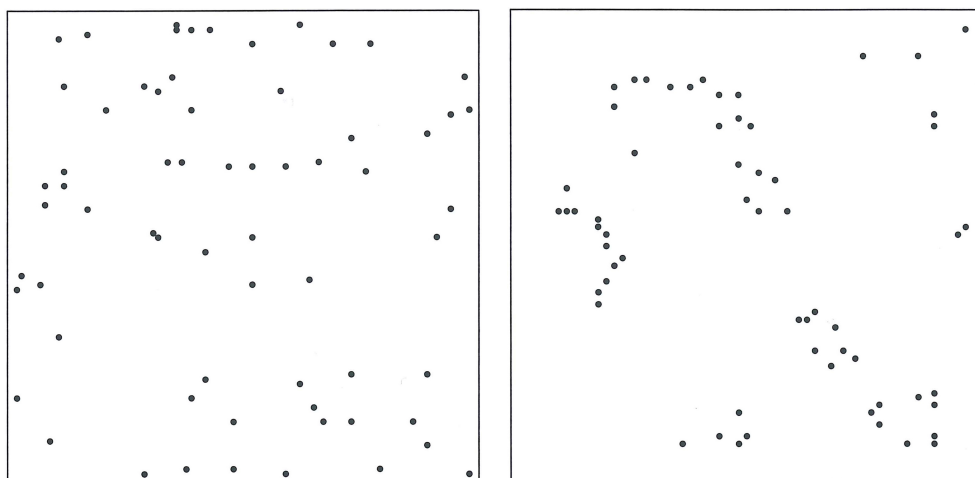
(a) Locations of 65 Japanese black pine saplings in a square of side-length 5.7 metres [78].

(b) Locations of 62 redwood seedlings in a square of side-length 23 metres [96][88].

Figure 2.1: Examples of random and clustered point patterns. Images obtained from [40].
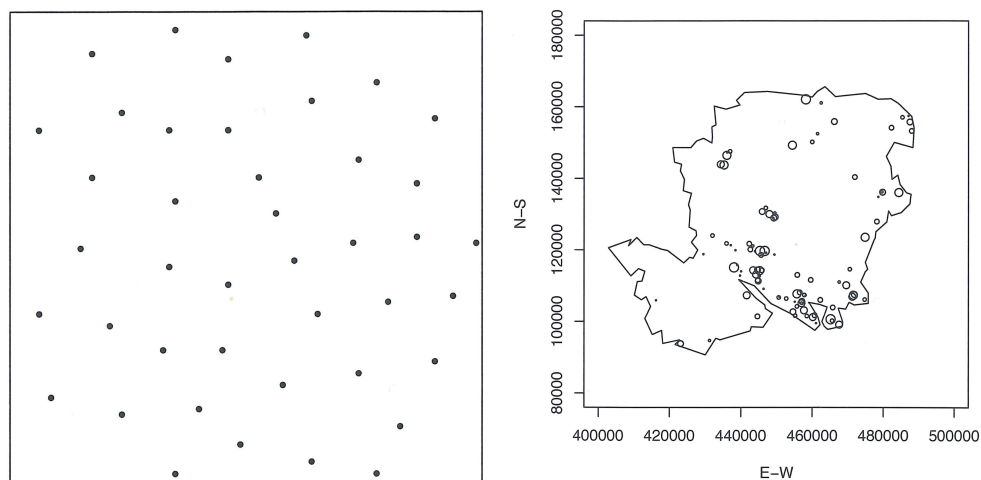
(a) Locations of 42 cell centres in a unit square [29][88].

(b) Locations of cases of non-specific gastrointestinal symptoms reported to NHS Direct Hampshire, UK, between 1 and 8 January 2001. The radius of each plotted circle codifies the reporting date.

Figure 2.2: Examples of regular and spatio-temporal point patterns. Images obtained from [40].

If points are found to be completely spatially random, using the methods in Section 2.2, we cannot sensibly fit a spatial model which takes into account a second-order trend. It is important to note that since second-order homogeneity refers to the interaction between points, we are in fact considering the correlation between these. In order to obtain an accurate measure of this correlation, we require sound knowledge of the process intensity. Without this it is likely that we will encounter problems of spurious correlation and confounding [10]. The evaluation of the intensity of the process will be considered in Section 2.3.

A difficulty that can arise in spatial point pattern analysis is so-called edge-effects. These effects occur when the region being observed is part of a larger region on which the underlying process operates. In these situations the points being observed may be interacting with points not being observed. There are various methods to adjust for these effects. Although our application in Chapter 3 of this mini-dissertation deals with complete point patterns and no edge irregularities are present, we will briefly consider the impact of applying edge corrections when investigating the first-order homogeneity of the process. [6, 36, 40, 89, 99]

2.2 Testing for Complete Spatial Randomness

An important step in modelling any observed point patterns is testing for CSR. Many tests are used to explore data and assist in formulating plausible alternatives for CSR. The informal combination of complementary tests is often useful in indicating the nature of departures from CSR, however, Cox et al [26] states that using multiple procedures as part of a diagnostic test only makes practical sense if the various tests examine different aspects of the pattern since a significant result in one test will not prevent a sensible interpretation for other tests. According to Diggle [40], tests for CSR should be seen as a natural starting point on the road to fitting stochastic spatial models. We will consider different hypotheses and their associated tests in the following sections. [40]

The spatial dependence of a point process considers the relationship between the number of events in pairs of subregions within a window W. This second-order property can be formally defined in terms of a mathematical limit as

$$\gamma\left(u_i,\, u_j\right) = \lim_{du_i,\, du_j \to 0} \left(\frac{E\left(n\left(\mathbf{du}_i\right) n\left(\mathbf{du}_j\right)\right)}{du_i du_j} \right)$$

where $\mathbf{du}$ is a small region around the point u, du is the area of this region, i is an index and $n\left(\mathbf{du}\right)$ is the number of observed points in this small region [44].

Classical techniques for investigating interpoint interaction are based on measuring the distances between points [7]. These techniques are referred to as distance methods and are discussed in Sections 2.2.4. There are three different types of distances we can consider, namely pairwise, nearest neighbour and empty space distances. We define these distances as follows

- **pairwise distances**, $s_{ij} = \|x_i - x_j\|$, considers all distinct pairs of points x_i and x_j in the pattern where $i \neq j$,

- **nearest neighbour distances**, $t_i = \min_{j \neq i} s_{ij}$, consider the distance from each point x_i to its nearest neighbour, and

- **empty space** or **void distances** , $d(u) = \min_i \|u - x_i\|$, considers the distance from a fixed reference location u in the window to the nearest data point.

2.2.1   Manual methods

An initial look at complete spatial randomness can be done using a semivariogram. A semivariogram, is a function describing the degree of spatial dependence in data and is defined as the variance of the difference between field values in two locations across realizations of the field (see Figure 2.3) [27]. Although many authors use the terms 'variogram' and 'semivariogram' interchangeably, the semivariogram is actually one half of the variogram [101]. This distinction does not always matter but could be crucial in some calculations. A graph of the semivariogram plotted against the separation distance yields valuable information regarding spatial variability. If observations close together are more alike, this graph starts at zero and increases as the separation distance increases, implying that pairwise differences increase and autocorrelation decreases, as the distance between points increase. As the separation distance increases, the semivariogram levels off to a constant value called a sill. Points which are further apart than this distance are considered spatially uncorrelated, having near constant variance in pairwise differences. This implies that a homogeneous point process will have constant variance in pairwise difference for all distances between points. In geostatistics, methods such as kriging require valid semivariograms which are often estimated by the empirical variogram. [101]

When comparing two spatial processes it is also useful to compare correlation instead of variance [101]. The correlogram, also known as an autocorrelation plot, is commonly used in time-series analysis to show the autocorrelations present in data. These can however also be applied to spatial data to check if points close together in space, as opposed to close together in time, exhibit some sort of dependence. If data is uncorrelated or random,
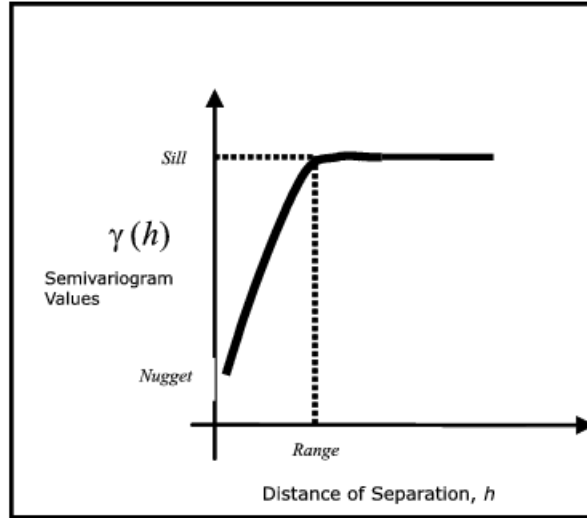
Figure 2.3: An example of a semivariogram. The semivariogram γ is plotted against the separation distance h. The range refers to the separation distance at which the semivariogram levels off, in other words the distance at which observations tend to become spatially uncorrelated. The nugget indicates if a spatial variable is continuous and the level of irregularity in the spatial variability. Image obtained from [43].

autocorrelations should be near zero with correlated data showing values significantly non-zero.

Moran's index, referred to as Moran's I [71], also provides a measure of spatial autocorrelation and is given by

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \left(X_i - \bar{X}\right) \left(X_j - \bar{X}\right)}{\sum_i \left(X_i - \bar{X}\right)^2}$$

for a point pattern where N is the number of spatial units, $\{X_i\}_{i=1}^N$ are the spatial units and w_{ij} are spatial weights. The weights are usually chosen as 1 for neighbouring spatial points and 0 otherwise, where the definition of a neighbour is appropriately chosen. Alternatively, weights can be assigned using distance decay functions such as covariogram contiguity which assigns weights based on the variogram model for the dataset [5, 73]. The index I ranges from -1 to 1, namely negative spatial autocorrelation to positive spatial autocorrelation, with a 0 value indicating CSR.

Another simple technique for assessing the correlation in a point pattern is based on dividing a rectangular observation window into quadrats of equal size and counting number of pairs of points within different quadrats [10]. If a pattern is both first-order homogeneous and CSR, then we would expect the total number of ordered pairs of points

within quadrats to be equal. In other words, if there are m quadrats containing $n_1, \ldots, n_m$ points respectively, then the jth quadrat should contain $n_j(n_j - 1)$ distinct ordered pairs of points. If the pattern has a total of n points, then there are $n(n-1)$ distinct ordered pairs of points. We can therefore calculate the fraction of all distinct ordered pairs of points in which both points fall in the same quadrat as $\frac{\sum_j n_j(n_j-1)}{n(n-1)}$. If the process is CSR, this fraction is expected to be $\frac{1}{m}$. The Morisita index is calculated as the ratio of the observed and expected fractions, namely

$$M = m\frac{\sum_j n_j\left(n_j - 1\right)}{n\left(n - 1\right)}.$$

This index should be close to 1 if the points are independent, less than 1 if the points are regular and greater than 1 if points are clustered. The Morisita index plot can be obtained by repeating this process for different sized quadrats and then plotting the Morisita index against the diameter of the quadrats. Figure 2.4 shows examples of these plots for random, clustered and regular patterns. It is important to note that this test can only be applied to a rectangular window and is based on the assumption that the point pattern has homogeneous intensity. If the pattern has inhomogeneous intensity then M can have large values because of the differences in intensity, rather than correlation between points. In addition, the derivation of the Morisita index is based on a Poisson process (the properties of this process are discussed in Section 2.2.2). If points are generated through some other process it is unclear how the value of this index should be interpreted. [10]

Fry plots can also be used to investigate the interspace distances between points. The technique was originally developed by Patterson [80, 81] for crystallography but was then independently reinvented for geophysics by Fry [42, 51]. These plots are created by printing the pattern on a piece of paper, then overlaying it with a transparency or tracing paper. This transparency should have a cross drawn in the middle, then position the cross over the first point in the pattern. Indicate the positions of all the points in the pattern on the transparency in their relative positions. It may happen that some points will not fit onto the transparency since they could be too far away but this is not a problem. This process is repeated by moving the cross to all data points and copying the positions of the other points onto the transparency. Examples of these Fry plots can be seen in Figure 2.5. The origin of the plot represents a typical point within the pattern and the points indicated on the plot represent the positions of nearby points relative to the typical point. If there is no obvious pattern, the plot indicates that points are CSR. If the center of the plot shows no points, it implies that all points are at least that minimum distance apart, meaning that the points have a regular distribution. If the center of the plot is crowded
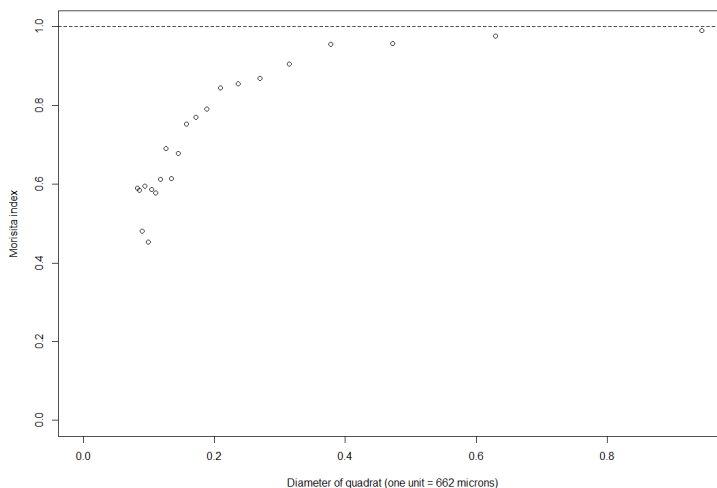
(a) Plot for a regular pattern created using the amacrine point pattern. [39]
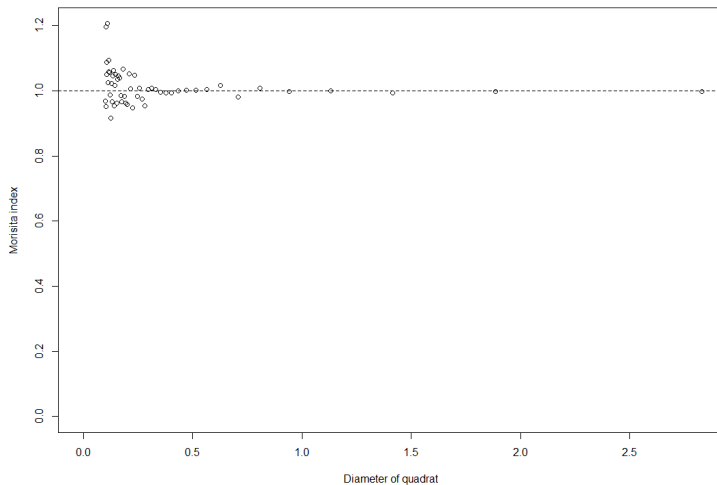
(b) Plot for a CSR pattern created from a simulated homogeneous point pattern.
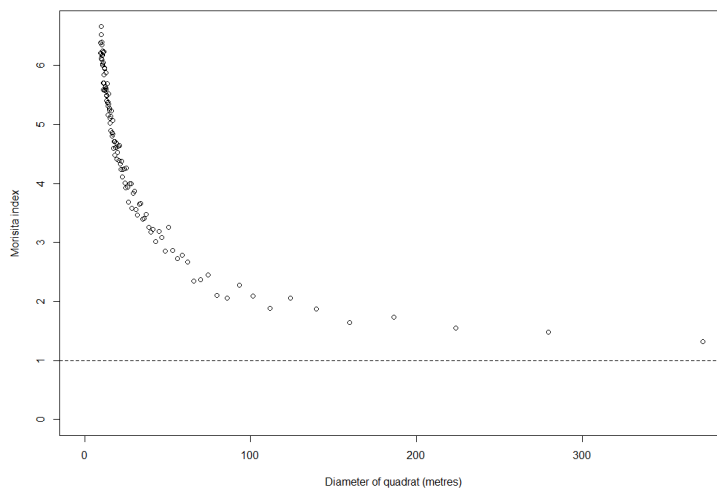
(c) Plot for a clustered pattern created using the `bei` point pattern. [23, 24, 54, 68]

Figure 2.4: Examples of Morisita index plots.
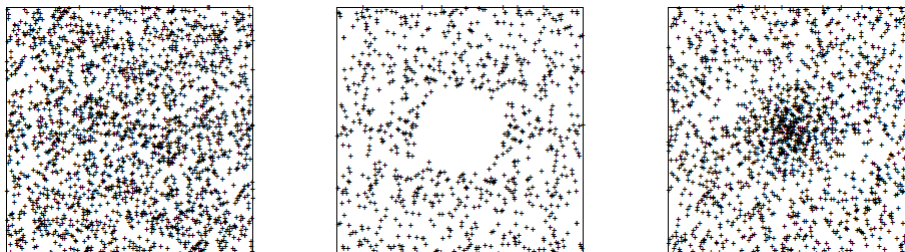
Figure 2.5: Examples of Fry plots for random (left), regular (middle) and clustered (right) patterns. Images obtained from [7].

with points, it implies that points tend to be close together, in other words a clustered pattern. Although these plots are easy to construct and interpret, they are based on the assumption that the process has homogeneous intensity. The interpretation of Fry plots have been very insightful in the field of geophysics, but their interpretation in other applications is very subjective and in order to extract useful information, these plots often need to be simplified, reduced or summarized.

2.2.2 Poisson Processes - a starting point

A point pattern that exhibits CSR can be modelled using a Poisson point process with intensity λ. Researchers usually focus on establishing that their data does not conform to this model and the Poisson point process therefore serves as a 'null model' in statistical analysis [7].

Definition 1. A homogeneous Poisson point process of intensity λ [7] has the following properties:

1. the number of points n falling within any region B is a Poisson random variable, with intensity parameter λ,

2. the expected number of points within a region B is equal to the intensity times the area of the region, $\lambda \times area(B)$,

3. if B_1 and B_2 are disjoint regions, then the number of points in each respective region, n_1 and n_2, are independent random variables, and

4. the n points in a region B are uniformly distributed in the region.

An important property of the Poisson distribution that should be considered is that a Poisson variable has a mean which is equal to its variance. The sample variance-to-mean ratio of the counts within subregions of the observation window can therefore be interpreted as a measure of over- or underdispersion of the counts, assuming these have a constant mean [10].

2.2.3 The inhomogeneous Poisson process

Definition 2. An inhomogeneous Poisson process of intensity $\lambda(u)$, $u \in \mathbb{R}^2$ where u is a spatial location, is a modification of the homogeneous Poisson process with the following properties:

1. the number of points n falling within any region B is a Poisson random variable,

2. the number of points, $N(\mathbf{X} \cap B)$, falling in a region B has expectation $E[N(\mathbf{X} \cap B)] = \int_B \lambda(u)\,du$,

3. if B_1 and B_2 are disjoint regions, then the number of points in each respective region, n_1 and n_2, are independent random variables, and

4. given that $N(\mathbf{X} \cap B) = n$, the n points are independently and identically distributed with common probability density $f(u) = {\lambda(u)}/{I}$, where $I = \int_B \lambda(u)\,du$.

2.2.4 Tests for second-order homogeneity

The F function

The F function of empty space distances was developed by Baddeley et al. [8]. If $\mathbf{X}$ is stationary point process, the cumulative distribution function of the empty space distance is given by

$$F(r) = P[d(u, \mathbf{X}) \leq r]$$

where u is an arbitrary reference location. For a first-order homogeneous process, this definition is not dependent on u [7]. The empirical distribution function of the observed empty space distances on a grid of locations u_j, $j = 1, ..., m$ is given by

$$F^*(r) = \frac{1}{m} \sum_j \mathbf{1}\left\{d(u_j, \mathbf{x}) \leq r\right\}$$

and is a negatively biased estimator of $F(r)$. This bias is a result of so-called edge effects and can be dealt with using a variety of corrections, all effectively forms of the Horvitz-Thompson estimator [53]. An unbiased estimator for $F(r)$ is given by

$$\hat{F}(r) = \sum_j e(u_j, r)\,\mathbf{1}\left\{d(u_j, x) \leq r\right\}$$

where $e(u_j, r)$ is an edge correction weight. Assuming an homogeneous point process, we can therefore compute an unbiased and reasonably accurate estimate of the empty space function F.

A useful benchmark for the interpretation of this estimate is the Poisson process. Note

that $d(u, \mathbf{X}) > r$ if and only if there are no points of $\mathbf{X}$ in the disc $b(u, r)$ of radius r, centered at u. A homogeneous Poisson process with intensity λ will have a mean number of $\mu = \lambda \times area(b(u, r)) = \lambda \pi r^2$ points falling in $b(u, r)$. Therefore, the probability of having no points in this region is given by $e^{-\mu} = e^{-\lambda \pi r^2}$. This means that for a Poisson process we have that

$$F_{pois}(r) = 1 - e^{-\lambda \pi r^2}.$$

To decide whether a point process is random, regular or clustered, we compare $\hat{F}(r)$ to the value obtained for $F_{pois}(r)$ by plugging in the estimated intensity $\hat{\lambda} = \frac{n(\mathbf{x})}{area(W)}$ where W is the observation window.

- If $\hat{F}(r) > F_{pois}(r)$ it suggests that the empty space distances in the observed point pattern are shorter than for a Poisson process which points to a regular spaced pattern.

- If $\hat{F}(r) < F_{pois}(r)$, a clustered pattern is indicated.

- If $\hat{F}(r) = F_{pois}(r)$, the pattern is said to exhibit complete spatial randomness and further spatial analysis is therefore not justified.

The G function

Unlike the F function which is based on empty space distances, the G function is based on nearest neighbour distances. The empirical distribution of these distances depends on the geometry of the sampling window W, as well as the characteristics of the point process $\mathbf{X}$. This again implies that corrections need to be made for edge effect bias.

The G function, developed by Baddeley et al. [8], states that if $\mathbf{X}$ is a first-order homogeneous point process, we can define the cumulative distribution of the nearest-neighbour distance for a typical point in the pattern as

$$G(r) = P(d(u, \mathbf{X} \setminus \{u\}) \leq r \mid u \in \mathbf{X})$$

where u is an arbitrary location and $d(u, \mathbf{X} \setminus \{u\})$ is the shortest distance from u to the point pattern $\mathbf{X}$ excluding u itself. For a stationary process, this definition does not depend on u itself.

The empirical distribution function of the observed nearest-neighbour distances is given by

$$G^*(r) = \frac{1}{n(\mathbf{x})} \sum_i \mathbf{1}\{t_i \leq r\}$$

and is a negatively biased estimator of $G(r)$ due to edge effects. There are many edge corrections available to deal with this bias, typically being weighted versions of the empirical cumulative distribution function,

$$\hat{G}(r) = \sum_i e\,(x_i, r)\,\mathbf{1}\,\{t_i \leq r\}$$

where $e\,(x_i, r)$ is an edge correction weight designed so that $\hat{G}(r)$ is approximately unbiased. A counterpart to the Kaplan-Meier estimator [9] is defined as follows. For $i = 1, ..., m$, let (y_i, d_i) with y_i the distance from each of the m sample points to the nearest other event in W and d_i denote the distance to the nearest point on the boundary of W. The estimator is then defined as

$$\hat{G}(r) = \frac{\#\,(y_i \leq r \text{ and } d_i > r)}{\#\,(d_i > r)}$$

where $\#$ is read as 'The number of points such that'.

The nearest-neighbour distance distribution function for a homogeneous Poisson point process of intensity λ is given by

$$G_{pois}(r) = 1 - e^{-\lambda \pi r^2}.$$

When comparing this function to the empty space function $F_{pois}(r)$, it can be seen that these functions are identical. We should however note that the interpretation of the G function differs from that of the F function.

- If $\hat{G}(r) > G_{pois}(r)$, this suggests that the observed nearest neighbour distances are shorter than for a Poisson process, suggesting a clustered pattern.

- If $\hat{G}(r) < G_{pois}(r)$ a regular pattern is suggested.

- If $\hat{G}(r) = G_{pois}(r)$, the pattern is said to exhibit complete spatial randomness and furher spatial analysis is therefore not justified.

The J function

The J function, a nonparametric measure of spatial interaction developed by van Lieshout and Baddeley [100], combines the G and F functions a useful manner as

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

for all $r \geq 0$ such that $F(r) < 1$. For a first-order homogeneous Poisson process, $F_{pois} = G_{pois}$, so that

$$J_{pois}(r) \equiv 1$$

for $r > 0$.

This function compares inter-event distances to distances from a fixed sampling point and can be used to measure both the strength and range of interactions. The function can be evaluated in closed form for more models than the F and G functions and contrary to these, a plot of the estimated J function can be interpreted without comparison to a Poisson point process. This function has been found to be a competitive choice, producing tests with similar power or more powerful alternatives than those of the F and G functions. However, the performance of tests based on the J function worsens drastically as the range of values taken into account grows, especially in tests using the Maximum Statistic. A simulation study investigating the power of tests based on the J function compared to tests based on the F and G functions was conducted by Thönnes et al. [97].

The J function can be estimated by

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

where $\hat{F}(r)$ and $\hat{G}(r)$ are defined as on pages 21 and 22. This estimator has been shown to be insensitive to edge effects [100].

The J function can be interpreted as follows.

- If $\hat{J}(r) > J_{pois}(r) \equiv 1$, a regular pattern is suggested.

- If $\hat{J}(r) < J_{pois}(r) \equiv 1$ a clustered pattern is suggested.

- If $\hat{J}(r) = J_{pois}(r) \equiv 1$, the pattern is said to exhibit complete spatial randomness.

The K function

When observing pairwise distances $s_{ij} = \|x_i - x_j\|$ in a point process, a bias occurs in favour of smaller distances. This bias is due to our inability to observe a pairwise distance greater than the diameter of the window W. The K function for a first-order homogeneous point process was defined by Ripley [88] in such a way that $\lambda K(r)$ is the expected number of points in the process within a distance r of a typical point in the process. The function is given by

$$K(r) = \tfrac{1}{\lambda} E\left[n\left(\mathbf{X} \cap b(u, r) \setminus \{u\} \right) \,|\, u \in \mathbf{X} \right].$$

For a homogeneous Poisson process, the fact that u is known to be a point of $\mathbf{X}$ does not affect other points in the process, thereby implying that $\mathbf{X} \setminus \{u\}$ is conditionally a Poisson process. This means that the expected number of points falling in $b(u, r)$ is given by $\lambda \pi r^2$. For a homogeneous Poisson process,

$$K_{pois}(r) = \pi r^2$$

irrespective of the intensity.

The empirical distribution function of the pairwise distances is of the general form

$$\hat{K}(r) = \frac{1}{\hat{\lambda}^2 area(W)} \sum_i \sum_{j \neq i} \mathbf{1}\{\|x_i - x_j\| \leq r\} e(x_i, x_j; r)$$

where $e(u, v, r)$ is an edge correction weight. As long as some edge correction is applied, the choice of estimator does not appear to be very important. The formulation of the edge correction depends on the shape of the study area W. More details about this can be found in the book by Diggle [36, 40].

To make an inference on the randomness of the point pattern, the estimate $\hat{K}(r)$ is compared to the K_{pois} function.

- If $\hat{K}(r) > K_{pois}(r)$, a clustered pattern is suggested.

- If $\hat{K}(r) < K_{pois}(r)$, a regular pattern is suggested.

- If $\hat{K}(r) = K_{pois}(r)$, the pattern is said to exhibit complete spatial randomness and further spatial analysis is therefore not justified.

The L function

The L function is merely a commonly used transformation of the K function and is defined as

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

In effect, the K function is transformed to the straight line $L_{pois}(r) = r$ which makes visual assessment much easier. The transformation approximately stabilises the variance of the estimator, which simplifies the assessment of deviations.

The L function is estimated by

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

and is interpreted as follows.

- If $\hat{L}(r) > L_{pois}(r)$, a clustered pattern is suggested.

- If $\hat{L}(r) < L_{pois}(r)$, a regular pattern is suggested.

- If $\hat{L}(r) = L_{pois}(r)$, the pattern is said to exhibit complete spatial randomness and further spatial analysis is therefore not justified.

In this section we considered various functions that can be used to evaluate second-order homogeneity in point patterns. The F function classifies a pattern as clustered, regular or random by considering empty space distances, in other words, the distance from a fixed reference location which may or may not be in the pattern, to the nearest observed point in the pattern. This function is useful as we gain more information by considering the cumulative distribution function of the empty space distances, $F(r)$, than we would from considering the average empty space distance as we are now considering the probability of observing a specified distance[10]. In the same way, we can gain more information regarding the distances between points and their closest neighbouring points by considering the cumulative distribution function of the nearest neighbour distances, $G(r)$, rather than the average of the nearest-neighbour distances. These empty space and nearest neighbour distribution functions, F and G, are combined to give the J function, a nonparametric measure of spatial interaction. This implies that the J function compares the surroundings of a point belonging to the process to the neighbourhood of a random point. The K and L functions consider spatial interaction of points by considering pairwise distances. The K function describes characteristics at different scales, both large and small. When considering pairwise distances, a bias occurs in favour of smaller distances since pairwise distances between points inside the observation window and points outside the observation window cannot be observed. The K function therefore considers the expected number of points in the process within a specified distance, r, of a typical point in the process. The L function merely transforms the K function into a straight line, making visual inspection easier.

2.2.5 Adjustment for the impact of first-order inhomogeneity

In Section 2.2.4 we discussed the F, G, J, K and L functions for first-order homogeneous point processes. The inhomogeneous counterparts of the F, G, J, K and L functions also hold and are available for implementation in the `spatstat` package of R [10, 11]. If a point process is not first-order homogeneous, deviations between the empirical and theoretical functions should not necessarily be seen as confirmation of interpoint interaction since it

could be attributable to variations in intensity [7]. An first-order inhomogeneous point process with intensity function $\lambda(u) \in \mathbb{R}^2$ has expected number of points in any region A given by $I = \int_A \lambda(u)\,du$. In addition, given n events $x_1, x_2, \ldots, x_n$ in a region A, the x_i's are an independent random sample from common distribution $f(u) = \lambda(u)/I$ on A.

2.3 First-order homogeneity

The hypothesis tests we now present, originally applied to one-dimensional Poisson processes for statistically testing first-order homogeneity, namely that of the intensity, $\lambda(t)$ where t is time. For the spatial application we consider $\lambda(u)$ where u is a spatial location.

In different contexts the average number of points can represent different concepts, for instance abundance of particles on a surface, density of light-sensitive cells in a retina, productivity of crops, risk of crimes, intensity of a lightning storm or prospectivity of undiscovered mineral deposits [10]. The standard generic term for this is intensity. A very important first step in analysing spatial data is determining whether the rates or intensities of K underlying spatial processes are the same [10, 21]. In this context, the K underlying spatial processes refers to subdividing the observation window into K quadrats and comparing the intensities of these. In practice, the intensity of a spatial point process could vary from place to place, in which case, the process is said to be first-order heterogeneous [44]. When calculating an intensity for a surface by dividing the total number of points by the area, we are in fact making a tentative assumption that the intensity is uniform across the area. This is often a working assumption in some kinds of analysis, but may be inappropriate. When intensity varies spatially it is effectively a function of spatial location and the intensity function can be estimated through statistical methods using observed data. [10]

Understanding the intensity of a process is often the aim of the scientific question under question. For instance, spatial variation in intensity could reflect the preference of wild animals for certain habitats or the natural segregation of certain plant species. However, when intensity is not the main focus of a study, it remains a crucial part of the analysis as intensity can easily be confused with other pattern characteristics such as clustering. For us to establish that a pattern is clustered we first need to eliminate alternative explanations such as spatially varying intensity. When it is known that intensity varies across a pattern, adjustments for this effect can be made to the tools used to investigate clustering. [10, 44]

Let $\mathbf{X}$ be a point pattern defined in two-dimensional space. In the homogeneous case, for any subregion B of a two-dimensional space, the intensity of a point pattern is taken

as the expected number of points of $\mathbf{X}$ falling in B and is proportional to the area of B:

$$E\left[n\left(\mathbf{X} \cap B\right)\right] = \lambda \left|B\right| \tag{2.1}$$

where $n\left(\mathbf{X} \cap B\right)$ is the number of points in $\mathbf{X}$ falling in the subregion B and $\left|B\right|$ is tbe area of the subregion B. The intensity, indicated as λ, can therefore be thought of as the average density of points per unit area where λ is a constant [7, 10, 44]. In the heterogeneous case, the expected number of points falling within a region B, with spatial varying intensity function of the process, $\lambda\left(u\right)$, is defined by Baddeley [7] as

$$E\left[n\left(\mathbf{X} \cap B\right)\right] = \int_{B} \lambda\left(u\right) du. \tag{2.2}$$

If Equation 2.2 holds, $\lambda\left(u\right)$ is called the intensity function of the point process. When points are concentrated along an edge/line, the point process does not have an intensity function. This is a common occurrence in seismology where earthquakes tend to occur on tectonic plate boundaries [10]. This implies that we can define the intensity of a point process as the mathematical limit

$$\lambda\left(u\right) = \lim_{du \to 0} \left(\frac{E\left(n\left(\mathbf{du}\right)\right)}{du}\right)$$

where $\mathbf{du}$ is a small region around the point u, du is the area of this region and $n\left(\mathbf{du}\right)$ is the number of observed points in this small region [44]. A point process is said to be stationary or first-order homogeneous if the intensity of the process, $\lambda\left(u\right) = \lambda$, is constant over the observation window W [44].

The techniques used to estimate intensity differ according to whether the process is first-order homogeneous or heterogeneous. According to Baddeley et al. [10], analysis often starts with the tentative assumption that a process is homogeneous and the validity of this assumption is then assessed. In point processes, the empirical density of points,

$$\hat{\lambda} = \frac{n\left(\mathbf{x}\right)}{\left|W\right|}, \tag{2.3}$$

is an unbiased estimate of the true intensity λ, assuming the point process has homogeneous intensity. In this equation, $\mathbf{x}$ is the observed point pattern observed in the window W, $n\left(\mathbf{x}\right)$ is the number of points in $\mathbf{x}$ and $\left|W\right|$ is the area of the window W. Additional assumptions are required in order to give a standard error for the intensity estimate. If we provisionally assume that the point process is Poisson then the observed number of points within W is a Poisson random variable with mean $\lambda\left|W\right|$. Since the mean and variance for

a Poisson random variable are equivalent, this implies that $\hat{\lambda}$ has variance given by

$$var\left(\hat{\lambda}\right) = \frac{var\left[n\left(\mathbf{X} \cap W\right)\right]}{|W|^2} = \frac{\lambda}{|W|}$$

leading to an estimate for the standard error of $\hat{\lambda}$ given by $\sqrt{\hat{\lambda}/|W|}$.

The intensity function can be estimated non-parametrically using techniques such as kernel smoothing, as recommended in the literature, or the method of quadrat counting [10, 11]. The `density.ppp` function available in the `spatstat` package in `R` [10, 11] performs kernel smoothing using an isotropic Gaussian kernel as default and yields a pixel image object which can indicate the direction and pattern of a possible trend. Quadrat counting is also implemented in `spatstat`, providing the number of observed points in a specified number of quadrats. Examples of both these methods can be found in [7]. These techniques are explained in the following sections followed by new suggested hypothesis tests for first-order homogeneity.

2.3.1 Quadrat counts

Quadrat counting, a simple visual technique that can be employed to check for first-order homogeneity, is employed by dividing the observation window W into sub-regions, called quadrats, and then counting the number of points within each quadrat. If a point process is homogeneous, the number of points in equal sized areas should be roughly equal. For simplicity we will assume that the m sub-regions, $B_1, \ldots, B_m$, have equal area, but this need not be the case. Let $n_j = n\left(\mathbf{x} \cap B_j\right)$ be the number of points in region B_j for $j = 1, \ldots, m$. If the intensity is homogeneous, these counts should be equal 'on average' since the counts in are unbiased estimators of the corresponding expected values $E\left[n\left(\mathbf{X} \cap B_j\right)\right]$. If a spatial trend is observed in the counts, this suggests that the intensity is inhomogeneous. [10]

Consider the point pattern shown in Figure 2.6. This pattern was simulated from a homogeneous Poisson point process with intensity $\lambda = 50$ generated in a square window with a length of 4 units. It can be seen from this pattern that homogeneous intensity does not imply that points are uniformly distributed, rather that the average number of points in areas of equal size are roughly the same. Figure 2.7 shows additional examples of point patterns with homogeneous intensity and Figure 2.8 shows examples of point patterns with inhomogeneous intensity.

Figure 2.9 shows the simulated homogeneous point pattern divided into quadrats of different sizes with the top figures showing the counts for the quadrats and the bottom fig-
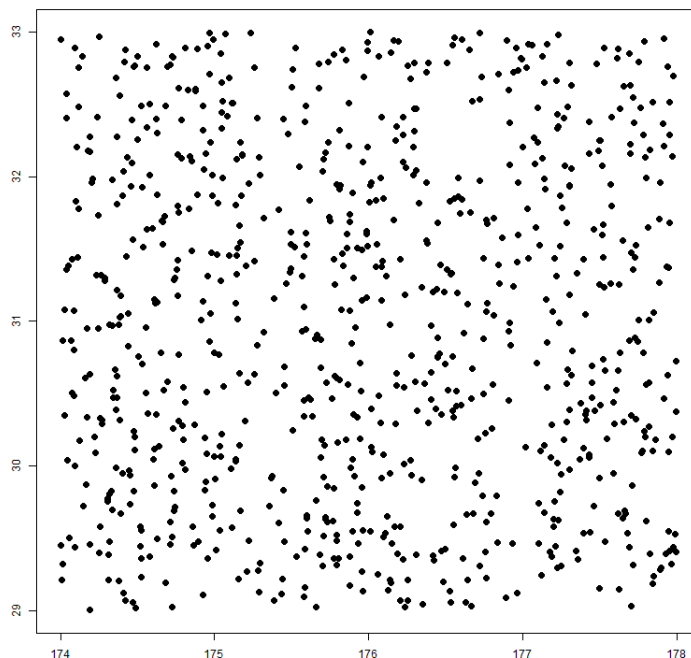
Figure 2.6: A point pattern simulated from a homogeneous Poisson point process with intensity $\lambda = 50$.
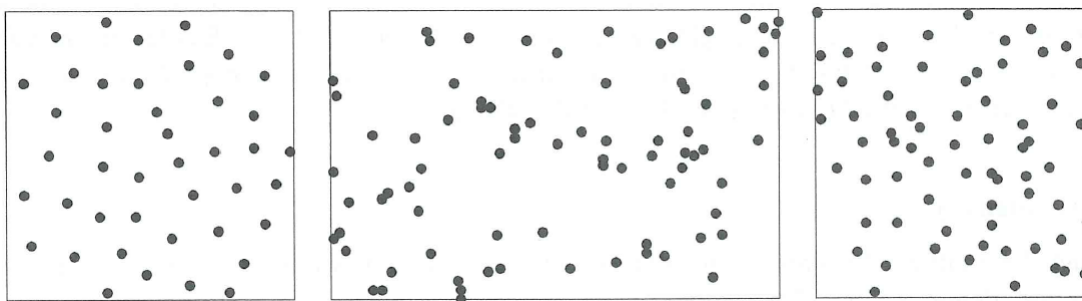
Figure 2.7: Examples of point patterns with roughly homogeneous intensity. The image on the left shows biological cell centres in a histological section [87, 88]. This image in the middle shows the locations of trees in a forest in New Zealand, excluding a five-foot border [65, 87]. The image on the right shows the location of Swedish Pine saplings [88, 95]. Images obtained from Baddeley [10].

Figure 2.8: Examples of point patterns with inhomogeneous intensity. The image on the left shows the locations of enterochromaffin-like cells in a histological section of gastric mucosa (Dr Thomas Bendtsen in [69]). The top of the image shows the interior of the stomach. The image on the right shows the locations of gorilla nesting sites in Kagwene Gorilla Sanctuary, Cameroon [10].

ures showing estimates for the intensity in each of the quadrats. These intensity estimates are calculated by dividing the quadrat counts by the quadrat size. The choice of the size of the quadrats is a tradeoff between bias and variability. While large quadrats reduce the relative error of the counts n_j, it also smooths out the variation in intensity within each of the quadrats. If we are willing to assume that the counts within the different quadrats are approximately independent variables with the same unknown distribution, and the quadrats are of equal size and shapes, the quadrat counts can be used to calculate the standard error for the overall estimate of intensity as

$$\sqrt{\frac{var\left(\hat{\lambda}\right)}{m-1}}.$$

Quadrats can have different shapes and sizes, however, should this happen the counts cannot be compared directly. Under the assumption of homogeneity, Equation 2.1 implies that the expected count in each quadrat is proportional to the area of the quadrat. This in turn implies that the average intensity given by Equation 2.3 is an unbiased estimator for the homogeneous intensity λ. We can therefore compare the estimated intensities in different sized and shaped quadrats in order to make an inference about the homogeneity of a point pattern. An example of this is shown in Figure 2.10.

It is however clear that visual inspection of these output can lead to subjective decisions regarding the homogeneity of the observed pattern. Hypothesis tests should be applied in order to make inferences in this regard. These suggested tests are discussed in Section 2.3.3.
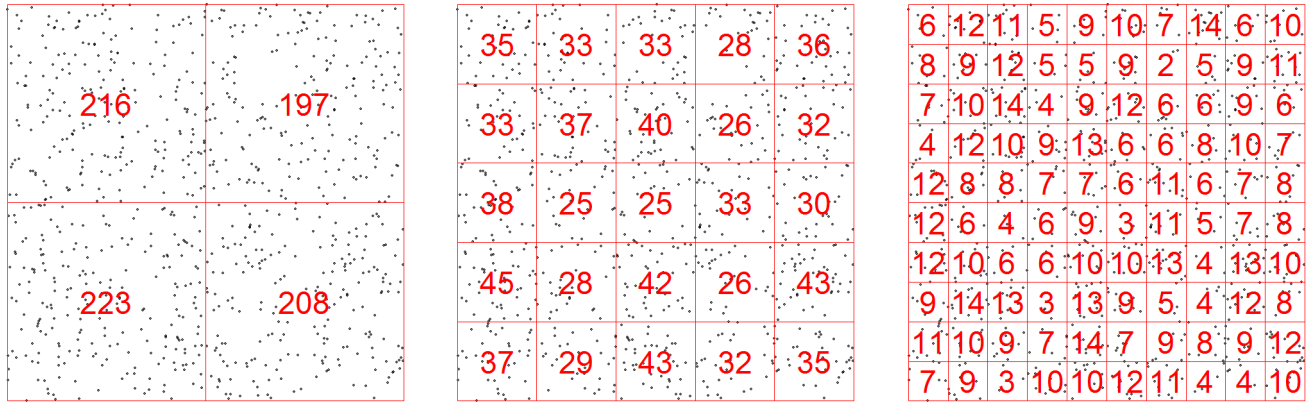
(a) Quadrat counts for 2×2, 5×5 and 10×10 grids.
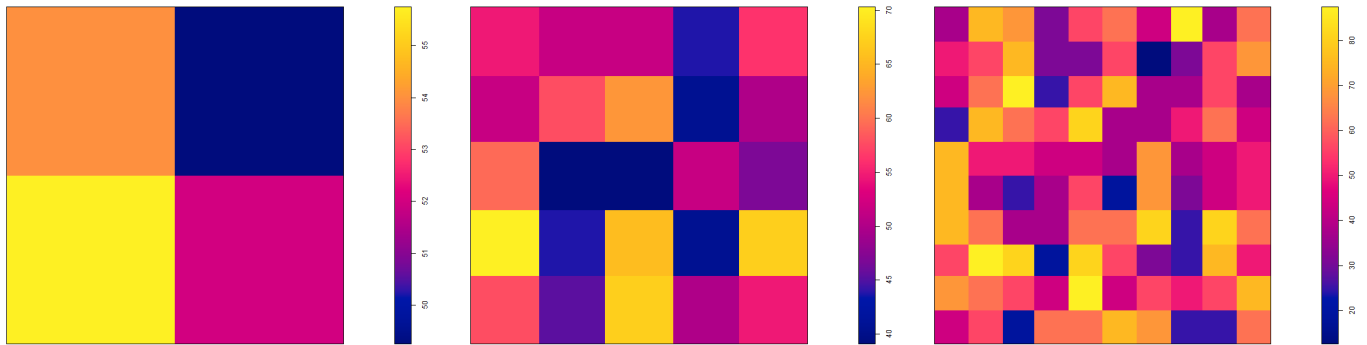
(b) Intensity estimates for 2×2, 5×5 and 10×10 grids.

Figure 2.9: The simulated pattern shown in Figure 2.6 divided into quadrats of various sizes.
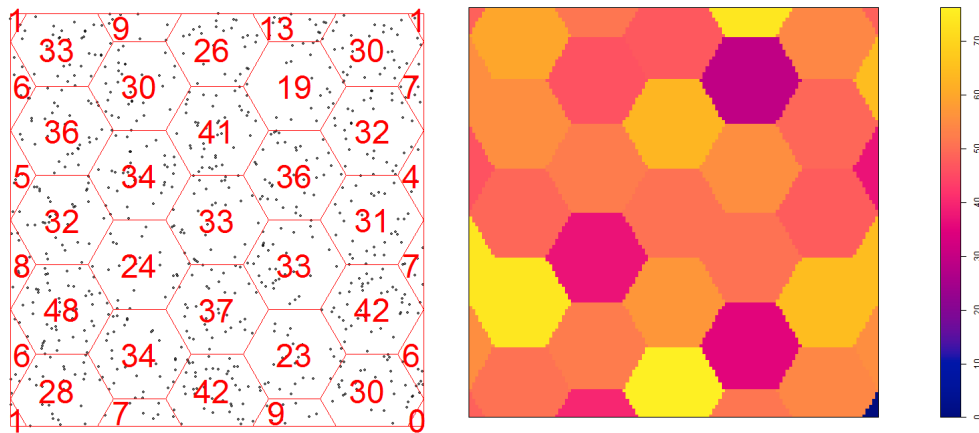
Figure 2.10: The point pattern in Figure 2.6 divided into quadrats of unequal sizes and shapes.

2.3.2 Kernel density estimation

Kernel smoothing is a nonparametric data smoothing technique that can be used to estimate the intensity function, $\lambda(u)$, of a point pattern. Waller et al. [101] compared fitting a kernel density estimate to mapping out events on a tabletop and then placing an identical mound of molding clay over each event. The mounds of clay will overlap for groups of events occurring close together, resulting in a taller pile of clay in those areas. When considering the entire surface, the clay represents a surface reflecting a nonparametric estimate of the intensity function. In effect, kernel density estimation centers a kernel function at each data point and then sums all these functions in order to create an overall smooth density estimate.

The kernel density estimator with kernel K is defined by

$$\hat{f}(u) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{u - X_i}{h}\right) \tag{2.4}$$

where h is the smoothing parameter, also known as the window width or bandwidth [93]. The n observed points are denoted by $X_1, \ldots, X_n$ and u is a location within the observed window. Equation 2.4 therefore implies that we will estimate the density at each point u in the observed window, regardless of whether or not a point was observed at u. This means that at locations with few observed points the density estimate will be low whilst locations with many observed points will have a much bigger density estimate. The shape of this density estimate at each location is given by the choice of the kernel function K. The choice of bandwidth parameter h will affect the width of the density estimate at each location. A larger bandwidth leads to more smoothing and the choice of bandwidth involves a tradeoff between variance and bias since as the bandwidth increases, the bias typically increases and the variance decreases [10]. The usual properties of kernel function as discussed by Silverman [93] are as follows:

- A kernel function must be non-negative if the density estimate is required to be a probability density function.

- A kernel function must be a valid probability density function.

- The kernel function must be continuous and differentiable so that $\hat{f}$ will inherit these properties.

In special circumstances the first property can be relaxed. For practical purposes, the kernel function is usually symmetric unless the researcher has prior knowledge to indicate

that an unsymmetric kernel would provide a better estimate. The usual kernel estimators of the intensity function $\lambda(u)$ are the uncorrected, uniformly corrected and Diggle's correction estimators, given respectively by

$$\hat{\lambda}^{(0)}(u) = \sum_{i=1}^{n} K(u - x_i), \tag{2.5}$$

$$\hat{\lambda}^{(U)}(u) = \frac{1}{e(u)} \sum_{i=1}^{n} K(u - x_i), \tag{2.6}$$

$$\hat{\lambda}^{(D)}(u) = \sum_{i=1}^{n} \frac{1}{e(x_i)} K(u - x_i), \tag{2.7}$$

where u is any spatial location inside the window W, $K(u)$ is the kernel function and $e(u) = \int_W K(u - v)\, dv$ is the correction that deals with the bias caused by edge effects [10]. The estimated intensity outside the observed window W is zero. Figure 2.11 shows the kernel estimates obtained for the estimators given in Equations 2.5 to 2.7 when applied to the simulated homogeneous point pattern shown in Figure 2.6 and Figure 2.12 shows the associated perspective plots of the estimates. These estimates were all obtained using a Gaussian kernel. The uncorrected estimate in Figure 2.11a shows a decline towards the boundary of W, commonly known as an edge effect. Points lying outside the boundary do not contribute to the intensity estimates near the boundary, leading to the decline. An uncorrected estimate should therefore only be used if it is known that all points have been observed and are contained within the observation window. When observing only some of the points in the pattern inside a window, we need to apply an edge correction. The uniformly corrected estimator is unbiased when the true intensity is homogeneous whilst Diggle's corrected estimator has a smaller mean square error leading to better overall performance. It should however be noted that kernel estimators are generally slightly biased since they smooth out the details of the intensity function [10].

The shape of the estimated density at each point u is given by the shape of the kernel function. Any probability density function can be used as a kernel. Kernels that appear regularly in the literature include the uniform, triangle, Epanechnikov, quartic/biweight, triweight and normal distribution density functions, the last being the most common choice [10, 93, 101]. The mathematical or computational properties may be better for some functions, however the differences between estimates based on different kernels are often small [101]. This is illustrated in Figure 2.13 by applying the kernels available in the `spatstat` package to the simulated point pattern in Figure 2.6. The Epanechnikov, qaurtic and Gaussian kernels give very similar estimates with only the disc kernel differing

(a) Uncorrected.   (b) Uniformly corrected.   (c) Diggle's corrected.
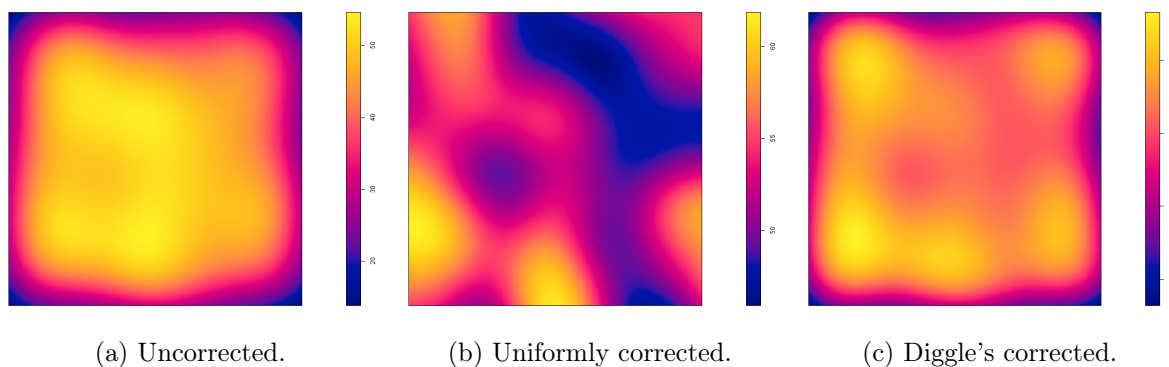
Figure 2.11: Density estimated of the simulated homogeneous point pattern using various edge corrections and a Gaussian kernel.

(a) Uncorrected.   (b) Uniformly corrected.   (c) Diggle's corrected.
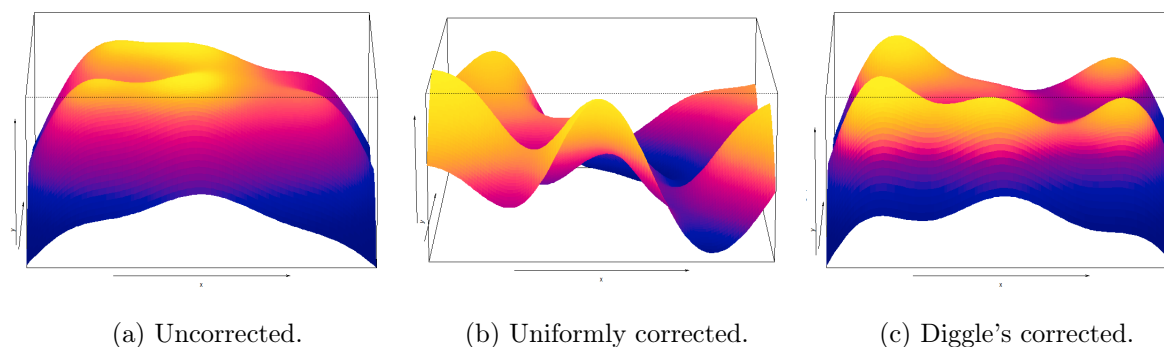
Figure 2.12: Perspective plots associated with the density estimates shown in Figure 2.11.

(a) Epanechnikov kernel.   (b) Quartic kernel.   (c) Disc kernel.
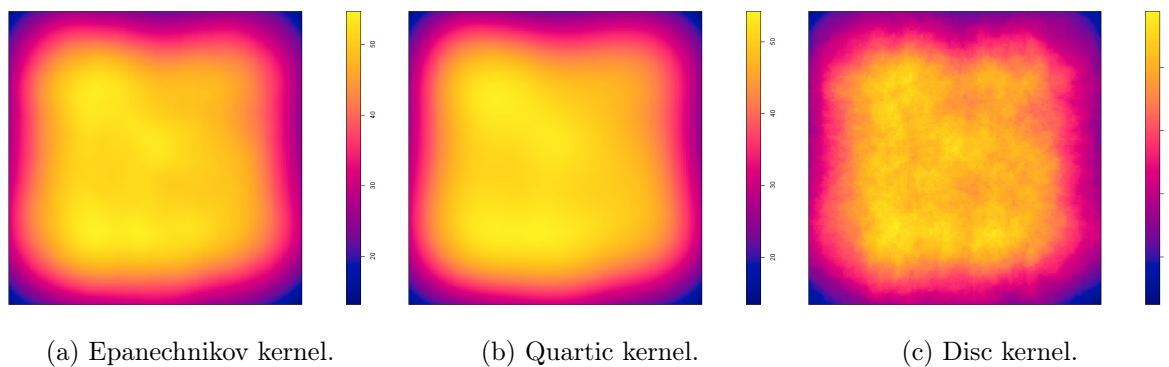
Figure 2.13: Density estimates for the simulated homogeneous point pattern in Figure 2.6 using different kernels.

noticeably.

The degree of smoothing is controlled by the kernel bandwidth. A small bandwidth leads to an undersmoothed intensity surface whilst a larger bandwidth leads to an over-smoothed surface. This is illustrated in Figure 2.14 where the bandwidth ranges from 0.1 to 1. The choice of bandwidth is not clear-cut but a very rough rule of thumb used to select the bandwidth is to set it equal to one-eighth of the shortest side length of the rectangular window, a rule which is often not satisfactory [10, 44]. Many bandwidth selection methods exist and are based on minimising error measures. These include Diggle and Berman's mean-square cross-validation method [14, 37], the likelihood cross-validation method [63] and Scott's rule of thumb [92]. Different bandwidth selection methods can give very different estimates, as it is often based on possibly inappropriate assumptions about the dependence between points. The mean square error cross-validation method for example assumes a Cox process whilst the likelihood cross-validation method assumes an inhomogeneous Poisson process. Optimal bandwidth selection is however not the focus of this mini-dissertation, but the interested reader is referred to [30, 38].

An indication of the accuracy of kernel estimates can be obtained by calculating standard errors and confidence intervals, however these require additional assumptions. Standard errors are justified by asymptotic theory but confidence intervals based on the standard error are notoriously unreliable since estimates of the intensity and estimates of the variance of the intensity are highly correlated [10]. In future research we will be investigating the application of the Kolmogorov-Smirnov test for the comparison of different kernel estimates [1]. This is however outside the scope of this mini-dissertation.

2.3.3 Hypothesis tests for first-order homogeneity

Currently in practice, first-order homogeneity is investigated through the visual inspection of kernel density plots. Two of the hypothesis tests discussed in this section, namely Pearson's χ^2 and the likelihood ratio test, are already available for use in the `spatstat` package in `R` [10, 11], however these tests are to our knowledge not widely applied in practice. The application of these tests also currently have software limitations that will be discussed in Section 3.3.2 of this mini-dissertation. In this section we elaborate on the theory of the aforementioned as well as three further hypothesis tests, newly extended in this mini-dissertation to the investigation of first-order homogeneity in point pattern data.

For data that is modeled by the Poisson distribution, count data is typically used to test for the equality of these intensities on K time intervals or spatial regions. The fixed observation periods or areas need not be of equal length/size, but are generally assumed to be independent. A number of studies have been performed to compare the size and
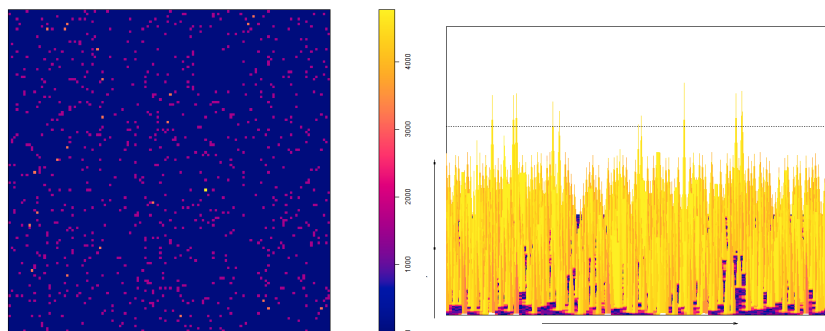
(a) Bandwidth of 0.01.
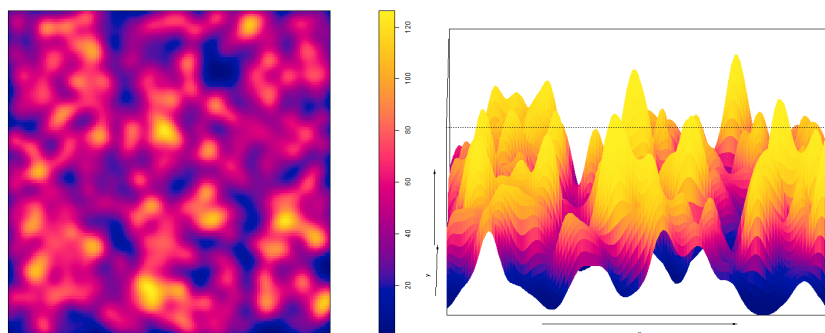
(b) Bandwidth of 0.1.
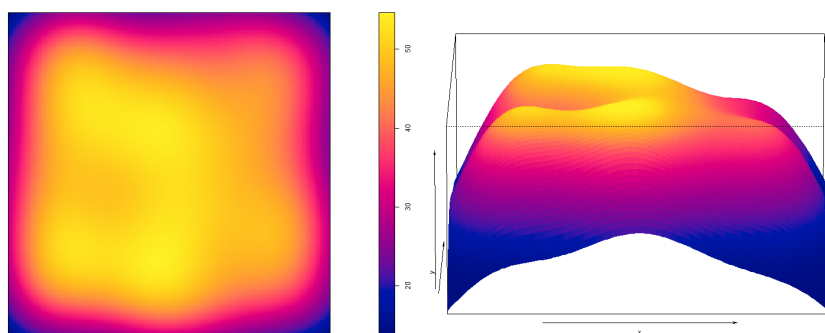
(c) Bandwidth of 0.5.
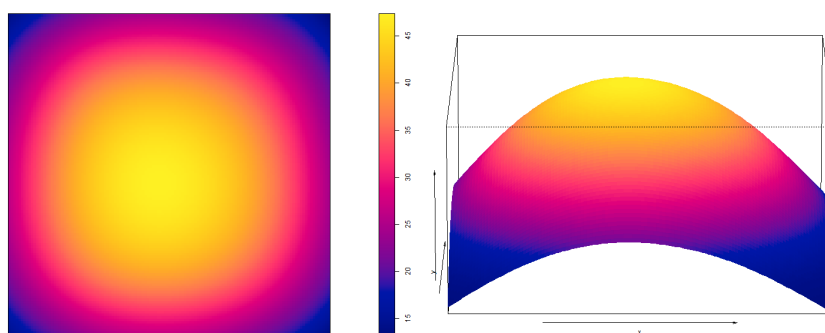
(d) Bandwidth of 1.

Figure 2.14: Comparison of density estimates and their perspective views for the simulated homogeneous point pattern in Figure 2.6 using different bandwidths.

x_1 points in area of size A_1	$\cdots$	$\cdots$
$\vdots$	$\ddots$	$\vdots$
$\cdots$	$\cdots$	x_K points in area of size A_K

Figure 2.15: Illustration of the division of a rectangular point pattern into quadrats containing x_i points in corresponding areas of size A_i.

power of numerous test statistics if $K = 2$ for both equal sampling frames (see [16, 35]) and unequal sampling frames (see [20, 76, 77]).

Points in a spatial pattern that are close together are however unlikely to be independent, violating this assumption. In some cases, points that are far apart may also be considered dependent. An example of this occurs when analysing weather patterns along a coastline or property damage along a fault line, when in one direction short distances will exhibit dependence, while in a different direction independence. Nonetheless, in some spatial patterns, points that are further apart could validly be considered independent. Where the researcher suspects that this is the case, the tests discussed in the following sections give valuable insight into the first-order homogeneity. Some level of intuition or judgement should be applied by the researcher when deciding whether the assumption of independence is justifiable, as well as deciding on the number of areas to compare.

Suppose we observe K independent random variables X_1, X_2, ..., X_K where each x_i is the observed number of occurrences in a spatial Poisson process with an unknown rate λ_i in a sampling frame with known area A_i. This concept is illustrated in Figure 2.15. The density function of X_i is given by

$$f\left(x_i\right) = \frac{e^{-\lambda_i A_i}\left(\lambda_i A_i\right)^{x_i}}{x_i!}, \qquad \texttt{for } i = 1,\ 2,\ \ldots,\ K.$$

The maximum likelihood estimate of λ_i is given by the empirical rate, $\hat{\lambda}_i = {x_i}/{A_i}$. The homogeneity test for K independent Poisson variates is given by $H_0 :\ \lambda_1 = \lambda_2 = \ldots = \lambda_K$ against $H_A :\ \lambda_i \neq \lambda_j$ for some $i,\ j \in [1,\ \ldots,\ K]$.

Under the assumption of independence, the likelihood function is given by

$$L\left(\lambda_1, \ldots, \lambda_K \mid x_1, \ldots, x_K\right) = e^{-\sum_{i=1}^{K} \lambda_i A_i} \frac{\prod_{i=1}^{K} \left(\lambda_i A_i\right)^{x_i}}{\prod_{i=1}^{K} x_i!}.$$

Under the null hypothesis, each X_i follows a Poisson distribution with parameter $\lambda_i = \lambda$, implying that the point pattern is first-order homogeneous. The maximum likelihood estimator of λ is given by

$$\hat{\lambda} = \frac{\sum_{i=1}^{K} x_i}{\sum_{i=1}^{K} A_i}.$$

Numerous statistics have been proposed to test for the equality of the rates of Poisson processes outside the field of spatial statistics, namely for time in one dimension. Here we extend these tests to be applicable in the spatial domain. In the following subsections we will be discussing these tests in the context of spatial point processes. These tests are closely linked to the the process of quadrat counts as the window W is divided into quadrats $B_1, ..., B_K$. The number of points within each quadrat, $x_1, ..., x_K$ is then counted and the size of each quadrat, $A_1, ..., A_K$, is noted. Under the null hypothesis, if the areas of each quadrat are equal, $A = A_1 = ... = A_K$, and the unknown intensity is λ, then the counts x_i are independent Poisson random variables with equal mean $A\lambda$ [10].

2.3.3.1 Parametric bootstrap tests

Sections 2.3.3.2 to 2.3.3.6 describe first-order homogeneity tests, all of which have χ^2-distributions that are valid only asymptotically. Tests using p-values calculated by these asymptotic distributions may encounter size distortions which can be addressed through the use of parametric bootstrap tests. These tests approximate the p-value through Monte Carlo resamples generated according to the homogeneity hypothesis [21, 33].

The procedure for approximating p-values is as follows. The test statistic, denoted τ, is calculated based on the observed counts. Next we draw a sample of K independent Poisson variates with mean $A_i\hat{\lambda}$, $i = 1, \ldots K$, where $\hat{\lambda}$ is the maximum likelihood estimate of the intensity, that is under H_0. The test statistic is calculated for this sample and is denoted by τ_1^*. This process is repeated R times so that we have $\tau_1^*, \ldots, \tau_R^*$. The bootstrap p-value as proposed by Davison and Hinkley [33] is

$$\text{bootstrap } p\text{-value} = \frac{\#\left\{i : \tau_i^* \geq \tau\right\} + 1}{R + 1}. \tag{2.8}$$

We reject the null hypothesis of homogeneity if the bootstrap p-value is less than or equal to the nominal significance level α. An alternative bootstrap p-value can be calculated

without adding 1 in the nominator and denominator in Equation 2.8 (see [31, 32]). In this mini-dissertation we will follow the original testing procedure.

Hope [52] illustrated that, in the case that there is no nuisance parameter, the power loss resulting from using parametric bootstrap tests is slight compared the corresponding uniformly most powerful test. This implies that R need not necessarily be large. Besag and Diggle [15] suggested that at $\alpha = 0.05$, $R = 99$ would be adequate. A simulation study conducted by Davison and Hinkley [33] showed that for $\alpha \geq 0.05$, the loss of power with $R = 99$ is not serious but that $R = 999$ should generally be a safe choice. In this mini-dissertation, we will compare the asymptotic p-values to the approximate p-values obtained from parametric bootstrap samples of sizes 99, 999 and 9999.

2.3.3.2 Pearson's χ^2 -test

Pearson's χ^2-test is arguably the most widely used χ^2-test. The test statistic was proposed by Potthoff et al. [83] and is given in the current context of a point pattern by

$$\chi^2_{\text{Pearson}} = \sum_{i=1}^{K} \frac{\left(x_i - \hat{\lambda}A_i\right)^2}{\hat{\lambda}A_i} = \frac{\sum_{i=1}^{K} A_i}{\sum_{i=1}^{K} x_i} \sum_{i=1}^{K} \frac{x_i^2}{A_i} - \sum_{i=1}^{K} x_i \tag{2.9}$$

This test can be applied in two different ways. Firstly, it can be used to test goodness-of-fit to the Poisson distribution assuming homogeneous intensity [17]. Secondly, it can be used to test for homogeneity by assuming independence [48]. Since we are interested in testing for first-order homogeneity we will be applying the second, namely the χ^2 test of uniformity. Under the null hypothesis, this test statistic has an asymptotic χ^2-distribution with $K - 1$ degrees of freedom. Traditionally, this approximation is acceptable if the expected number of points in each of the quadrats is at least 5. The power of the test depends on the size of the quadrats and is optimal when these are neither very large nor very small [10].

This test can be performed in the `spatstat` package using the function `quadrat.test`. The p-value for this test can be calculated in `spatstat` using the χ^2 approximation as well as Monte Carlo simulations. Figure 2.16 shows plots of the χ^2 test performed on the simulated homogeneous point pattern shown in Figure 2.6 using different grid sizes. In these plots we can see the expected number of points per quadrat and it should be noted that each is at least 5. The test statistics and parametric bootstrap p-values for these tests are shown in Table 2.1, indicating that the pattern has homogeneous intensity. Although `spatstat` can calculate Monte Carlo p-values, the ones given in this table were calculated using the approach described in Section 2.3.3.1 using 9999 samples. For all grid

Grid size	Test statistic	Asymptotic p-value	Bootstrap p-value		
2×2	1.7725	0.6209	0.6178	0.6182	0.6262
5×5	24.78	0.4178	0.4238	0.4138	0.4151
10×10	102.92	0.3737	0.3764	0.3873	0.3845

Table 2.1: Comparison of the asymptotic and three parametric bootstrap p-values obtained for the χ^2 test applied to the simulated homogeneous point pattern shown in Figure 2.6.

sizes the asymptotic and bootstrap p-values are similar, leading to the conclusion that the asymptotic distribution is "acceptable".

Baddeley [10] however pointed out the limitations of this test. Firstly, the alternative hypothesis that the point pattern has inhomogeneous intensity merely negates the null hypothesis of a homogeneous intensity. There are two possible reasons why a point process may fail to be a homogeneous point process, namely that the process does in fact not have homogeneous intensity, or that it violates the property of independence between points. The test can alternatively be viewed as a test of the independence property, assuming the intensity is homogeneous.

2.3.3.3   Likelihood ratio test statistic

The likelihood ratio test can be used to test both simple and composite hypotheses, by comparing the likelihood of the parameter(s) under the null to the likelihood of the parameter(s) under the alternative, given an observed dataset. Since the likelihood function is explicitly known, the likelihood ratio is obtained as

$$\Lambda = \frac{L\left(\hat{\lambda}, \ \ldots, \ \hat{\lambda} \mid x_1, \ \ldots, \ x_K\right)}{L\left(\hat{\lambda}_1, \ \ldots, \ \hat{\lambda}_K \mid x_1, \ \ldots, \ x_K\right)}.$$

It is convention to take that $0/0 = 0$ in case the sum of the x_i's equals 0. The likelihood ratio test statistic was originally proposed by Neyman and Pearson in 1928 [74] with the asymptotic distribution of the test derived by Wilks in 1938 [105]. Based on this test statistic, Chiu and Wang [21] developed a likelihood ratio test statistic to test for the equivalence of Poisson intensities, namely

$$LR = -2\Lambda = 2\left(\sum_{i=1}^{K} x_i \ln \frac{x_i}{A_i} - \sum_{i=1}^{K} x_i \ln \frac{\sum_{i=1}^{K} x_i}{\sum_{i=1}^{K} A_i}\right),$$

Top-left grid (2×2): each cell shows observed count (top left), expected count (top right), Pearson residual (bottom).

216 211 0.34	197 211 -0.96
223 211 0.83	208 211 -0.21

Top-right grid (5×5):

35 33.8 0.21	33 33.8 -0.13	33 33.8 -0.13	28 33.8 -0.99	36 33.8 0.39
33 33.8 -0.13	37 33.8 0.56	40 33.8 1.1	26 33.8 -1.3	32 33.8 -0.3
38 33.8 0.73	25 33.8 -1.5	25 33.8 -1.5	33 33.8 -0.13	30 33.8 -0.65
45 33.8 1.9	28 33.8 -0.99	42 33.8 1.4	26 33.8 -1.3	43 33.8 1.6
37 33.8 0.56	29 33.8 -0.82	43 33.8 1.6	32 33.8 -0.3	35 33.8 0.21

Bottom grid (10×10):

6 8.4 -0.84	12 8.4 1.2	11 8.4 0.88	5 8.4 -1.2	9 8.4 0.19	10 8.4 0.54	7 8.4 -0.5	14 8.4 1.9	6 8.4 -0.84	10 8.4 0.54
8 8.4 -0.15	9 8.4 0.19	12 8.4 1.2	5 8.4 -1.2	5 8.4 -1.2	9 8.4 0.19	2 8.4 -2.2	5 8.4 -1.2	9 8.4 0.19	11 8.4 0.88
7 8.4 -0.5	10 8.4 0.54	14 8.4 1.9	4 8.4 -1.5	9 8.4 0.19	12 8.4 1.2	6 8.4 -0.84	6 8.4 -0.84	9 8.4 0.19	6 8.4 -0.84
4 8.4 -1.5	12 8.4 1.2	10 8.4 0.54	9 8.4 0.19	13 8.4 1.6	6 8.4 -0.84	6 8.4 -0.84	8 8.4 -0.15	10 8.4 0.54	7 8.4 -0.5
12 8.4 1.2	8 8.4 -0.15	8 8.4 -0.15	7 8.4 -0.5	7 8.4 -0.5	6 8.4 -0.84	11 8.4 0.88	6 8.4 -0.84	7 8.4 -0.5	8 8.4 -0.15
12 8.4 1.2	6 8.4 -0.84	4 8.4 -1.5	6 8.4 -0.84	9 8.4 0.19	3 8.4 -1.9	11 8.4 0.88	5 8.4 -1.2	7 8.4 -0.5	8 8.4 -0.15
12 8.4 1.2	10 8.4 0.54	6 8.4 -0.84	6 8.4 -0.84	10 8.4 0.54	10 8.4 0.54	13 8.4 1.6	4 8.4 -1.5	13 8.4 1.6	10 8.4 0.54
9 8.4 0.19	14 8.4 1.9	13 8.4 1.6	3 8.4 -1.9	13 8.4 1.6	9 8.4 0.19	5 8.4 -1.2	4 8.4 -1.5	12 8.4 1.2	8 8.4 -0.15
11 8.4 0.88	10 8.4 0.54	9 8.4 0.19	7 8.4 -0.5	14 8.4 1.9	7 8.4 -0.5	9 8.4 0.19	8 8.4 -0.15	9 8.4 0.19	12 8.4 1.2
7 8.4 -0.5	9 8.4 0.19	3 8.4 -1.9	10 8.4 0.54	10 8.4 0.54	12 8.4 1.2	11 8.4 0.88	4 8.4 -1.5	4 8.4 -1.5	10 8.4 0.54

Figure 2.16: χ^2 test for homogeneity performed on the point pattern shown in Figure 2.6 using various grid sizes. Within each quadrat the observed count (top left), expected count (top right) and the Pearson residual (bottom) is shown. The Pearson residual gives a comparison of the observed and expected counts relative to the square root of the expected count. The magnitude of the residual indicates the degree to which the two differ. Roughly speaking, a residual above 2 in absolute value indicates a lack of fit.

Grid size	Test statistic	Asymptotic p-value	Bootstrap p-value
2×2	1.7805	0.7616	0.6254
5×5	24.7	0.8444	0.4348
10×10	108.67	0.4757	0.2927

Table 2.2: Comparison of the asymptotic and parametric bootstrap p-values obtained for the likelihood ratio test applied to the simulated homogeneous point pattern shown in Figure 2.6.

which under the convention that $0 \ln 0 = 0$, has an asymptotic χ^2-distribution with $K - 1$ degrees of freedom. Rao and Chakravarti [85] considered the particular case where the time periods (in our case area) are of length to 1, however Chiu and Wang [21] were the first to discuss the more general case.

This test can also be performed in the `spatstat` package using the argument `CR = 0` in the `quadrat.test` function. The p-value for this test can be calculated in `spatstat` using the χ^2 approximation as well as Monte Carlo simulations. The test statistics and p-values for these tests are shown in Table 2.2, indicating that the pattern has homogeneous intensity.

2.3.3.4 Score test

In 1948, the score test was introduced by Rao as an alternative to likelihood ratio test [84]. This test is simpler than the likelihood ratio test as it only requires estimation under the null hypothesis. This is particularly applicable when the success of an experiment depends on disproving the null hypothesis. The test is computationally quite simple but was only put to serious use in the 1980s when its use was popularized in the field of econometrics [13]. In the context of point processes, this test statistic is given by

$$SC = u\left(\hat{\lambda}, \ldots, \hat{\lambda}\right)' I\left(\hat{\lambda}, \ldots, \hat{\lambda}\right)^{-1} u\left(\hat{\lambda}, \ldots, \hat{\lambda}\right) = \left(\frac{\sum_{i=1}^{K} x_i}{\sum_{i=1}^{K} A_i}\right)^2 \sum_{i=1}^{K} \frac{A_i^2}{x_i} - \sum_{i=1}^{K} x_i$$

where

$$u\left(\lambda_1, \ldots, \lambda_K\right) = \left(\frac{\partial \ln L\left(\lambda_1, \ldots, \lambda \mid x_1, \ldots, x_K\right)}{\partial \lambda_1}, \ldots, \frac{\partial \ln L\left(\lambda_1, \ldots, \lambda \mid x_1, \ldots, x_K\right)}{\partial \lambda_K}\right)',$$

$$I\left(\lambda_1, \ldots, \lambda_K\right) = \left[-\frac{\partial^2 \ln L\left(\lambda_1, \ldots, \lambda \mid x_1, \ldots, x_K\right)}{\partial \lambda_i \partial \lambda_j}\right]_{i,j=1,\ldots,K}.$$

Grid size	Test statistic	Asymptotic p-value	Bootstrap p-value
2×2	1.799674	0.615006	0.6122
5×5	25.2935	0.389955	0.4601
10×10	149.7024	0.000761	0.3002

Table 2.3: Comparison of the asymptotic and parametric bootstrap p-values obtained for the score test applied to the simulated homogeneous point pattern shown in Figure 2.6.

A general approach to avoiding the problem of division by zero is to add 0.5 to every x_i whenever one of the observed x_i is zero. However in the spatial context, we are only interested in comparing intensities for areas where points have been observed, thereby making this step of the test avoidable. Under the homogeneity hypothesis, SC has an asymptotic χ^2-distribution with $K-1$ degrees of freedom. Ng and Cook [75] investigated score tests of homogeneity for Poisson processes with equal observation periods. Chiu et al. [21] performed a comparative study to consider the homogeneity of $K \geq 3$ Poisson count data with different observation periods.

A comparison of the asymptotic and bootstrap p-values for this test is given in Table 2.3. Considering a 2×2 and a 5×5 grid, the asymptotic p-values indicate a homogeneous pattern, but fails using a 10×10 grid. Although the asymptotic and bootstrap p-values are similar for a 2×2 grid, these differ considerably when increasing the grid size, bringing into doubt the validity of the asymptotic distribution. For all grid sizes, the bootstrap p-values correctly indicate that the pattern has homogeneous intensity.

2.3.3.5 VT test

In 1966, Potthoff et al. [83] proposed two test statistics that deal with the homogeneity of Poisson count data with different observation periods. The test statistic was developed in the context of time intervals, t_i, but we will consider the case of areas, A_i. The first of these is the VT test statistic based on the statistic

$$V = \left(\sum_{i=1}^{K} A_i \right) \sum_{i=1}^{K} \frac{x_i \, (x_i - 1)}{A_i}$$

where

$$\frac{V - \left(\sum_{i=1}^{K} x_i \right) \left(\sum_{i=1}^{K} x_i - 1 \right)}{\sqrt{2 \, (K-1) \left(\sum_{i=1}^{K} x_i \right) \left(\sum_{i=1}^{K} x_i - 1 \right)}}$$

Grid size	Test statistic	Asymptotic p-value	Bootstrap p-value
2×2	1.774614	0.621133	0.6178
5×5	24.211839	0.417349	0.4235
10×10	92.438459	0.372661	0.3765

Table 2.4: Comparison of the asymptotic and parametric bootstrap p-values obtained for the VT test applied to the simulated homogeneous point pattern shown in Figure 2.6.

follows the standard normal distribution asymptotically under the null hypothesis. A more mathematically refined approximation can be found as

$$VT = eV + f \sim \chi^2_{\nu_1} \tag{2.10}$$

with degrees of freedom $\nu_1 = e^2 (K-1) \left(\sum_{i=1}^{K} x_i \right) \left(\sum_{i=1}^{K} x_i - 1 \right)$. The constants in Equation 2.10 are given by

$$e = \frac{2(K-1)}{\left(\sum_{i=1}^{K} A_i \right) \left(\sum_{i=1}^{K} \frac{1}{A_i} \right) - 3K + 2 + 2(K-1) \left(\sum_{i=1}^{K} x_i - 2 \right)}$$

and

$$f = e \left[(K-1) e - 1 \right] \left(\sum_{i=1}^{K} x_i \right) \left(\sum_{i=1}^{K} x_i - 1 \right)$$

and are determined in such a way that VT and $\chi^2_{\nu_1}$ have the same first three moments.

Table 2.4 shows the asymptotic and bootstrap p-values obtained when applying the VT test to the simulated homogeneous point pattern. All grid sizes lead to similar p-values and indicate a homogeneous intensity present in the pattern.

2.3.3.6 UT test

The second test statistic proposed by Potthoff et al. [83] is also an asymptotic χ^2-test which has been proven to be the locally most powerful test against a selected alternative if λ is known [21, 83]. As with the VT test, we will consider areas, not time intervals. The UT test statistic is based on the statistic

$$U = \sum_{i=1}^{K} x_i^2 - \sum_{i=1}^{K} x_i - 2\lambda \sum_{i=1}^{K} A_i x_i.$$

and

$$\frac{U + \lambda^2 \sum_{i=1}^{K} A_i^2}{\sqrt{2\lambda^2 \sum_{i=1}^{K} A_i^2}} \tag{2.11}$$

follows the standard normal distribution asymptotically under the null hypothesis.

A more mathematically refined approximation can be found as

$$UT = gU + h \sim \chi^2_{\nu_2} \tag{2.12}$$

with degrees of freedom $\nu_2 = g^2 \lambda^2 \sum_{i=1}^{K} A_i^2$. The constants in Equation 2.12 are given by

$$g = \frac{\sum_{i=1}^{K} A_i^2}{\sum_{i=1}^{K} \frac{A_i^2}{2} + \lambda \sum_{i=1}^{K} A_i^3}$$

and

$$h = g\left(g+1\right)\lambda^2 \sum_{i=1}^{K} A_i^2$$

and are determined in such a way that UT and $\chi^2_{\nu_2}$ have the same first three moments. Potthoff and Whittinghill [83] proposed an alternative estimator, λ^*, for the case where λ is unknown. This estimator is the value of λ that minimizes Equation 2.11 and is given by

$$\lambda^* = \sqrt{\frac{\sum_{i=1}^{K} x_i^2 - \sum_{i=1}^{K} x_i}{\sum_{i=1}^{K} A_i^2}}.$$

Chiu and Wang [21] confirmed through a simulation study that the UT test is sometimes too conservative in small experiments and shows a marked loss of power in such a situation. This handicap is not shared by the Pearson's χ^2 and VT tests.

Finally, Table 2.5 shows the comparison of the asymptotic and bootstrap p-values obtained for the UT test. For a 2×2 grid, there is a considerable difference evident in these p-values, however as the grid size is increased, the bootstrap and asymptotic p-values become more similar. As with the score test, we have reason to doubt the validity of the asymptotic distribution, however all p-values correctly indicate a homogeneous intensity in the pattern.

Grid size	Test statistic	Asymptotic p-value	Bootstrap p-value
2×2	1.757402	0.777848	0.6177
5×5	24.05834	0.474219	0.4236
10×10	91.89836	0.399748	0.3764

Table 2.5: Comparison of the asymptotic and parametric bootstrap p-values obtained for the UT test applied to the simulated homogeneous point pattern shown in Figure 2.6.

2.4 Conclusion

In this chapter we described two different types of homogeneity, namely first-order and second-order. We also described tests and functions that are currently employed in assessing second-order homogeneity, also known as tests for complete spatial randomness. We described the methods currently employed for assessing first-order homogeneity and introduced hypothesis tests as a more rigorous form of assessment. Visualisation of the density fitted using kernel density estimation is considered as the current method for assessing the spatial intensity of a point process. We should note that it is also possible to employ adaptive kernel estimation, a method in which the bandwidth varies according to the density of observed points within an area. Here, in order to avoid smoothing out too much detail, the bandwidth is reduced in areas that are densely populated since we have more detailed information available on the variation in intensity [18, 44].

In this chapter, methods for testing first-order homogeneity were assessed using a simulated homogeneous point pattern. The results over various grid sizes and window choices were discussed.

In the next chapter we will employ these tests on real-life spatial point pattern data.

Chapter 3

Application

3.1 Data description and setting of analysis

Our census data was collected in Mara province, Northern Tanzania[1,2]. Data is available for a total 34253 households in 78 villages. The smallest village consists of 102 households and the largest has 1630 households. Table 3.1 shows an overview of this data.

Number of Villages	78
Total number of Households	34253
Minimum number of households	102
Maximum number of households	1630
Average number of households	439.14
Median number of households	370.5

Table 3.1: Overview of village descriptives

Figures 3.1 show the locations of households in four randomly selected villages. The data has spatial mark information of the number of cats and dogs in each household as the data was collected with the intention of developing an effective rabies vaccination schedule for rural villages. This mini-dissertation however only focuses on the homogeneity checks for the household spatial distributions, not on the creation of a sampling design, which could be investigated in future work. A summary of the number of households within each of the selected villages is given in Table 3.2. It should be noted that when calculating the distance between two coordinates, r, a measurement of $r = 0.01$ is approximately equivalent to a distance of 1.1km.

---

[1]http://www.gla.ac.uk/researchinstitutes/bahcm/staff/katiehampson/
[2]http://www.katiehampson.com/#intro

Village	Number of households
Machochwe	713
Mesaga	512
Morotonga	501
Ring'wani	234

Table 3.2: Number of households (points) within the selected villages.
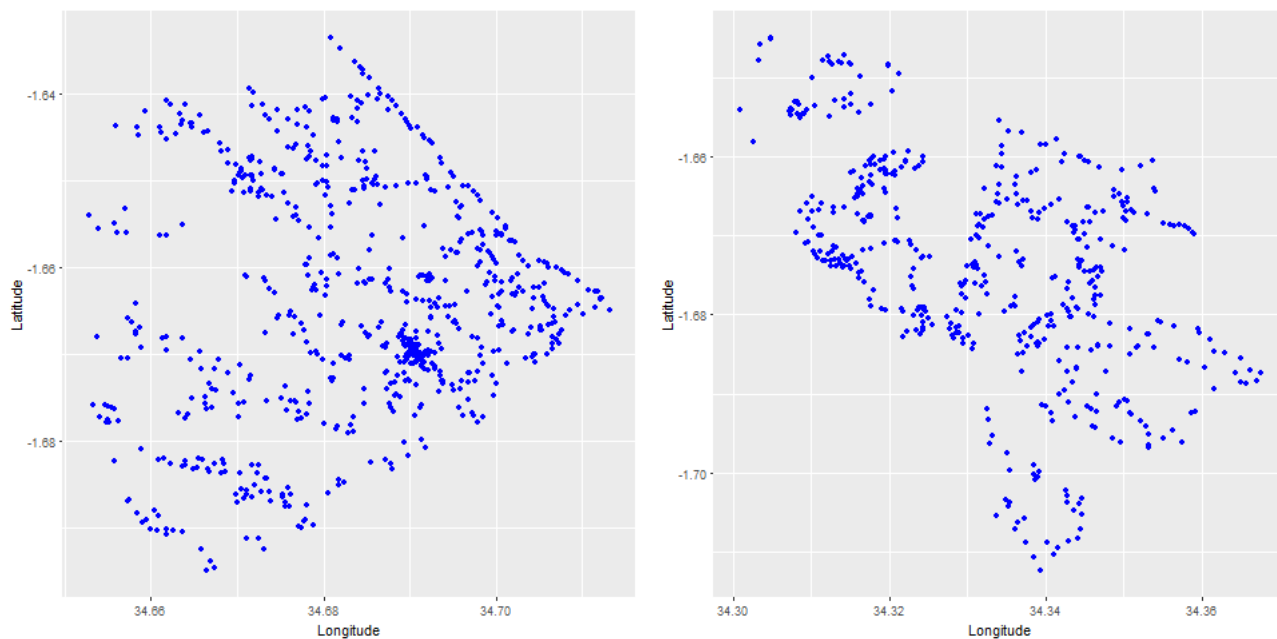
(a) Locations of households in Machochwe (left) and Mesaga (right).
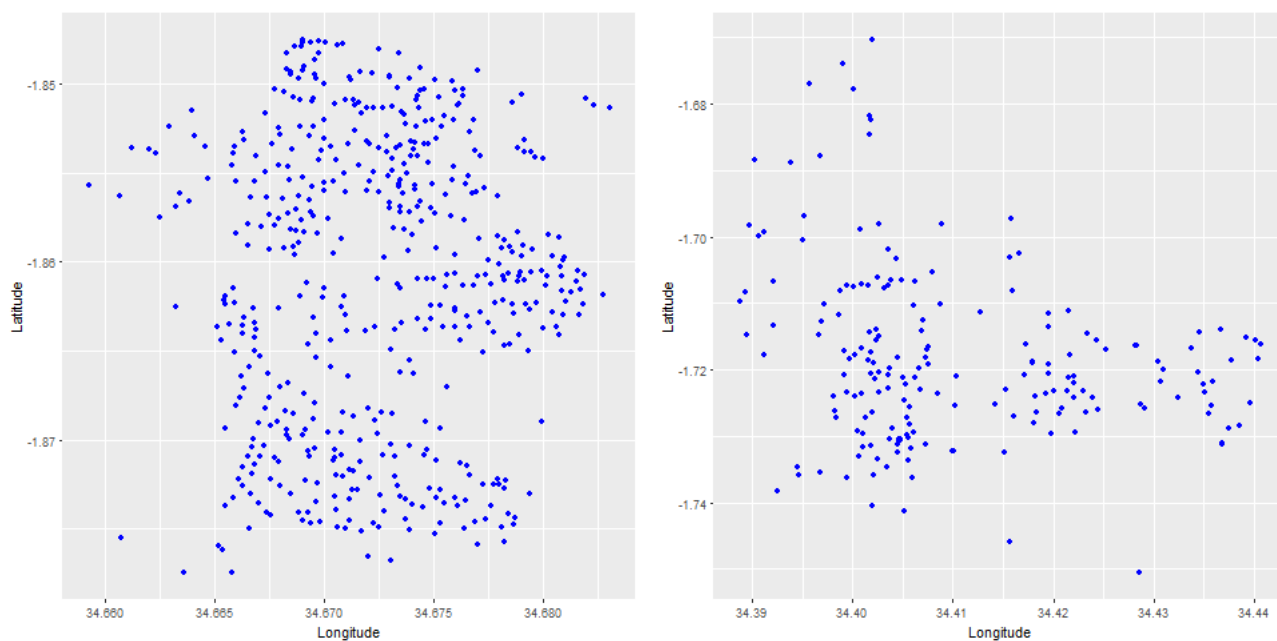
(b) Locations of households in Morotonga (left) and Ring'wani (right).

Figure 3.1: Locations of households for four villages in Mara province, Northern Tanzania.

We will also be considering the effect that the choice of window has on the various tests. Tests will be performed on both rectangular windows as well as convex hull windows, shown in Figures 3.2 and 3.3 respectively.
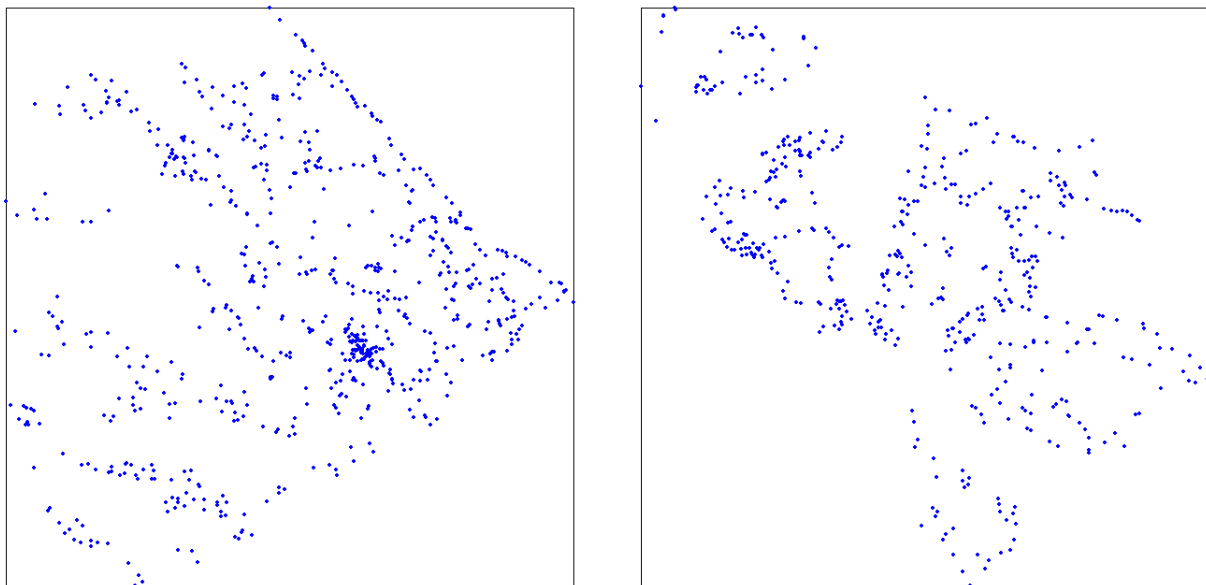
(a) Machochwe (left) and Mesaga (right).
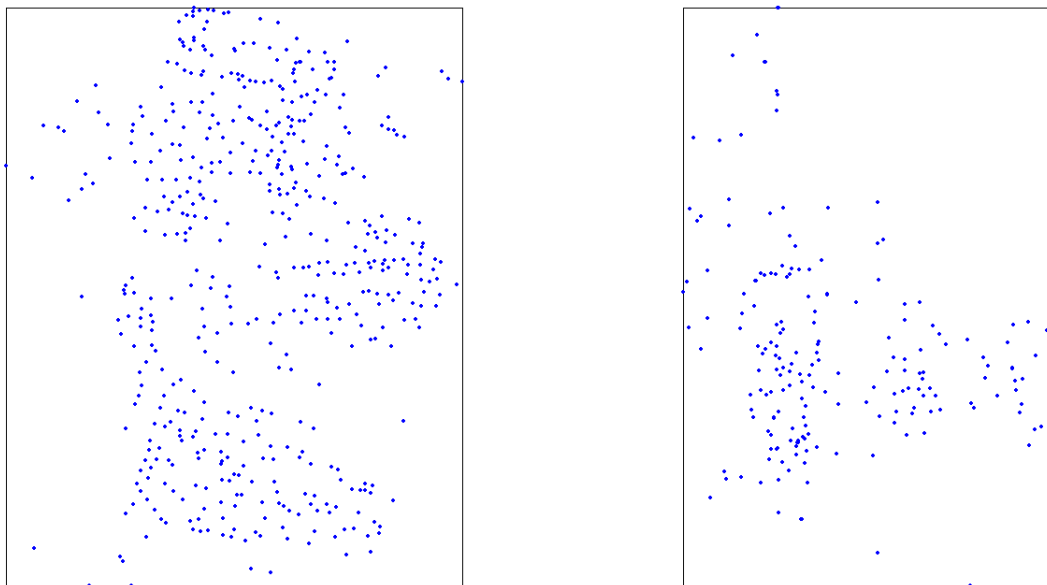
(b) Morotonga (left) and Ring'wani (right).

Figure 3.2: Rectangular windows for four villages in Mara province, Northern Tanzania.
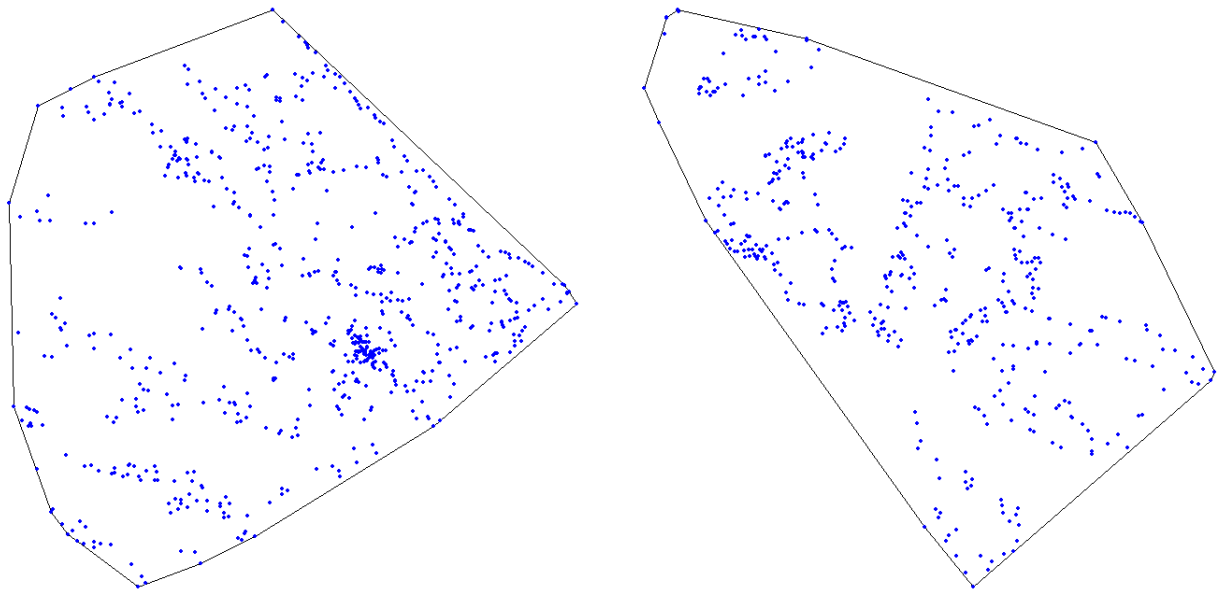
(a) Machochwe (left) and Mesaga (right).
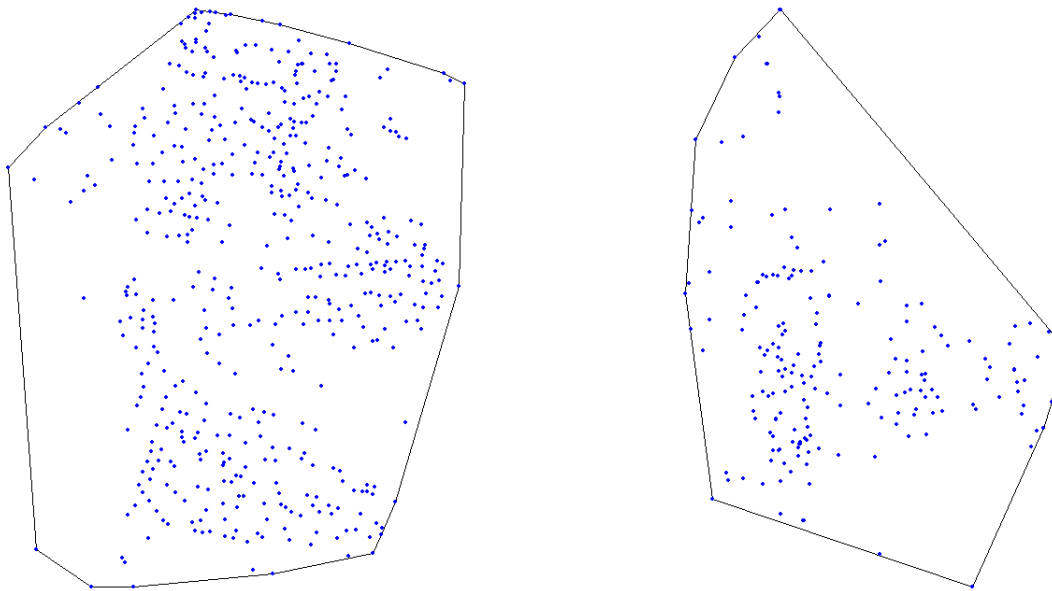
(b) Morotonga (left) and Ring'wani (right).

Figure 3.3: Convex hull windows for four villages in Mara province, Northern Tanzania.

3.2 Second-order homogeneity

To illustrate the effect of the homogeneous CSR tests on inhomogeneous patterns, first the homogeneous F, G, J, K and L functions are applied to the data.
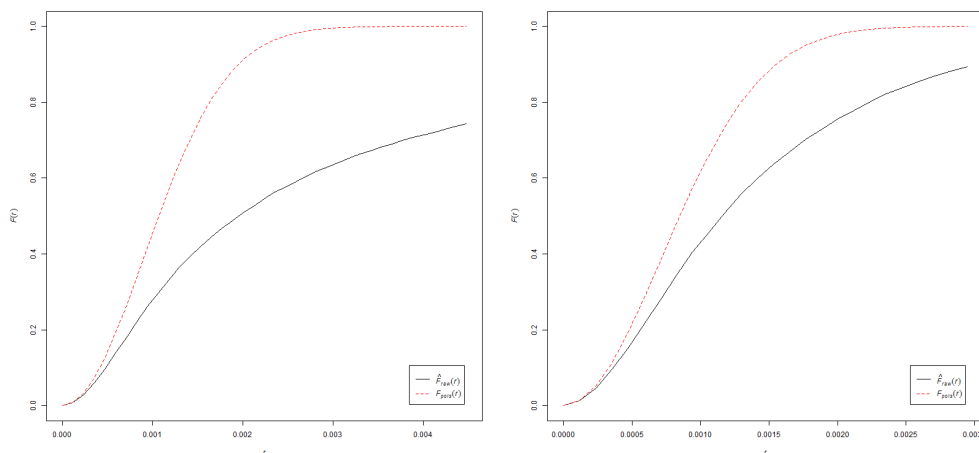
(a) Machochwe
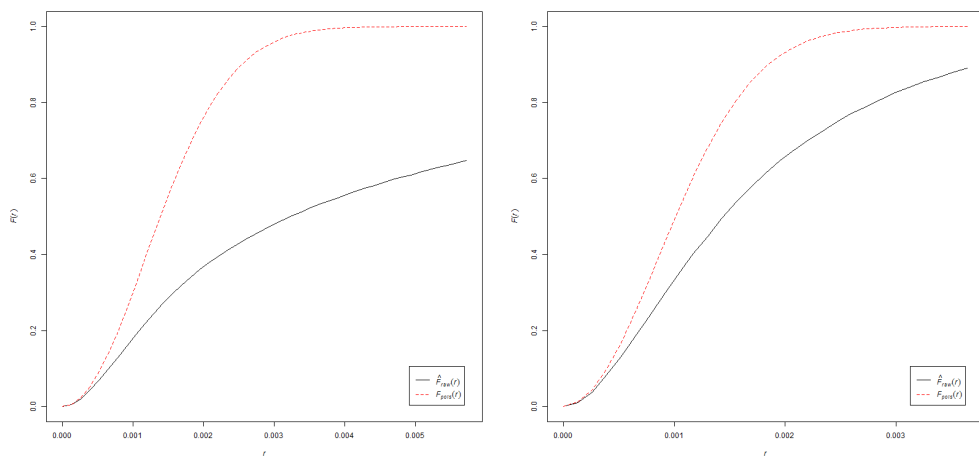
(b) Mesaga

Figure 3.4: Examples of household locations analysed by the F function using rectangular (left) and convex hull (right) windows.

Consider the locations of the households in each of the villages as a realisation of a point process. Using the `spatstat` package in R [10, 11], the F function was used to analyse each of the point patterns for randomness. The results for the four selected villages can be seen in Figures 3.4 and 3.5 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the function $F(r)$ is found on the y-axis. Since we are observing all points in a pattern (the locations of all households in each village are

known) and not merely points in a sampling window, we need not consider edge effects. The graphs show the values for a theoretical Poisson point process, $F_{pois}(r)$ as well as the uncorrected F function of the data, $\hat{F}_{raw}(r)$. For all villages and both window types, it is clear that $\hat{F}(r) < F_{pois}(r)$ for all values of r, indicating that points are clustered.
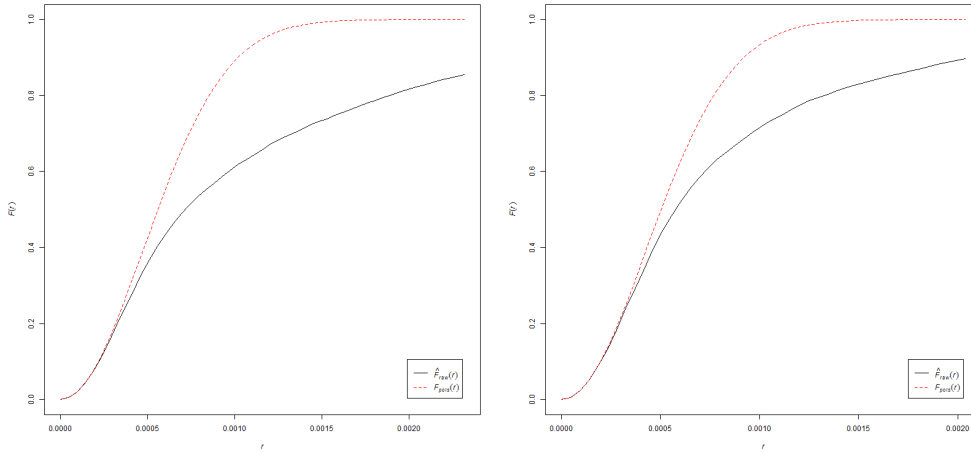
(a) Morotonga
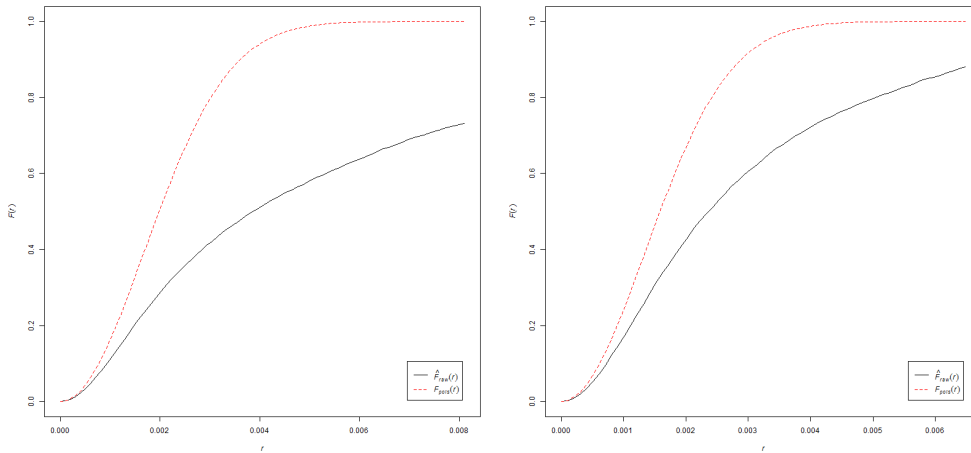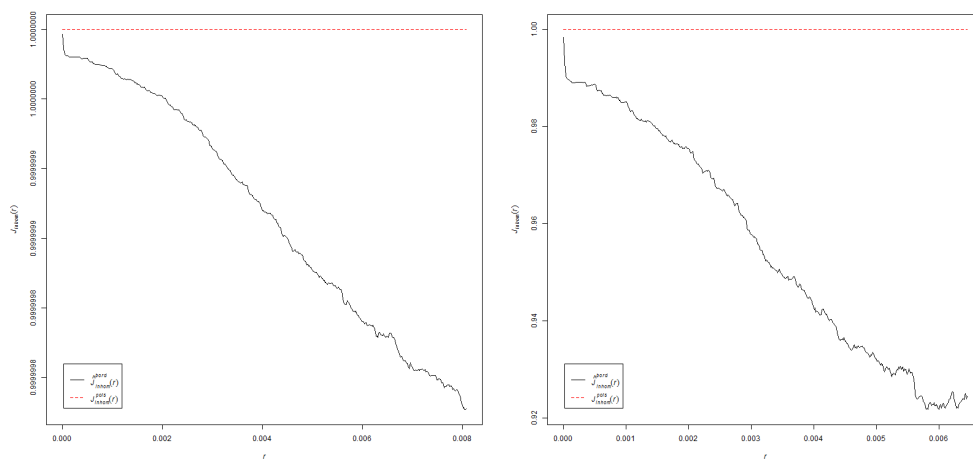

(b) Ring'wani

Figure 3.5: Examples of household locations analysed by the F function using rectangular (left) and convex hull (right) windows.

The household locations were also analysed using the G function. The results for four selected villages can be seen in Figures 3.6 and 3.7 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the function $G(r)$ is found on the y-axis. The graphs show the values for a theoretical Poisson point process, $G_{pois}(r)$ as well as the uncorrected G function of the data, $\hat{G}_{raw}(r)$. For all villages, it is clear that $\hat{G}(r) > G_{pois}(r)$ for small values of r. This indicates that points form a clus-

tered pattern at close distances but that points are random when further apart. As with the F function, the results are consistent for both window types, although the distance up to which clustering occurs differs. For Machochwe we observe clustering up to a distance of approximately 300m ($r \approx 0.00275$) using a rectangular window or approximately 220m ($r \approx 0.002$) using a convex hull window; for Mesaga we observe clustering up to a distance of approximately 385m ($r \approx 0.0035$) and approximately 275m ($r \approx 0.0025$) using rectangular and convex hull windows respectively; and for Ring'wani we observe clustering up to a distance of approximately 605m ($r \approx 0.0055$) and approximately 330m ($r \approx 0.003$) using rectangular and convex hull windows respectively. For Morotonga, the values of $\hat{G}(r)$ are close to $G_{pois}(r)$, indicating that the household locations are distributed randomly.
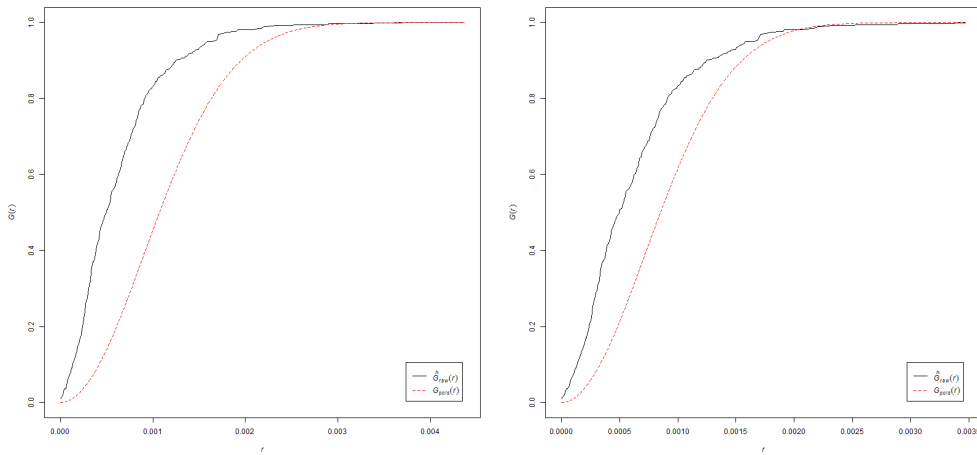
(a) Machochwe
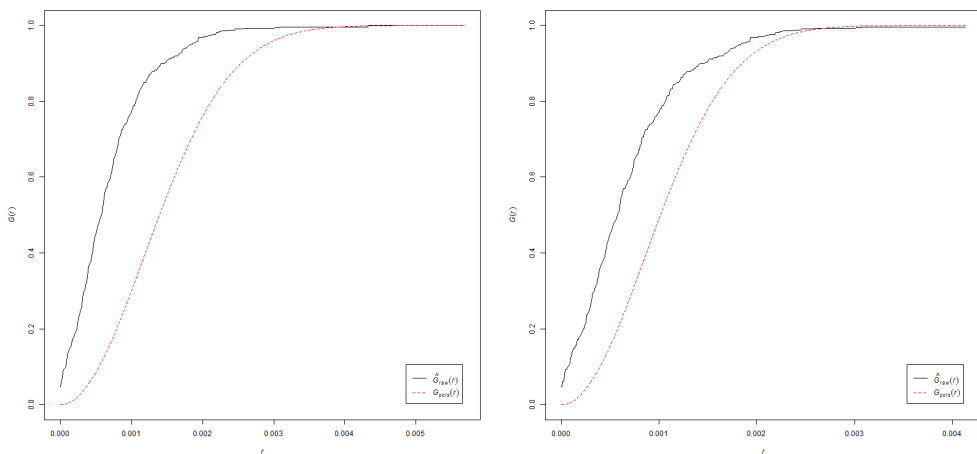
(b) Mesaga

Figure 3.6: Examples of household locations analysed by the G function using rectangular (left) and convex hull (right) windows.
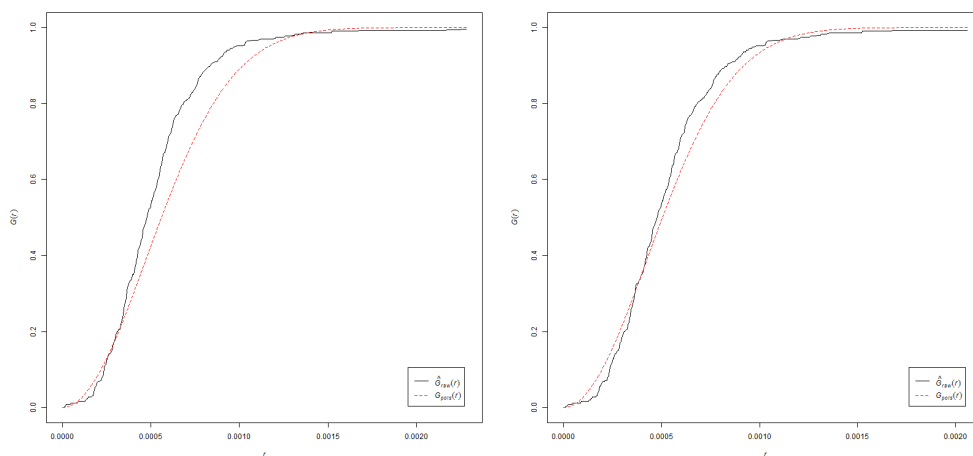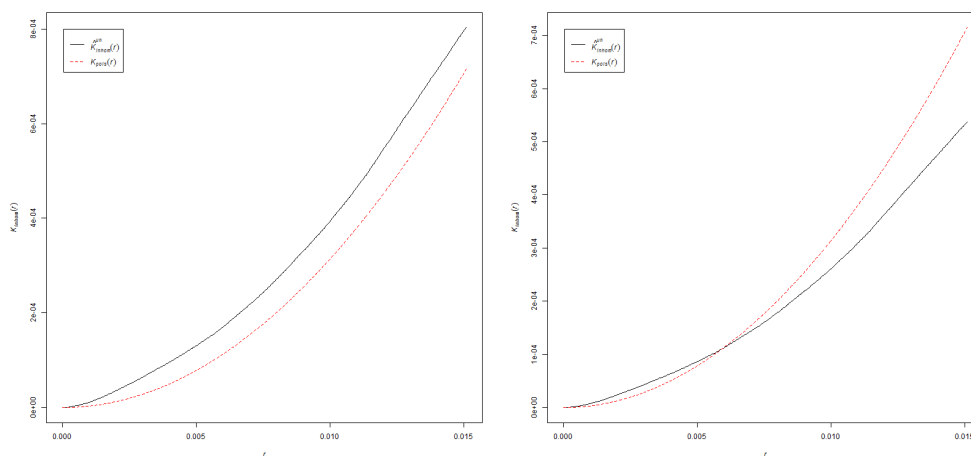
(a) Morotonga
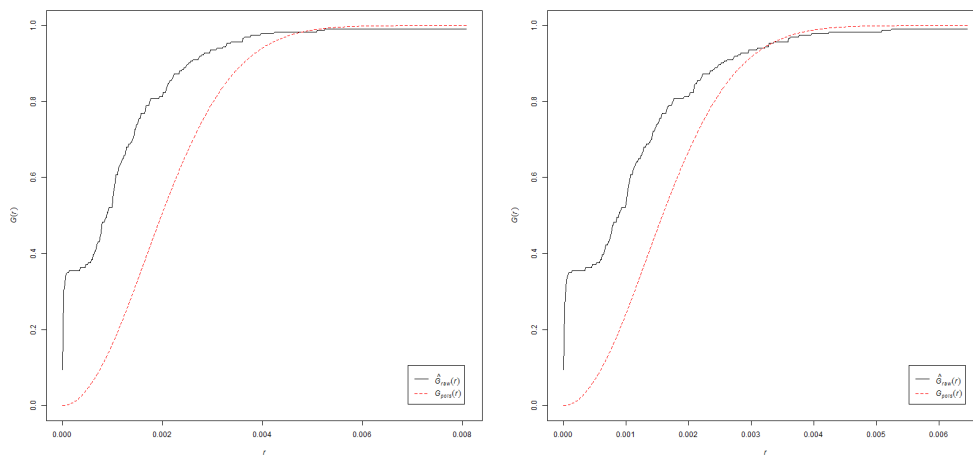
(b) Ring'wani
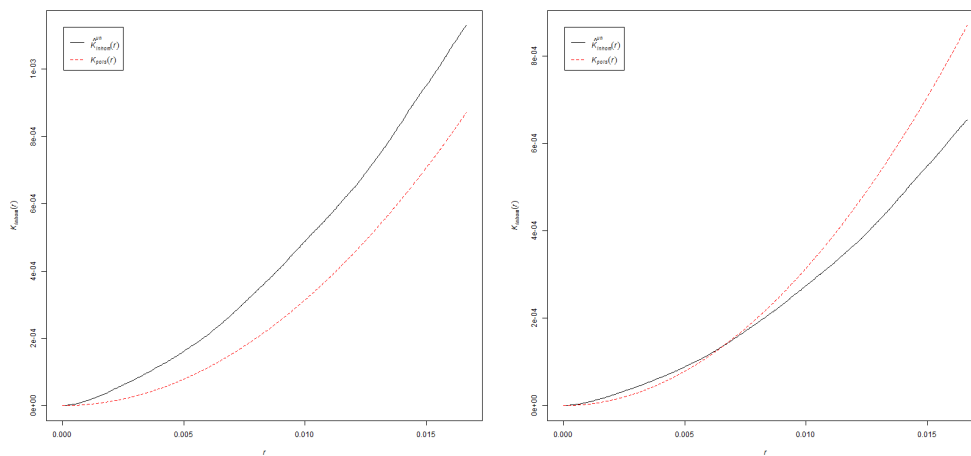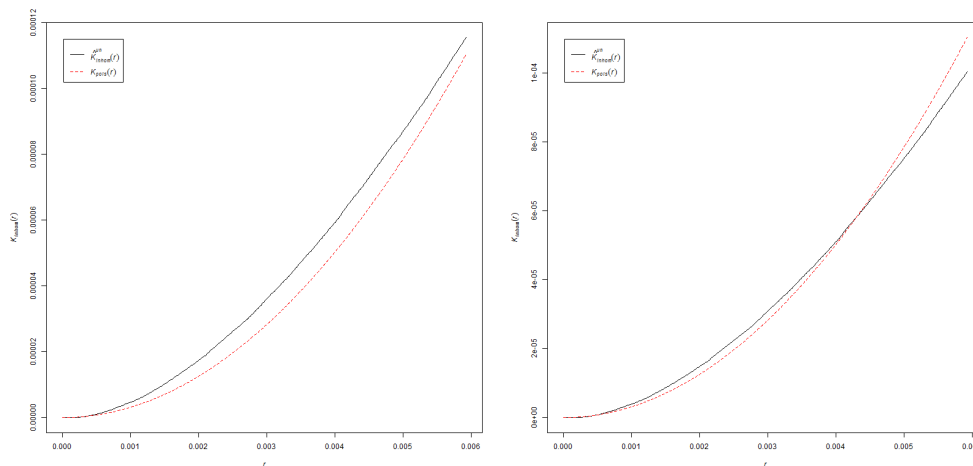
Figure 3.7: Examples of household locations analysed by the G function using rectangular (left) and convex hull (right) windows.
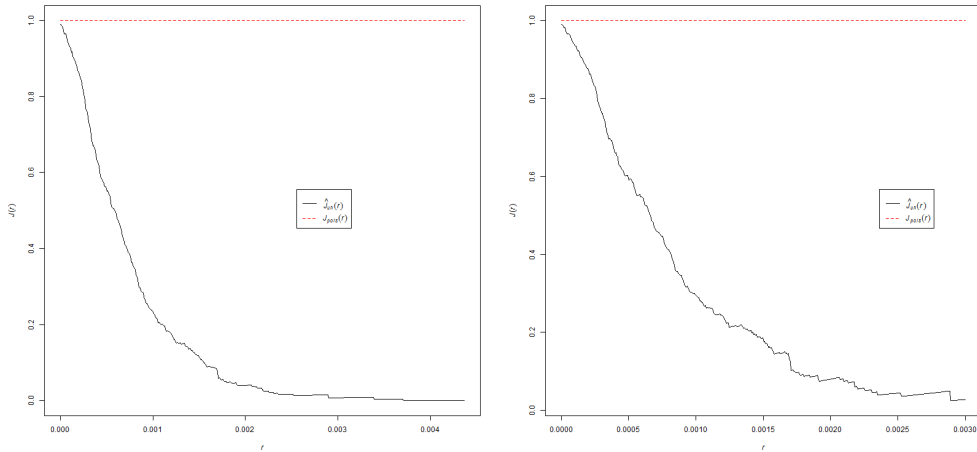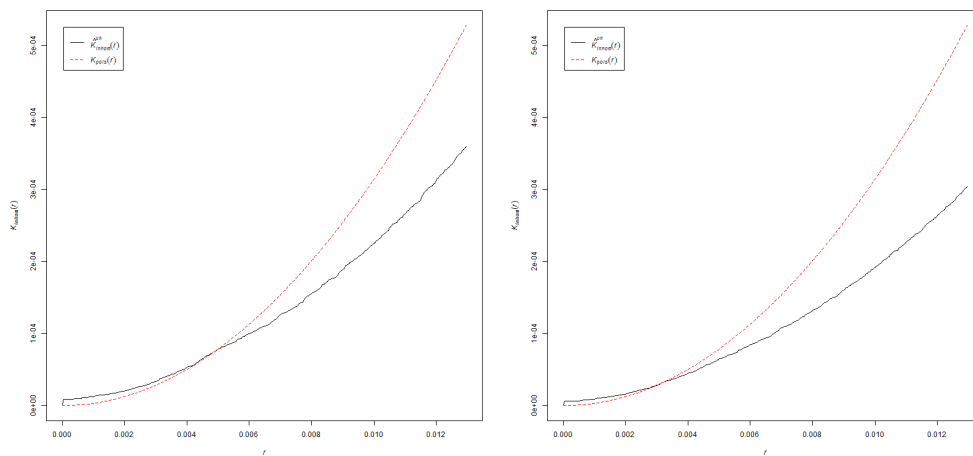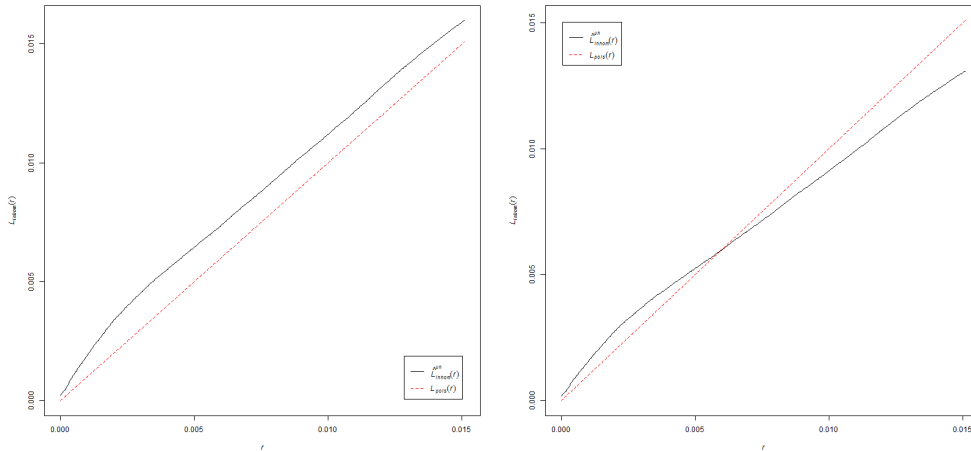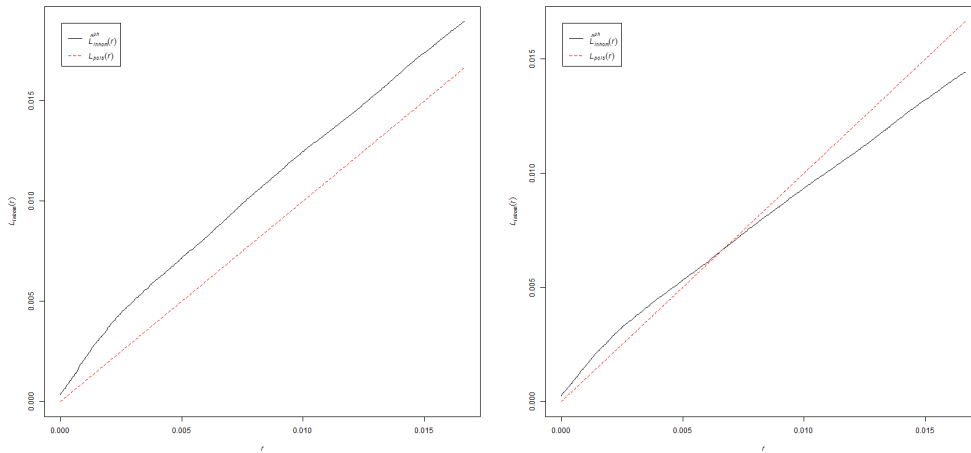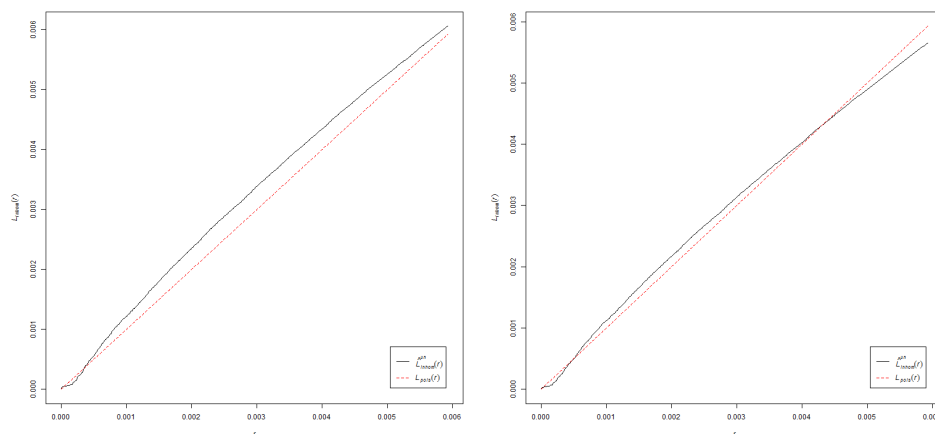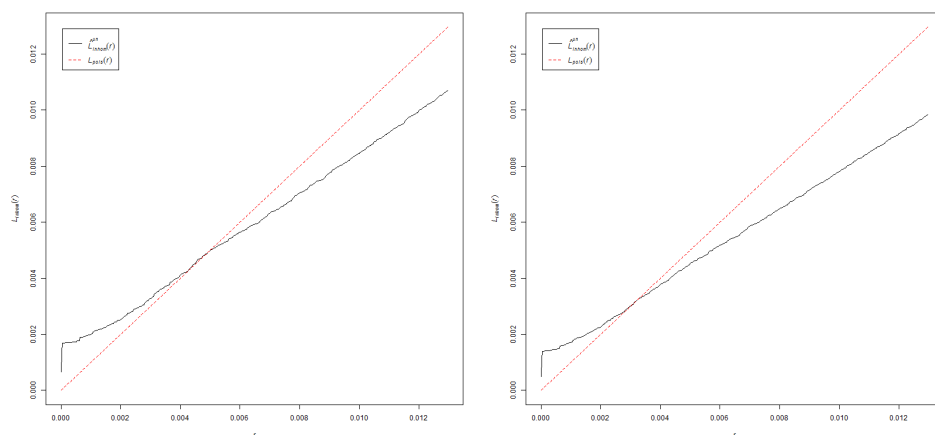
(a) Machochwe
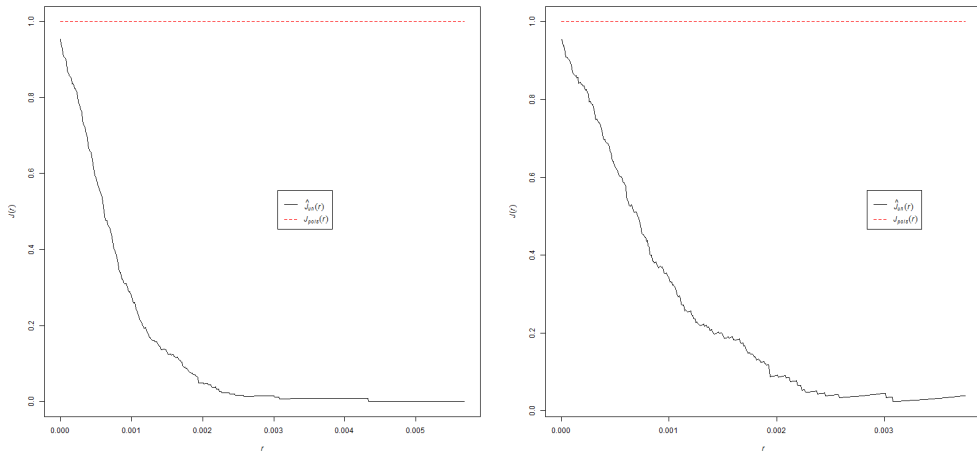
(b) Mesaga

Figure 3.8: Examples of household locations analysed by the J function using rectangular (left) and convex hull (right) windows.

Next we analysed the locations using the J function. The results for four selected villages can be seen in Figures 3.8 and 3.9 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the function $J(r)$ is found on the y-axis. The graphs show the values for a theoretical Poisson point process, $J_{pois}(r)$ as well as the uncorrected J function of the data, $\hat{J}_{raw}(r)$. For all villages except Morotonga, it is clear that $\hat{J}(r) < J_{pois}(r) \equiv 1$ for all values of r. This again indicates that points form a clustered pattern. The results are again consistent for both window types. For Morotonga we observe that $\hat{J}(r) > J_{pois}(r)$ up to approximately 32m ($r \approx 0.000292$) using a rectangular window and approximately 38m ($r \approx 0.000346$) using a convex hull window, indicating a regular pattern. Beyond this distance, $\hat{J}(r) < J_{pois}(r)$, indicating
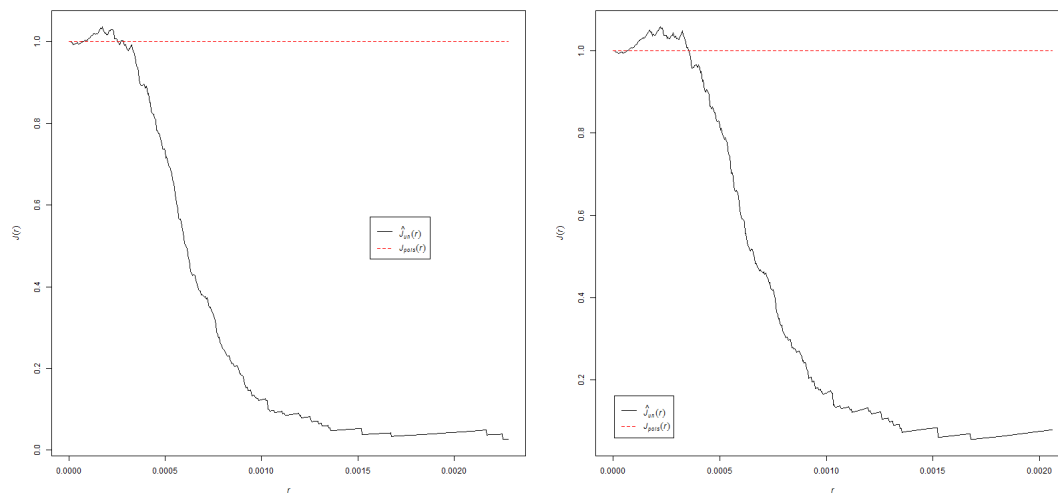
(a) Morotonga
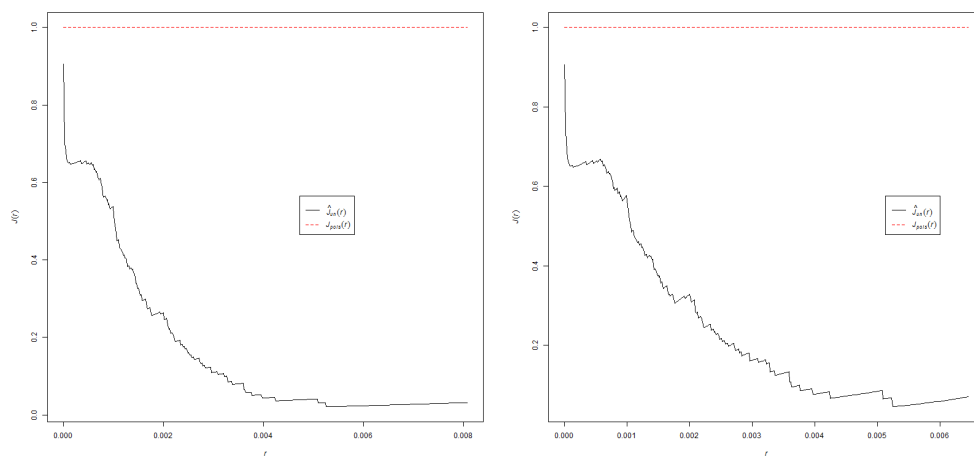
(b) Ring'wani

Figure 3.9: Examples of household locations analysed by the J function using rectangular (left) and convex hull (right) windows.

that households are clustered.
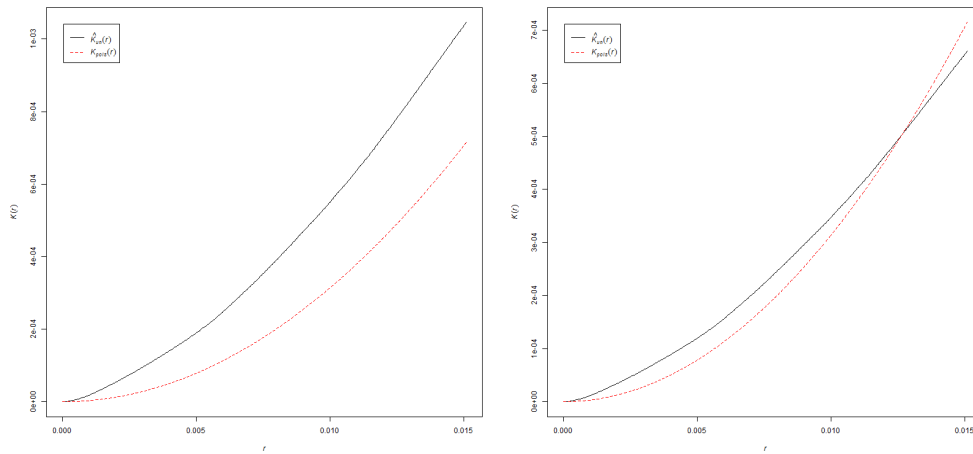
(a) Machochwe
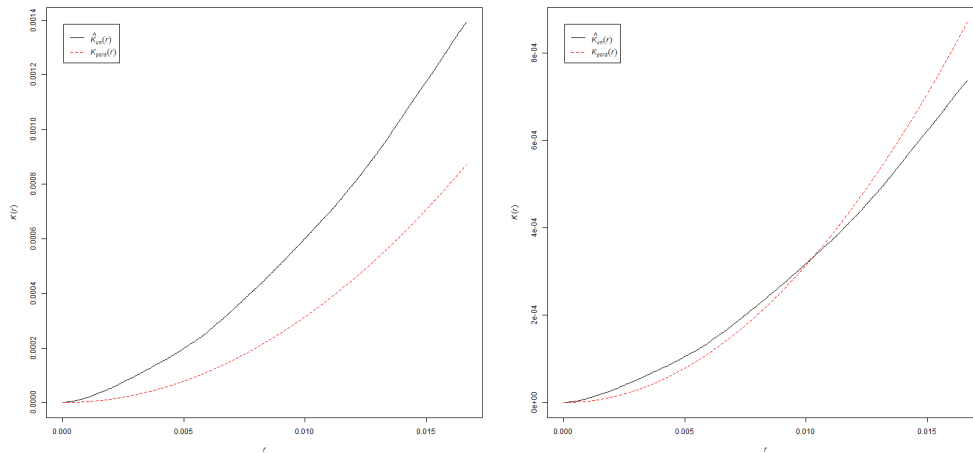
(b) Mesaga

Figure 3.10: Examples of household locations analysed by the K function using rectangular (left) and convex hull (right) windows.

The locations were then analysed using the K function. The results for four selected villages can be seen in Figures 3.10 and 3.11 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the function $K(r)$ is found on the y-axis. The graphs show the values for a theoretical Poisson point process, $K_{pois}(r)$ as well as the uncorrected K function of the data, $\hat{K}_{raw}(r)$. For all villages, it is clear that $\hat{K}(r) > K_{pois}(r)$ for all values of r when using a rectangular window. This indicates that points form a clustered pattern. Using a convex hull window the same result is obtained for Ring'wani, however for Machochwe, Mesaga and Morotonga, the K function indicates a random pattern.
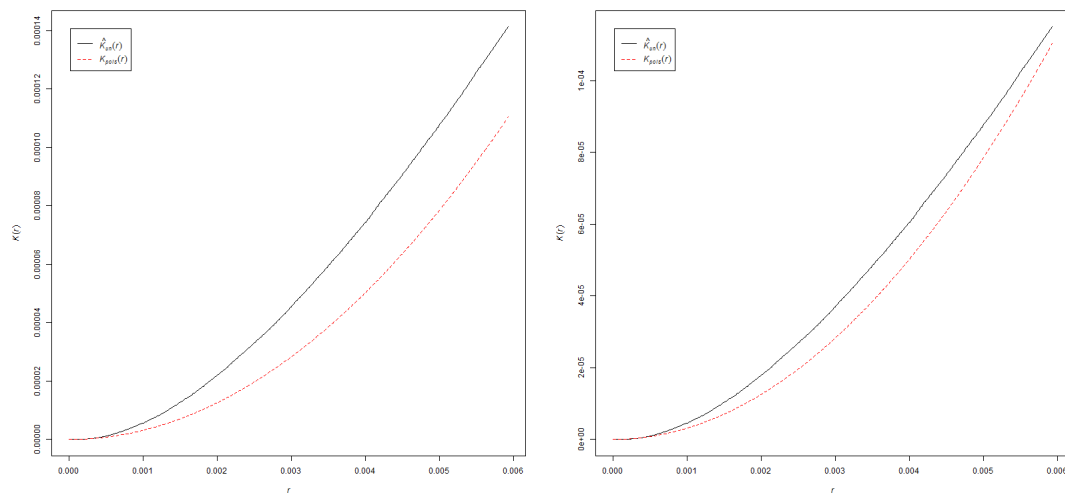
(a) Morotonga
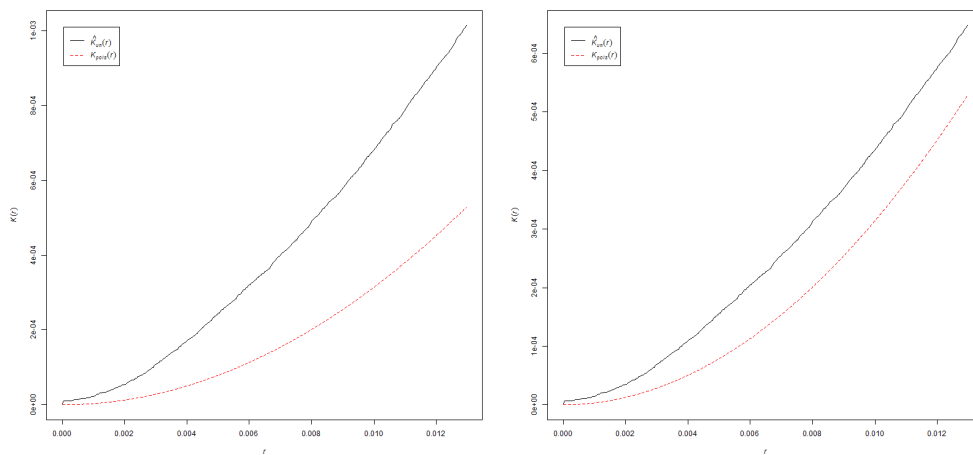
(b) Ring'wani

Figure 3.11: Examples of household locations analysed by the K function using rectangular (left) and convex hull (right) windows.
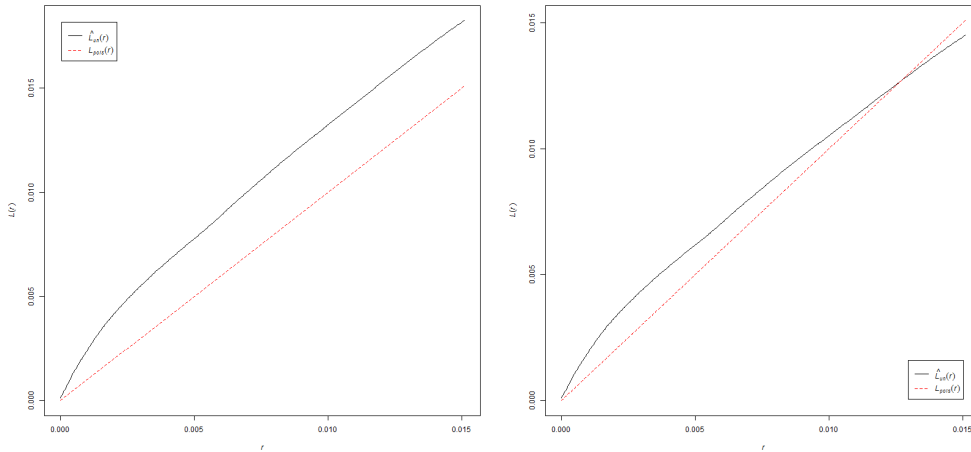
(a) Machochwe
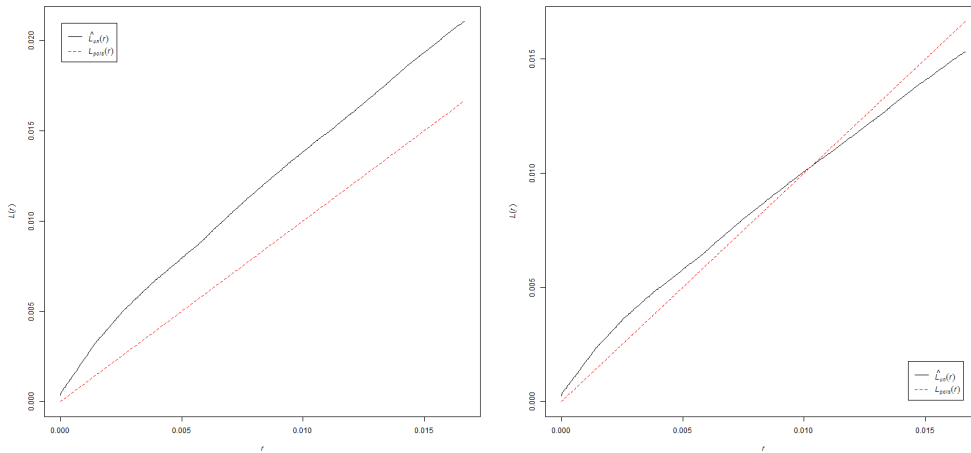
(b) Mesaga

Figure 3.12: Examples of household locations analysed by the L function using rectangular (left) and convex hull (right) windows.

Finally, the locations were analysed using the L function. The results for four selected villages can be seen in Figures 3.12 and 3.13 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the function $L(r)$ is found on the y-axis. The graphs show the values for a theoretical Poisson point process, $L_{pois}(r)$ as well as the uncorrected L function of the data, $\hat{L}_{raw}(r)$. For all villages, it is clear that $\hat{L}(r) > L_{pois}(r)$ for all values of r when using a rectangular window. This indicates that points form a clustered pattern. As with the K function, when using a convex hull window the L function indicates a clustered pattern for Ring'wani but a random pattern for Machochwe, Mesaga and Morotonga.

The F, G, J, K and L functions used thus far assume homogeneity, namely that the
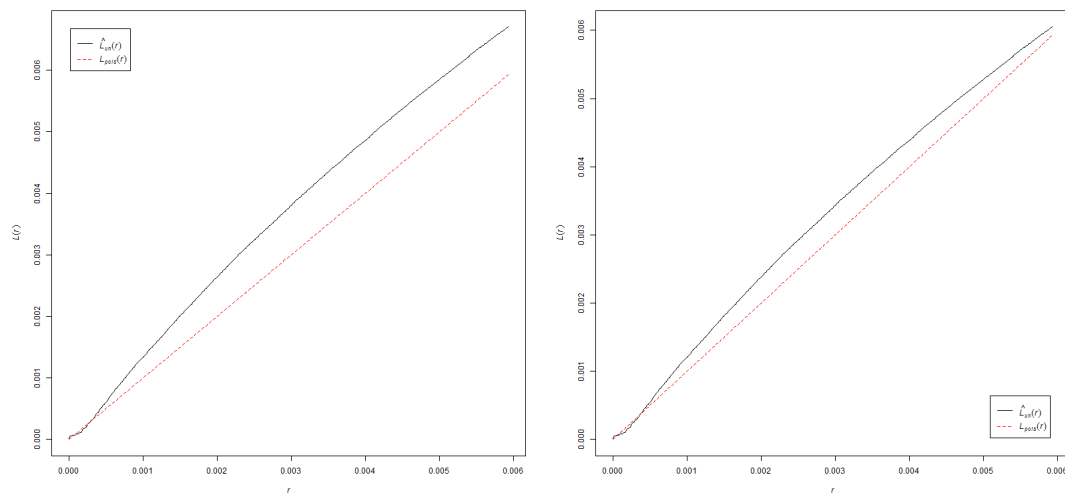
(a) Morotonga
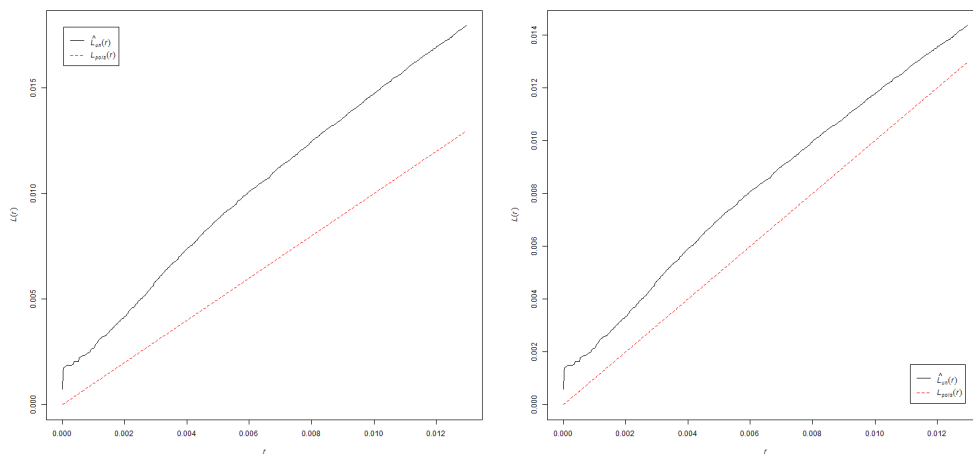
(b) Ring'wani

Figure 3.13: Examples of household locations analysed by the L function using rectangular (left) and convex hull (right) windows.
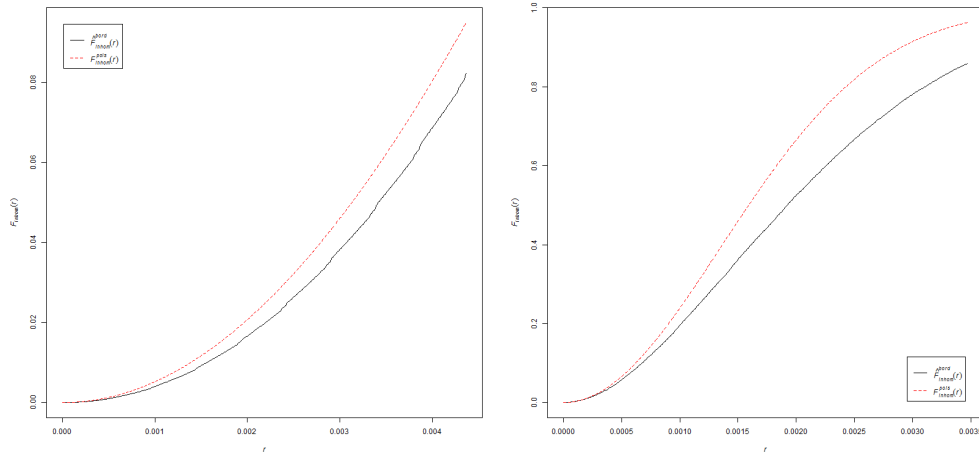
(a) Machochwe
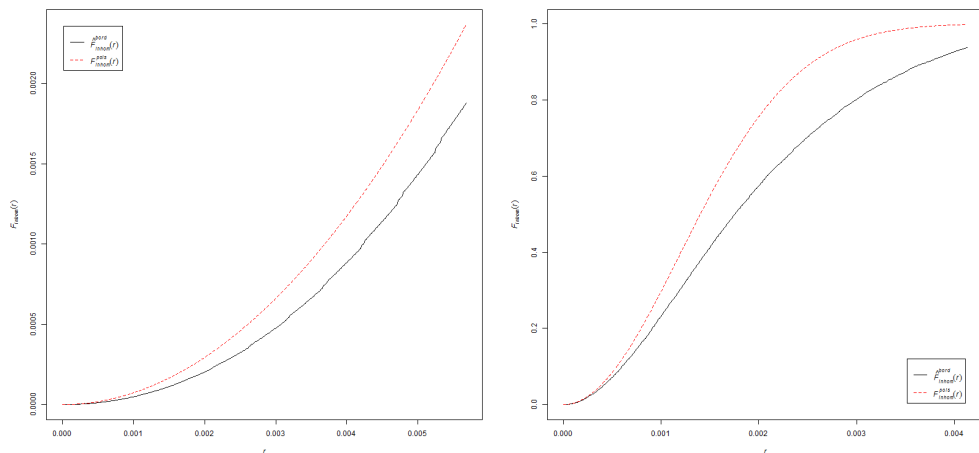
(b) Mesaga

Figure 3.14: Examples of household locations analysed by the inhomogeneous F function using rectangular (left) and convex hull (right) windows.

intensity of points remains constant across the pattern. We will however now consider the results when household locations are analysed using the inhomogeneous F, G, J, K and L functions.

The results for the inhomogeneous F function for four selected villages can be seen in Figures 3.14 and 3.15 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the inhomogeneous function $F(r)$ is found on the y-axis. As with the homogeneous functions, we need not consider edge effects since we are observing all points in the pattern. However, spatstat does not allow for an uncorrected inhomogeneous F function[3] to be calculated and automatically calculates the

[3]http://www.inside-r.org/packages/cran/spatstat/docs/Finhom
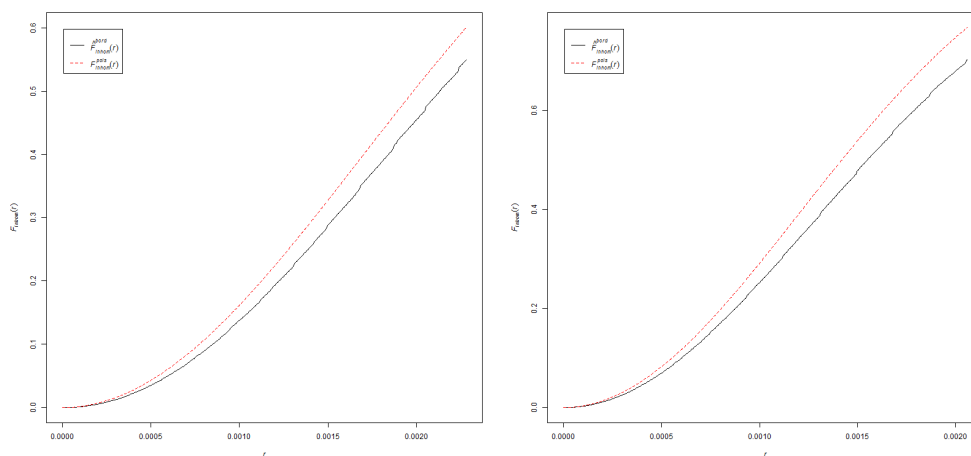
(a) Morotonga
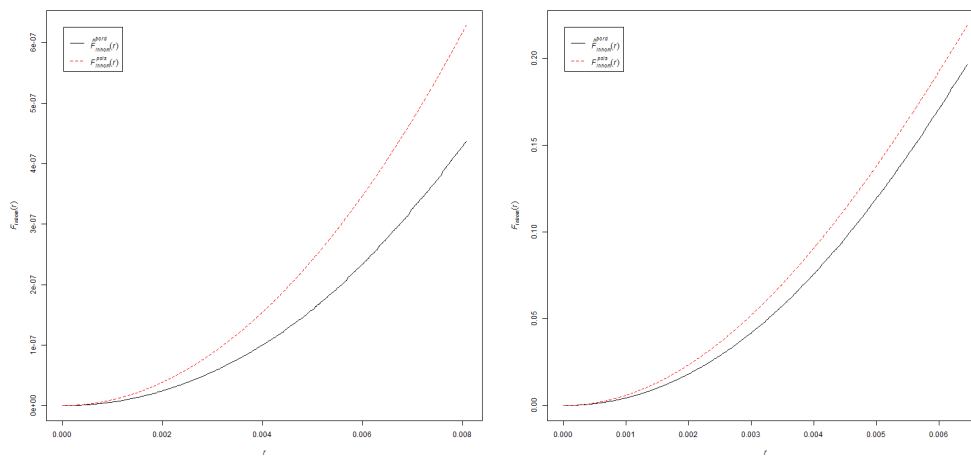
(b) Ring'wani

Figure 3.15: Examples of household locations analysed by the inhomogeneous F function using rectangular (left) and convex hull (right) windows.
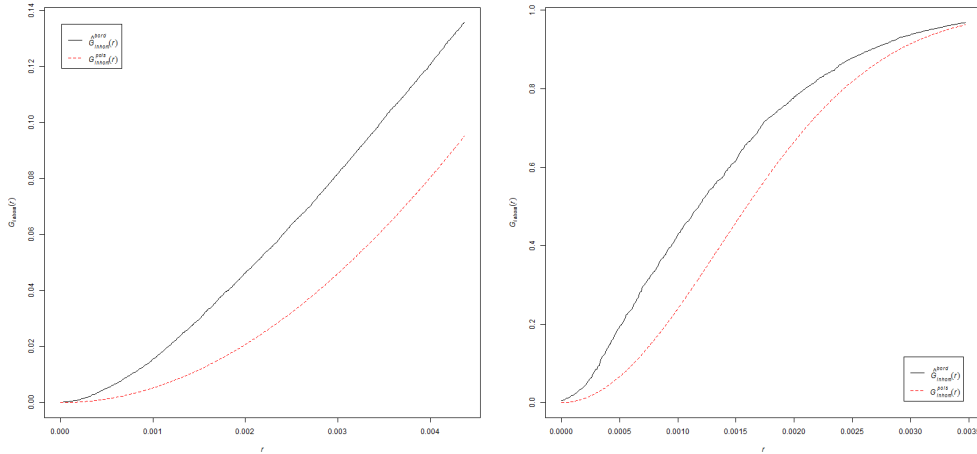
(a) Machochwe
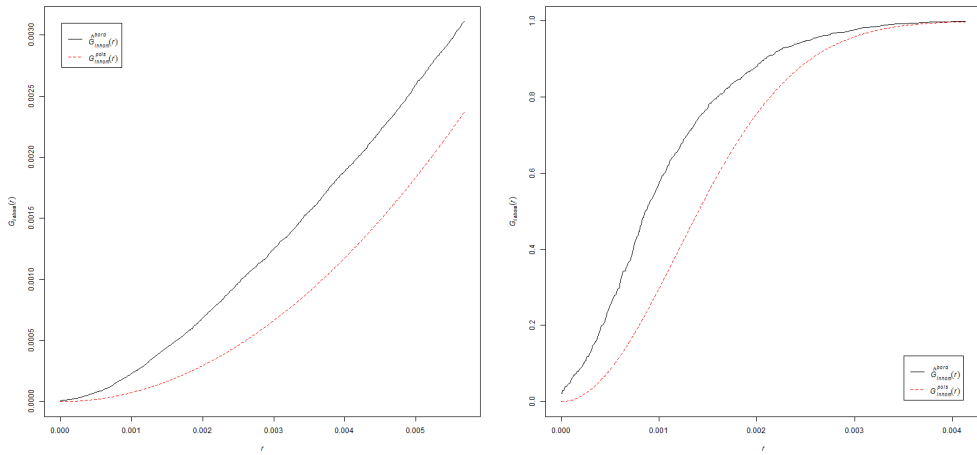
(b) Mesaga

Figure 3.16: Examples of household locations analysed by the inhomogeneous G function using rectangular (left) and convex hull (right) windows.

function with the border method correction. Since we have observed the entire pattern, this correction will not affect our conclusion. The graphs show the values for a theoretical inhomogeneous Poisson point process, $F_{pois}(r)$ as well as the corrected F function of the data, $\hat{F}_{bord}(r)$. For Machochwe and Mesaga using both window types, as well as Ring'wani using a rectangular window, it is clear that $\hat{F}_{bord}(r) < F_{pois}(r)$ for all values of r. This indicates that points form a clustered pattern. For Morotonga using both window types, as well as Ring'wani using a convex hull window $\hat{F}_{bord}(r)$ is only slightly less than $F_{pois}(r)$ for all values of r, indicating a random pattern.

The results for the inhomogeneous G function for four selected villages can be seen in Figures 3.16 and 3.17 for both rectangular and convex hull windows. The parameter r is
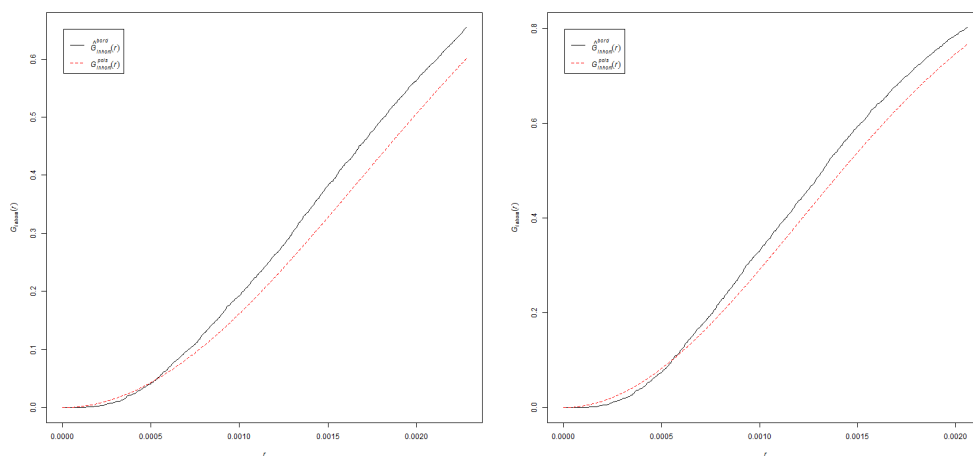
(a) Morotonga
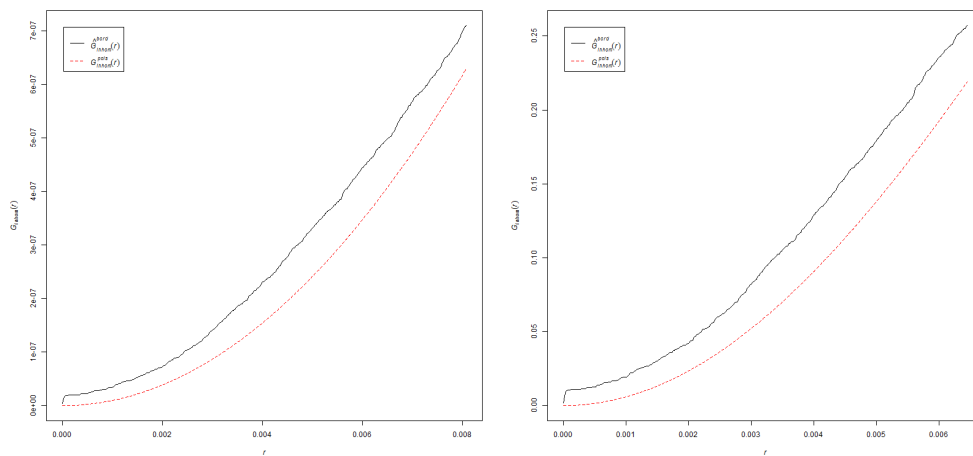
(b) Ring'wani

Figure 3.17: Examples of household locations analysed by the inhomogeneous G function using rectangular (left) and convex hull (right) windows.

represented on the x-axis and the value of the inhomogeneous function $G(r)$ is found on the y-axis. As with the inhomogeneous F function, `spatstat` does not allow for an uncorrected inhomogeneous G function[4] but the correction made will not affect our conclusion. The graphs show the values for a theoretical inhomogeneous Poisson point process, $G_{pois}(r)$ as well as the corrected G function of the data, $\hat{G}_{bord}(r)$. Using a rectangular window, for Machochwe, Mesaga and Ring'wani it is clear that $\hat{G}_{bord}(r) > G_{pois}(r)$ for all values of r. This indicates that points form a clustered pattern. For Morotonga, $\hat{G}_{bord}(r) \approx G_{pois}(r)$ indicating a random pattern up to approximately 65m ($r \approx 0.000591$), with $\hat{G}_{bord}(r) > G_{pois}(r)$ for $r > 0.000591$, indicating weak clustering beyond this distance. Using a convex hull window, for Machochwe and Mesaga, $\hat{G}_{bord}(r) > G_{pois}(r)$ for $r < 0.0035$, indicating a clustered pattern up to approximately 385m. Beyond this distance, $\hat{G}_{bord}(r) \approx G_{pois}(r)$, indicating a random pattern. For Ring'wani, $\hat{G}_{bord}(r) > G_{pois}(r)$ for all values of r, indicating a clustered pattern. For Morotonga, the result obtained using a convex hull window is consistent with the rectangular window result, indicating a random pattern.

The results for the inhomogeneous J function for four selected villages can be seen in Figures 3.18 and 3.19 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the inhomogeneous function $J(r)$ is found on the y-axis. As with the inhomogeneous F and G functions, `spatstat` does not allow for an uncorrected inhomogeneous J function[5] to be calculated and automatically calculates the function with the border method correction. Since we have observed the entire pattern, this correction will not affect our conclusion. The graphs show the values for a theoretical inhomogeneous Poisson point process, $J_{pois}(r) \equiv 1$ as well as the corrected J function of the data, $\hat{J}_{bord}(r)$. For all villages except Morotonga, it is clear that $\hat{J}_{bord}(r) < J_{pois}(r)$ for all values of r. This indicates that points form a clustered pattern. For Morotonga, $\hat{J}_{bord}(r) > J_{pois}(r)$ for r smaller than approximately 0.0005 and $\hat{J}_{bord}(r) < J_{pois}(r)$ for r>0.0005, indicating a weak regular pattern up to approximately 55m and clustering beyond this distance. The results are again consistent for both window types.

The results for the inhomogeneous K function for four selected villages can be seen in Figures 3.20 and 3.21 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the inhomogeneous function $K(r)$ is found on the y-axis. Unlike the inhomogeneous F, G and J functions, `spatstat` does allow for an uncorrected inhomogeneous J function to be calculated. The graphs show the values for a theoretical inhomogeneous Poisson point process, $K_{pois}(r)$ as well as the uncorrected K function of the data, $\hat{K}_{un}(r)$. Using a rectangular window, for Machochwe, Mesaga and

---

[4]http://www.inside-r.org/packages/cran/spatstat/docs/Ginhom
[5]http://www.inside-r.org/packages/cran/spatstat/docs/Jinhom
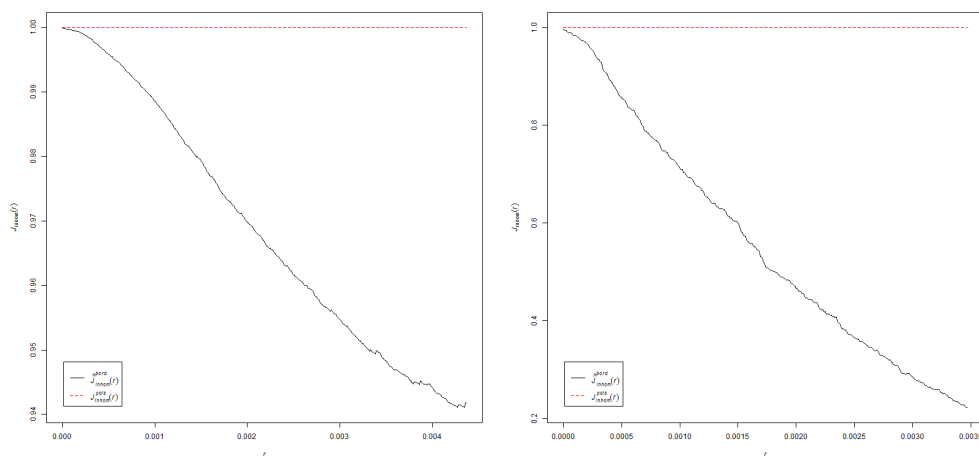
(a) Machochwe
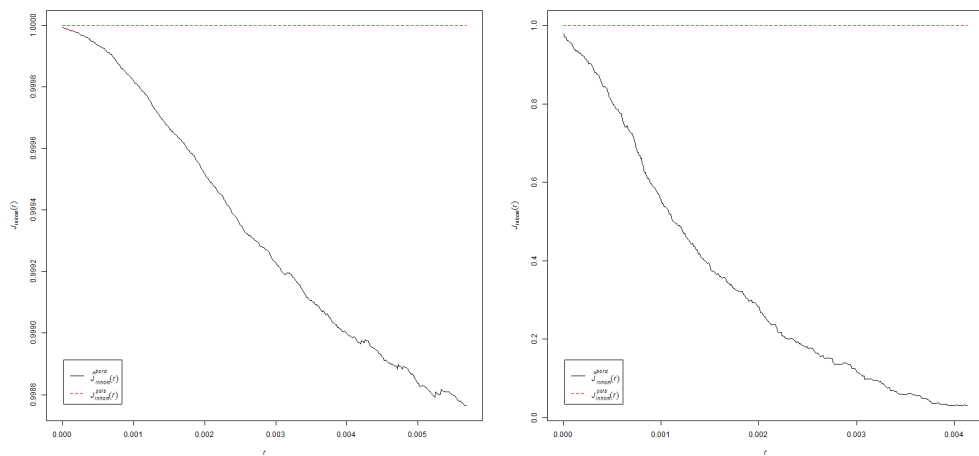
(b) Mesaga

Figure 3.18: Examples of household locations analysed by the inhomogeneous J function using rectangular (left) and convex hull (right) windows.
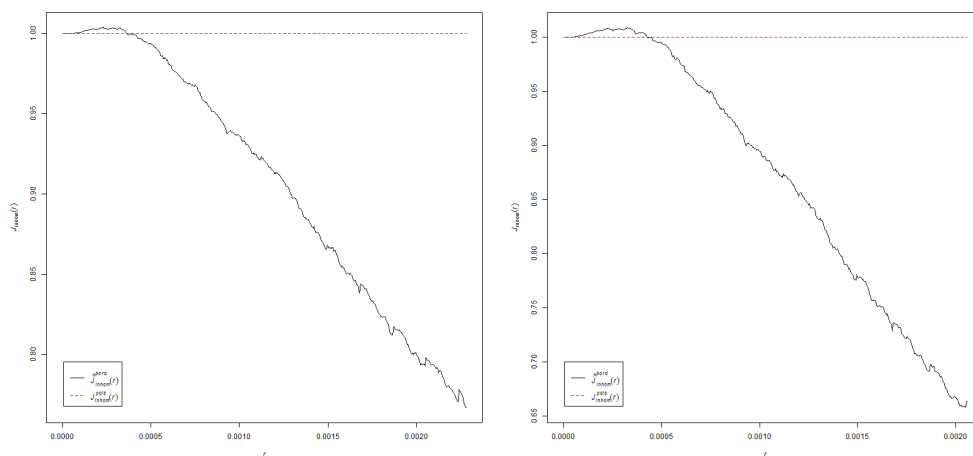
(a) Morotonga

(b) Ring'wani

Figure 3.19: Examples of household locations analysed by the inhomogeneous J function using rectangular (left) and convex hull (right) windows.

(a) Machochwe

(b) Mesaga

Figure 3.20: Examples of household locations analysed by the inhomogeneous K function using rectangular (left) and convex hull (right) windows.

(a) Morotonga

(b) Ring'wani

Figure 3.21: Examples of household locations analysed by the inhomogeneous K function using rectangular (left) and convex hull (right) windows.

(a) Machochwe

(b) Mesaga

Figure 3.22: Examples of household locations analysed by the inhomogeneous L function using rectangular (left) and convex hull (right) windows.

Morotonga, it is clear that $\hat{K}(r) > K_{pois}(r)$ for all values of r. This indicates that points form a clustered pattern, although for Morotonga the clustering appears to be weak. For Ring'wani, $\hat{K}(r) \approx K_{pois}(r)$ up to $r \approx 0.005$, thereby indicating a random pattern up to a distance of approximately 550m, and $\hat{K}(r) < K_{pois}(r)$ for $r > 0.005$, thereby indicating a regular pattern beyond this distance. A similar result is obtained for Ring'wani when using a convex hull window, but here the regular pattern becomes evident at a distance of approximately 360m when $r \approx 0.00325$. For Machochwe, Mesaga and Morotonga, using a convex hull window, we again observe a random pattern for small values of r and a regular pattern beyond this. For Machochwe, Mesaga and Morotonga, this occurs at approximately 750m ($r \approx 0.0068$), 800m ($r \approx 0.0073$) and 495m ($r \approx 0.0045$) respectively.

(a) Morotonga

(b) Ring'wani

Figure 3.23: Examples of household locations analysed by the inhomogeneous L function using rectangular (left) and convex hull (right) windows.

The results for the inhomogeneous L function for four selected villages can be seen in Figures 3.22 and 3.23 for both rectangular and convex hull windows. The parameter r is represented on the x-axis and the value of the inhomogeneous function $L(r)$ is found on the y-axis. The graphs show the values for a theoretical inhomogeneous Poisson point process, $L_{pois}(r)$ as well as the uncorrected L function of the data, $\hat{L}_{un}(r)$. Using a rectangular window, for Machochwe and Mesaga we observe that $\hat{L}(r) > L_{pois}(r)$ for all values of r, indicating that points form a clustered pattern. For Morotonga the values of $\hat{L}(r) > L_{pois}(r)$ for all values of r, thereby indicating weak clustering since the differences between $\hat{L}(r)$ and $L_{pois}(r)$ values are small. For Ring'wani, $\hat{L}(r) > L_{pois}(r)$ up to $r \approx 0.004$ (roughly translating to 440m), thereby indicating a clustered pattern up to this distance and $\hat{L}(r) < L_{pois}(r)$ for $r > 0.005$ (roughly translating to 550m), thereby indicating a regular point pattern. Using a convex hull window, we obtain similar results for Ring'wani with a clustered pattern up to approximately 275m ($r \approx 0.0025$) and regular pattern beyond approximately 365m ($r \approx 0.0033$). For Morotonga, $\hat{L}(r) \approx L_{pois}(r)$ for all values of r, indicating a random pattern. For Machochwe and Mesaga we observe weak clustering up to distances of approximately 630m ($r \approx 0.00576$) and 715m ($r \approx 0.0065$) respectively, and a regular pattern beyond these distances.

From the results given in this section it is clear that the choice of an appropriate window can have an effect on the outcome of these second-order homogeneity tests. In the case where a natural window does not exist, it is therefore important to consider the choice of this boundary carefully. In our case study we are specifically considering household within a village and therefore, the definition of a village could be considered for this purpose. Empty spaces that do not contain households naturally occur within villages, for instance, a village could contain a village green, sports fields, streams or rivers, etc. Some general guidelines regarding the optimal window choice are provided in [86, 90, 94]. It is also clear that the outcome of second-order homogeneity tests are affected by the assumption of first-order homogeneity. As an example, consider the outcome of the F function for Morotonga, using either a rectangular or convex hull window. In both cases, the homogeneous F function indicates a clustered pattern while the inhomogeneous F function indicates a random pattern. These discrepancies highlight the importance of correctly identifying whether a pattern is first-order homogeneous or heterogeneous.

3.3 First-order homogeneity

3.3.1 Density plots

Examination of Figures 3.1 and 3.3 suggest that the point process could be inhomogeneous.

The intensity of points was thus estimated for each village using kernel smoothing via the `density.ppp` function of `R`. The output for the four example villages can be seen in Figures 3.24, 3.26, 3.28 and 3.30. These estimates were obtained using a Gaussian kernel with bandwidth roughly equal to one-eighth of the shortest side length of the rectangular window. The estimates were obtained on both rectangular and convex hull windows, without edge corrections. Visual inspection suggests that all of the villages have inhomogeneous intensity and that the choice of window leads to subtle differences in the intensity estimates.

Although the entire pattern was observed, for completeness we also considered the effect of including edge corrections on the estimates. These estimates are shown in Figures 3.25, 3.27, 3.29 and 3.31. As in Section 2.3.2, it is clear that the choice of edge correction has a significant effect on the density estimates.

In the following section we consider a more rigorous approach for confirming inhomogeneity through the hypothesis tests defined in Section 3.3.2.
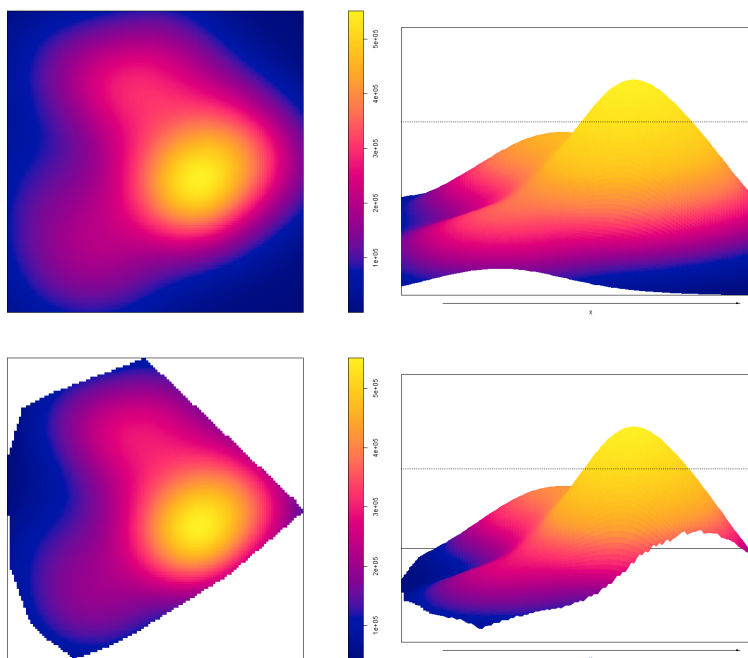
Figure 3.24: Uncorrected kernel density estimates of point intensity (left) and perspective view of density estimates (right) for Machochwe using rectangular (top) and convex hull (bottom) windows.
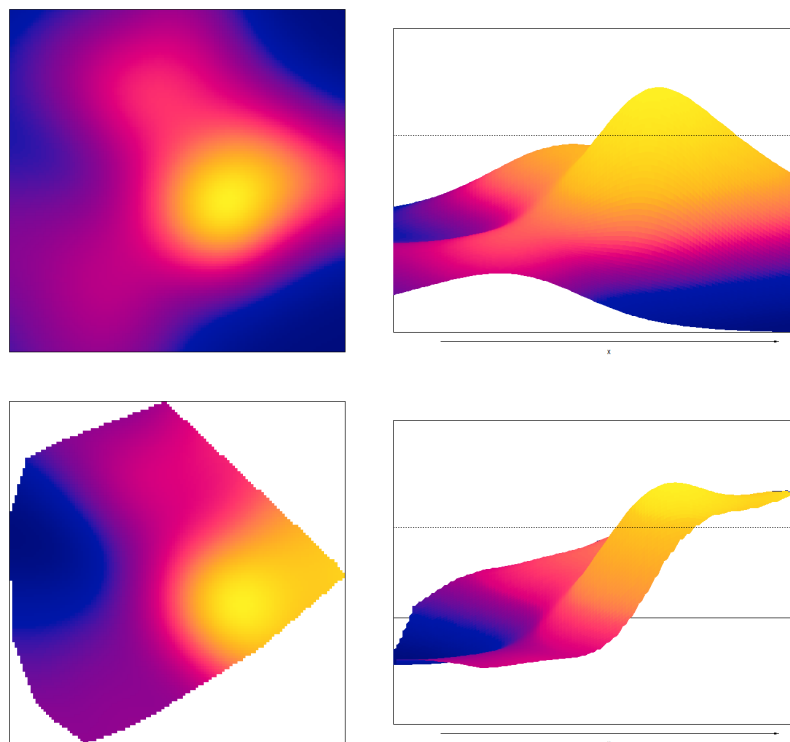
(a) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.
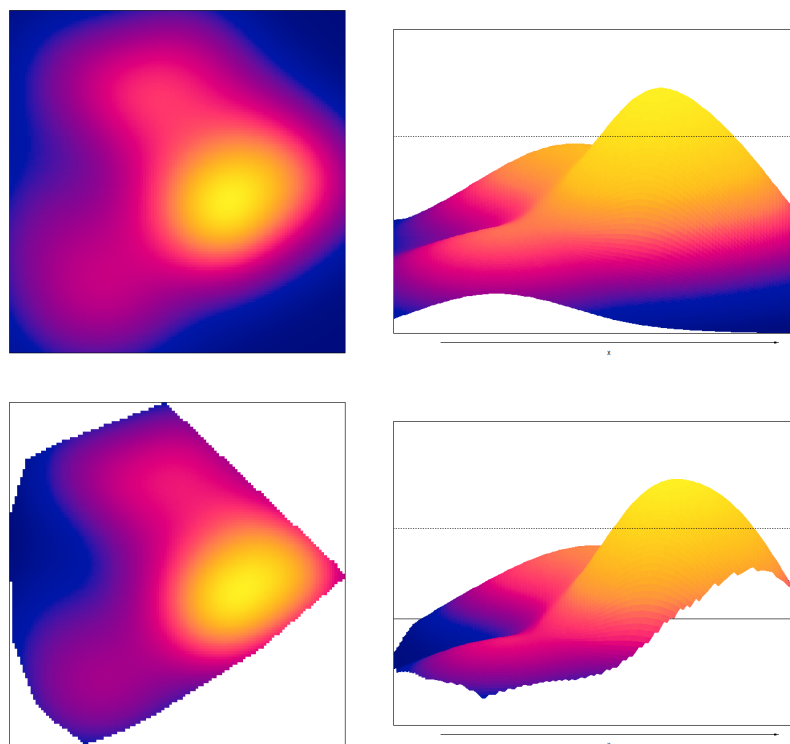
(b) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.

Figure 3.25: Uniformly corrected (a) and Diggle's corrected (b) kernel estimates for Machochwe.
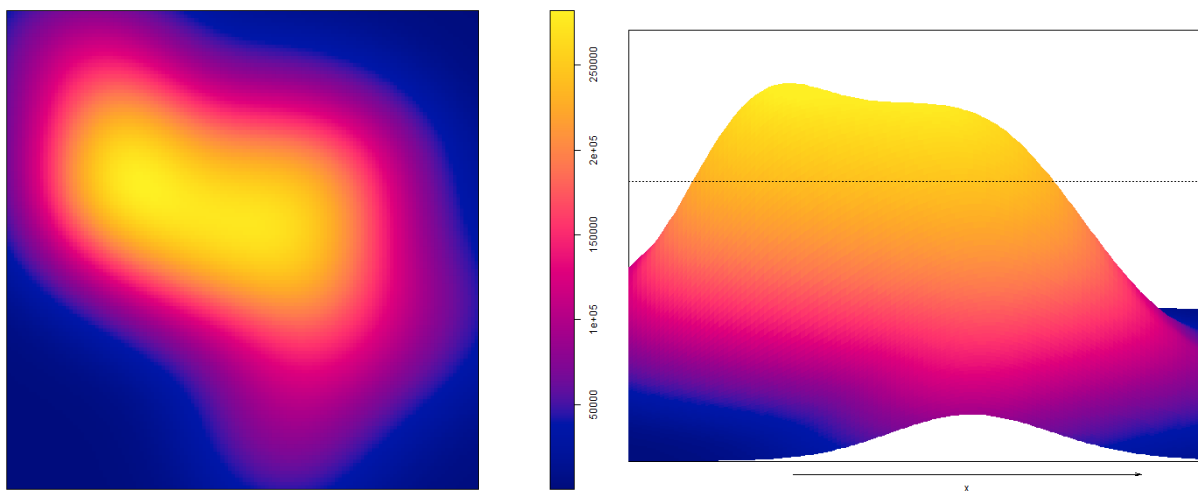
(a) Density estimate (left) and perspective view of density estimate (right)
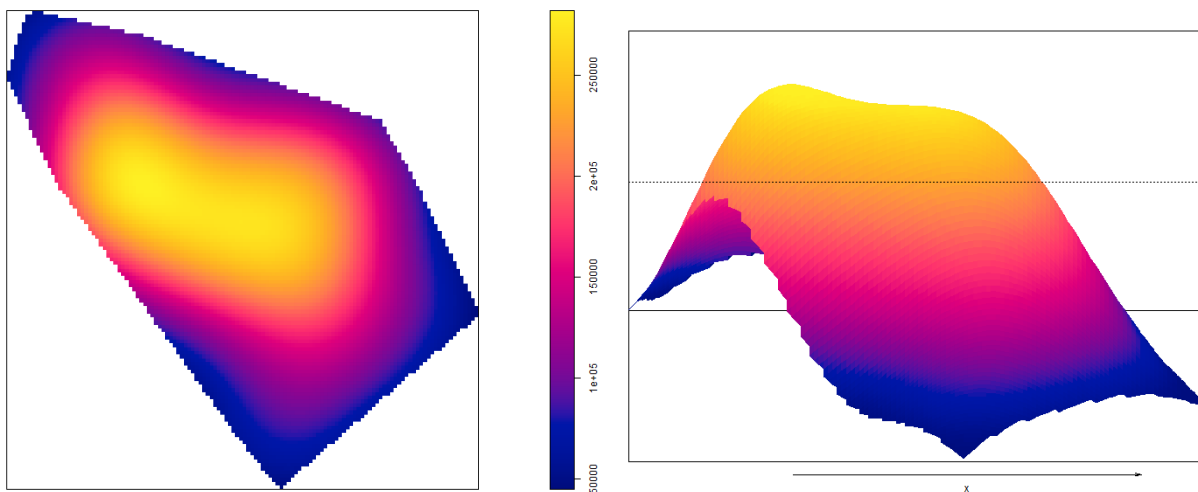
(b) Density estimate (left) and perspective view of density estimate (right)

Figure 3.26: Kernel estimation of point intensity for Mesaga using rectangular (top) and convex hull (bottom) windows.
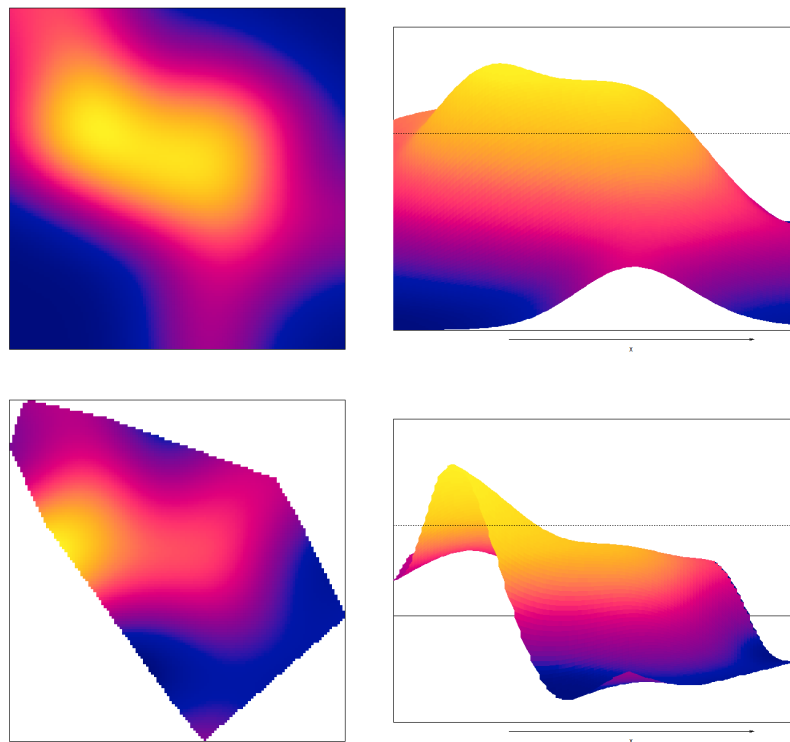
(a) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.
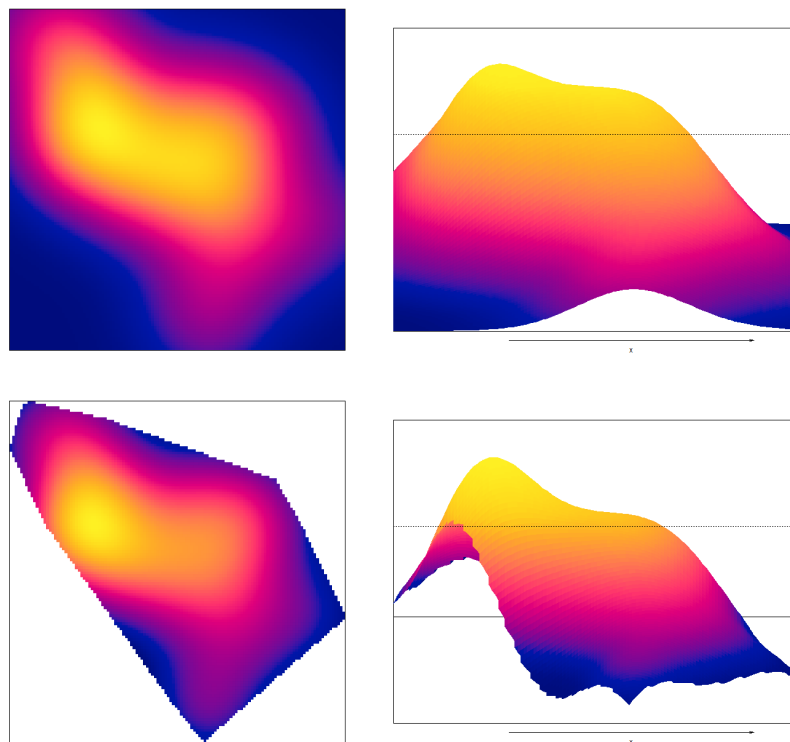
(b) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.

Figure 3.27: Uniformly corrected (a) and Diggle's corrected (b) kernel estimates for Mesaga.
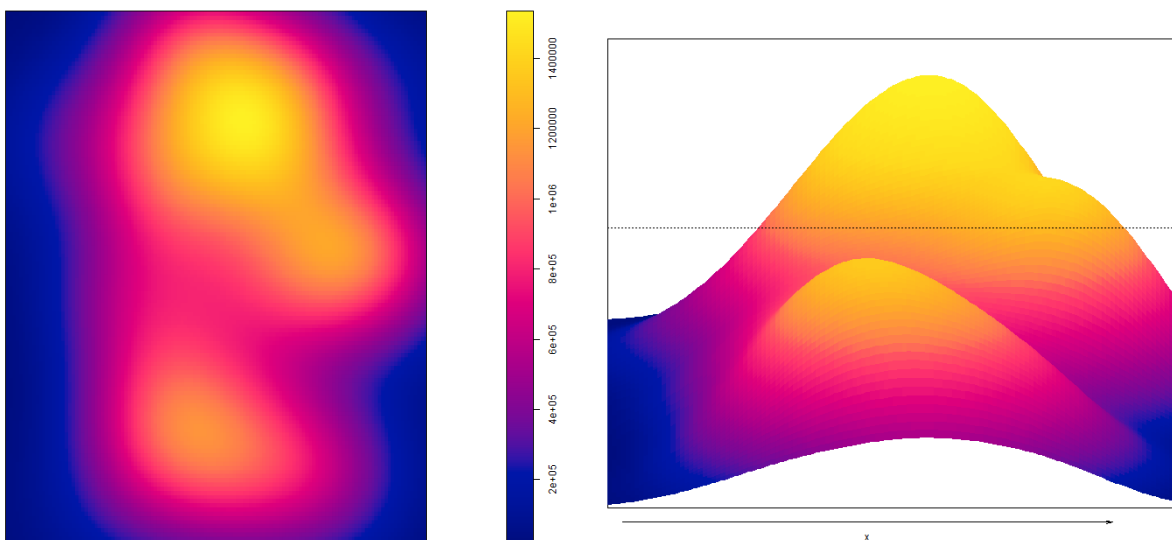
(a) Density estimate (left) and perspective view of density estimate (right)
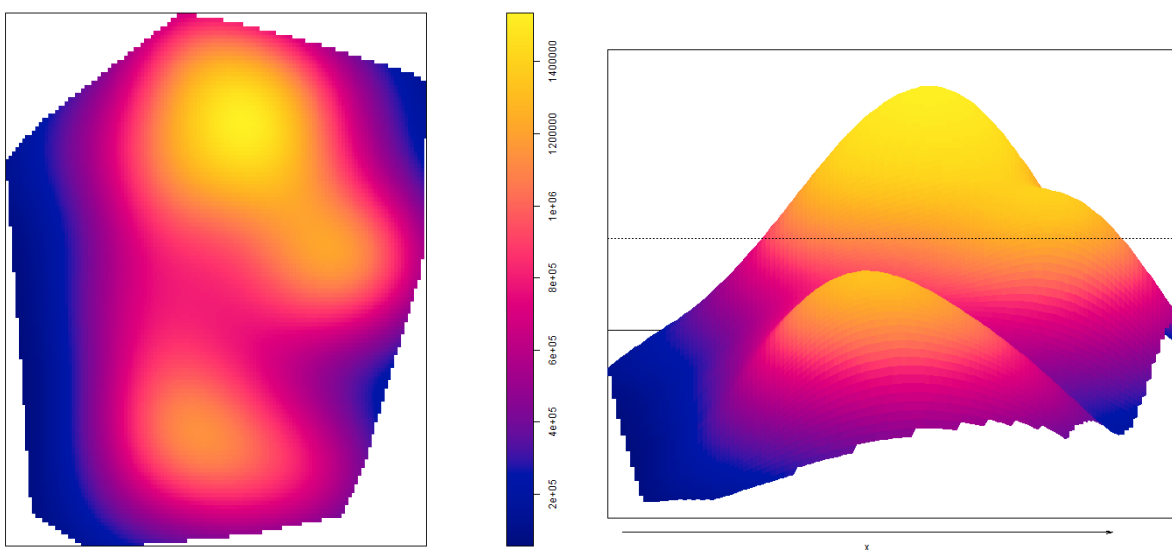
(b) Density estimate (left) and perspective view of density estimate (right)

Figure 3.28: Kernel estimation of point intensity for Morotonga using rectangular (top) and convex hull (bottom) windows.
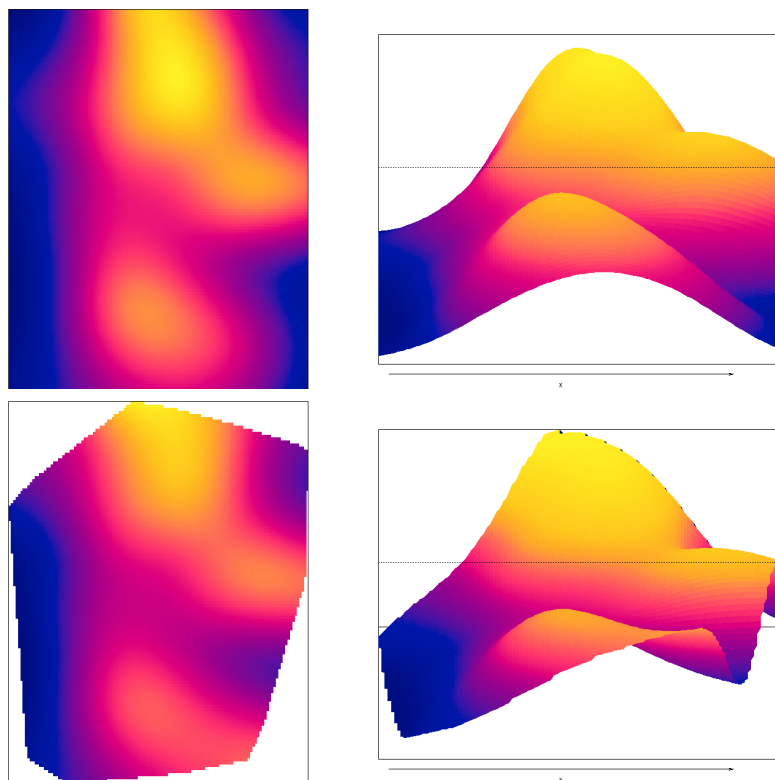
(a) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.
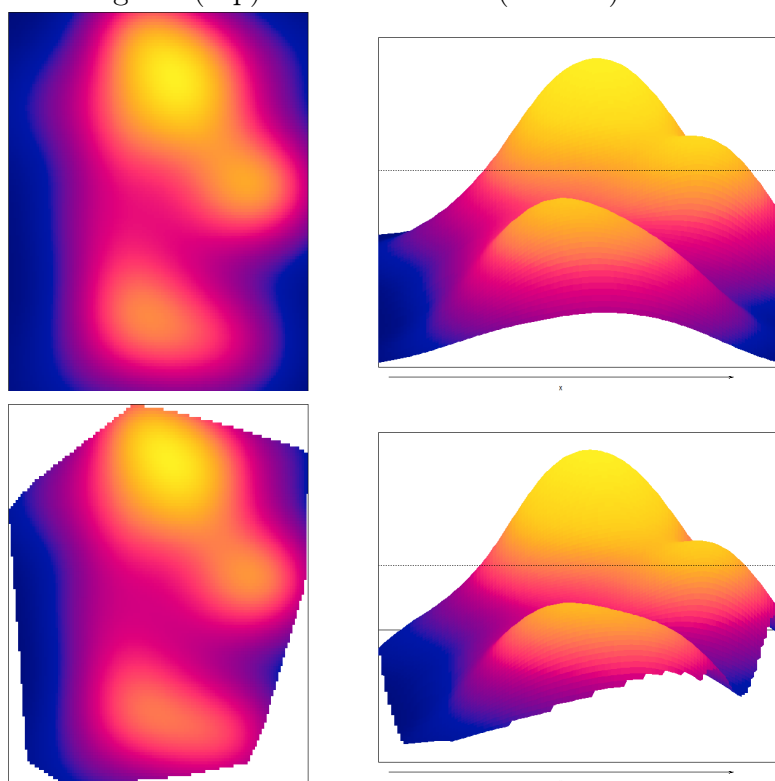
(b) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.

Figure 3.29: Uniformly corrected (a) and Diggle's corrected (b) kernel estimates for Morotonga.
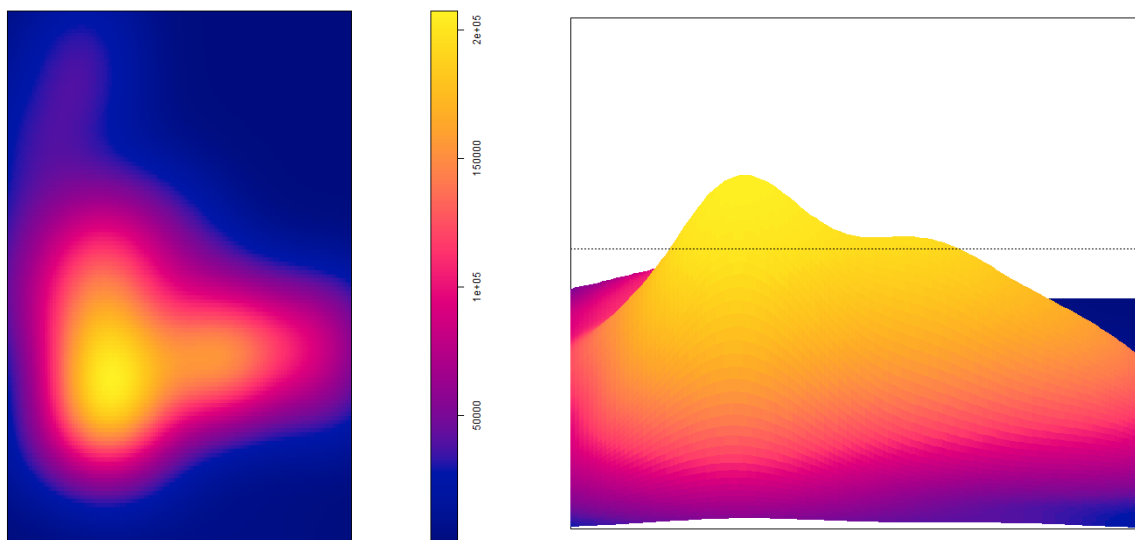
(a) Density estimate (left) and perspective view of density estimate (right)
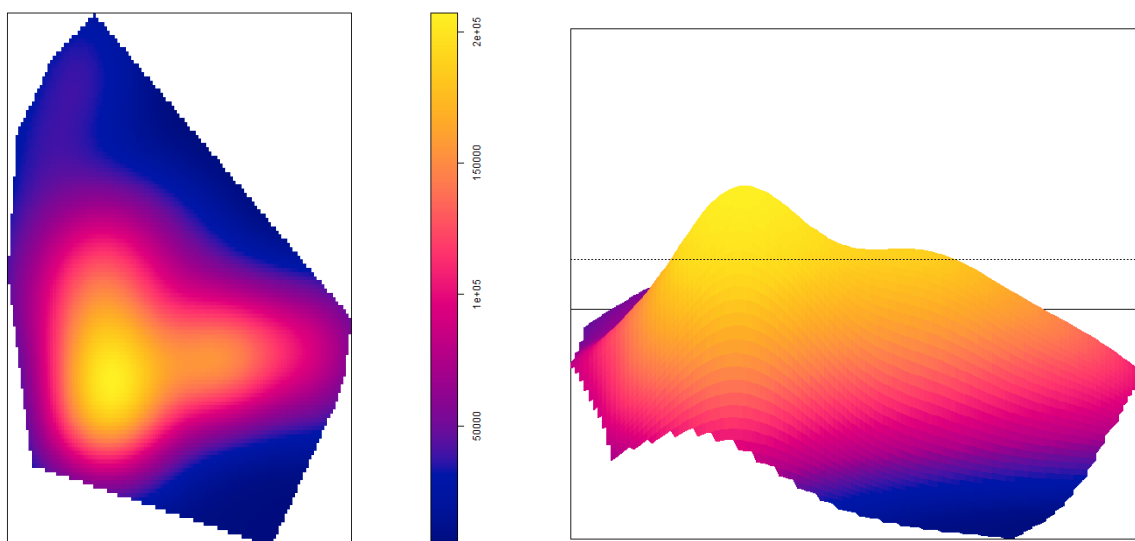
(b) Density estimate (left) and perspective view of density estimate (right)

Figure 3.30: Kernel estimation of point intensity for Ring'wani using rectangular (top) and convex hull (bottom) windows.
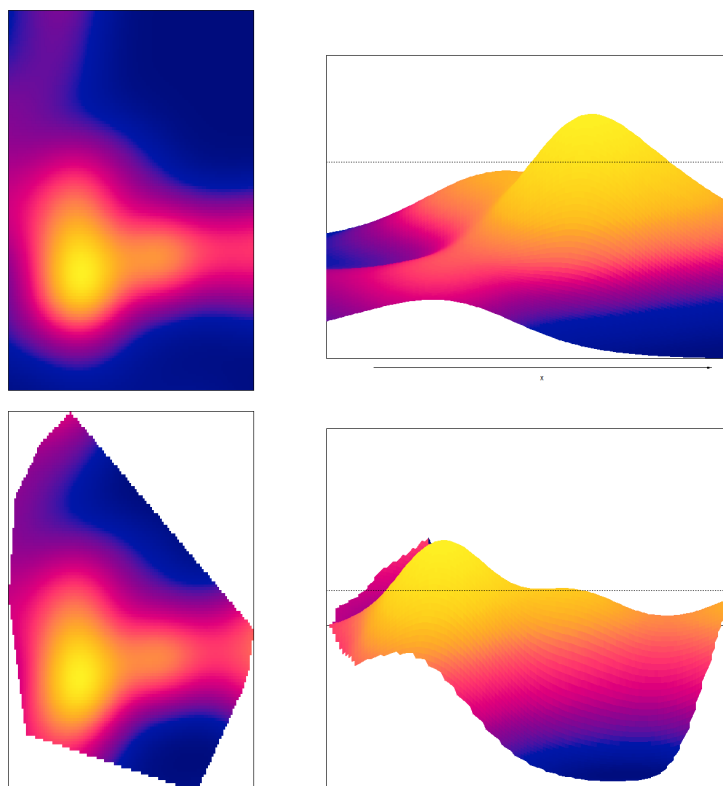
(a) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.
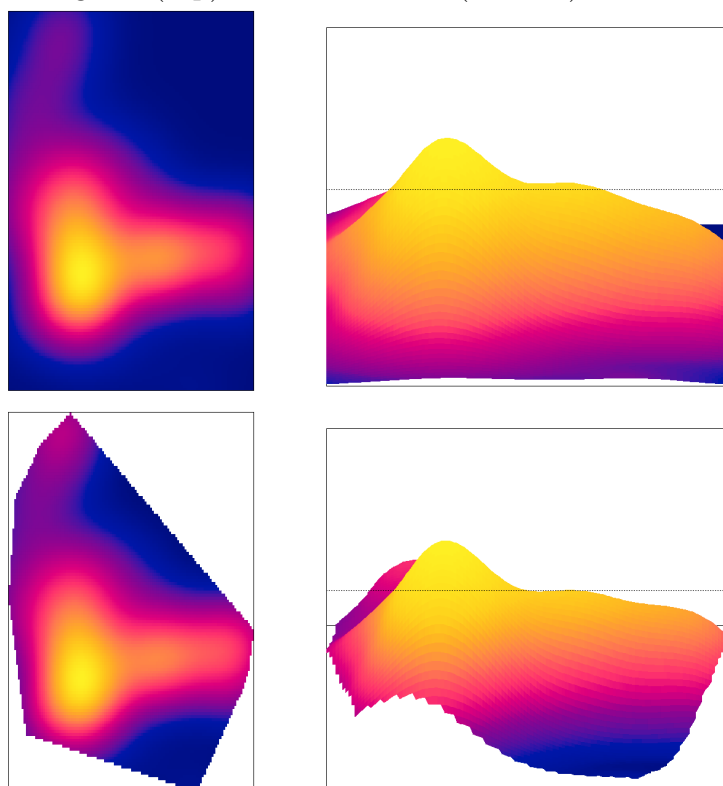
(b) Density estimate (left) and perspective view of density estimate (right) for rectangular (top) and convex hull (bottom) windows.

Figure 3.31: Uniformly corrected (a) and Diggle's corrected (b) kernel estimates for Ring'wani.

3.3.2 Testing for first-order homogeneity in household locations

We now look at the village data and test for first order homogeneity. To test for first-order homogeneity within each village, each village was divided into $n \times n$ grids, where $n = 3, 5, 7$. Only blocks containing points were used to estimate the intensity. Figures 3.32, 3.33, 3.34 and 3.35 show each of the villages divided into quadrats along with the number of households falling into each of the quadrats. Table 3.3 contains the number of non-empty quadrats as well as the maximum likelihood estimates of the intensity for each of the villages and each of the grid sizes and window types. The aim of excluding blocks with no points from calculations is to only include the actual village area and to exclude the empty area outside the village borders so as to not underestimate the intensity. For this reason, the intensity estimate for the convex hull windows will always be higher than the intensity of the rectangular windows. We should however also note that this will exclude empty space within the village if the grid becomes too fine, leading to the intensity being overestimated.
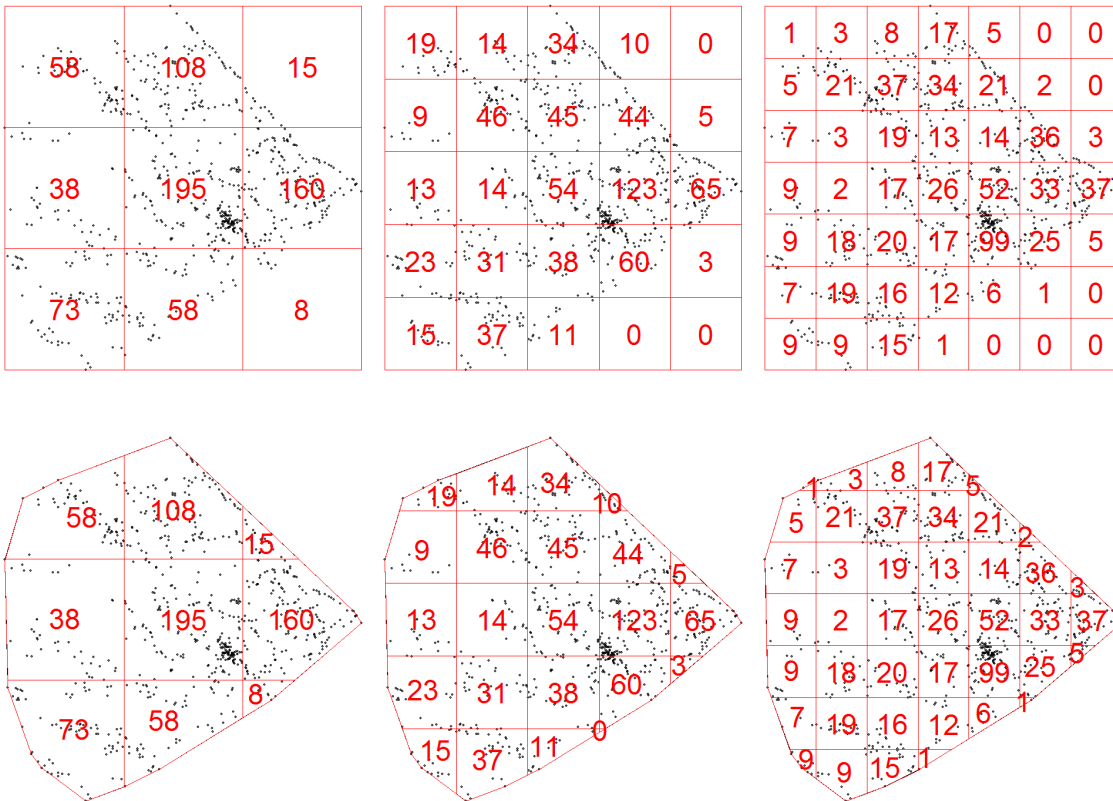
Figure 3.32: Quadrat counts for different grid sizes grids for Machochwe.
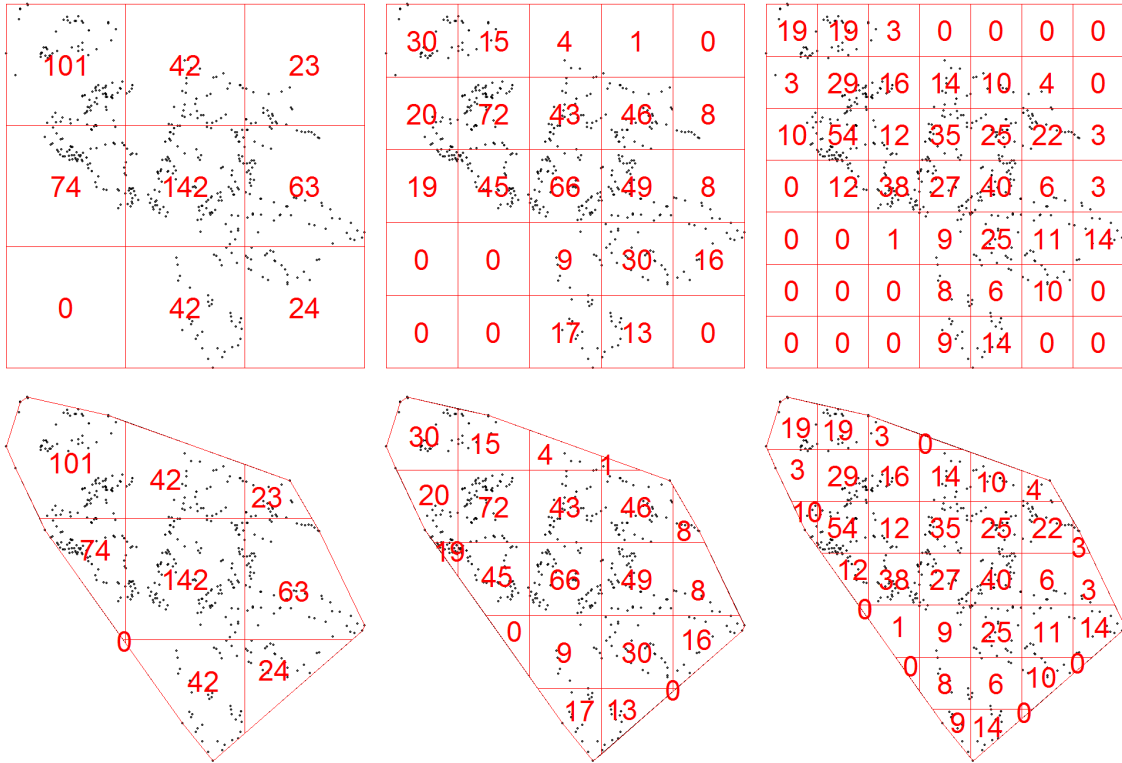
Figure 3.33: Quadrat counts for different grid sizes grids for Mesaga.
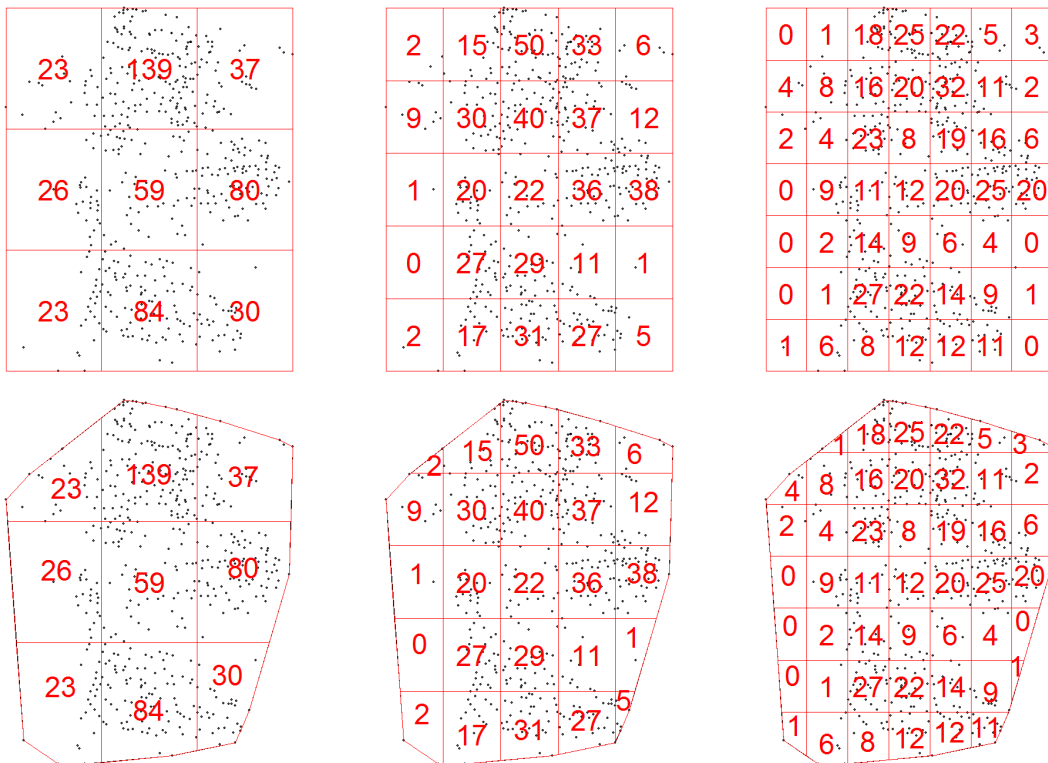
Figure 3.34: Quadrat counts for different grid sizes grids for Morotonga.
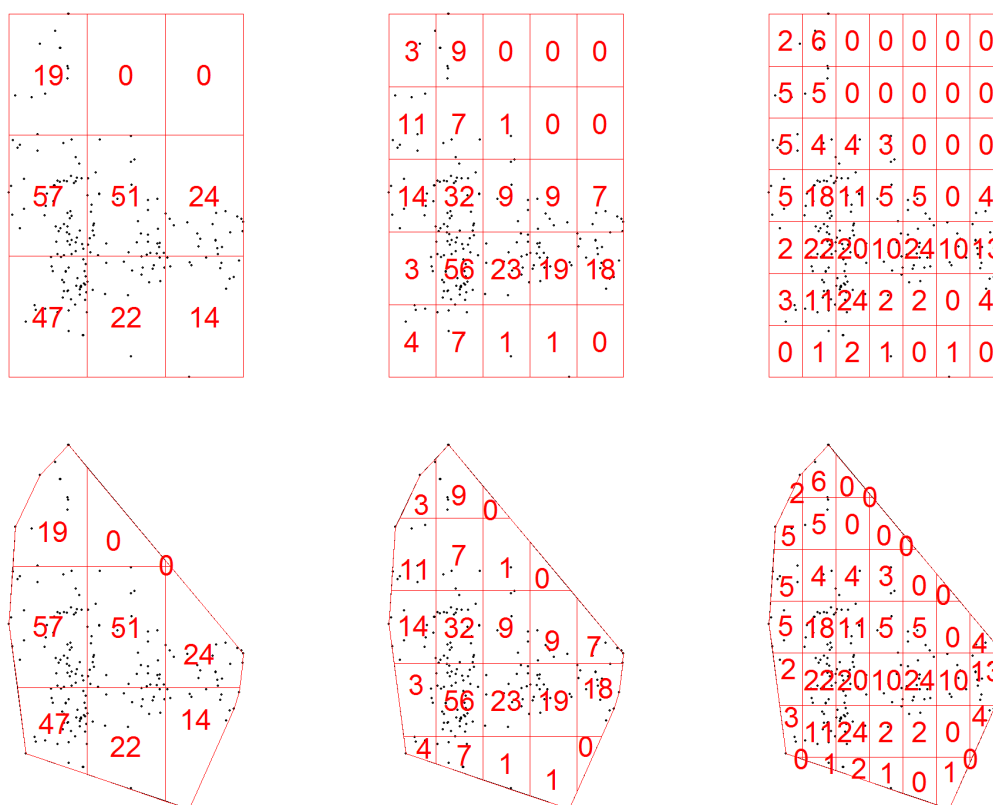
Figure 3.35: Quadrat counts for different grid sizes grids for Ring'wani.

Village	n	Number of quadrats	Non-empty quadrats	$\hat{\lambda}_{Rec}$	$\hat{\lambda}_{CH}$
Machochwe	3	9	9	192324.2	304596.9
	5	25	22	218550.2	304641.6
	7	49	42	224378.2	304596.9
Mesaga	3	9	8	127839.8	222634.1
	5	25	19	149520.2	218858.3
	7	49	32	174004.1	217824.8
Morotonga	3	9	9	704702.3	865131.6
	5	25	24	734064.9	898689.2
	7	49	43	803032.9	929458.8
Ring'wani	3	9	7	72275.9	118552.8
	5	25	19	73966.6	91578.2
	7	49	31	88855.4	108845.9

Table 3.3: Maximum likelihood estimates for λ obtained for different $n \times n$ grids

The χ^2 and likelihood ratio tests can be performed using the `quadrat.test` function available in the `spatstat` package in `R`. It should however be noted that this built-in function should be used with caution. During the application phase of this mini-dissertation we discovered some irregularities in the test statistics obtained. Although the built-in functions correctly calculated the test statistic when using a rectangular window with no empty quadrats, discrepancies occurred when there were empty quadrats or when using a convex hull window. When empty quadrats are present, the areas for these are still included in the calculation of the statistic and no option is available to exclude these. We were unable to determine the exact source of the discrepancy in the test statistic calculations when using a convex hull window. Secondly, by default the function performs a two-sided hypothesis test. In this context we consider only one-sided tests, specifically upper-tailed tests, as a large absolute difference between observed and expected counts is an indication of a deviation from the null. Thirdly, when specifying in the function that a one-sided test must be performed, the p-value given in the output is for a lower-tailed test and needs to be adjusted as we are performing an upper-tailed test. Due to these concerns, the formulae for all tests and p-values in this chapter were hard-coded.

Next we compared the number of households in the K non-empty blocks to see if the intensities are different in different areas. Tables 3.4 and 3.5 show a summary of the results for Pearson's χ^2 test for four villages using rectangular and convex hull windows respectively. For a rectangular window and all grid sizes, all asymptotic and bootstrap p-values were found to be significant. However when considering a convex hull window, using a 3×3 grid, both Mochochwe and Mesaga have insignificant bootstrap p-values. In these cases the tests do not identify the patterns as first-order heterogeneous. For these two cases, the bootstrap and asymptotic p-values give contradictory results.

A summary of the results obtained for the four villages using the likelihood ratio test can be seen in Tables 3.6 and 3.7 for rectangular and convex hull windows respectively. As with Pearson's χ^2 test, for a rectangular window and all grid sizes, all asymptotic and bootstrap p-values are significant and indicate first-order heterogeneity. However, for a convex hull window and a 3×3 grid, the bootstrap p-values do not identify the patterns as first-order heterogeneous for Machochwe and Mesaga. In these two cases the bootstrap and asymptotic p-values give contradictory results.

For the score test using a rectangular window, all asymptotic and bootstrap p-values indicate that, within all villages, we have first-order heterogeneity. Using a convex hull window and a 3×3 grid, the bootstrap p-values do not identify the patterns as first-order heterogeneous for Machochwe, Mesaga and Ring'wani. Again the asymptotic p-values give contradictory results for these three cases. For 5×5 and 7×7 grids, all asymptotic and

bootstrap p-values are significant. The results for our four example villages are given in Tables 3.8 and 3.9 for rectangular and convex hull windows respectively.

The results obtained from the VT test are again in line with previous results, showing asymptotic and bootstrap p-values close or equal to zero when using a rectangular window, and contradictory asymptotic and bootstrap p-values for Machochwe and Mesaga when using a convex hull window. In this case, all villages show significant asymptotic and bootstrap p-values for 5×5 and 7×7 grid sizes. The results for our four example villages are given in Tables 3.10 and 3.11 for rectangular and convex hull windows respectively.

For the UT test, the results for our four example villages are given in Tables 3.12 and 3.11 for rectangular and convex hull windows respectively. As with all previous tests using a rectangular window, all asymptotic and bootstrap p-values are significant. However for Mesaga, using a convex hull window and a 3×3 grid, we obtain contradictory bootstrap and asymptotic p-values, with bootstrap p-values again not identifying the heterogeneity. All other tests however identified this heterogeneity for all grid sizes and all villages.

Based on both the simulation study in Section 2.3.3, as well as the real-world application it is clear that we have reason to doubt the accuracy of the asymptotic p-values. It is clear that this is particularly true when the pattern is divided into a smaller number of quadrats and the window is not rectangular. The optimal choice of grid size as well as the reliability of using asymptotic p-values warrants further investigation.

In this section we showed that the choice of an appropriate window can also affect the outcome of the hypothesis tests, especially if a smaller grid size is used. In the case where a natural window does not exist, it is therefore important to consider the choice of this boundary carefully. Some guidelines are provided in [86, 90, 94].

The setting in which we performed these analysis necessitated that we divide the pattern into blocks. If some natural subdivision exists, this could be used instead. For instance, if we wanted to compare reported crime intensities in different wards, the wards are a natural subdivision that can be used since ward sizes and the number of reported crimes per ward will be known.

3×3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	411.419	0	0.01	0.001	0.0001
Mesaga	8	1	184.781	0	0.01	0.001	0.0001
Morotonga	9	0	222.251	0	0.01	0.001	0.0001
Ring'wani	7	2	55.453	3.75×10^{-10}	0.01	0.001	0.0001

5×5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	486.941	0	0.01	0.001	0.0001
Mesaga	19	6	294.250	0	0.01	0.001	0.0001
Morotonga	24	1	230.641	0	0.01	0.001	0.0001
Ring'wani	19	6	262.761	0	0.01	0.001	0.0001

7×7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	754.055	0	0.01	0.001	0.0001
Mesaga	32	17	310.292	0	0.01	0.001	0.0001
Morotonga	43	6	247.852	0	0.01	0.001	0.0001
Ring'wani	31	18	198.145	0	0.01	0.001	0.0001

Table 3.4: $\chi^2_{Pearson}$ test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	416.427	0	0.01	0.001	0.0001
Mesaga	8	1	175.073	0	0.01	0.001	0.0001
Morotonga	9	0	200.181	0	0.01	0.001	0.0001
Ring'wani	7	2	55.798	3.2×10^{-10}	0.01	0.001	0.0001

5×5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	425.433	0	0.01	0.001	0.0001
Mesaga	19	6	294.244	0	0.01	0.001	0.0001
Morotonga	24	1	270.851	0	0.01	0.001	0.0001
Ring'wani	19	6	209.028	0	0.01	0.001	0.0001

7×7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	579.589	0	0.01	0.001	0.0001
Mesaga	32	17	288.031	0	0.01	0.001	0.0001
Morotonga	43	6	262.676	0	0.01	0.001	0.0001
Ring'wani	31	18	172.772	0	0.01	0.001	0.0001

Table 3.6: LR test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	534.931	0	0.70	0.707	0.7080
Mesaga	8	1	16386.267	0	0.23	0.179	0.1547
Morotonga	9	0	150.981	0	0.01	0.001	0.0001
Ring'wani	7	2	23819.173	0	0.01	0.001	0.0002

5×5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	300.166	0	0.01	0.001	0.0001
Mesaga	19	6	196.686	0	0.01	0.001	0.0001
Morotonga	24	1	167.665	0	0.01	0.001	0.0001
Ring'wani	19	6	186.717	0	0.01	0.001	0.0001

7×7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	481.922	0	0.01	0.001	0.0001
Mesaga	32	17	203.572	0	0.01	0.001	0.0001
Morotonga	43	6	195.987	0	0.01	0.001	0.0001
Ring'wani	31	18	134.811	3.11×10^{-15}	0.01	0.001	0.0016

Table 3.5: $\chi^2_{Pearson}$ test statistics and p-values obtained using a convex hull window.

3×3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	411.322	0	0.11	0.057	0.0656
Mesaga	8	1	303.970	0	0.24	0.186	0.1655
Morotonga	9	0	140.277	0	0.01	0.001	0.0001
Ring'wani	7	2	572.555	0	0.01	0.002	0.0004

5×5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	275.237	0	0.01	0.001	0.0001
Mesaga	19	6	161.624	0	0.01	0.001	0.0001
Morotonga	24	1	195.180	0	0.01	0.001	0.0001
Ring'wani	19	6	165.114	0	0.01	0.001	0.0001

7×7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	383.434	0	0.01	0.001	0.0001
Mesaga	32	17	204.914	0	0.01	0.001	0.0001
Morotonga	43	6	207.566	0	0.01	0.001	0.0001
Ring'wani	31	18	126.239	9.04×10^{-14}	0.01	0.001	0.0001

Table 3.7: LR test statistics and p-values obtained using a convex hull window.

3 × 3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	1087.011	0	0.01	0.001	0.0001
Mesaga	8	1	219.705	0	0.01	0.001	0.0001
Morotonga	9	0	225.126	0	0.01	0.001	0.0001
Ring'wani	7	2	67.281	1.47×10^{-12}	0.01	0.001	0.0001

5 × 5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	818.736	0	0.01	0.001	0.0001
Mesaga	19	6	1052.824	0	0.01	0.001	0.0001
Morotonga	24	1	1330.757	0	0.01	0.001	0.0001
Ring'wani	19	6	530.716	0	0.01	0.001	0.0001

7 × 7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	1510.263	0	0.01	0.001	0.0001
Mesaga	32	17	636.222	0	0.01	0.001	0.0001
Morotonga	43	6	764.609	0	0.01	0.001	0.0001
Ring'wani	31	18	291.704	0	0.01	0.001	0.0001

Table 3.8: SC test statistics and p-values obtained using a rectangular window.

3 × 3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	663.535	0	0.35	0.394	0.3899
Mesaga	8	1	94.291	0	0.14	0.107	0.1121
Morotonga	9	0	160.692	0	0.01	0.001	0.0001
Ring'wani	7	2	146.629	0	0.17	0.211	0.1843

5 × 5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	400.035	0	0.01	0.001	0.0001
Mesaga	19	6	242.406	0	0.01	0.001	0.0001
Morotonga	24	1	1077.170	0	0.01	0.001	0.0001
Ring'wani	19	6	525.883	0	0.01	0.001	0.0001

7 × 7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	654.905	0	0.01	0.001	0.0001
Mesaga	32	17	489.498	0	0.01	0.001	0.0001
Morotonga	43	6	514.706	0	0.01	0.001	0.0001
Ring'wani	31	18	201.143	0	0.01	0.001	0.0001

Table 3.9: SC test statistics and p-values obtained using a convex hull window.

3 × 3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	410.528	0	0.01	0.001	0.0001
Mesaga	8	1	184.393	0	0.01	0.001	0.0001
Morotonga	9	0	221.548	0	0.01	0.001	0.0001
Ring'wani	7	2	55.296	3.79×10^{-10}	0.01	0.001	0.0001

5 × 5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	481.279	0	0.01	0.001	0.0001
Mesaga	19	6	290.297	0	0.01	0.001	0.0001
Morotonga	24	1	226.128	0	0.01	0.001	0.0001
Ring'wani	19	6	255.113	0	0.01	0.001	0.0001

7 × 7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	734.448	0	0.01	0.001	0.0001
Mesaga	32	17	301.786	0	0.01	0.001	0.0001
Morotonga	43	6	237.506	0	0.01	0.001	0.0001
Ring'wani	31	18	186.538	0	0.01	0.001	0.0001

Table 3.10: VT test statistics and p-values obtained using a rectangular window.

3 × 3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	524.981	0	0.70	0.703	0.7051
Mesaga	8	1	4875.603	0	0.23	0.187	0.1632
Morotonga	9	0	150.740	0	0.01	0.001	0.0001
Ring'wani	7	2	12286.505	0	0.01	0.004	0.0008

5 × 5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	289.761	0	0.01	0.001	0.0001
Mesaga	19	6	186.177	0	0.01	0.001	0.0001
Morotonga	24	1	161.977	0	0.01	0.001	0.0001
Ring'wani	19	6	180.195	0	0.01	0.001	0.0001

7 × 7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	434.115	0	0.01	0.001	0.0001
Mesaga	32	17	195.464	0	0.01	0.001	0.0001
Morotonga	43	6	185.493	0	0.01	0.001	0.0001
Ring'wani	31	18	122.802	1.1×10^{-14}	0.01	0.001	0.0001

Table 3.11: VT test statistics and p-values obtained using a convex hull window.

3×3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	364.700	0	0.01	0.001	0.0001
Mesaga	8	1	170.471	0	0.01	0.001	0.0001
Morotonga	9	0	201.809	0	0.01	0.001	0.0001
Ring'wani	7	2	52.368	4.02×10^{-9}	0.01	0.001	0.0001

5×5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	423.546	0	0.01	0.001	0.0001
Mesaga	19	6	260.528	0	0.01	0.001	0.0001
Morotonga	24	1	208.175	0	0.01	0.001	0.0001
Ring'wani	19	6	213.137	0	0.01	0.001	0.0001

7×7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	618.262	0	0.01	0.001	0.0001
Mesaga	32	17	272.510	0	0.01	0.001	0.0001
Morotonga	43	6	220.854	0	0.01	0.001	0.0001
Ring'wani	31	18	165.873	0	0.01	0.001	0.0001

Table 3.12: UT test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	9	0	249.612	0	0.01	0.001	0.0001
Mesaga	8	1	45.851	1.47×10^{-8}	0.09	0.108	0.1136
Morotonga	9	0	139.212	0	0.01	0.001	0.0001
Ring'wani	7	2	109.004	0	0.02	0.003	0.0034

5×5 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	22	3	266.451	0	0.01	0.001	0.0001
Mesaga	19	6	106.396	1.11×10^{-16}	0.01	0.001	0.0001
Morotonga	24	1	139.091	0	0.01	0.001	0.0001
Ring'wani	19	6	159.835	0	0.01	0.001	0.0001

7×7 grids:

Village	K	$n^2 - K$	Test statistic	Asymptotic	R=99	R=999	R=9999
Machochwe	42	7	408.684	0	0.01	0.001	0.0001
Mesaga	32	17	180.299	0	0.01	0.001	0.0001
Morotonga	43	6	172.218	0	0.01	0.001	0.0001
Ring'wani	31	18	115.646	4.22×10^{-14}	0.01	0.001	0.0001

Table 3.13: UT test statistics and p-values obtained using a convex hull window.

3.4 Considering the independence assumption in hypothesis tests

The hypothesis tests applied to the point patterns in Section 3.3.2 assume that counts used in the calculation of the test statistics are obtained from areas that are independent of each other. This assumption of independence is unlikely to be valid and was discussed in Sections 2.3.3 and 2.3.3.2. In this section we consider alternative methods of applying these tests such that the independence assumption is no longer violated.

As a first alternative, the counts from every second quadrat is used in the calculation of the test statistic. We perform these tests by first taking every second quadrat into account, starting at the first (odd-numbered quadrats - see Figures 3.36 and 3.37), and then calculating test statistics using every second quadrat, starting at the second (even-numbered quadrats - see Figures 3.36 and 3.39). As in Section 3.3.2, only non-empty quadrats were included in the calculation of the test statistics. Since the quadrats that are used in the calculations are not directly adjacent to each other, it is unlikely that the number of points in one quadrat is dependent on the number of points in another quadrat. The results obtained for each of the five tests for the odd- and even-numbered quadrats using a rectangular window are compared in Tables 3.14, 3.18, 3.22, 3.26 and 3.30. The results obtained for each of the five tests for the odd- and even-numbered quadrats using a convex hull window are compared in Tables 3.15, 3.19, 3.23, 3.27 and 3.31. Although the results for both the even- and odd-numbered quadrats are given for these test, it is important to note that the test statistics should not be compared directly as the number of quadrats used in the calculations of these may not be the same. In addition, even if the number of quadrats are the same, the area of the quadrats included are unlikely to be then same if a convex hull window is used. For these tests the bootstrap p-values are obtained as follows. First a homogeneous Poisson point pattern is generated on the same window as the original pattern using the maximum likelihood estimate obtained for the intensity of the original pattern. The quadrats in the simulated pattern corresponding to those used to calculate the test statistic for the original pattern, are then used to calculated the first bootstrap test statistic. For instance, if odd-numbered quadrats are used for the calculation of the original test statistic, then the odd-numbered quadrats in the simulated pattern are used for the calculation of the bootstrap test statistic. This process of generating patterns from a homogeneous Poisson point process and calculating test statistics is then repeated 99, 999 and 9999 times after which the bootstrap test statistics are compared to the original test statistic and the bootstrap p-value is calculated using Equation 2.8 on page 39.
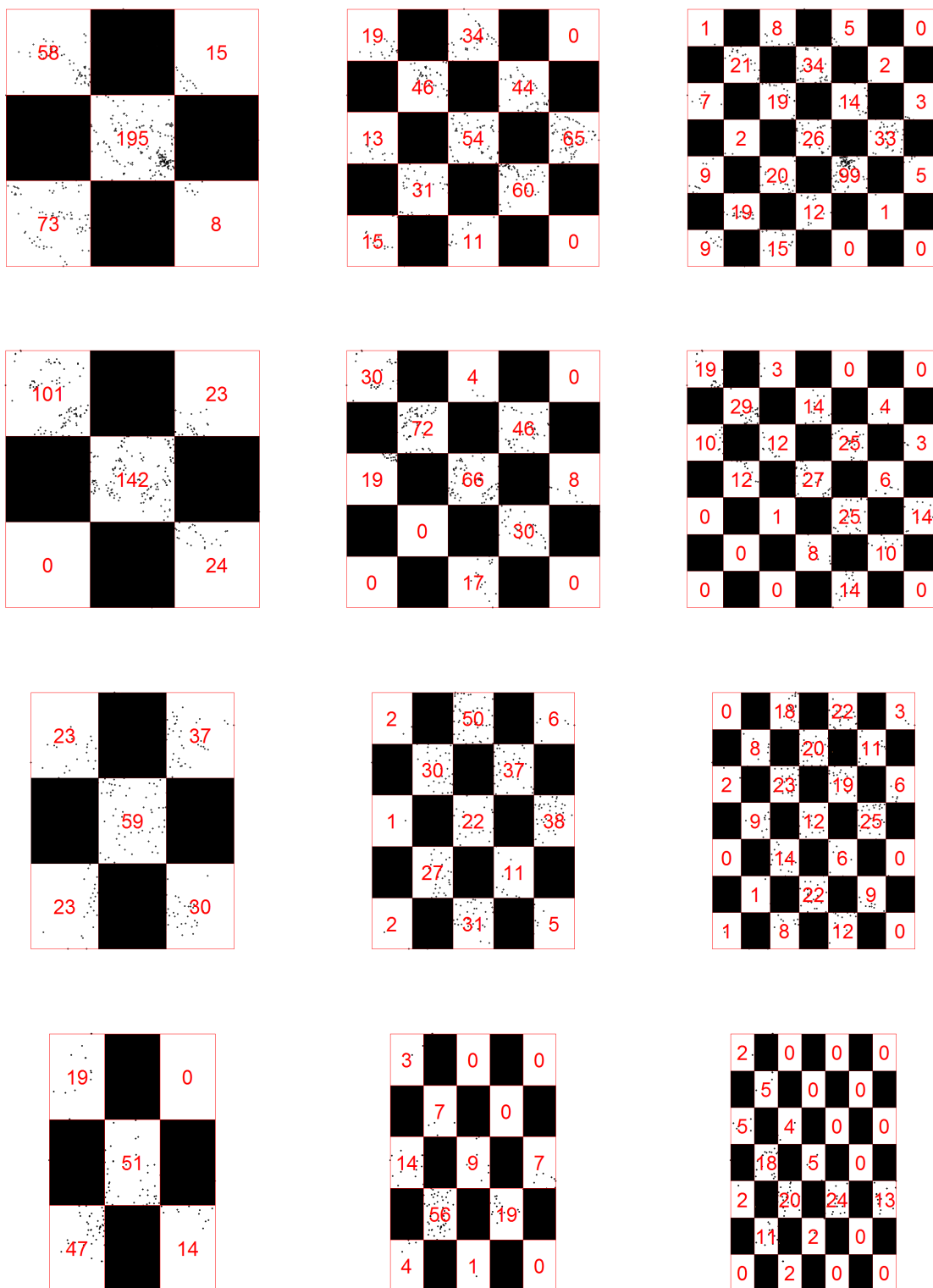
Figure 3.36: Quadrats used for the calculation of test statistics using a rectangular window if odd-numbered quadrats are included.
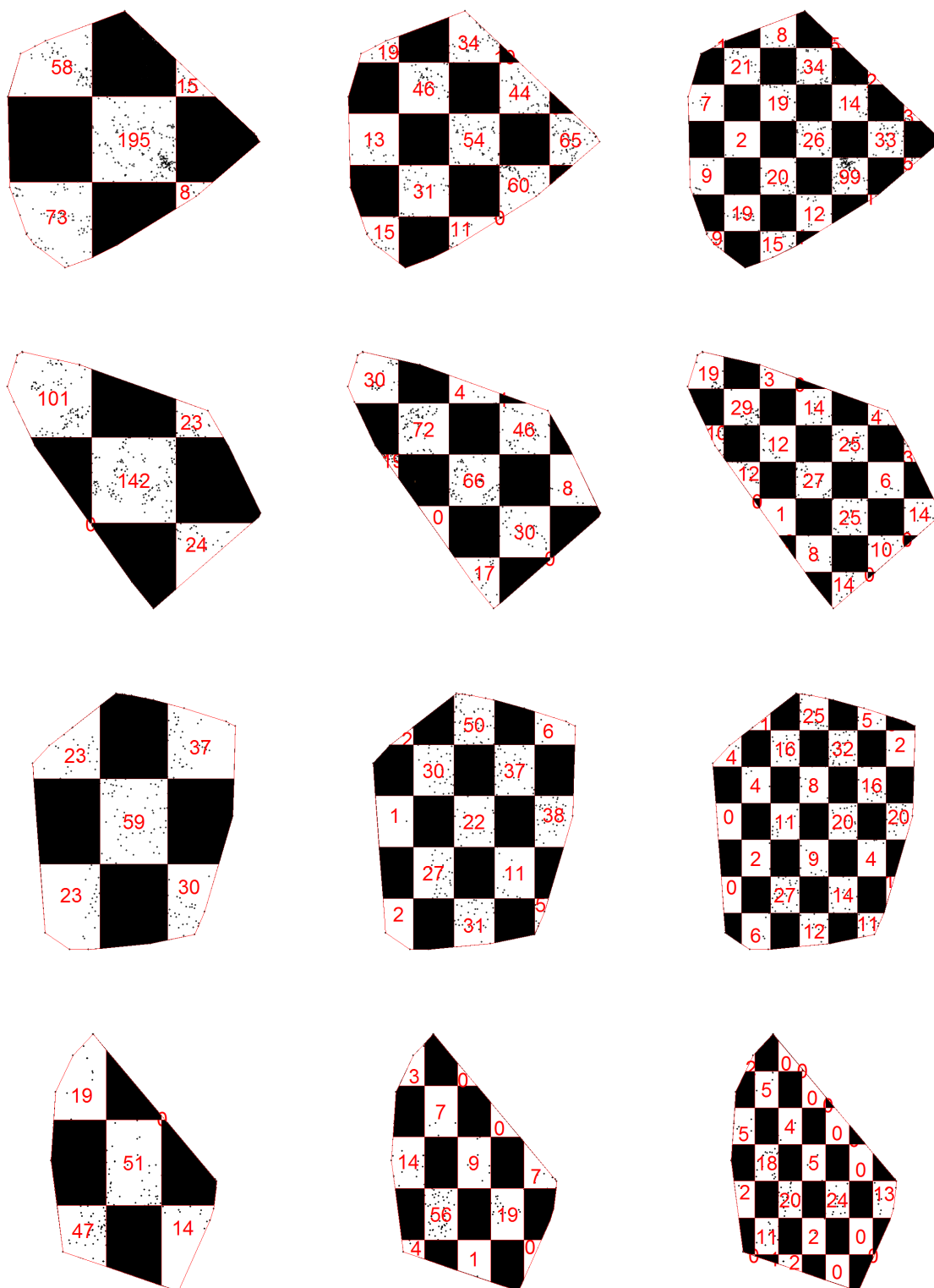
Figure 3.37: Quadrats used for the calculation of test statistics using a convex hull window if odd-numbered quadrats are included.
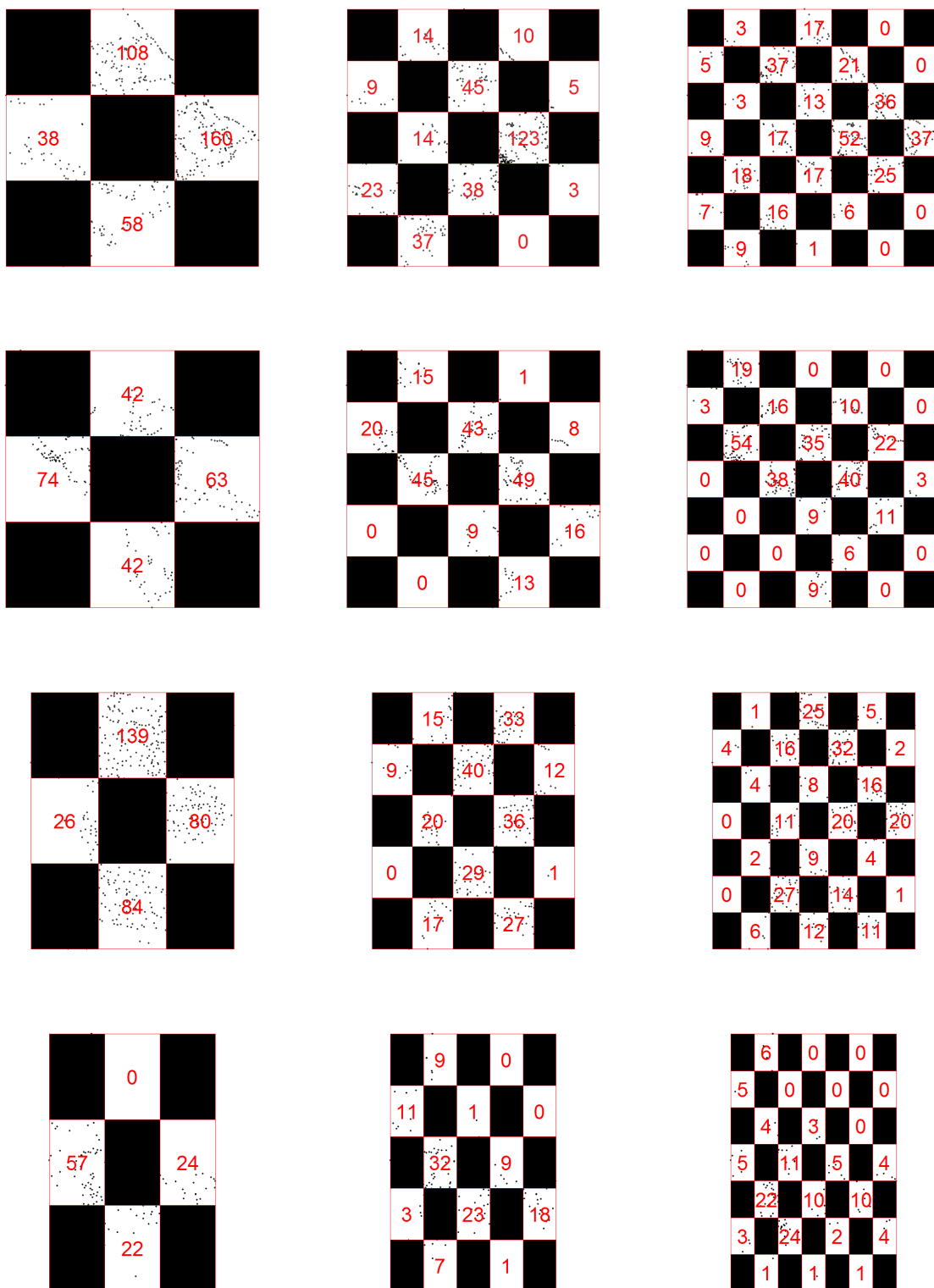
Figure 3.38: Quadrats used for the calculation of test statistics using a rectangular window if even-numbered quadrats are included.
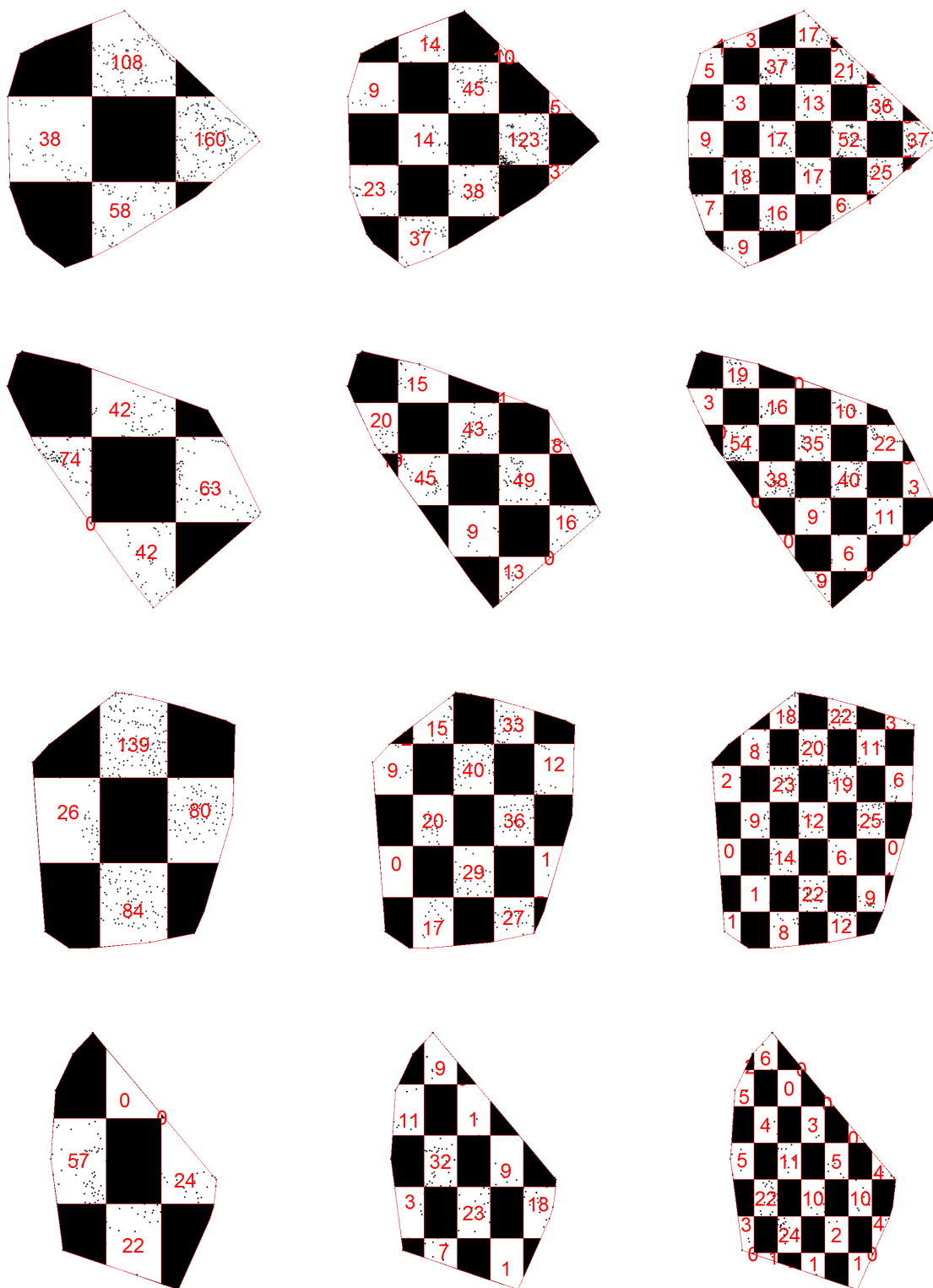
Figure 3.39: Quadrats used for the calculation of test statistics using a convex hull window if even-numbered quadrats are included.

A second approach for dealing with the independence violation involves randomly sampling a subset of the non-empty quadrats and using these to calculate a test statistic. This process is repeated 99 times, each time drawing a different sample of quadrats. Summaries of the results obtained using this approach using a rectangular window are given in Tables 3.16, 3.20, 3.24, 3.28 and 3.32 with the summaries for convex hull windows are given in Tables 3.17, 3.21, 3.25, 3.29 and 3.33. These summaries are based on taking 99 random samples of 50% of the non-empty quadrats and calculating a test statistic for each of these. For each of these test statistics, bootstrap p-values are calculated as follows. As before, point patterns are generated from a homogeneous Poisson process on the same window as the original pattern. A sample of 50% of the non-empty quadrats is then used to calculate the bootstrap test statistic. This process is repeated first 99 and then 999 times to obtain two bootstrap p-values for each of the original test statistic values using Equation 2.8 on page 39. The asymptotic and bootstrap p-values for each of the 99 test statistics are summarized to show the minimum, maximum, average and median p-value. The standard deviation along with an empirical 95% confidence interval for the p-values is also shown.

For the χ^2 test using the first approach, from now on referred to as the alternating quadrats approach, on a rectangular window and for all grid sizes, it is clear from Table 3.14 that all p-values, both asymptotic and bootstrap, are significant and indicate a heterogeneous intensity. This corresponds to the results obtained in Table 3.4 on page 87. Using the second approach, from now on referred to as the bootstrap approach, the χ^2 test indicates a heterogeneous pattern for all villages using 5×5 and 7×7 grids as indicated in Table 3.16 with bootstrap and asymptotic p-values showing little to no variation. For a 3×3 grid, the test indicates a heterogeneous pattern for Machochwe and Mesaga. For Morotonga and Ring'wani, the p-values vary considerably and empirical confidence intervals are wide. However, in both cases the average and median p-values still indicate a heterogeneous intensity.

Next we consider using a convex hull window and calculating the test statistics using the alternating quadrats approach. For Mesaga using a 3×3 grid, both the even and odd quadrats lead to insignificant p-values, thereby incorrectly indicating a homogeneous intensity. For all other villages using a 3×3 grid we obtain contradictory results for the odd- and even-numbered quadrats. Except for the same contradictory results obtained for Mesaga on a 5×5 grid, all tests performed on all villages using 5×5 and 7×7 grids show significant p-values and identify the heterogeneity in the pattern. The comparison of the odd- and even-numbered quadrat tests implies that the choice of which quadrats to include in the calculation of the test statistic can have a significant impact on the

outcome of the test. The p-values for the bootstrap approach on convex hull windows are summarised in Table 3.17. For a 3×3 grid, the p-values obtained for different test statistics vary considerably and in most cases empirical confidence intervals are wide and include both significant and insignificant p-values. For example, Ring'wani has an empirical 95% confidence interval given as (0.01, 0.944) obtained from 99 Monte Carlo simulations and a 95% confidence interval given as (0.001, 0.891) obtained from 999 Monte Carlo simulations. The high variability in the p-values obtained suggests that the choice of quadrats that are included in the calculation of the test statistic has a significant impact on the conclusion reached by the test. However, when using a 5×5 or a 7×7 grid, the variability in the p-values obtained is reduced and results from tests become more consistent.

For we consider the likelihood ratio (Table 3.18), score (Table 3.22), VT (Table 3.26) and UT tests (Table 3.30) using a rectangular window and the alternating quadrats approach. We observe similar results as with the χ^2 test as all tests have significant p-values for all grid sizes. This is consistent with the results obtained in Section 3.3.2. As with the bootstrap approach discussed for the χ^2 test, we observe that specifically Morotonga and Ring'wani show highly variable p-values when a 3×3 grid is used. The p-values obtained for larger grids are more stable with only Ring'wani having some large p-values for the LR and SC tests. However, in both these cases the empirical confidence intervals are narrow and contain only significant p-values. Summaries of the bootstrap p-values obtained for rectangular windows corresponding to the likelihood ratio, score, VT and UT tests can be seen in Tables 3.20, 3.24, 3.28 and 3.32 respectively.

Using convex hull windows and the alternating quadrats approach, we again see similar results to the χ^2 test. Using a 3×3 grid, the choice of using the odd- or even-numbered quadrats have a significant effect on the outcomes of the tests as can be seen in Tables 3.19, 3.23, 3.27 and 3.31. For larger grid sizes and all tests the p-values are found to be significant and the tests correctly identify the heterogeneity in the intensity. The only exception to this occurred with the UT test on a 5×5 grid for Mesaga using odd-numbered quadrats. Summaries for p-values obtained for the likelihood ratio, score, VT and UT tests using the bootstrap approach are given in Tables 3.21, 3.25, 3.29 and 3.33 respectively. As with the χ^2 test, p-values obtained through randomly sampling non-empty quadrats from a 3×3 grid are highly variable and often fail to pick up the heterogeneity in intensity across the patterns. However, for larger grid sizes the p-values stabilise and in cases where large p-values still occur, the average and median p-values are significant. In most of these cases, the 95% empirical confidence intervals contain only significant p-values, except for the SC and VT tests performed on Ring'wani using a 7×7 grid. It should however be noted that Ring'wani is the smallest village analysed, consisting of only 234 households.

3×3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	324.453	98.330	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	4	4	144.069	13.805	0	0.00318	0.01	0.02	0.001	0.003	0.0001	0.0036
Morotonga	5	4	25.907	77.723	3.30×10^{-5}	1.11×10^{-16}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	4	3	32.878	22.205	3.42×10^{-7}	1.30×10^{-5}	0.01	0.01	0.001	0.001	0.0001	0.0003

5×5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	107.097	402.156	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	145.240	122.142	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	13	11	163.626	69.138	0	6.50×10^{-11}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	172.350	80.737	0	1.15×10^{-13}	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	22	20	554.319	207.504	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	98.983	169.080	1.37×10^{-13}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	21	22	99.474	149.168	1.56×10^{-12}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	82.230	111.810	1.55×10^{-13}	5.55×10^{-16}	0.01	0.01	0.001	0.001	0.0001	0.0001

Table 3.14: $\chi^2_{Pearson}$ test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	354.415	154.435	0	0	0.98	0.01	0.994	0.001	0.9912	0.0001
Mesaga	4	4	13558.231	20.964	0	0.00011	0.2	0.95	0.24	0.946	0.2711	0.9566
Morotonga	5	4	9.272	78.199	0.05464	1.11×10^{-16}	0.07	0.01	0.087	0.001	0.0798	0.0001
Ring'wani	4	3	23367.020	63.147	0	1.94×10^{-14}	0.01	0.34	0.001	0.364	0.0001	0.3655

5×5 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	199.084	102.799	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	17.145	175.354	0.02863	0	0.08	0.01	0.095	0.001	0.0951	0.0001
Morotonga	13	11	95.474	71.219	4.22×10^{-15}	2.58×10^{-11}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	55.405	124.972	3.68×10^{-9}	0	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	21	21	148.491	333.335	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	98.879	101.226	1.43×10^{-13}	9.99×10^{-16}	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	22	21	107.973	87.774	1.09×10^{-13}	1.81×10^{-10}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	60.498	71.651	1.83×10^{-8}	1.12×10^{-8}	0.02	0.01	0.005	0.004	0.0038	0.0015

Table 3.15: $\chi^2_{Pearson}$ test statistics and p-values obtained using a convex hull window.

3×3 grid:

	Machochwe - $m = 4$			Mesaga - $m = 4$			Morotonga - $m = 4$			Ring'wani - $m = 4$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	5.9×10^{-9}	0.01	0.001
Maximum	0.01226165	0.01	0.001	0.015	0.01	0.017	0.28469	0.31	0.274	0.41156	0.37	0.419
Average	0.00012386	0.01	0.001	0.00016	0.01	0.00133	0.00856	0.01939	0.00981	0.00428	0.01384	0.02558
Median	0	0.01	0.001	1.9×10^{-14}	0.01	0.001	3.5×10^{-12}	0.01	0.001	3.6×10^{-6}	0.01	0.001
Std deviation	0.00123234	0	0	0.00151	0	0.0022	0.04268	0.04673	0.04363	0.04135	0.03518	0.09274
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	5.9×10^{-9}	0.01	0.001
$P_{97.5}$	5.01×10^{-8}	0.01	0.001	0.00029	0.01	0.0021	0.18584	0.191	0.19685	0.00212	0.01	0.4062

5×5 grid:

	Machochwe - $m = 11$			Mesaga - $m = 10$			Morotonga - $m = 12$			Ring'wani - $m = 10$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	1.08×10^{-10}	0.01	0.001	2.4×10^{-7}	0.01	0.001	3×10^{-7}	0.01	0.001	0.00723	0.01	0.001
Average	1.94×10^{-12}	0.01	0.001	2.4×10^{-8}	0.01	0.001	3×10^{-9}	0.01	0.001	8.6×10^{-5}	0.01	0.001
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Std deviation	1.28×10^{-11}	0	0	2.4×10^{-8}	0	0	3×10^{-8}	0	0	0.00073	0	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	9.24×10^{-12}	0.01	0.001	7.7×10^{-13}	0.01	0.001	5×10^{-10}	0.01	0.001	0.00024	0.01	0.001

7×7 grid:

	Machochwe - $m = 21$			Mesaga - $m = 16$			Morotonga - $m = 22$			Ring'wani - $m = 16$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	1.5×10^{-14}	0.01	0.001	2.1×10^{-8}	0.01	0.001	3×10^{-7}	0.01	0.001	0.00263	0.01	0.001
Average	1.5×10^{-16}	0.01	0.001	2.4×10^{-10}	0.01	0.001	3×10^{-9}	0.01	0.001	3×10^{-5}	0.01	0.001
Median	0	0.01	0.001	0	0.01	0.001	1.1×10^{-16}	0.01	0.001	8.4×10^{-15}	0.01	0.001
Std deviation	1.5×10^{-15}	0	0	2.1×10^{-9}	0	0	3×10^{-8}	0	0	0.00027	0	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0	0.01	0.001	7.1×10^{-10}	0.01	0.001	6.1×10^{-10}	0.01	0.001	9.6×10^{-6}	0.01	0.001

Table 3.16: Summary of the asymptotic and bootstrap p-values obtained for $\chi^2_{Pearson}$ test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a rectangular window.

3 × 3 grid:

	Machochwe - $m = 4$			Mesaga - $m = 4$			Morotonga - $m = 4$			Ring'wani - $m = 4$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.08	0.091	0	0.01	0.016	0	0.01	0.001	0	0.01	0.001
Maximum	0.235564	0.96	0.822	0.146441	0.99	0.989	0.051503	0.09	0.175	0.664397	1	0.988
Average	0.00476	0.444646	0.446202	0.012377	0.398081	0.433141	0.001739	0.012626	0.007879	0.020982	0.312929	0.357081
Median	0	0.43	0.441	0	0.17	0.166	3.4×10^{-9}	0.01	0.001	0	0.01	0.554
Std deviation	0.03331	0.167754	0.158406	0.035311	0.356109	0.369686	0.006655	0.010161	0.026754	0.114371	0.345833	0.348588
$P_{2.5}$	0	0.1315	0.1218	0	0.02	0.023	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0.0000482	0.7655	0.7263	0.141	0.97	0.9698	0.0142	0.0355	0.1088	0.378	0.944	0.891

5 × 5 grid:

	Machochwe - $m = 11$			Mesaga - $m = 10$			Morotonga - $m = 12$			Ring'wani - $m = 10$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.001	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	8.05×10^{-8}	0.01	0.001	0.040453	0.04	0.003	0.0000157	0.02	0.01	0.003193	0.02	0.035
Average	1.38×10^{-9}	0.01	0.001	0.000618	0.010505	0.001061	3.48×10^{-7}	0.010202	0.001343	0.000105	0.010202	0.001657
Median	0	0.01	0.001	4.3×10^{-14}	0.01	0.001	8.21×10^{-13}	0.01	0.001	2.22×10^{-16}	0.01	0.001
Std deviation	8.56×10^{-9}	0	0	0.004146	0.003607	0.000345	1.81×10^{-6}	0.001414	0.001334	0.000479	0.001414	0.003634
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	8.12×10^{-9}	0.01	0.001	0.00456	0.01	0.0021	3.36×10^{-6}	0.01	0.0046	0.00193	0.01	0.0072

7 × 7 grid:

	Machochwe - $m = 21$			Mesaga - $m = 16$			Morotonga - $m = 22$			Ring'wani - $m = 16$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	3.52×10^{-13}	0.01	0.001
Maximum	2.47×10^{-8}	0.01	0.001	0.000228	0.01	0.002	0.0000438	0.01	0.002	0.006522	0.04	0.047
Average	3.13×10^{-10}	0.01	0.001	2.93×10^{-6}	0.01	0.00101	4.8×10^{-7}	0.01	0.00101	0.00012	0.010303	0.002636
Median	2.52×10^{-9}	0.01	0.001	7.83×10^{-14}	0.01	0.001	9.49×10^{-13}	0.01	0.001	9.47×10^{-9}	0.01	0.001
Std deviation	0	0	0	0.0000233	0	0.000101	0.0000044	0	0.000101	0.000794	0.003015	0.007577
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.21×10^{-12}	0.01	0.001
$P_{97.5}$	1.66×10^{-9}	0.01	0.001	0.0000109	0.01	0.001	8.58×10^{-7}	0.01	0.001	0.000311	0.01	0.0317

Table 3.17: Summary of the asymptotic and bootstrap p-values obtained for $\chi^2_{Pearson}$ test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a convex hull window.

3×3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	304.948	98.958	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	4	4	152.008	13.719	0	0.00331	0.01	0.02	0.001	0.003	0.0001	0.0032
Morotonga	5	4	23.803	85.082	8.75×10^{-5}	0	0.01	0.01	0.001	0.001	0.0002	0.0001
Ring'wani	4	3	34.650	21.020	1.44×10^{-7}	2.73×10^{-5}	0.01	0.01	0.001	0.001	0.0001	0.0003

5×5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	113.799	304.553	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	149.787	124.884	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	13	11	187.92	82.204	0	1.85×10^{-13}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	126.675	80.918	0	1.06×10^{-13}	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	22	20	377.157	201.928	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	104.487	162.716	1.30×10^{-14}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	21	22	111.464	150.893	1.07×10^{-14}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	77.624	91.309	1.17×10^{-11}	3.53×10^{-12}	0.01	0.01	0.001	0.001	0.0001	0.0001

Table 3.18: LR test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	271.422	134.987	0	0	0.88	0.01	0.905	0.001	0.9064	0.0001
Mesaga	4	4	261.946	22.790	0	4.47×10^{-5}	0.13	0.95	0.156	0.924	0.1659	0.9285
Morotonga	5	4	9.677	86.139	0.04623	0	0.06	0.01	0.079	0.001	0.0693	0.0001
Ring'wani	4	3	520.825	51.649	0	6.09×10^{-12}	0.01	0.53	0.001	0.497	0.0005	0.5153

5×5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	172.902	102.006	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	19.716	141.593	0.01146	0	0.03	0.01	0.026	0.001	0.0243	0.0001
Morotonga	13	11	114.442	80.354	0	4.28×10^{-13}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	58.764	104.898	8.14×10^{-10}	0	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	21	21	139.300	244.134	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	104.968	97.316	1.05×10^{-14}	5.44×10^{-15}	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	22	21	111.663	95.864	2.36×10^{-14}	6.87×10^{-12}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	61.085	63.209	1.43×10^{-8}	3.08×10^{-7}	0.01	0.01	0.001	0.001	0.0001	0.0001

Table 3.19: *LR* test statistics and *p*-values obtained using a convex hull window.

3×3 grid:

	Machochwe - $m = 4$			Mesaga - $m = 4$			Morotonga - $m = 4$			Ring'wani - $m = 4$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	2.3×10^{-9}	0.01	0.001
Maximum	0.01004	0.01	0.001	0.014	0.01	0.017	0.29322	0.33	0.284	0.39217	0.36	0.4
Average	0.0001	0.01	0.001	0.00015	0.01	0.00132	0.00907	0.02	0.01052	0.00404	0.01354	0.02454
Median	0	0.01	0.001	4.6×10^{-14}	0.01	0.001	5.9×10^{-11}	0.01	0.001	4.3×10^{-6}	0.01	0.001
Std deviation	0.00101	0	0	0.00141	0	0.00213	0.045	0.04959	0.04316	0.03941	0.03318	0.09274
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	8.9×10^{-9}	0.01	0.001
$P_{97.5}$	1.3×10^{-8}	0.01	0.001	0.00023	0.01	0.0021	0.19861	0.212	0.21855	0.00127	0.01	0.38965

5×5 grid:

	Machochwe - $m = 11$			Mesaga - $m = 10$			Morotonga - $m = 12$			Ring'wani - $m = 10$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	1.3×10^{-9}	0.01	0.001	1.7×10^{-8}	0.01	0.001	2×10^{-9}	0.01	0.001	0.01073	0.05	0.001
Average	1.4×10^{-11}	0.01	0.001	1.7×10^{-10}	0.01	0.001	2.1×10^{-11}	0.01	0.001	0.00011	0.0104	0.001
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Std deviation	1.3×10^{-10}	0	0	1.7×10^{-9}	0	0	2×10^{-10}	0	0	0.00108	0.00402	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	3.6×10^{-12}	0.01	0.001	9.3×10^{-13}	0.01	0.001	5.6×10^{-12}	0.01	0.001	0.000042	0.01	0.001

7×7 grid:

	Machochwe - $m = 21$			Mesaga - $m = 16$			Morotonga - $m = 22$			Ring'wani - $m = 16$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	2.6×10^{-15}	0.01	0.001	3.1×10^{-8}	0.01	0.001	7.6×10^{-9}	0.01	0.001	0.00216	0.02	0.001
Average	2.6×10^{-17}	0.01	0.001	3.3×10^{-10}	0.01	0.001	7.9×10^{-11}	0.01	0.001	0.000039	0.0101	0.001
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	2.9×10^{-12}	0.01	0.001
Std deviation	2.6×10^{-16}	0	0	3.1×10^{-9}	0	0	7.6×10^{-10}	0	0	0.00025	0.00101	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0	0.01	0.001	9.6×10^{-11}	0.01	0.001	2.4×10^{-11}	0.01	0.001	0.0003	0.01	0.001

Table 3.20: Summary of the asymptotic and bootstrap p-values obtained for LR test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a rectangular window.

3 × 3 grid:

	Machochwe - m = 4			Mesaga - m = 4			Morotonga - m = 4			Ring'wani - m = 4		
	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999
Minimum	0	0.07	0.038	0	0.02	0.042	0	0.01	0.001	0	0.01	0.001
Maximum	0.229904	0.96	0.814	0.142102	0.99	0.989	0.047075	0.07	0.173	0.65949	1	0.988
Average	0.004645	0.408182	0.417909	0.010672	0.395455	0.43	0.001779	0.012727	0.008071	0.020996	0.31303	0.356879
Median	0	0.37	0.392	0	0.16	0.161	6.23×10^{-9}	0.01	0.001	0	0.01	0.554
Std deviation	0.03251	0.198311	0.200491	0.032029	0.343919	0.35847	0.006452	0.009127	0.027258	0.113518	0.346426	0.348588
$P_{2.5}$	0	0.14	0.1118	0	0.0345	0.04645	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0.0000288	0.77	0.724	0.129	0.97	0.971	0.0192	0.04	0.111	0.378	0.949	0.891

5 × 5 grid:

	Machochwe - m = 11			Mesaga - m = 10			Morotonga - m = 12			Ring'wani - m = 10		
	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	1.83×10^{-7}	0.01	0.001	0.012234	0.03	0.003	5.46×10^{-6}	0.01	0.009	0.000699	0.01	0.02
Average	3.35×10^{-9}	0.01	0.001	0.000295	0.010303	0.001051	7.78×10^{-8}	0.01	0.001121	0.0000207	0.01	0.001313
Median	0	0.01	0.001	6.93×10^{-12}	0.01	0.001	4.44×10^{-16}	0.01	0.001	8.16×10^{-14}	0.01	0.001
Std deviation	2.12×10^{-8}	0	0	0.001477	0.002238	0.000299	5.65×10^{-7}	0	0.000836	0.000102	0	0.002013
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	2.05×10^{-8}	0.01	0.001	0.00375	0.01	0.00155	2.93×10^{-7}	0.01	0.002	0.000279	0.01	0.003

7 × 7 grid:

	Machochwe - m = 21			Mesaga - m = 16			Morotonga - m = 22			Ring'wani - m = 16		
	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.41×10^{-13}	0.01	0.001
Maximum	1.62×10^{-7}	0.01	0.001	0.0000734	0.01	0.001	0.000023	0.01	0.002	0.007095	0.03	0.068
Average	2.75×10^{-9}	0.01	0.001	1.38×10^{-6}	0.01	0.001	2.58×10^{-7}	0.01	0.00101	0.000183	0.010303	0.002909
Median	1.82×10^{-8}	0.01	0.001	1.67×10^{-14}	0.01	0.001	7.13×10^{-14}	0.01	0.001	3.58×10^{-8}	0.01	0.001
Std deviation	0	0	0	8.45×10^{-6}	0	0	2.31×10^{-6}	0	0.000101	0.000987	0.002238	0.008986
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	6.62×10^{-12}	0.01	0.001
$P_{97.5}$	1.83×10^{-8}	0.01	0.001	0.0000151	0.01	0.001	2.01×10^{-7}	0.01	0.001	0.00174	0.01	0.0267

Table 3.21: Summary of the asymptotic and bootstrap p-values obtained for *LR* test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a convex hull window.

3×3 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	760.533	125.129	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	4	4	246.601	14.064	0	0.00282	0.01	0.02	0.001	0.007	0.0001	0.0046
Morotonga	5	4	246.601	144.964	0.00017	0	0.01	0.01	0.001	0.001	0.0007	0.0001
Ring'wani	4	3	45.913	20.377	5.92×10^{-10}	3.76×10^{-5}	0.01	0.02	0.001	0.001	0.0001	0.0015

5×5 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	171.650	542.905	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	343.692	528.274	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	13	11	826.150	480.694	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	257.258	265.432	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	22	20	937.965	551.884	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	278.576	305.868	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	21	22	363.362	398.585	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	126.140	152.903	0	0	0.01	0.01	0.002	0.001	0.0009	0.0001

Table 3.22: SC test statistics and p-values obtained using a rectangular window.

3 × 3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	586.978	142.651	0	0	0.51	0.01	0.508	0.001	0.4919	0.0001
Mesaga	4	4	45.038	29.214	9.08×10^{-10}	2.02×10^{-6}	0.03	0.85	0.030	0.815	0.0351	0.8081
Morotonga	5	4	11.146	148.941	0.02497	0	0.06	0.01	0.065	0.001	0.0534	0.0001
Ring'wani	4	3	102.736	45.250	0	1.49×10^{-10}	0.05	0.76	0.079	0.724	0.0882	0.7467

5 × 5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	247.332	147.599	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	35.019	215.436	2.65×10^{-5}	0	0.03	0.01	0.010	0.001	0.0105	0.0001
Morotonga	13	11	698.457	391.753	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	184.639	353.178	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001

7 × 7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	21	21	242.106	412.974	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	316.056	141.718	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	22	21	229.687	284.008	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	104.624	96.307	1.11×10^{-16}	4.28×10^{-13}	0.01	0.01	0.002	0.003	0.0006	0.0017

Table 3.23: *SC* test statistics and *p*-values obtained using a convex hull window.

3 × 3 grid:

	Machochwe - $m = 4$			Mesaga - $m = 4$			Morotonga - $m = 4$			Ring'wani - $m = 4$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.5×10^{-13}	0.01	0.001
Maximum	0.00481	0.01	0.001	0.00976	0.01	0.015	0.29883	0.36	0.3	0.3748	0.34	0.362
Average	0.000049	0.01	0.001	0.0001	0.01	0.00129	0.00967	0.02111	0.01176	0.00348	0.01384	0.02256
Median	0	0.01	0.001	1.1×10^{-16}	0.01	0.001	5.1×10^{-13}	0.01	0.001	3.6×10^{-7}	0.01	0.001
Std deviation	0.00048	0	0	0.00098	0	0.00186	0.04771	0.05449	0.05269	0.03391	0.03318	0.0831
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.5×10^{-13}	0.01	0.001
$P_{97.5}$	2.2×10^{-9}	0.01	0.001	0.000036	0.01	0.00265	0.2063	0.243	0.2491	0.00162	0.0155	0.34975

5 × 5 grid:

	Machochwe - $m = 11$			Mesaga - $m = 10$			Morotonga - $m = 12$			Ring'wani - $m = 10$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	6.7×10^{-10}	0.01	0.001	8.2×10^{-13}	0.01	0.001	0	0.01	0.001	0.00847	0.09	0.008
Average	6.8×10^{-12}	0.01	0.001	8.3×10^{-15}	0.01	0.001	0	0.01	0.001	0.000086	0.01091	0.00115
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Std deviation	6.8×10^{-11}	0	0	8.3×10^{-14}	0	0	0	0	0	0.00085	0.00144	0.00079
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	6.1×10^{-17}	0.01	0.001	0	0.01	0.001	0	0.01	0.001	5.3×10^{-13}	0.01	0.002

7 × 7 grid:

	Machochwe - $m = 21$			Mesaga - $m = 16$			Morotonga - $m = 22$			Ring'wani - $m = 16$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	0	0.01	0.001	5.4×10^{-15}	0.01	0.002	0	0.01	0.002	0.000013	0.05	0.011
Average	0	0.01	0.001	6.6×10^{-17}	0.01	0.00101	0	0.01	0.00101	1.6×10^{-7}	0.01182	0.00132
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Std deviation	0	0	0	5.6×10^{-16}	0	0.0001	0	0	0.0001	1.3×10^{-6}	0.00645	0.00138
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	6.6×10^{-7}	0.0355	0.004

Table 3.24: Summary of the asymptotic and bootstrap p-values obtained for SC test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a rectangular window.

3×3 grid:

	Machochwe - $m = 4$			Mesaga - $m = 4$			Morotonga - $m = 4$			Ring'wani - $m = 4$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.057	0	0.01	0.008	0	0.01	0.001	0	0.01	0.001
Maximum	0.213347	0.96	0.804	0.12905	0.99	0.988	0.031591	0.07	0.176	0.641824	0.99	0.988
Average	0.00431	0.394747	0.431667	0.007522	0.392929	0.431131	0.001629	0.012525	0.008071	0.020486	0.214444	0.247838
Median	0	0.37	0.435	2.86×10^{-8}	0.3	0.407	8.68×10^{-11}	0.01	0.001	0	0.05	0.166
Std deviation	0.030169	0.195727	0.169582	0.026233	0.303531	0.322995	0.006027	0.00962	0.026745	0.110489	0.283204	0.301118
$P_{2.5}$	0	0.059	0.0888	6.99×10^{-16}	0.02	0.012	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	3.15×10^{-6}	0.761	0.724	0.12	0.96	0.97	0.028	0.0455	0.0981	0.371	0.954	0.892

5×5 grid:

	Machochwe - $m = 11$			Mesaga - $m = 10$			Morotonga - $m = 12$			Ring'wani - $m = 10$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	2.73×10^{-8}	0.01	0.003	0.007846	0.02	0.01	3.45×10^{-7}	0.03	0.027	3.57×10^{-9}	0.01	0.004
Average	2.78×10^{-10}	0.01	0.00103	0.000102	0.010505	0.001253	3.5×10^{-9}	0.010202	0.001556	3.86×10^{-11}	0.01	0.001101
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Std deviation	2.75×10^{-9}	0	0.000224	0.0008	0.002201	0.001181	3.47×10^{-8}	0.00201	0.002775	3.6×10^{-10}	0	0.000463
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	4.6×10^{-11}	0.01	0.001	0.000568	0.02	0.0041	9.01×10^{-11}	0.01	0.005	5.3×10^{-14}	0.01	0.003

7×7 grid:

	Machochwe - $m = 21$			Mesaga - $m = 16$			Morotonga - $m = 22$			Ring'wani - $m = 16$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	5.92×10^{-9}	0.02	0.003	1.75×10^{-6}	0.02	0.01	2.19×10^{-8}	0.02	0.004	0.000378	0.05	0.128
Average	8.46×10^{-11}	0.010101	0.001061	3.11×10^{-8}	0.010202	0.001182	2.22×10^{-10}	0.010101	0.001091	4.7×10^{-6}	0.013636	0.007556
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	9.99×10^{-16}	0.01	0.001
Std deviation	6.12×10^{-10}	0.001005	0.000314	2.17×10^{-7}	0.001414	0.000973	2.21×10^{-9}	0.001005	0.00038	0.0000381	0.008138	0.01823
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	6.62×10^{-10}	0.01	0.002	2.82×10^{-8}	0.01	0.0002	3.57×10^{-14}	0.01	0.002	0.0000215	0.04	0.0614

Table 3.25: Summary of the asymptotic and bootstrap p-values obtained for SC test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a convex hull window.

3×3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	324.912	98.601	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	4	4	144.567	13.868	0	0.00313	0.01	0.02	0.001	0.003	0.0001	0.0036
Morotonga	5	4	25.971	77.960	3.21×10^{-5}	1.11×10^{-16}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	4	3	33.131	22.847	3.12×10^{-7}	1.16×10^{-5}	0.01	0.01	0.001	0.001	0.0001	0.0003

5×5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	106.330	398.941	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	144.430	120.915	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	13	11	161.268	68.279	0	7.37×10^{-11}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	170.060	79.116	0	1.54×10^{-13}	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	22	20	541.901	203.006	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	96.049	166.426	2.08×10^{-13}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	21	22	95.949	143.833	2.39×10^{-12}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	79.310	105.635	2.45×10^{-12}	1.55×10^{-15}	0.01	0.01	0.001	0.001	0.0001	0.0001

Table 3.26: VT test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	346.091	154.534	0	0	0.98	0.01	0.993	0.001	0.9907	0.0001
Mesaga	4	4	2457.568	20.828	0	0.00012	0.16	0.95	0.203	0.945	0.2217	0.9567
Morotonga	5	4	9.287	78.430	0.05394	1.11×10^{-16}	0.07	0.01	0.086	0.001	0.0798	0.0001
Ring'wani	4	3	8309.308	63.366	0	1.85×10^{-14}	0.01	0.33	0.002	0.348	0.0006	0.3527

5×5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	193.302	98.860	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	17.207	163.904	0.02468	0	0.04	0.01	0.041	0.001	0.0386	0.0001
Morotonga	13	11	91.974	70.021	7.99×10^{-15}	3.32×10^{-11}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	54.427	121.813	3.75×10^{-9}	0	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	21	21	129.483	306.990	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	93.887	99.292	3.36×10^{-13}	1.33×10^{-15}	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	22	21	102.112	84.261	2.01×10^{-13}	2.46×10^{-10}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	55.487	66.423	3.77×10^{-8}	1.50×10^{-8}	0.01	0.01	0.001	0.001	0.0002	0.0001

Table 3.27: VT test statistics and p-values obtained using a convex hull window.

3 × 3 grid:

	Machochwe - m = 4			Mesaga - m = 4			Morotonga - m = 4			Ring'wani - m = 4		
	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	5.1×10^{-9}	0.01	0.001
Maximum	0.01214	0.01	0.001	0.01475	0.01	0.0157	0.28465	0.31	0.273	0.41238	0.37	0.415
Average	0.00012	0.01	0.001	0.00016	0.01	0.00133	0.00854	0.01939	0.00971	0.00429	0.01384	0.02541
Median	0	0.01	0.001	1.8×10^{-14}	0.01	0.001	3.1×10^{-12}	0.01	0.001	3.3×10^{-6}	0.01	0.001
Std deviation	0.00122	0	0	0.00148	0	0.0022	0.04262	0.04673	0.04316	0.04144	0.03619	0.09634
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	5.1×10^{-9}	0.01	0.001
$P_{97.5}$	4.7×10^{-8}	0.01	0.001	0.00028	0.01	0.0021	0.18541	0.191	0.19475	0.00208	0.0155	0.4031

5 × 5 grid:

	Machochwe - m = 11			Mesaga - m = 10			Morotonga - m = 12			Ring'wani - m = 10		
	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	1.2×10^{-10}	0.01	0.001	2.6×10^{-7}	0.01	0.001	3.1×10^{-7}	0.01	0.001	0.00725	0.01	0.001
Average	2.2×10^{-12}	0.01	0.001	2.6×10^{-9}	0.01	0.001	3.2×10^{-9}	0.01	0.001	0.000087	0.01	0.001
Median	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Std deviation	1.4×10^{-11}	0	0	2.6×10^{-8}	0	0	3.2×10^{-8}	0	0	0.00073	0	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	1×10^{-11}	0.01	0.001	8.5×10^{-13}	0.01	0.001	5.6×10^{-10}	0.01	0.001	0.00025	0.01	0.001

7 × 7 grid:

	Machochwe - m = 21			Mesaga - m = 16			Morotonga - m = 22			Ring'wani - m = 16		
	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999	Asymptotic	R = 99	R = 999
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	2.9×10^{-14}	0.01	0.001	2.6×10^{-8}	0.01	0.001	3.6×10^{-7}	0.01	0.001	0.00276	0.01	0.001
Average	2.9×10^{-16}	0.01	0.001	3×10^{-10}	0.01	0.001	3.6×10^{-9}	0.01	0.001	0.000032	0.01	0.001
Median	0	0.01	0.001	0	0.01	0.001	2.2×10^{-16}	0.01	0.001	2×10^{-14}	0.01	0.001
Std deviation	2.9×10^{-15}	0	0	2.6×10^{-9}	0	0	3.6×10^{-8}	0	0	0.00028	0	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0	0.01	0.001	8.9×10^{-10}	0.01	0.001	8.2×10^{-10}	0.01	0.001	0.000012	0.01	0.001

Table 3.28: Summary of the asymptotic and bootstrap p-values obtained for VT test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a rectangular window.

3 × 3 grid:

	Machochwe - $m = 4$			Mesaga - $m = 4$			Morotonga - $m = 4$			Ring'wani - $m = 4$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.08	0.094	0	0.05	0.066	0	0.01	0.001	0	0.01	0.001
Maximum	0.241597	0.96	0.831	0.140931	0.99	0.988	0.048558	0.09	0.181	0.663244	1	0.988
Average	0.004882	0.445354	0.446273	0.01161	0.398687	0.43397	0.001754	0.012828	0.008121	0.02086	0.312525	0.356606
Median	0	0.43	0.441	0	0.14	0.119	3.19×10^{-9}	0.01	0.001	0	0.01	0.554
Std deviation	0.034163	0.167705	0.158038	0.033504	0.351593	0.364816	0.006523	0.010502	0.027749	0.114179	0.345565	0.348074
$P_{2.5}$	0	0.137	0.12465	0	0.05	0.07145	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0.0000575	0.761	0.726	0.135	0.97	0.967	0.015	0.04	0.112	0.377	0.944	0.887

5 × 5 grid:

	Machochwe - $m = 11$			Mesaga - $m = 10$			Morotonga - $m = 12$			Ring'wani - $m = 10$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	2.37×10^{-7}	0.01	0.001	0.028136	0.03	0.003	0.0000259	0.02	0.011	0.005443	0.02	0.03
Average	3.66×10^{-9}	0.01	0.001	0.000434	0.010303	0.001061	4.68×10^{-7}	0.010202	0.001323	0.000132	0.010202	0.001727
Median	0	0.01	0.001	8.04×10^{-14}	0.01	0.001	1.02×10^{-12}	0.01	0.001	2.22×10^{-16}	0.01	0.001
Std deviation	2.43×10^{-8}	0	0	0.002885	0.002238	0.000345	2.76×10^{-6}	0.001414	0.001384	0.000658	0.001414	0.003434
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	2.15×10^{-8}	0.01	0.001	0.00327	0.01	0.0021	4.33×10^{-6}	0.01	0.00455	0.00192	0.01	0.0102

7 × 7 grid:

	Machochwe - $m = 21$			Mesaga - $m = 16$			Morotonga - $m = 22$			Ring'wani - $m = 16$		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	6.63×10^{-13}	0.01	0.001
Maximum	1.35×10^{-7}	0.01	0.001	0.000232	0.01	0.002	0.0000374	0.01	0.002	0.006409	0.04	0.116
Average	2.03×10^{-9}	0.01	0.001	3.3×10^{-6}	0.01	0.00101	4.17×10^{-7}	0.01	0.00101	0.000124	0.010303	0.004566
Median	1.43×10^{-8}	0.01	0.001	1.14×10^{-13}	0.01	0.001	1.43×10^{-12}	0.01	0.001	1.43×10^{-8}	0.01	0.001
Std deviation	1.43×10^{-8}	0	0	0.000024	0	0.000101	3.76×10^{-6}	0	0.000101	0.000806	0.003015	0.017567
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	2.28×10^{-12}	0.01	0.001
$P_{97.5}$	1.4×10^{-8}	0.01	0.001	0.0000216	0.01	0.001	9.03×10^{-7}	0.01	0.001	0.000405	0.01	0.0602

Table 3.29: Summary of the asymptotic and bootstrap p-values obtained for VT test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a convex hull window.

3×3 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	271.571	92.429	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	4	4	129.558	13.544	0	0.0084	0.01	0.02	0.001	0.003	0.0001	0.0036
Morotonga	5	4	24.898	73.574	0.00013	3.77×10^{-15}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	4	3	30.959	21.335	2.74×10^{-6}	8.15×10^{-5}	0.01	0.01	0.001	0.001	0.0001	0.0003

5×5 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	100.435	321.424	1.11×10^{-16}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	130.464	108.396	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	13	11	143.507	64.424	0	9.06×10^{-10}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	134.301	69.513	0	2.76×10^{-11}	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		R=99		R=999		R=9999	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	22	20	427.731	182.750	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	89.502	148.512	6.72×10^{-12}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	21	22	90.115	130.697	4.93×10^{-11}	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	70.253	92.334	2.61×10^{-10}	1.06×10^{-12}	0.01	0.01	0.001	0.001	0.0001	0.0001

Table 3.30: UT test statistics and p-values obtained using a rectangular window.

3×3 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	5	4	125.490	117.495	0	0	0.44	0.01	0.438	0.001	0.4378	0.0001
Mesaga	4	4	9.491	21.960	0.01144	0.0001	0.03	0.86	0.053	0.846	0.0493	0.8399
Morotonga	5	4	8.346	74.596	0.08467	2.22×10^{-15}	0.07	0.01	0.074	0.001	0.0627	0.0001
Ring'wani	4	3	64.598	44.087	3.04×10^{-14}	6.48×10^{-10}	0.01	0.42	0.001	0.424	0.0002	0.4382

5×5 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	11	11	181.632	79.753	0	7.06×10^{-14}	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	9	10	11.822	90.834	0.08533	1.11×10^{-16}	0.05	0.01	0.065	0.001	0.0537	0.0001
Morotonga	13	11	79.362	57.982	6.25×10^{-13}	7.10×10^{-9}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	9	10	48.524	104.912	3.66×10^{-8}	0	0.01	0.01	0.001	0.001	0.0001	0.0001

7×7 grids:

Village	K		Test statistic		Asymptotic		$R=99$		$R=999$		$R=9999$	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Machochwe	21	21	119.015	277.882	0	0	0.01	0.01	0.001	0.001	0.0001	0.0001
Mesaga	18	14	82.335	95.101	6.75×10^{-12}	4.55×10^{-15}	0.01	0.01	0.001	0.001	0.0001	0.0001
Morotonga	22	21	95.819	75.946	7.56×10^{-13}	4.06×10^{-9}	0.01	0.01	0.001	0.001	0.0001	0.0001
Ring'wani	13	18	52.278	61.598	1.18×10^{-7}	3.92×10^{-8}	0.01	0.01	0.001	0.001	0.0001	0.0001

Table 3.31: UT test statistics and p-values obtained using a convex hull window.

3 × 3 grid:

	Machochwe - m = 4			Mesaga - m = 4			Morotonga - m = 4			Ring'wani - m = 4		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	9.4×10^{-8}	0.01	0.001
Maximum	0.0284	0.01	0.001	0.03457	0.01	0.017	0.43311	0.32	0.281	0.5796	0.37	0.438
Average	0.00029	0.01	0.001	0.00039	0.01	0.00133	0.01375	0.0196	0.01009	0.0622	0.01384	0.02642
Median	0	0.01	0.001	4×10^{-13}	0.01	0.001	9.8×10^{-11}	0.01	0.001	2.4×10^{-5}	0.01	0.001
Std deviation	0.00285	0	0	0.00348	0	0.0022	0.06787	0.04776	0.045	0.05806	0.03619	0.10034
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1×10^{-7}	0.01	0.001
$P_{97.5}$	3.8×10^{-7}	0.01	0.001	0.000997	0.01	0.0021	0.30557	0.1965	0.2042	0.00586	0.0155	0.41465

5 × 5 grid:

	Machochwe - m = 11			Mesaga - m = 10			Morotonga - m = 12			Ring'wani - m = 10		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	1.4×10^{-9}	0.01	0.001	1.9×10^{-6}	0.01	0.001	1.3×10^{-6}	0.01	0.001	0.01417	0.02	0.001
Average	2.5×10^{-11}	0.01	0.001	1.9×10^{-8}	0.01	0.001	1.3×10^{-8}	0.01	0.001	0.00019	0.0101	0.001
Median	0	0.01	0.001	0	0.01	0.001	1.1×10^{-16}	0.01	0.001	0	0.01	0.001
Std deviation	1.7×10^{-10}	0	0	1.9×10^{-7}	0	0	1.3×10^{-7}	0	0	0.00144	0.00101	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	8.7×10^{-11}	0.01	0.001	1.4×10^{-11}	0.01	0.001	4.4×10^{-9}	0.01	0.001	0.00093	0.01	0.001

7 × 7 grid:

	Machochwe - m = 21			Mesaga - m = 16			Morotonga - m = 22			Ring'wani - m = 16		
	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$	Asymptotic	$R = 99$	$R = 999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.1×10^{-16}	0.01	0.001
Maximum	1.9×10^{-12}	0.01	0.001	2.1×10^{-7}	0.01	0.001	1.5×10^{-6}	0.01	0.001	0.00595	0.02	0.001
Average	2×10^{-14}	0.01	0.001	2.5×10^{-9}	0.01	0.001	1.6×10^{-8}	0.01	0.001	0.000073	0.0101	0.001
Median	0	0.01	0.001	2.1×10^{-8}	0.01	0.001	1.6×10^{-14}	0.01	0.001	5.6×10^{-12}	0.01	0.001
Std deviation	2×10^{-13}	0	0	0	0	0	1.5×10^{-7}	0	0	0.00061	0.00101	0
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.5×10^{-15}	0.01	0.001
$P_{97.5}$	0	0.01	0.001	9×10^{-9}	0.01	0.001	6.6×10^{-9}	0.01	0.001	0.000069	0.01	0.001

Table 3.32: Summary of the asymptotic and bootstrap p-values obtained for UT test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a rectangular window.

3 × 3 grid:

	Machochwe - m = 4			Mesaga - m = 4			Morotonga - m = 4			Ring'wani - m = 4		
	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$
Minimum	0	0.01	0.001	1.23×10^{-8}	0.02	0.02	0	0.01	0.001	0	0.01	0.001
Maximum	0.230004	0.94	0.818	0.317694	0.97	0.986	0.081244	0.07	0.152	0.837393	1	0.988
Average	0.004648	0.296364	0.288838	0.042748	0.486768	0.509273	0.004745	0.013232	0.008576	0.026995	0.130707	0.152758
Median	0	0.2	0.209	0.002827	0.58	0.683	5.31×10^{-8}	0.01	0.001	4.08×10^{-13}	0.01	0.005
Std deviation	0.032524	0.238736	0.238708	0.078641	0.35139	0.353046	0.014989	0.010674	0.027657	0.144112	0.270239	0.304328
$P_{2.5}$	0	0.01	0.001	4.03×10^{-8}	0.03	0.0238	0	0.01	0.001	0	0.01	0.001
$P_{97.5}$	0.0000757	0.74	0.726	0.271	0.957	0.94	0.0535	0.051	0.113	0.48	0.931	0.856

5 × 5 grid:

	Machochwe - m = 11			Mesaga - m = 10			Morotonga - m = 12			Ring'wani - m = 10		
	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$
Minimum	0	0.01	0.001	3.22×10^{-15}	0.01	0.001	0	0.01	0.001	0	0.01	0.001
Maximum	0.0000377	0.01	0.016	0.172865	0.14	0.004	0.0000983	0.01	0.006	0.00651	0.01	0.032
Average	7.89×10^{-7}	0.01	0.001152	0.002375	0.011515	0.00104	4.4×10^{-6}	0.01	0.001101	0.000188	0.01	0.001646
Median	0	0.01	0.001	5.51×10^{-8}	0.01	0.001	2.05×10^{-10}	0.01	0.001	1.14×10^{-13}	0.01	0.001
Std deviation	4.48×10^{-6}	0	0.001508	0.017447	0.013121	0.000317	0.0000172	0	0.000562	0.000863	0	0.003382
$P_{2.5}$	0	0.01	0.001	7.41×10^{-14}	0.01	0.001	5×10^{-17}	0.01	0.001	0	0.01	0.001
$P_{97.5}$	4.23×10^{-6}	0.01	0.001	0.0108	0.0155	0.001	0.000069	0.01	0.002	0.00315	0.01	0.007

7 × 7 grid:

	Machochwe - m = 21			Mesaga - m = 16			Morotonga - m = 22			Ring'wani - m = 16		
	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$	Asymptotic	$R=99$	$R=999$
Minimum	0	0.01	0.001	0	0.01	0.001	0	0.01	0.001	1.5×10^{-11}	0.01	0.001
Maximum	1.69×10^{-6}	0.01	0.001	0.000415	0.01	0.001	0.0000564	0.01	0.001	0.009722	0.02	0.054
Average	2.07×10^{-8}	0.01	0.001	8.54×10^{-6}	0.01	0.001	6.68×10^{-7}	0.01	0.001	0.000145	0.010101	0.002727
Median	1.71×10^{-7}	0.01	0.001	9.82×10^{-13}	0.01	0.001	1.55×10^{-11}	0.01	0.001	7.17×10^{-8}	0.01	0.001
Std deviation	0	0	0	0.0000524	0	0	5.68×10^{-6}	0	0	0.001038	0.001005	0.008405
$P_{2.5}$	0	0.01	0.001	0	0.01	0.001	1.69×10^{-15}	0.01	0.001	4.22×10^{-11}	0.01	0.001
$P_{97.5}$	8.74×10^{-8}	0.01	0.001	0.0000587	0.01	0.001	1.88×10^{-6}	0.01	0.001	0.000328	0.01	0.0307

Table 3.33: Summary of the asymptotic and bootstrap p-values obtained for UT test calculated based on m quadrats obtained through bootstrap sampling of approximately 50% of the non-empty quadrats using a convex hull window.

3.5 Conclusion

We started this chapter with a basic descriptive study of our dataset. The households locations in 4 randomly selected remote villages in Northern Tanzania were described as point patterns.

Second-order homogeneity was investigated using both the homogeneous and inhomogeneous F, G, J, K and L functions and the results were compared for rectangular and convex hull windows.

A current approach to checking for first-order homogeneity in point patterns is to construct a density plot of intensities using kernel density estimation. This method was applied to each of the villages with plots of the estimates indicating varying intensity across the patterns. For completeness we also showed the impact of including edge corrections in the estimation of the density.

This initial visual check was confirmed statistically significant by Pearson's χ^2, likelihood ratio, score, VT and UT tests, using different grid sizes on both rectangular and convex hull windows. When applied to rectangular windows, the tests identified the first-order heterogeneity for all grid sizes. This is however likely because some of the non-empty quadrats included areas which are not part of the village and therefore do not contain any points. When convex hull windows are applied to the patterns, many of these open areas were excluded from the observation window. In some cases the tests had trouble identifying the first-order heterogeneity when using a too small grid (3×3), but as the grid size was increased (5×5 and 7×7) the tests gave consistent results. In these specific cases for a small grid size, the asymptotic tests were not well-behaved, possibly due to the small number of quadrats as well as the violation of the independence assumption between quadrats.

Two possible solutions for the independence violation were investigated, namely calculating the test statistic using regular-spaced non-adjacent quadrats or by randomly sampling a subset of the quadrats for inclusion in the calculation of the test statistic. For the first approach, two cases were compared, firstly by including every non-empty odd-numbered quadrat and secondly by including every non-empty even-numbered quadrat. Since the areas included in the calculations are different for the two scenarios, the test statistics should not be compared directly. The asymptotic tests were also not well-behaved, but a comparison could be made using bootstrap p-values. It was shown that, when using regular-spaced non-adjacent quadrats, the choice of starting quadrat can have a significant effect on the outcome of the test when using a small grid size and a convex hull window. Here, contradictory conclusions were obtained for the two scenarios. This

shows that the choice of which quadrats to include in the calculation of the test statistic can have a major impact on the outcomes of the tests.

The second solution involved randomly selecting a subset of quadrats to include in the calculation of the test statistic, after which bootstrap p-values were obtained as before. This process was repeated 99 times and the resulting p-values were used to set up empirical 95% confidence intervals. For small grid sizes, these intervals tend to be wide, including both significant and insignificant p-values and no conclusion can be reached. When the grid size was increased, p-values stabilised and the empirical confidence intervals became more accurate. The only exception occurred for Ring'wani, a pattern that contains significantly fewer points than the other patterns.

We conclude that the first option of dealing with the independence violation, namely including non-adjacent quadrats in the calculation of the test statistics, although computationally simple is not ideal as the results for this are too dependent on the choice of quadrats to include. The second option, namely obtaining bootstrap samples of quadrats for inclusion in the test statistic calculations, provides us with a better measure of the first-order homogeneity within the pattern. These tests, although more computationally intensive, provides us with an indication of how p-values differ when different quadrats are included in the calculation of the test statistics. When p-values are highly variable, these empirical confidence intervals could also give us an indication of how appropriate the choice of grid size is. In our application, we showed how a small grid size can negatively influence the results of homogeneity tests and that larger grid sizes could potentially be more informative. However, the number of points within the pattern should also be taken into account as a too large grid could also influence the tests negative, as shown for Ring'wani. This is in line with the statement made by Baddeley et al. [10] that grid sizes should not be too small or too large.

Chapter 4

Final thoughts

4.1 Discussion of our contribution

In this mini-dissertation we developed three new hypothesis tests for stationarity in the context of spatial point patterns and compared their results for tests currently available in the `spatstat` package in R. These tests however assume that counts included in the calculation of the test statistics are independent, an assumption which is unlikely to be valid for spatial data. Two possible solutions for the independence violation were investigated, namely calculating the test statistic using regular-spaced non-adjacent quadrats or by randomly sampling a subset of the quadrats for inclusion in the calculation of the test statistic. For the first solution, two cases were compared, firstly by including all non-empty odd-numbered quadrats and secondly by including all non-empty even-numbered quadrats. A comparison was made using parametric bootstrap p-values. It was shown that, when using regular-spaced non-adjacent quadrats, the choice of which quadrats to include can have a significant effect on the outcome of the tests with contradictory conclusions obtained for the two scenarios. This shows that the choice of which quadrats to include in the calculation of the test statistic can have a major impact on the outcomes of the tests.

The second solution involved randomly selecting a subset of quadrats to include in the calculation of the test statistics, after which parametric bootstrap p-values were obtained as before. This process was repeated 99 times and the resulting p-values were used to set up empirical 95% confidence intervals. For small grid sizes, these intervals tend to be wide, including both significant and insignificant p-values and no conclusion can be reached. When the grid size was increased, p-values stabilised and the empirical confidence intervals became more accurate. The only exception occurred for Ring'wani, a pattern that contains significantly fewer points than the other patterns.

121

We conclude that the first option of dealing with the independence violation, namely including non-adjacent quadrats in the calculation of the test statistics, although computationally simple is not ideal as the results for this are too dependent on the choice of quadrats to include. The second option, namely obtaining random samples of quadrats for inclusion in the test statistic calculations, provides us with a better measure of the stationarity within the pattern.

The χ^2 and likelihood ratio tests can be performed using the `quadrat.test` function available in the `spatstat` package in `R`. This built-in function should however be used with caution. During the application phase of this article we discovered several irregularities in the test statistics obtained. Although the built-in functions correctly calculated the test statistic when using a rectangular window with no empty quadrats, discrepancies occurred if there were empty quadrats or when using a convex hull window. In the presence of empty quadrats, their areas are still included in the calculation of the statistic and no option is available to exclude these. We were unable to determine the exact source of the discrepancy in the test statistic calculations when using a convex hull window. In addition, small areas where the expected number of points is less than 5 are not excluded from the calculation of the test statistic. Secondly, by default the function performs a two-sided hypothesis test. In this context we consider only one-sided tests, specifically upper-tailed tests, as a large absolute difference between observed and expected counts is an indication of a deviation from the null. Thirdly, when specifying in the function that a one-sided test must be performed, the p-value given in the output is for a lower-tailed test and needs to be adjusted as we are performing an upper-tailed test. Due to these concerns, the formulae for all tests and p-values in this article were hard-coded. Finally, the calculation of the test statistics and p-values based on random samples of quadrats is not currently available in the `spatstat` package.

4.2   Conclusion

In conclusion, in this mini-dissertation we have

- discussed briefly second-order homogeneity in order to indicate to the reader the impact of not evaluating first-order homogeneity correctly,

- discussed two first-order homogeneity hypothesis tests currently available, but seldomly used, in point pattern analysis software, namely Pearson's χ^2 and the likelihood ratio tests,

- elaborated on the shortcomings and limitations of these tests available in point pattern analysis software,

- proposed three more first-order homogeneity tests, not currently employed in point pattern analysis, and

- proposed two methods of dealing with the violation of the first-order homogeneity tests.

In this mini-dissertation we briefly described the context and development of spatial statistics, spatial sampling and point patterns. Before selecting an appropriate sampling design in the spatial context, it is important to know whether the data is first- and second-order homogeneous. In our application we considered the household locations in four randomly selected remote villages as an unmarked point pattern. In future work we will be considering the villages as marked patterns, taking into account the locations of households as well as the presence of animals at each household location.

For completeness, second-order heterogeneity was briefly investigated using the homogeneous and inhomogeneous F, G, J, K and L functions on both rectangular and convex hull windows.

Currently the method of kernel density estimation is used to construct density plots which can be used to visually and subjectively infer on first-order homogeneity. This technique was applied to our point patterns and the impact of using different windows and including edge corrections was illustrated. We proposed the use of hypothesis tests, developed for the comparison of K Poisson intensities from independent samples, in the spatial setting. In our application, the K samples were obtained as non-empty quadrats within a grid, effectively comparing intensities across the entire point pattern, assuming independence between quadrats. Tobler's law of geography states that observations that are close together are more likely to be similar than observations that are further apart, which implies that the counts in adjacent quadrats may not be independent but counts in quadrats further apart may be. It was found that the asymptotic tests are not always well-behaved, possibly due to the violation of the independence assumption and/or the choice of grid size and window. Further investigation into the validity of asymptotic distributions is required.

Two possible solutions for dealing with the violation of the independence assumption were considered. The first involved calculating test statistics using non-adjacent quadrats, spaced at regular intervals, however the results from these tests were highly influenced by the choice of quadrats to include. Secondly, we considered calculating test statistics based on random samples of quadrats and setting up empirical confidence intervals for

the p-values. In addition, this approach gave us insight into the appropriateness of using different grid sizes.

The choice of window was shown to also have an impact on the accuracy of test results, both for first- and second-order homogeneity. The choice of an appropriate window also warrants further investigation. The interested reader is referred to [86, 90, 94] for guidelines on how to optimise window selection.

Future work

In future work we will focus on

- further investigating the validity and limitations of using asymptotic distributions for first-order homogeneity tests,

- exploring additional solutions to the violation of the independence assumption in first-order homogeneity tests,

- the optimal selection of grid size for these first-order homogeneity tests,

- optimal window selection for first- and second-order homogeneity tests, and

- criteria for choosing the most appropriate first-order homogeneity test, given a point pattern.

Bibliography

[1] Kolmogorov-Smirnov test. International Encyclopedia of Statistical Science. Pages 718 - 720. (Online article - Accessed 12 July 2017).

[2] Spatial join (analysis), ArcGIS Desktop Help 9.2. (Accessed 13 July 2017).

[3] B. Alijani, J. O'Brien, and B. Yarnal. Spatial analysis of precipitation intensity and concentration in Iran. *Theoretical and Applied Climatology*, 94(1-2):107–124, 2008.

[4] D. R. Anderson. The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin*, 29(4):1294–1297, 2001.

[5] L. Anselin. *Spatial Econometrics: Methods and Models*. NATO Asi Series. Series E, Applied Sciences. Springer, 1988.

[6] A. Baddeley. *Stochastic Geometry: Likelihood and Computation*, volume 80 of *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*, chapter 2, pages 37–78. CRC Press, 1998.

[7] A. Baddeley. Analysing spatial point patterns in R. Technical Report Version 4.1, CSIRO and University of Western Australia, December 2010.

[8] A. Baddeley, I. Bárány, R. Schneider, and W. Weil. *Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13-18, 2004*, chapter Spatial point processes and their applications, pages 1–75. Springer-Verlag Berlin Heidelberg, 2007.

[9] A. Baddeley and R. D. Gill. Kaplan-Meier estimators of distance distributions for spatial point processes. *Annals of Statistics*, 25(1):263–292, 1997.

[10] A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015.

[11] A. Baddeley and R. Turner. Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005.

[12] T. C. Bailey and A. C. Gatrell. *Interactive spatial data analysis*, volume 413. Longman Scientific & Technical Essex, 1995.

[13] A. K. Bera and Y. Bilias. Rao's score, Neyman's C(α) and Silvey's LM tests: an essay on historical developments and some new results. *Journal of Statistical Planning and Inference*, 97(1):9–44, 2001.

[14] M. Berman and P. Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 81–92, 1989.

[15] J. Besag and P. J. Diggle. Simple Monte Carlo tests for spatial pattern. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(3):327–333, 1977.

[16] D. J. Best. The difference between two Poisson expectations. *Australian Journal of Statistics*, 17(1):29–33, 1975.

[17] G. E. Blackman. A study by statistical methods of the distribution of species in grassland associations. *Annals of Botany*, 49(196):749–777, 1935.

[18] C. Brunsdon. Estimating probability surfaces in gis: an adaptive technique. In *Proceedings, European Conference on Geographical Information Systems*, 1991.

[19] C. Cambardella and D. Karlen. Spatial analysis of soil fertility parameters. *Precision Agriculture*, 1(1):5–14, 1999.

[20] S. N. Chiu. Parametric bootstrap and approximate tests for two Poisson variates. *Journal of Statistical Computation and Simulation*, 80(3):263–271, 2010.

[21] S. N. Chiu and L. Wang. Homogeneity tests for several Poisson populations. *Computational Statistics & Data Analysis*, 53(12):4266 – 4278, 2009.

[22] W. Cochran. *Sampling techniques*. Wiley, New York, 3rd edition, 1977.

[23] R. Condit. *Tropical forest census plots: methods and results from Barro Colorado Island, Panama and a comparison with other plots*. Springer Science & Business Media, 1998.

[24] R. Condit, S. P. Hubbell, and R. B. Foster. Changes in tree species abundance in a neotropical forest: impact of climate change. *Journal of tropical ecology*, 12(2):231–256, 1996.

[25] D. Cox and P. Lewis. *The statistical analysis of series of events*. John Wiley and Sons, 1966.

[26] D. R. Cox, E. Spjøtvoll, S. Johansen, W. R. van Zwet, J. F. Bithell, O. Barndorff-Nielsen, and M. Keuls. The role of significance tests [with discussion and reply]. *Scandinavian Journal of Statistics*, 4(2):49–70, 1977.

[27] N. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley, 2015.

[28] N. Cressie and C. Wikle. *Statistics for Spatio-Temporal Data*. CourseSmart Series. Wiley, 2011.

[29] F. H. C. Crick and P. A. Lawrence. Compartments and polyclones in insect development. *Science*, 189(4200):340–347, 1975.

[30] O. Cronie and M. van Lieshout. Bandwidth selection for kernel estimators of the spatial intensity function. *arXiv preprint arXiv:1611.10221*, 2016.

[31] R. Davidson and J. G. MacKinnon. Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School*, 66(1):1–26, 1998.

[32] R. Davidson and J. G. MacKinnon. Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*, 51(7):3259–3281, 2007.

[33] A. Davison and D. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[34] J. De Gruijter, D. J. Brus, M. F. Bierkens, and M. Knotters. *Sampling for natural resource monitoring*. Springer Science & Business Media, New York, 2006.

[35] K. Detre and C. White. 294. note: The comparison of two Poisson-distributed observations. *Biometrics*, 26(4):851–854, 1970.

[36] P. Diggle. *Statistical Analysis of Spatial Point Patterns*. Academic Press, 1983.

[37] P. Diggle. A kernel method for smoothing point process data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34(2):138–147, 1985.

[38] P. Diggle and J. S. Marron. Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association*, 83(403):793–800, 1988.

[39] P. J. Diggle. Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *Journal of Neuroscience Methods*, 18(1-2):115 – 125, 1986.

[40] P. J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, volume 128 of *Monographs on Statistics and Applied Probability*. CRC Press, Taylor & Francis Group, 3rd edition, 2014.

[41] P. Fine, K. Eames, and D. L. Heymann. 'Herd immunity': A rough guide. *Clinical Infectious Diseases*, 52(7):911–916, 2011.

[42] N. Fry. Random point distributions and strain measurement in rocks. *Tectonophysics*, 60(1-2):89–105, 1979.

[43] V. Gandhi. *Semivariogram Modeling*, pages 1042–1046. Springer US, Boston, MA, 2008.

[44] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British geographers*, 21(1):256–274, 1996.

[45] R. C. Geary. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3):115–146, 1954.

[46] J. A. González, F. J. Rodríguez-Cortés, O. Cronie, and J. Mateu. Spatio-temporal point process statistics: A review. *Spatial Statistics*, 18:505–544, 2016.

[47] M. F. Goodchild and R. P. Haining. GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science*, 83(1):363–385, 2004.

[48] P. Greig-Smith. The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany*, 16(2):293–316, 1952.

[49] R. Haining. *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, 2003.

[50] K. Hampson, J. Dushoff, S. Cleaveland, D. T. Haydon, M. Kaare, C. Packer, and A. Dobson. Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biology*, 7(3):462–471, 2009.

[51] S. Hanna and N. Fry. A comparison of methods of strain determination in rocks from southwest dyfed (pembrokeshire) and adjacent areas. *Journal of Structural Geology*, 1(2):155–162, 1979.

[52] A. C. A. Hope. A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(3):582–598, 1968.

[53] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[54] S. P. Hubbell. Diversity of canopy trees in a neotropical forest and implications for conservation. *Tropical rain forest: ecology and management*, pages 25–41, 1983.

[55] P. V. K. Iyer. The first and second moments of some probability distributions arising from points on a lattice and their application. *Biometrika*, 36(1/2):135–141, 1949.

[56] M. Jerrett, R. T. Burnett, R. Ma, C. A. Pope, D. Krewski, K. B. Newbold, G. Thurston, Y. Shi, N. Finkelstein, E. E. Calle, and M. J. Thun. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*, 16(6):727–736, 2005.

[57] D. L. Knobel, S. Cleaveland, P. G. Coleman, E. M. Fèvre, M. I. Meltzer, M. E. G. Miranda, A. Shaw, J. Zinsstag, and F.-X. Meslin. Re-evaluating the burden of rabies in Africa and Asia. *Bulletin of the World Health Organization*, 83(5):360–368, 2005.

[58] C. Kraamwinkel, I. Fabris-Rotelli, G. Fosgate, D. L. Knobel, and K. Hampson. A study on the apparent randomness of an animal sample. In *South African Statistical Association 2014 Conference Proceedings*, November 2014.

[59] D. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.

[60] D. G. Krige. A statistical approach to some mine valuation and allied problems on the Witwatersrand. M.Sc. in Engineering, University of the Witwatersrand, 1951.

[61] C. E. Krumm, M. M. Connor, and M. W. Miller. Relative vulnerability of chronic wasting disease infected mule deer to vehicle collisions. *Journal of Wildlife Diseases*, 41(3):503–511, 2005.

[62] Y. Lee and M. McCracken. Spatial adjustment of retail activity: A spatial analysis of supermarkets in metropolitan denver, 1960-1980. *Journal of Regional Analysis and Policy*, 12(2), 1982.

[63] C. Loader. *Local Regression and Likelihood*. Statistics and Computing. Springer New York, 2006.

[64] S. L. Lohr. *Sampling: Design and Analysis*. Brooks/Cole, Cengage Learning, 2nd edition, 2010.

[65] A. F. Mark and A. E. Esler. An assessment of the point-centred quarter method of plotless sampling in some New Zealand forests. *Proceedings (New Zealand Ecological Society)*, (17):106–110, 1970.

[66] B. Matérn. *Spatial variation*. Lecture notes in statistics. Springer-Verlag, 1986.

[67] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.

[68] J. Møller and R. Waagepetersen. Modern spatial point process modelling and inference (with discussion). *Scandinavian Journal of Statistics*, 34:643–711, 2007.

[69] J. Møller and R. P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.

[70] P. A. P. Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251, 1948.

[71] P. A. P. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[72] M. K. Morters, T. J. McKinley, O. Restif, A. J. K. Conlan, S. Cleaveland, K. Hampson, H. R. Whay, I. M. Damriyasa, and J. L. N. Wood. The demography of free-roaming dog populations and applications to disease and population control. *Journal of Applied Ecology*, 51(4):1096–1106, 2014.

[73] J. Negreiros, M. Painho, F. Aguilar, and M. Aguilar. A comprehensive framework for exploratory spatial data analysis: Moran location and variance scatterplots. *International Journal of Digital Earth*, 3(2):157–186, 2010.

[74] J. Neyman and E. S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika*, 20A(1/2):175–240, 1928.

[75] E. T. M. Ng and R. J. Cook. Adjusted score tests of homogeneity for Poisson processes. *Journal of the American Statistical Association*, 94(445):308–319, 1999.

[76] H. Ng, K. Gu, and M. Tang. A comparative study of tests for the difference of two Poisson means. *Computational Statistics & Data Analysis*, 51(6):3085 – 3099, 2007.

[77] H. K. T. Ng and M.-L. Tang. Testing the equality of two Poisson means using the rate ratio. *Statistics in Medicine*, 24(6):955–965, 2005.

[78] M. Numata. Forest vegetation in the vacinity of choshi. coastal flora and vegetation at choshi, chiba prefecture iv. *Bulletin of Choshi Marine Laboratory, Chiba University*, 3:28–48 [in Japanese], 1961.

[79] S. M. Nusser, W. R. Clark, D. L. Otis, and L. Huang. Sampling considerations for disease surveillance in wildlife populations. *The Journal of Wildlife Management*, 72(1):52–60, 2008.

[80] A. L. Patterson. A fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review*, 46(5):372, 1934.

[81] A. L. Patterson. A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift fuer Kristallographie-Crystalline Materials*, 90(1-6):517–542, 1935.

[82] W. J. Platt, G. W. Evans, and S. L. Rathbun. The population dynamics of a long-lived conifer (pinus palustris). *The American Naturalist*, 131(4):491–525, 1988.

[83] R. F. Potthoff and M. Whittinghill. Testing for homogeneity: II. the Poisson distribution. *Biometrika*, 53(1/2):183–190, 1966.

[84] C. R. Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1):50–57, 1948.

[85] C. R. Rao and I. M. Chakravarti. Some small sample tests of significance for a Poisson distribution. *Biometrics*, 12(3):264–282, 1956.

[86] J. Rasson, M. Rémon, and F. Henry. Finding the edge of a poisson forest with inside and outside observations: The discriminant analysis point of view. In *From Data to Knowledge*, pages 94–101. Springer, 1996.

[87] B. Ripley. *Spatial Statistics*. Wiley, New York, 1981.

[88] B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):172–212, 1977.

[89] B. D. Ripley. *Statistical Inference for Spatial Processes*. Cambridge University Press, 1991.

[90] B. D. Ripley and J.-P. Rasson. Finding the edge of a poisson forest. *Journal of Applied Probability*, 14(3):483–491, 1977.

[91] S. S. Rosenstock, D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. Landbird counting techniques: Current practices and an alternative. *The Auk*, 119(1):46–53, 2002.

[92] D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics. Wiley, 2009.

[93] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1986.

[94] C. J. Stone. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, pages 1285–1297, 1984.

[95] L. Strand. A model for stand growth. In *IUFRO Third Conference Advisory Group of Forest Statisticians*, volume 72, page 3, 1972.

[96] D. J. Strauss. A model for clustering. *Biometrika*, 62(2):467–475, 1975.

[97] E. Thönnes and M.-C. van Lieshout. A comparative study on the power of van Lieshout and Baddeley's J-function. *Biometrical Journal*, 41(6):721–734, 1999.

[98] W. R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240, 1970.

[99] M. N. M. van Lieshout. A J-function for inhomogeneous point processes. *Statistica Neerlandica*, 65(2):183–201, 2011.

[100] 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.

[101] L. Waller and C. Gotway. *Applied Spatial Statistics for Public Health Data*. Wiley Series in Probability and Statistics. Wiley, 2004.

[102] J. Wang, R. Haining, and Z. Cao. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. *International Journal of Geographical Information Science*, 24(4):523–543, 2010.

[103] J.-F. Wang, A. Stein, B.-B. Gao, and Y. Ge. A review of spatial sampling. *Spatial Statistics*, 2:1 – 14, 2012.

[104] P. Whittle. On stationary processes in the plane. *Biometrica*, 41:434–449, 1954.

[105] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

[106] V. Wirz, M. Schirmer, S. Gruber, and M. Lehning. Spatio-temporal measurements and analysis of snow depth in a rock face. *The Cryosphere*, 5(4):893, 2011.