Universidad San Francisco de Quito

College of Graduate Studies

Spatial Point Patterns Analysis for the Distribution of Dominant Canopy Trees (of commercial value) within Dense Humid Forests in Mai-Ndombe Area of the Democratic Republic of Congo

Myriam Elizabeth Saavedra López

A thesis submitted in partial fulfillment of the requirements for the Master of Applied Mathematics

at Universidad San Francisco de Quito

Quito, June 2011

Universidad San Francisco de Quito

Colegio de Posgrado

Análisis Espacial de Patrones Puntuales de la Distribución de árboles dominantes con valor comercial en el Dosel de los Bosques denso y húmedos en el Área Mai-Ndombe en la República Democrática del Congo.

Myriam Elizabeth Saavedra López

Tesis de grado presentada como requisito para la obtención del título de Magister en Matemáticas Aplicadas

Quito, junio de 2011
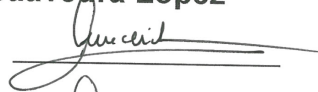
San Francisco of Quito University

College of Graduate Studies
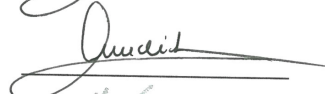
APPROBATION SHEET OF THESIS

Spatial Point Patterns Analysis for the Distribution of Dominant
Canopy Trees (of commercial value) within Dense Humid Forests in
Mai-Ndombe Area of the Democratic Republic of Congo

Myriam Elizabeth Saavedra López

Gonzalo Mendieta, Ph.D.
Director de Tesis

Gonzalo Mendieta, Ph.D.
Miembro del Comité de Tesis

Bwangoy Bankanza, Ph.D.
Codirector y Miembro del Comité de Tesis

Carlos F. Mena M., Ph.D.
Miembro del Comité de Tesis

Carlos Jiménez, Ph.D.
Director de la Maestría en Matemáticas Aplicadas

Fernando Romo, M.Sc.
Decano del Colegio de Ciencias e Ingeniería

Víctor Viteri Breedy, Ph.D.
Decano del Colegio de Postgrados

Quito, June 2011

Universidad San Francisco de Quito

Colegio de Posgrado

HOJA DE APROBACION DE TESIS
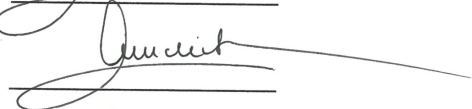
Análisis Espacial de Patrones Puntuales de la Distribución de árboles dominantes con valor comercial en el Dosel de los Bosques denso y húmedos en el Área Mai-Ndombe en la República Democrática del Congo.
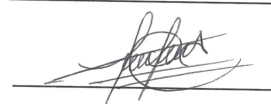
Myriam Elizabeth Saavedra López

Gonzalo Mendieta, Ph.D.
Director de Tesis

Gonzalo Mendieta, Ph.D.
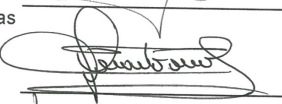Miembro del Comité de Tesis

Bwangoy Bankanza, Ph.D.
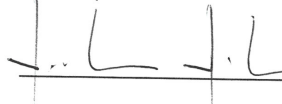Codirector y Miembro del Comité de Tesis

Carlos F. Mena M., Ph.D.
Miembro del Comité de Tesis

Carlos Jiménez, Ph.D.
Director de la Maestría en Matemáticas Aplicadas

Fernando Romo, M.Sc.
Decano del Colegio de Ciencias e Ingeniería

Víctor Viteri Breedy, Ph.D.
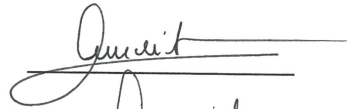Decano del Colegio de Postgrados

Quito, junio del 2011

Acknowledgements

I would like to express my sincere thanks to my director, Gonzalo Mendieta, PhD, Academic Dean of Universidad San Francisco de Quito, for his help, support and correction of my thesis. I will like to also express my thanks to Bwangoy Bankanza, PhD student at South Dakota State University and Congo Program Manager for Ecosystem Restoration Associates, for the thesis idea, his help in giving us acces to the the facilitation the data set, the first bibliographic references and sharing his in-depth knowledge and insight of the s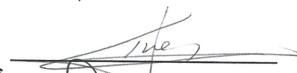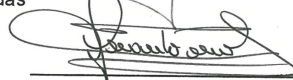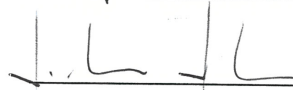tudy area (primary, equatorial lowland, forests of the Congo Basin). Carlos Mena, PhD was very helpful providing sereval recommendations that have included in this work. Finally, I am grateful to John Kendall, RPF and Chief Climate Forester of Ecosystems Restoration Associates, for teaching me basic knowledge about forest ecology, silviculture and management and for his patience in helping me with the edition my this work.

I appreciate all the knowledge which was transmitted to me by each and everyone of my teachers of the masters program at USFQ. It one way or another their teachings have contributed to the writing of this thesis. I am also very appreciative for the unlimited library access granted to me by Simon Fraser University library which helped to enhance my understanding of the topic.

I also would like to thank my family and friends for their encouragement and support, especially my husband who has supported me in key aspects of this work.

This thesis is dedicated to those struggling to obtain personal goals in the midst of adversity and, by doing so, help to confirm their faith in God and in themselves.

Abstract

The Congo Basin is the second largest rainforest after of the Amazon. This forest has more than 28 species of dominant trees (very large), of substantial commercial value, such as Wenge and Bosee-Clair. In one part of the Congo Basin forest biome, situated in the Mai-Ndombe District of the Democratic Republic of Congo, the government, forest communities and environmental service companies are searching for better ways to conserve and sustainably manage forest resources. This implies a knowledge of how much volume (i.e referred to as a 'benefit') one could sustainably extract without significantly changing the forest structure and composition. This 'benefit' can be described as the difference between some hypothetical uniform random distribution of trees and observed tree aggregation. This thesis attempts to provide some answers that can lead forest managers to better qualify and quantify this 'benefit', by determining the spatial distribution of dominant commercially valuable trees via the development and use of mathematical-algorithm-based measurements which show the level of tree aggregation (also referred to as clustering). For this purpose we have used spatial point patterns methodologies using tools available in the R package for statistical analysis, mathematical and forest criterions, and knowledge of the geomorphology of the selected Mai Ndombe District study area.

Resumen

La cuenca del Congo es el segundo bosque lluvioso más grande, luego de la Amazonía, él mismo que contiene más de 28 especies de árboles dominantes (muy largos) de apreciable valor comercial como son Wenge y Bosse-Clair. Una parte de este bosque está ubicado en el Distrito de Mai-Ndombe, República Democrática del Congo, donde tanto el gobierno, las comunidades del bosque, como compañías de servicio ambiental (nacionales e internacionales) desean encontrar un mecanismo que permita una tala sostenible de recursos forestales, es decir, remover solamente el *beneficio* que se puede obtener del bosque. Este beneficio es definido como la diferencia entre una hipotética distribución aleatoria uniforme de los árboles y una agregación observada. Entonces, esta tesis es una respuesta que ayudará al desarrollo de este mecanismo, determinando la distribución espacial que tienen estos árboles y desarrollando un algoritmo que mida cierto nivel de agregación o agrupamiento de árboles. Para ésto, se aplicó metodologías de patrones espaciales puntuales usando herramientas disponibles en el paquete R para análisis estadístico, criterios matemáticos y forestales, y conocimiento de geomorfología área de estudio la cual está ubicada en el distrito de Mai-Ndombe.

Contents

1 SECTION

1.1 Introduction

Forests have an important impact for the earth's global ecosystem equilibrium because they release oxygen necessary for the respiration of organisms and they absorb CO2 which is the main greenhouse gas that contributes to global warming. Forests are also important for community livelihood as they provide food, wood for construction and energy, (amongst other values) and they provide a huge set of critical habitat for a wide and very diverse range of fauna. The forest canopy acts as an interface between the forest and the atmosphere. It also plays an important role in determining the structure and properties of the understory and the forest floor.

Very large trees with heights greater than 30 m have been shown to be highly influential in the upper vegetative canopy structure and they also are highly deterministic of forest ecology and forest biomass regimes. In particular, they account for a large portion of the total forest above-ground biomass (Neef et al. 2005, 1; Saalovara et al 2005) and have been recognized as reproductively dominant. Very large canopy trees are therefore an important driving factor of forest succession, biodiversity and ecological stability.

One way to know about the composition and structure of the forest is to analyze forest inventory data using statistical methods in order to determine the distribution of tree species. Knowledge of species distribution represents a significant tool for foresters and environmentalists, as well as for governments, forest companies and forest dependent communities. Such knowledge can be very useful in guiding sustainable management of forest resources.

1.2 Background and Justification

Sustainable forest resource logging and management requires removal of a volume of timber that does not exceed the equivalent to the expected benefit[1]generated by projected tree growth over the course of the harvesting/logging cycle. In the case of the Democratic Republic of the Congo (the acronym is DRC, see appendix A for location) this can also be hypothesized as the difference between some hypothetical uniform random distribution of trees and observed aggregation. *In other words, if there are aggregations of trees in a forest stand, trees that are an excess volume (i.e.by selectively harvesting specific individual trees) can be removed without potentially altering the fundamental forest*

---

[1]'benefit' herein refers to the cumulative net annual increase in merchantable forest stemwood/timber

structure, composition and ecology. 'Excess' tree removal creates forest openings, which are sometimes referred to as canopy gaps. This gap creation leads to a succession[2]. Succession regeneration ensures the continuance of such 'benefits' in the future. Naturally regenerated trees in canopy gaps can be harvested in subsequent logging cycles (in a selective forest management regime, a logging cycle consists of a series of harvesting entries wherein the 'benefit' merchantable volume is progressively extracted). Unfortunately, such 'best management' practice has yet to be conducted in the Congo, mostly because of lack of information and knowledge, hence the importance of this study and others like it.

Characterizing and quantifying aggregation and their related 'benefits' is difficult in the fields of forestry ecology and management. It is not easy to obtain naturally-occurring, randomly distributed, tree data, organized in harvesting blocks, and replicate such data. Classical statistics methodologies are often not appropriate statistical models for such types of analyses (Cressie 1993, 578-579). There is a need for new statistical models and approaches that address new questions.

A significant percentage of data used in resource assessment, environmental monitoring (e.g. for global warming), and medicinal imaging, is spatial in nature. Additionally, in the field of forestry, recent studies have showed encouraging results from spatial point patterns analysis of tree populations in humid, wet tropical and dry forests of Africa (Linares Palomino, 2005, 317-318). Application of these statistical models in analyzing Spatial Point Patterns vis-a-vis the study of tree distribution sheds light needed to find answers to some fundamental questions related to: the statistical distribution function for a tree population set, the level aggregation, the nature of relationship between individuals (i.e. inhibition, repulsion, dominance, etc.) Such applications may help in ascertaining the presence and/or non presence of external factors, such as human intervention (i.e. logging) or natural events (ex. tree mortality and tree competition within forested management units) that can explain observed tree distribution patterns. This enhanced knowledge could lead to better prediction and characterization of future forest composition, species spatial distribution and forest stand structure that could, in turn, better inform best management practices.

1.3 Hypothesis and Context

The upper canopy layer of the dense humid forests in the Lake Mai-Ndombe area that dominate the study sites from which the data set was collected. This area is characterized by a relatively high abundance of *Milletia laurentii* (Family: Fabaceae), commercially known as Wenge. Wenge is one of 28 other dominant, upper canopy, species that have commercial value (see Appendix B) in the project area. The results of the spatial distribution analysis of these trees from a mathematical perspective

---

[2]'succession' herein refers to nature regeneration of trees of the same dominant canopy species

can potentially solve many questions about their ecological characteristics. Previous studies had considered possible aggregations and this study seeks to shed light on a specific form of aggregation, a possible tree grouping referred to as clustering. Considerable knowledge about spatial analysis has been developed by many statisticians or mathematicians, such as Diggle, Cressie, Illian and others. This study used such knowledge, along with the R software implementation package developed by Baddeley and Turner (2005).

The spatial analysis of forest distribution, in the Lake Mai-Ndombe District, Democratic Republic of Congo, from a forestry perspective has, as its primary objective: an enhanced understanding of tree distribution (i.e. aggregation/clustering, random or uniform distribution) and ecological neighborhood and interaction characteristics (see Appendix C). Such analytical results will help increase the overall knowledge of the composition and structure of equatorial forest ecosystems (characterized by high species diversity) and tree growth models. However, these analytical results will need to be ground validated to be useful in *decision support systems* for foresters and ecologists vis-a-vis natural resource management considerations, such as optimal, ecologically sustainable, selective tree extraction and timber harvesting rates.

1.4 Proposal

The purpose of this study is to apply recent spatial point pattern analytical tools and methodologies to data sets that represent a 100% sample of all commercially valuable trees within a specific set of management units located in the dense, humid, 'terra firma' forests of the Lake Mai-Ndombe District in the Democratic Republic of the Congo so as to: 1 - generally characterize the spatial distribution these commercially valuable trees 2 - specifically explore a first level of aggregation/clustering of Wenge (*Milletia laurentii*) and Bosse-Clair (*Guarea Cedrata*) trees.

1.5 Outline

Section 2 starts by providing the pertinent basic mathematic and forest definitions. Sub-section 2.1 presents mathematic definitions for such terms as 'spatial point process, 'point patterns', 'complete spatial randomness' (CSR) and others similar terms that lie within the domain of Spatial Statistical Analyses. Sub-section '2.2.' offers forest definitions, such as 'tree crown spread', 'tree canopy', and 'tree aggregation', which assist the reader in understanding the origin and characteristics of the data sets used in this thesis study.

In section 3, the spatial statistical analysis methods and statistics tests which were applied in the data set are described. Sub-section 3.1 presents the Quadrat Counts Method and Chi-Quadrat test for CSR, which are used to investigate the intensity of the spatial points. Sub-section 3.2 describes

the Monte Carlos Test, which is used in the Simulation Envelopes Methods (Sub-section 3.3). This latter method is fundamental for the spatial analysis because it is applied over all the spatial distance methods. Sub-section 3.4 describes in detail each of the spatial distance methods (F function, G function and J function) for the First Moment Characteristic (intensity). Sub-section 3.5 conceptually complements sub-section 3.4 with the description of the distance methods: K Function and Pair Correlation g Function, which both prove the Second Moment Characteristic of the spatial point patterns. Sub-section 3.6., provides definitions for the corrected estimates, which are used within the distance methods to correct the existing edge-errors when the spatial point patterns are analyzed. Sub-section 3.7 describes in general how and why modeling or model fitting is done for the studied spatial points.

Section 4 contains the critical part of the thesis. It summarizes the key spatial statistic analyses of the Bimpe Concession data set on the western edge of the Congo Basin Equatorial Forest Biome. This overall analysis begins in sub-section 4.1 with a description of the data. The subsequent sub-sections are generated by following the spatial statistics methods described in section 3, using the implementation of these methods in the spatstat package developed by Baddeley in the statistic R Tool. Firstly, the spatial point process intensity of all the blocks which form the Bimpe Concession was investigated (Sub-section 4.2). With the observations and preliminary conclusions obtained in sub-section 4.2, Block I was the first area subjected to the spatial exploration and analysis. Subsequently, all the distance methods for the First and Second Moments were applied to Block I (Sub-section 4.3.). Then, using these last results with the basic forest and mathematic definitions, sub-section 4.3.5 describes the fit models applied to Block I.

Subsequently, sub-section 4.4 describes the analysis of the First and Second Moment Characteristics for H, R1 and Z2 blocks via the application of the J and K functions, focusing on the functionality of the methods and taking into account pertinent forest knowledge of the data set. Sub-section 4.5 then describes the same process in sub-section 4.4, but, in this case, for J, Q and Z1 Blocks. At the end of these sub-sections, conclusions are presented about the spatial distribution of the Bimpe Concession data set. The only step remaining was the analysis of the aggregation/clustering of these spatial points, which is described in sub-section 4.6. Firstly, this analysis consisted of an exploration of aggregation (Sub-section 4.6.1) using the knowledge of the preceding sub-sections. Secondly, a description of the potential level of aggregation (sub-sections 4.6.2 and 4.6.3) for all the commercially valuable trees was made. Thirdly, this exercise was repeated specifically for Wenge and Bosse-Clair species trees only.

Section 5 presents the conclusions and recommendation of the thesis in summary form followed by the bibliography, appendices, listed of tables, list of figures and nomenclature.

2 SECTION

2.1 Basic Mathematical Definitions

A stochastic process is a collection of random variables $\{X(t) \mid t \in T\}$ indexed by a set T. It is common to take T to be a subset of the real numbers, say $\{1, 2, 3....\}$ or $[0, \infty)$ but the index set can also be points in a higher dimensional space (Ripley 1981, 9).

A spatial point process is a stochastic process. Let X is a finite random subset of a given bounded region W under d-dimensional Euclidean space, and a realization of such a process is a spatial point pattern x $= \{x_1, ..., x_n\}$ of $n \geq 0$ points contained in W. We say that the point process is defined on W, and we write $x = \emptyset$ for the empty point pattern. The number of points $n(X)$, is a random variable $N(B) = n(X_B)$ for subsets $B \subseteq W$, where $X_B = X \bigcap B$ (Møller, Jesper and Rasmus, Waagepetersen P. 2006, 7).

In the notation of the spatial point are often numbered but does not imply any meaningful order of the points which are characterized by the point patterns (Illian et al, 2008, 23).

The term process is usually associated with an evolution over time and that is why physicists and engineers might prefer the term *point field* to refer to process where the index space is not time.

The spatial point processes can be a *stationary process*, if all probability statements about the process in any region A of the plane are invariant under arbitrary translation of A, and/or an *isotropic process* when the process is stationary under rotations about a fixed point, with homogeneity, about any point (same invariance holds under rotation), i.e. if there are no directional effects (Diggle 1983, 46; Ripley 1988, 7).

The point patterns of a data set in a study can be Complete Spatial Randomness, which means that the spatial points are follow the Uniform Poisson Distribution (rather unusual in nature), clustering, which means presence of aggregation in the study data and regularly which means the presence of inhibition in the studied data. According to Diggle (1983, 2-3), the nature of the pattern generated by a biological process can be affected by the physical scale on which the process is observed. At a sufficiently large scale a realization would show heterogeneity, which will tend to produce aggregated patterns. At a smaller scale, the realization can be a less pronounced variation and the major determinant of the pattern may then be the nature of the interactions amongst the points themselves.

The homogeneous planar Poisson point process, according to Diggle (1983, 50), subsequently referred to without qualification as the Poisson process, is the cornerstone on which the theory of spatial point processes is built. The Poisson process is defined by the following postulates, which are the same used in the hypothesis of Complete Spatial Randomness.

1. For some $\lambda > 0$, and any finite planar region A, $N(A)$ as a Poisson distribution with mean $\lambda |A|$.

2. Given $N(A) = n$, the n events in A, form an independent random sample from the uniform distribution on A.

To demonstrate that **PP1** and **PP2** are self-consistent, one first establish

1. For any two disjoint regions A and B, $N(A)$ and $N(B)$ are independent.

The demonstration can be seen in page 50 and 51 from Diggle (1983). The postulates **PP1** and **PP2** consequently define a spatial point process as well.

The first moment or intensity (υ) of point processes is the analogue of the expected value of a random variable. The Campbell's formula (See Baddeley et al 2007, 28) is an important result for the intensity according to Baddeley. Then, let X be a point process on d-dimensional Euclidean space W (or on any locally compact metric space W). Let

$$\upsilon(B) = \mathbb{E}[N_x(B)], \quad B \subset W \tag{1}$$

where $\mathbb{E}$ is the expected value and υ on W is a measure, called the *intensity measure* of X, provided $\upsilon(B) < \infty$ for all compact B (See Baddeley et al, 2007, 26). If X is a stationary point process, the intensity measure needs to be proportional to Lebesgue measure υ. Then, considering a constant λ as the *intensity* so

$$\mathbb{E}[N_X(B)] = \lambda \upsilon(B) \tag{2}$$

(Ripley 1988, 22).

Second moment (dependence and interaction) of Point processes: As the intensity υ has an analogue in the classic statistic, the Second Moment of point processes has analogues which are the variance and covariance of random variables. Let X be a point process on W, then $X \times X$ is a point process on $W \times W$ consisting of all ordered pairs (x, x') of points $x, x' \in X$. The intensity measure υ_2 of $X \times X$ is a measure $W \times W$ satisfying

$$\upsilon_2(A \times B) = \mathbb{E}[N_X(A) N_X(B)] \tag{3}$$

, called the *second moment measure* of X (See Baddeley et al 2007, 32).

According to Ripley, if X is a stationary point process under translation and rotation and using the formula (2), the second moment is reduced as (1988, 22):

$$\mathbb{E}[N_X(A) N_X(B)] = \lambda \upsilon(A \cap B) + \lambda^2 \int_0^\infty \upsilon_r(A \times B) \, dK(r) \tag{4}$$

and $\upsilon_r(A \times B) = \int \sigma_r(\{y - x | x \in A, y \in B, d(x, y) = r\}) \, d\upsilon(x)$ where σ_r is the uniform probability

on the surface of the sphere in d-dimension Euclidean space of radius r. The crucial part of formula (4) is that the second-moment description which is reduced to an increasing function $K()$ on $(0, \infty)$. Taking $K(0) = 0$, it is possible to have $\lambda K(r) = \mathrm{E}[\#\text{of points with distance} \leq r | \text{point at } x]$ using Palm probabilities (Baddeley et al 2007, 44-45) to formalize the conditioning on an event of probability zero. Thus $K()$ is a sort of inter-point distance distribution function, although $K(r) \to \infty$ as $r \to \infty$.

2.2 Basic Forest Definitions

Except when otherwise stated the definitions below are from (Ministry of Forestry and Range, 2008)

Forest aggregation: According Stephenson (1987, 1), there are two possible definitions for Forest Aggregation:

1. where there are patches of trees exhibiting relatively homogeneous age and species composition;

2. "Groups [of plants] that have a high degree of homogeneity relative to their growth-form, species, height, dbh (density by hectare), and number of layers, - definition by Bonnicksen and Stone's (1982)".

The Congo Basin covers approximately three million square kilometers, which holds the world's second largest tropical forest of approximately two million square kilometers. The Basin's forest zone includes parts of six central African countries: Cameroon, Central African Republic, Congo, Democratic Republic of the Congo, Equatorial Guinea, and Gabon, listed alphabetically. In the Congo Basin the central area, also known as the "Cuvette Centrale" is characterized by extensive forested and non forested wetlands. The climate in some areas is warm and humid with two wet and two dry seasons. The mean temperature is approximately 25 ^{0}C. The average rainfall is about 1800 mm per year in 115 days. The result is an equatorial rainforest in which some species such as the Wenge may shed their leaves during the major dry season (Bwangoy et al. 2010, 74)

A **canopy** is the more or less continuous cover of branches and foliage formed collectively by the crowns of adjacent trees (11).

Climate change is defined as "An alteration in measured quantities (e.g., precipitation, temperature, radiation, wind, and cloudiness) within the climate system that departs significantly from previous average conditions and is seen to endure, bringing about corresponding change in ecosystems and socio-economic activity" (13).

Co-dominant trees are average-to-fairly large trees with medium-sized crowns that form the forest canopy. These trees receive full light from above but are crowded on the sides. They are healthy trees of a species that is ecologically suitable for the site, and commercially valuable (Tarr and Stewards, 2009) Note that commercial value is included in the definition here specifically due to the nature of this data set.

A **Crop tree** is a healthy tree of a species that is ecologically suitable for the site and commercially valuable (18).

Crown is "The live branches and foliage of a tree" and **crown-class** is "A tree classification category referring to a tree's relative height, foliage density, and ability to intercept light. Crown-class is an indicator of past growth performance and calls attention to crop trees that could benefit from future thinning and harvest operations. There are four classifications of trees: co-dominant trees, dominant trees, intermediate trees and suppressed or overtopped trees."(Tarr and Stewards, 2009)

"**The dominant trees** are larger-than-average trees with broad, well-developed crowns. These trees receive direct sunlight from all sides and above but are crowded on the sides" (Tarr and Stewards, 2009)

A **forest** is "A complex community of plants and animals in which trees are the most conspicuous members and where the tree crown density- the amount of compactness of foliage in the tree tops- is greater than 10 percent" (36).

Forestry is the profession embracing the science, art, and practice of creating, managing, using and conserving forests and associated resources for human benefit and in a sustainable manner to meet desired goals, needs, and values (40).

3 SECTION

Remember the definition of Point Poisson Process, this stochastic process is defined as the cornerstone of the theory of spatial point processes. Then for any spatial analysis to be successful, it usually begins by the CSR test has a null hypothesis that considers that the spatial distribution for the studied data set is uniform Poisson.

According to Illian (2008, 57-58), the simulation of the Poisson processes is easy to understand. However, the derivation of the summary characteristics is technically challenging in some places. Moreover, the method that has been implemented allows a comparison of empirical characteristics with theoretical ones and facilitates the general understanding of the various summary characteristics. A literature review of spatial statistic material by several authors who described and/or implemented these methods and tests and their application pertaining to the data sets studied in this thesis was carried out. The following methods, estimates and tests were selected to be part of the Spatial Statistic Analysis of the spatial points (trees) of the Bimpe Concession in the Mpata Mbalu forest located in the western shore of Lake Mai-Ndombe, in the Democratic Republic of the Congo (in the west central part of the country).

3.1 Quadrat-counts Method and Chi-square Test for CSR

The most elementary methods to study the intensity of a spatial points process based on counting events in regions is *Quadrat-counts*, which is based on dividing the domain A (the total region) into non-overlapping regions (quadrats) $A_1, ..., A_k$ of equal size such that $A_1 \cup ... \cup A_k = A$. Most frequently, the domain is assumed to have a rectangle boundary which is partitioned into r rows and c columns. If n_k is the number of events in the quadrat k and $\bar{n} = n/(rc)$ is the expected number of events in any quadrat under CSR, then the standard Pearson Chi-square statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_k - \bar{n})^2}{\bar{n}} \tag{5}$$

which is obtained in the Chi-quadrat test goodness-of-fit. The hypothesis of this test considered the n points distributed uniformly and independently in the domain A, i.e., the quadrat counts are independent Poisson variants with common mean. The reference distribution for χ^2 is χ^2_{rc-1}. Although X^2 is written as a Chi-square statistic χ^2 for a contingency table, the double summation is used to emphasize the row-column partition of the domain. Furthermore, no additional degree of freedom is lost to the estimation of $\bar{n}$, since n is known in a mapped point pattern. If the pattern is CSR, then the ratio of sample variance and the sample mean should be approximately 1. X^2, is thus also referred to as the index dispersion

$$ID = X^2/(rc - 1) \tag{6}$$

according to Diggle (1983, 33). The goodness-of-fit test based on quadrat counts is simple and the Chi-square approximation performs well provided that the expected number of events per quadrat exceeds 1 and $rc > 6$. It is, however, very much influenced by the selection of the quadrat size.

3.2 Monte Carlo Tests

According to Schabenberger et al (2005, 87-88), the hypothesis of Monte Carlo test considers an observed patterns $Z \in W$ could be the realization of a points process model ψ. A Q statistic test is chosen which can be evaluated for the observed pattern and for any realization simulated under the model ψ. Let x_0 denote the realized value of the statistic test for the observed pattern. Then generate n realizations of ψ and calculate their respective test statistics: $x_1 = x(\psi_1), ..., x_n = x(\psi_n)$. The statistic x_0 is combined with these and the set of $x + 1$ values is ordered (ranked). Depending on the hypothesis and the choice of Q, either small or large values of Q will be inconsistent with the model ψ. If ψ is rejected as a data-generating mechanism for the observed pattern when $x_0 \leq x_{(k)}$ or $x_0 \geq x_{(n+1-k)}$, where $x_{(k)}$ denotes the kth smallest value, then this is a two-sided test with significance

level

$$\alpha = 2k/(n+1) \tag{7}$$

Advantages:

- The p-values of the tests are exact in the sense that no approximation of the distribution of the test statistic is required.

- The p-values are inexact in the sense that the number of possible realizations under ψ is typically infinite. At least the number of realizations will be so large that enumeration is not possible.

- The number n of simulations must be chosen sufficiently large. For a 5% level lest $n = 99$ and for a 1% level test $n = 999$ have been recommended. As long as the model ψ can be simulated, the observed pattern can be compared against complex point processes by essentially the same procedure. Simulation tests thus provide great flexibility.

Disadvantages:

- Simulations of test have several critical choices at the user side, such as the number of simulations and the test statistic.

- Diggles (1993) cautions against "data dredging", the selection of a non-sensible statistic test for the sake of rejecting a particular hypothesis. Even if a sensible statistic test is chosen, the results of test simulations may not agree. The power of this procedure is also difficult to establish, in particular, when applied to test for point patterns. The alternative hypothesis for which the power is to be determined is not clear.

3.3 Simulation Envelopes

According to Schabenberger (2005, 88-89), a Monte Carlos Test calculates a single statistic test for the observed pattern and each of the simulated patterns. Often, it is used with functions of the point patterns. For example, let r_i denote the distance from event s_i to the nearest other event and let $I(r_i \leq r)$ denote the indicator function with return 1, whether $r_i \leq r$. Then $\hat{G}(r) = \frac{1}{n} \sum_{i=1}^{n} I(r_i \leq r)$ is an estimate of the distribution function of nearest-neighbour event distances and can be calculated for any value of r. With a clustered pattern, we expect an excess number of short nearest-neighbour distances (compared to a CSR pattern). The method for obtaining simulation envelopes is similar to that used for a Monte Carlos test, but instead of evaluating a single test statistic for each simulation, a function such as $\hat{G}(r)$ is computed. Let $\hat{G}_0(r)$ denote the empirical distribution function based on the observed point pattern. Calculate $\hat{G}_1(r), ..., \hat{G}_g(r)$ from g point patterns simulated under CSR (or any other hypothesis of interest). Calculate the percentiles of the investigated function from

the g simulations. For example, upper and lower 100% simulation envelopes are given by $\hat{G}_l(r) = \min\limits_{i=1,\dots,g} \left\{ \hat{G}_i(r) \right\}$ and $\hat{G}_u(r) = \max\limits_{i=1,\dots,g} \left\{ \hat{G}_i(r) \right\}$.

Finally, a panel is produced which plots $\hat{G}_0(r)$, $\hat{G}_l(r)$, and $\hat{G}_u(r)$ against the theoretical distribution function $G(r)$, or, if $G(r)$ is not attainable, against the average empirical distribution function from the simulation,

$$\bar{G}(r) = \frac{1}{n} \sum_{i=1}^{g} \hat{G}_i(r) \tag{8}$$

3.4 Distance Methods for the First Moment Characteristic of a Point Process

3.4.1 Empty space function F

Let X be a stationary point process in the d-dimensions Euclidean space. The contact distance function or empty space function $F()$ is the cumulative distribution function (c.d.f.) of the distance $d(u, X)$ from a fixed point u to the nearest point of X. That is

$$F(r) = P(d(u, X) \leq r) = P(N_X(b(u, r)) > 0) \tag{9}$$

where $N_X(b(u, r))$ is a number of events inside the ball b with radius r and center u. (Baddeley et al 2007, 22)

For a completely spatial random (Poisson) point process, the empty space function $F(r)$ is determined as

$$F(r) = 1 - \exp\left(-\lambda \pi r^2\right) \tag{10}$$

where λ is the expected number of points per unit area (or intensity). According to Baddeley and Rolf Turner (Estimate the empty space function F), an estimate of F derived from a spatial point pattern dataset can be used in exploratory data analysis and formal inference about the pattern (Cressie 1991; Diggle 1983; Ripley, 1988). The estimate of F is a useful statistic summarizing the sizes of gaps in the pattern in the analysis.

If $X = \{x_1, \dots x_n\}$ is a spatial point process in a region W, then an unbiased estimator of F is

$$\hat{F}(r) = \frac{1}{\lambda_d(w)} \int_w I\{d(u, X) \leq r\} du \tag{11}$$

where $I\{\}$ is the indicator function. It returns 1 when the condition is true and 0 otherwise. $\hat{F}(r)$ is an unbiased estimator of F for each r fixed value. For inferential purposes, an estimate of F is usually compared to the theoretical value of $F(r)$ and the deviations between both curves may suggest spatial clustering or spatial regularity patterns. See the following illustration:

Illustration of Empty Space Function F
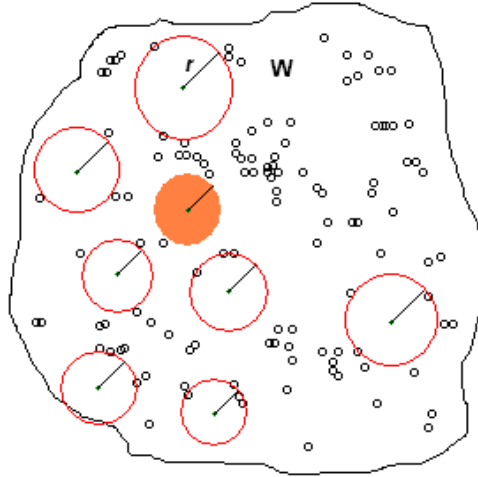
The light-brown ball shows wholly empty

3.4.2   The nearest neighbour distance function G

According to Baddeley et al (2004, 51-52) the nearest neighbour distance distribution is a related concept to the Empty Space Function F. Then, let X be a stationary point process in the d-dimensional Euclidean space. The nearest neighbour function G is the cumulative distribution function of the distance $d(x, X\backslash x)$ from a typical point $x \in X$ to the nearest other point of X. That is:

$$G(r) = \mathbb{P}^x\left(d\left(x, X\backslash\{x\}\right) \le r\right) = \mathbb{P}^x\left(N_X\left(b\left(x,r\right)\backslash\{x\}\right) > 0\right) \tag{12}$$

where $N_X\left(b\left(x,r\right)\backslash\{x\}\right)$ is the numbers of event inside the ball b with radius r and center x and $\mathbb{P}^x$ a probability. Using the Campbell-Mecke formula, the formula (9) is written as

$$G(r) = \frac{\mathbb{E}\left[\sum_{x \in X \cap W} I\left\{d\left(x, X\backslash\{x\}\right) \le r\right\}\right]}{\mathbb{E}\left[X\left(W\right)\right]} \tag{13}$$

where W is a window which contains the spatial points. See the folowing illustration:

Illustration of Nearest Neighbour Distance Function G
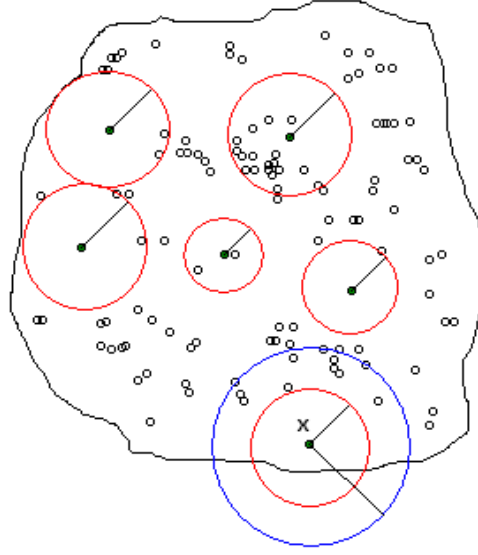
The blue ball b with center in 'x' and readius 'r' shows how the number of nearest neighbours of 'x' increases when 'r' increases.

By stationary, we means that this does not depend on x. For a stationary Poisson process in the d-dimensional Euclidean space, since $X^x \equiv XU\{x\}$, then

$$G(r) = 1 - \exp\left(-\lambda \pi_d r^d\right) \tag{14}$$

and in this case,

$$G(r) \equiv F(r) \tag{15}$$

3.4.3 The function J

The Function J is an combination of the Empty Space Function $F()$ and the Nearest Neighbour Distance Function $G()$, which is defined as follows: Let X be a stationary point process in d-dimensional Euclidean space, then the Function J of X is defined as:

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \forall r \geq 0,\ F(r) < 1 \tag{16}$$

For a uniform Poisson process the Function J is $J(r) \equiv 1$ using the equivalence (15).

This function has good properties with respect to many operations on point processes. One could be the follwing: suppose X and Y are independent stationary point processes, with intensities: v_X and v_Y. Then the superposition $X \bigcup Y$ has function J which is: $J_{X \bigcup Y}(r) = \frac{v_X}{v_X + v_Y} J_X(r) + \frac{v_Y}{v_X + v_Y} J_Y(r)$ where $J_X(r)$ is J function of X and $J_Y(r)$ is J function of Y (Baddeley et al, 2007, p.55-56).

3.5 Inter-point Distance Methods for Second Moment Characteristic of Point Processes

According to Illian et al (2008, 214) , Ripley's K function, Besag's L-function and the pair correlation function g have often been regarded as the functional summary characteristics that are the most important tools for the analysis of point patterns. In the spatial analysis of this thesis were the spacial analysis of this thesis used Ripley's K function and the pair correlation function g following the recommendation of Illian et al (2008, 218) and the result of the exploration of the spatial tool implemented in the package spatstat from the statistic software R.

3.5.1 The reduced second moment κ and the Ripley K function

Regarding the second moment of a stationary point process (section 2.1.9 and formula 4) when $K(0) = 0$, it is possible to have

$$\lambda K(r) = \mathbb{E}\left[N_X\left(b\left(x, r\right)\right) \setminus \{x\}\right] \tag{17}$$

where x is the typical not counted point. In order to estimate the functions, let n be the number of points in the window W and $N_X\left(b\left(x_i, r\right) \setminus \{x_i\}\right)$ be the number of points of N within distance r from points x_i, excluding x_i itself. Then an estimator of $\lambda K(r)$ is

$$\bar{n}(r) = \frac{1}{n} \sum_{i=1}^{n} N_X\left(b\left(x_i, r\right) \setminus \{x_i\}\right) \tag{18}$$

following the Palm distribution concept, remember that in a stationary point process the position of the center does not matter. Then it is possible to read formula (17) as

$$\lambda K(r) = \mathbb{E}_0\left[N_X\left(b\left(0, r\right)\right) \setminus \{0\}\right] \tag{19}$$

Finally, this formula is divided by λ to have $K(r)$ (the local point density fluctuations) alone from the global point density λ. Hence, the popular summary characteristic K function is:

$$K(r) = \mathbb{E}_0\left[N_X\left(b\left(0, r\right)\right) \setminus \{0\}\right] / \lambda \tag{20}$$

for $r \geq 0$. Using the Campbell-Mecke formula in (17), one has

$$\lambda K(r) = \frac{\mathbb{E}\left[\sum_{x \in X \cap W} X\left(b\left(x, r\right) \setminus \{x\}\right)\right]}{\mathbb{E}\left[X\left(W\right)\right]} \tag{21}$$

and the following illustration shows how works the K function.

Illustration of Ripley´s Function K
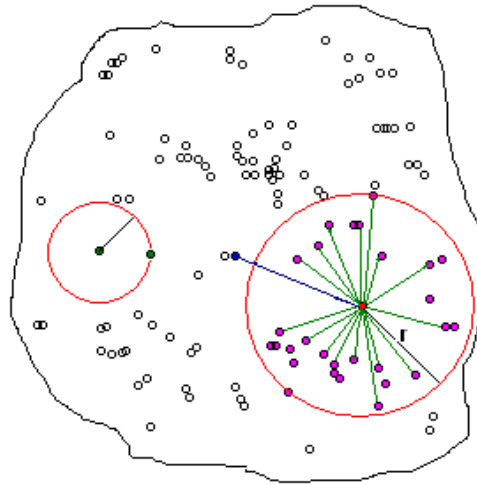
The big ball is 'b', and have 27 spatial points. The center point 'x' is red and the blue point does not interact with the red point when the distance-radiu is 'r'.

In general, one can expect $K(r)$ proportional to r^d; only for small r interesting deviations are observed. The function has a simple form in the Poisson process case this is $K(r) = \pi r^2$ for a planar case and this $K(r) = \pi r^d$ for a general d-dimensional case.

The shape of $K(r)$ relative to that of the Poisson process provides valuable information on the point process distribution. Hence, when the data are not a Poisson process there are two interpretations of $K(r)$: If $K(r) > \pi r^2$ or $K(r) > \pi r^d$ then there are expectations of a cluster process (presence of aggregation), and if $K(r) < \pi r^2$ or $K(r) < \pi r^d$ then there are expectations of a regular process (presence of inhibition or repulsion) (Illian et al, 2008, 214-216).

The *reduced second moment measure* κ is defined by $\lambda\kappa(B) = E_0(N_X(B\setminus\{0\}))$ where there is a reduction of sets from two to one. This justifies the word 'reduced'. This mathematic expression means that $\lambda\kappa(B)$ is equal to the mean number of points of N_X in the set B conditional on the typical point of N_X begining at zero.

3.5.2   The pair correlation function g

According to Illian (2008, 219-220), the function g is considered the best methods to give information about the second moment summary characteristic because it contains the same information and is an easy way to understanding the $K()$ function and $J()$ function. The Pair Correlation $g(r)$ function is proportional to the derivative of $K(r)$ with respect to r. Hence, considering the planar case the pair correlation function is

$$g(r) = \frac{K^{'}(r)}{2\pi r} \tag{22}$$

for $r \geq 0$ and in the general d-dimensional case, $g\left(r\right) = \frac{K^{'}(r)}{d\pi_d r^{d-1}}$ for $r \geq 0$.

When r is large, the $\lim_{r\to\infty} g\left(r\right) = 1$, means there is independent between the two compared points for large r. To be more precise, for this equation, the point process must have some distributional properties (called mixing properties).

If there is a finite distance r_{corr} with $g\left(r\right) = 1$ for $r \geq r_{\mathrm{corr}}$ then r_{corr} is called *range of correlation*. This means that there are no correlations between point positions at larger distances.

It is clear that in the Poisson process or CSR case $g\left(r\right) = 1$ for $r \geq 0$, i.e. which means that the location of any point is entirely independent of the locations of the other points. Again, in typical non-Poisson cases a characteristic behaviour of $g\left(r\right)$ may be found in the cluster process if $g\left(r\right) > 1$ and regular process if $g\left(r\right) < 1$, in particular for small radii in both cases.

The Second Moment Characteristics are related to statistical concepts such as spatial autocorrelation and semi-variogram because all these functions study the distances (space) between variables (spatial data) and what is it happening between these variables when these distances are short or long (interaction and/or dependence). Therefore all these functions support the first law of geography that "everything is related to everything else, but near things are more related than distant things" (Waldo Tobler). See the following descriptions:

1. Spatial Autocorrelation, according to Zuur, states that pairs of subjects (points) that are close to each other are more likely to have values that are more similar, and pair of subjects far apart forms each other are more likely to have values that are less similar. It is important to mention that the Function Ripley's K is one of the indices of spatial autocorrelation.

2. According to Cressei (1993, 58), semi-variogram has been called the variogram divided by 2, by Matheron (1962), therefore they are commonly referenced. The semi-variogram is a plot of semivariance as a function of distance. The semi-variance measures the dissimilarity of subjects within a single variable, compared to covariance which measures the similarity of one or more variables. It is not normalized and values are not as constrained as are most correlation coefficients (Zuur).

3.6   Corrected Estimates

According to Illian et al (2008, 180-183), the spatial statistics analysis with stationary point processes faces a difficult problem at its window edges: data are given for a bounded observation window W only, but the pattern is (implicitly) assumed to be infinite and the summary characteristic to be estimated is defined independently of W and should not show any traces of W. However, natural

estimators of the summary characteristic would need information from outside W, in particular for unbiased or ratio-unbiased estimators.

Ripley (1998) presents the border method which is available when estimating a c.d.f. and has quite wide applicability. However, accordingly the method used in this thesis for the "Border Corrected Estimate" of the $G\left(\right)$ and $F\left(\right)$ distance methods, and the inter-point method $K\left(\right)$ is the reduced-sample estimator developed by Baddeley et al.

3.6.1 Kaplan-Meier estimators

The content of this sub-section is a summary using of the paper from Baddeley and Gill (1997), focusing on the application to these Kaplan-Meier estimators. Also, there are some ideas taken from Balakrisman (2010, 481-490) in order to better understand the concept of survival data.

The estimation of $F\left(\right)$, $G\left(\right)$ and $K\left(\right)$ is hampered by edge effects arising because the point process is observed within a bounded window W. Essentially the distance from a given reference point to the nearest point of the process is censored by its distance to the boundary of W. The edge effects increase in proportion to the dimension of the space or the distances r. This problem of estimation has a clear analogy with the estimation of a survival function based on a sample of randomly censored survival times. The Kaplan-Meier estimator made this analogy and has various large-sample optimality properties. Before describing the Kaplan-Meier estimators, will be presented preliminary definitions will be presented that are necessary for understanding these estimates.

3.6.1.1 Survival data *Survival data* is defined as follows: Let $X_1, ..., X_n$ be a set of nonnegative random variables (i.i.d.) representing the smaller of the failure time of interest (observed times) with distribution function $F\left(t\right) = \mathbb{P}\left[X_l \leq t\right]$ and $S_{(t)} = 1 - F_{(t)}$ a survival function. Let $C_1, ..., C_n$ be a set of nonnegative random variables (i.i.d.) independent of all X_l with distribution function $H\left(\right)$. Under the right censoring one observes only $D_l = I\left\{X_l \leq C_l\right\}$, the indicator of the event $\left\{X_l \leq C_l\right\}$, and $Z_l = X_l \wedge C_l$, the $\min\left(X_l, C_l\right)$. This X_l is completely observable if and only if $X_l \leq C_l$, i.e. $D_l = 1$. Then $\left(Z_1, D_1\right), ..., \left(Z_n, D_n\right)$ is a sample of censored survival times Z_l.

3.6.1.2 Kaplan-Meier estimator of the empty space function F Every point x in the window W contributes one possibly censored observation of the distance from an arbitrary point in space to the point process X. The analogy with survival times is to regard $d\left(x, X\right)$ as the distance (time) to failure and $d\left(x, \partial W\right)$ as the censoring distance, where ∂W denotes the boundary of W. The observation is censored if $d\left(x, \partial W\right) < d\left(x, X\right)$.

From the data $X \bigcap W$ it is possible to compute $d\left(x, X \bigcap W\right)$ and $d\left(x, \partial W\right)$ for each $x \in W$.

Regarding that $\min\left(d\left(x,X\right),d\left(x,\partial W\right)\right) = \min\left(d\left(x,X\bigcap W\right),d\left(x,\partial W\right)\right)$ (Baddeley and Gill, 1997, 36), then it can indeed be observed that $\min\left(d\left(x,X\right),d\left(x,\partial W\right)\right)$

and $I\left\{d\left(x,X\right) \leq d\left(x,\partial W\right)\right\}$ for each $x \in W$. Then the set $\left\{x \in W : \min\left(d\left(x,X\right),d\left(x,\partial W\right)\right) \geq r\right\}$ can be thought of as the set of points "at

risk of failure at distance r" and $\left\{x \in W : d\left(x,X\right) = r,\ d\left(x,\partial W\right) \geq r\right\}$ are the "observed failures at distance r".

Geometrically, these two sets are the closures of:

$\left\{x \in W : d\left(x,W^c\right) > r\right\} \setminus \left\{x \in R^d : d\left(x,X\right) \leq r\right\}$ and

$\partial\left(\left\{x \in R^d : d\left(x,X\right) \leq r\right\}\right) \bigcap \left\{x \in W : d\left(x,W^c\right) > r\right\}$ respectively.

The following panel illustrated the analogy with survival times:

Illustration of Analogy the spatial distances with survival times
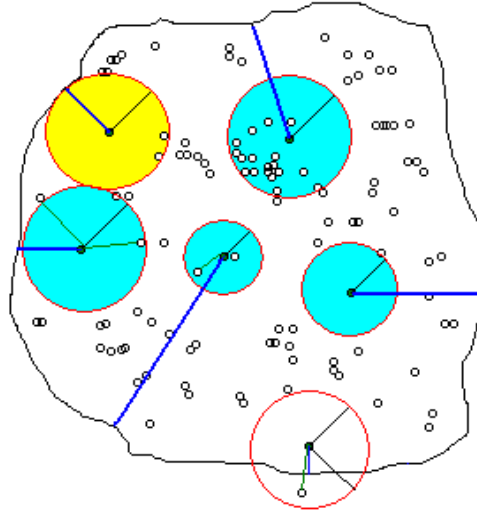
d(x,X) is the green lines, d(x, δW) is blue lines and the black is the radius ´r´ of each ball.

Now, let X be an almost surely stationary point process and W be defined in the d-dimensional Euclidean space, a fixed compact set. Based on the data $X \bigcap W$, the Kaplan-Meier estimator $\hat{F}$ of the Empty Space Function F of this point process, is defined by:

$$\hat{F} = 1 - \prod_0^r \left(1 - d\hat{\Lambda}\left(s\right)\right) = 1 - \exp\left(-\hat{\Lambda}\left(r\right)\right) \tag{23}$$

where $\hat{\Lambda}\left(r\right)$ is defined by:

$$\hat{\Lambda}\left(r\right) = \int_0^r \frac{\left|\partial\left(\left\{x \in R^d : d\left(x,X\right) \leq s\right\}\right) \bigcap \left\{x \in W : d\left(x,W^c\right) > s\right\}\right|_{k-1}}{\left|\left\{x \in W : d\left(x,W^c\right) > s\right\} \setminus \left\{x \in R^d : d\left(x,X\right) \leq s\right\}\right|_k} ds \tag{24}$$

where $\left|.\right|_{k-1}$ denotes $k-1$ dimensional Hausdorff measure (surface area).

The reduced-sample estimator of the Empty Space Function F which is the border correction is:

$$\hat{F}^{rs} = \frac{\left|\{x \in W : d(x, W^c) > r\} \bigcap \{x \in R^d : d(x, X) \leq r\}\right|_k}{\left|\{x \in W : d(x, W^c) > r\}\right|_k} \tag{25}$$

The K-M estimator of the formula (23) of F is based on the continuum of observations generated by all $x \in W$. It is see that, the estimator is a proper distribution function and is even absolutely continuous, with hazard rate $\hat{\delta}(r)$ equal to the fraction of the integral in the formula (24) evaluated at r and using the k dimensional Hausdorff measure for the integral.

3.6.1.4 Kaplan-Meier estimator of the nearest neighbour function G Observe that the function $G()$ has special continuity properties, in contrast to $F()$; in fact G may degenerate, as in the case of a randomly translated lattice.

Let $X \bigcap W = \{x_1, ..., x_m\}$ be the observed point pattern. For each point $x_i \in X$ observed in the window W, one has the censored distance from x_i to the nearest other point of X is $s_i = d(x_i, X \setminus \{x\})$ and the censored distance by its distance to ∂W is $c_i = d(x_i, \partial W)$. Then

$$G(r) = 1 - \prod_s \left(1 - \frac{\#\{i : s_i = s, s_i \leq c_i\}}{\#\{i : s_i \geq s, c_i \geq s\}}\right) \tag{26}$$

where s in the product ranges over the finite set $\{s_i\}$. Hence, the new reduced-sample estimator of $G(r)$ is:

$$\hat{G}_1(r) = \frac{\#\{i : s_i \leq r, c_i \geq r\}}{\#\{i : b_i \geq r\}} \tag{27}$$

and the modification obtained by replacing $X(\{x \in W : d(x, W^c) > r\})$ by an estimate of its expectation is:

$$\hat{G}_2(r) = \frac{|W|_k}{n} \frac{\#\{i : s_i \leq r, c_i \geq r\}}{\left|\{x \in W : d(x, W^c) > r\}\right|_k} \tag{28}$$

3.6.1.5 Kaplan-Meier estimator of the Ripley's function K $K(r)$ was defined in formula (20). Equivalently it is $\lambda K(r) = \sum_{n=0}^{\infty} G_n(r)$ where

$$G_n(r) = \mathbb{P}\{X(b(0, r)) > n | 0 \in X\} \tag{29}$$

is the distribution function of the distance from a typical point of X to the nth nearest point. For each G_n, it is possible to form a Kaplan-Meier estimator, since that this distance is censored at the boundary just as before. The sequence of K-M estimators always satisfies the natural stochastic ordering of the distance distributions.

3.6.1.6 Advantages and disadvantages The distance from a fixed point to the nearest point of the process is right-censored by its distance to the boundary of the window, it is a first advantage.

The resulting estimators have a ratio-unbiasedness property that is standard in spatial statistics, it is a second advantage.

The estimators are strongly consistent when there are independent replications or when the sampling window becomes large, it is a third advenatage.

A disadventage is according to Baddeley & Gill, that the Kaplan-Meier estimator for the functions $F()$, $G()$ and $K()$ appears to be substantially more efficient than the simple border correction (reduced-sample) under simulations. However, the K-M estimator of function K is less efficient than sophisticated edge corrections (Ripley Isotropic Estimator) but it is offset by its ease implementation in arbitrary windows.

3.6.2   Additional corrected estimates for inter-point distance methods

According to Ripley (2008, 31), there are a two problems with the estimator of $K(r)$ formula 16. First: the bias for small λ is unavoidable; it stems from the need for two points to compute any inter-point distance information. Second: the bias from edge correction is both more serious and avoidable. Ripley presents six different edge corrector estimates to solve this but only three (using in the applied methods) are describe a follows:

3.6.2.1 Ripley isotropic correction estimate of K function This correction is based on isotropy. Consider just one pair of spatial points (x, y). If the ball $b(x, d(x, y))$ is not wholly contained within W, then there could be unobserved points at distance $d(x, y)$ from x.

It is this that causes the underestimation in $\hat{K}_0(r)$, and to compensate one counts the pair (x, y) more than once, in fact $k(x, y)$ times where $\frac{1}{k(x,y)} = \frac{|\partial b(x,d(x,y)) \cap W|}{|\partial b(x,d(x,y))|}$ is the proportion of the perimeter of the circle which is within W and $\partial b(x, d(x, y))$ is all the points which are in the ball $b(x, d(x, y))$. Note that in most cases $k(x, y) = 1$ since x will not be within distance r of W^c, at least for small r. Define

$$\hat{K}_2(r) = \frac{1}{\lambda^2 a} \sum k(x, y) I[0 < d(x, y) \le r] \tag{30}$$

to be the unbiased estimator with a window W convex and $r \le r_0$ the circum-radius of W. Considering (4) $\mathbb{E}\left[\lambda^2 a K_2(r)\right] = \int k(x, y) I[0 < d(x, y) \le r] d\mu_2(x, y)$

one has $\mathbb{E}\left[\lambda^2 a K_2(r)\right] = \lambda^2 \int_0^\infty \left\{\int k(x, y) I[0 < d(x, y) \le r] dv_r(x, y)\right\} dK(s)$

or $\mathbb{E}\left[\lambda^2 a K_2(r)\right] = \lambda^2 \int_0^t a \, dK(s) = \lambda^2 a K(r)$ for $r \le r_o$

The effect of the correction factor k is to rescale the σ_s terms in ν_s. This is possible if $W \cap \partial b(x, s)$ has positive length for each $x \in W$. For a convex set this will hold for $s \le r_0$.

Unbiasedness up to r_0 will usually suffice; for a disc it is the radius and for a rectangle half the length of a diagonal. Nevertheless, Ohser (1983) extended the range of unbiasedness by redefining $\hat{K}_2'(r)$ as:

$$\hat{K}_2'(r) = \frac{1}{\lambda^2} \sum \frac{k(x,y) I[0 < d(x,y) \leq r]}{\nu\{x|\partial b(x, d(x,y)) \cap W \neq \emptyset\}} \tag{31}$$

This rescales the ν_s term to a provided $\nu\{x|\partial b(x,s) \cap W \neq \emptyset\} > 0$, and so $\hat{K}_2'(r)$ is unbiased for $t \leq t_1$, the diameter of W, the diameter of W which is convex. Note that $\hat{K}_2'$ and K_2 agree for $r \leq r_0$ (Ripley 1988, 32-33).

3.6.2.2 Translation-corrected estimate of K function Ripley (1988, 34), Ohser and Stoyan (1981) also proposed a correction based on the proportion of translations of $(x,y)/x \in W \wedge y \in W$. Call this $\theta_E(x-y)$ and note that it is symmetric, $\theta(-h) = \theta(h)$.

Then $\hat{K}_4(t) = \frac{1}{\lambda_2 a} \sum \theta(x-y)^{-1}[0 < d(x,y) \leq t]$ which is again unbiased.

3.7 Fiting Models

As mentioned in the beginning of section 2.3., the spatial statistics analysis is started by applying some methods under the homogeneous Poisson process. The purpose of this was to find the distribution of studied patterns which characterize the data. When the CSR hypothesis is rejected, this means the pattern could be clustered, or regular, or both. Fitting some alternatives (parametric) models is a way to provide more definition and clarity. After a model has been fitted, diagnostic tests should be performed to assess its goodness-of-fit. Finally, inference for the estimated parameters is often needed in response to a specific scientific question. A current theoretical distribution for the estimators can be difficult to obtain, in which case approximations may be necessary. At the very least, a bias and variance should be obtained. If the point patterns are not CSR there are three more options: clustering, regular or a combination of both. In the literature, several spatial models were found to model spatial data sets, accordingly, these three options:

- in case of clustered distribution one can use an inhomogeneous Poisson process or a Cox process or a Poisson cluster process,

- in case of regular distribution, one can use a simple inhibition process

- The Markov point process is a model which incorporates both patterns.

4   SECTION

4.1   Data Description

4.1.1   Selection of the study area

Three sites located in the Democratic Republic of the Congo were initially considered for spatial point pattern data analysis: Bimpe (see Appendix A), Lisala and Lukolela, all of which are located within the Congo Forest Basin (that comprises 45% of the Democratic Republic of Congo). However, due to differences in structure and composition of the forests in the three areas, a final selection of one site was made based on recommendations from experts who took into consideration key factors such as forest intactness and planned future works. The Mpata Mbalu forest, within the Bimpe forest concession, was chosen for the following reasons: (1) the timber extraction has been suspended in the concession and (2) the concession has been allocated for conservation. This ensures that the analysis conducted herein will be used for conservation management purposes (Bwangoy, personal communication). The position of trees, recorded using a Global Positioning System (GPS) in spherical coordinates (latitude, longitude), were converted to projected coordinate systems (Universal Traverse Mercator – UTM - zones 33 and 34). The data was provided in a Geopanelic Information System (GIS) format as Environment Systems Research Institute (ESRI) shape file format. All files were uploaded into the R package wherein analyses were done on the attributes of each area

The shapefiles contain the (x, y) coordinates representing locations of each commercially valuable dominant canopy tree which are the principal data for this study. They also contain attributes pertaining to each tree such as: species name, block name and others attributes, all of which are stored in the shapefile (*dbf* subfile). To facilitate analysis in R, the tree (x, y) coordinates were joined with the attributes in the "dbf" file.

In personal communication with the forester, (Bwangoy Bankanza) who designed the forest inventory and supervised and carried out quality control checks at the end, it was ascertained that the error never exceeded 5% (i.e. missed trees did not exceed 5% and were normally in the 1 to 2% range). The crews were well trained and supervised.

4.1.2   Description of the study site (Bimpe Concession)

The Bimpe Concession is located in the Lake Mai Ndombe District on the western shore of the Lake Mai Ndombe. The terrain is flat with nearly 50% of the area being swamp forest (See appendix A). The climax forests are both dense, humid evergreen and semi-deciduous (some of the trees shed their leaves during the dry season and some do not). Both forest climax types are species diverse (over 50 different species per hectare). These forests are part of the Congo lowland equatorial forest biome.

The study area excluded flooded forest (also referred to as swamp forests) and considered only 'terra firma' forest (those forests that are not flooded during the rainy season).

The climate is humid with an average annual rainfall of approximately 1800 mm per year recorded in Selenge, a community located in the mid-section of the Bimpe forest concession and just north of the actual study area. The region is characterized by two rainy seasons (one major[3] and one minor) and two dry seasons (one major and one minor). The major rainy season starts in September and ends in November and the minor one starts in March and ends in May of each year. The major dry season starts in June and ends in August, while the minor dry season starts in December and ends in February of each year. Soils are poor with little organic matter in the mineral horizons. Tree roots are shallow near the surface (upon which one finds the litter-humus soil layer, which is also referred to as the forest floor) so to be able to uptake nutrients from the rapidly decomposing organic matter found in the forest floor (i.e. known as detritus coming from litter fall). The commercially valuable trees are usually very large dominants, with heights often exceeding 30 m.

The data points consist of all inventoried (100% sample) commercially valuable trees. These trees are of species that have recognized commercial valuable. They are ground- marked for harvesting, with their (x, y) coordinates recorded along with the species and diameter at breast height attributes. Some of the tree species may be clustered as opposed to randomly

dispersed due to seed dispersion vectors and other ecological and geomorphological factors.

4.1.3 Description of forest blocks to be analyzed

The Bimpe data set is distributed across seven blocks (Figure 1) situated on *terra firma* forests of commercial value surrounded by swamp forest areas. This study began analyzing all the blocks as a complete undivided data set, but the first observations created the necessity to continue the spatial analysis on a block by block basis. The analysis first took a mathematical approach vis-à-vis the characterization of the spatial distribution of commercially valuable trees. It was focused so that eventually some aspects of the mathematical analysis could have a forestry dimension, which, in turn, could lead to possible forestry applications. For each block, the attributes (i.e. (x, y) coordinates, species, diameter at breast height, etc... as contained in a dbf file) from the existing data set at the time of collection were used. Using the R statistical package, we generated discrete area polygons from the existing forest blocks. Each polygon has a discrete set of points. The polygon borders were drawn as closely as possible to the actual Mpata-Mbalu terra firma forest management block boundaries (part of the Bimpe forest concession) and each polygon contains all of the data set trees of these forest management blocks (it should be noted that Block I boundaries were subsequently modified as part of this study).

---

[3]The major rainy season is the one wherein most of the annual precipitation occurs

The words tree(s) and point(s) are used interchangeably in the analysis to describe the spatial point patterns of this data set.

The main unit of measure of forest area is the hectare or $10,000\ m^2$. Other units used in the study were either fractions (i.e. 50X50 meter quadrats or $\frac{1}{4}$ of a hectare) or multiples (i.e. 2 hectares) of that standard unit of measure, with the object to see whether the results from the spatail statistics methods (QuadratCount) was or not dependent of this unit of measure using. All main variables are usually expressed as units per hectares (ex. density is expressed in trees per hectare)

In this study the standard unit of area measure was the 'hectare', whose symbol is 'ha'. The following are the main reasons it is used:

1. The Democratic Republic of Congo adopted the metric system in 1910 and the hectare is metric unit of measure. The hectare is the measurement unit of area the most widely used throughout the world. In the European Union and others countries (DRC), it is the legal unit of measure in domains concerned with land ownership, planning, and management, including law (land deeds), *agriculture, forestry* and town planning.

2. One of the principal variables was crown diameter (range was 6 to 30 meters - according to Asner et al, 2002, 486)) and the 1 hectare quadrat (100m X 100m) is a suitable size for visualizing tree crowns spatial distribution at various densities/intensities.

Study results need to use units of measure that are in standard use by forest ecologists and managers.
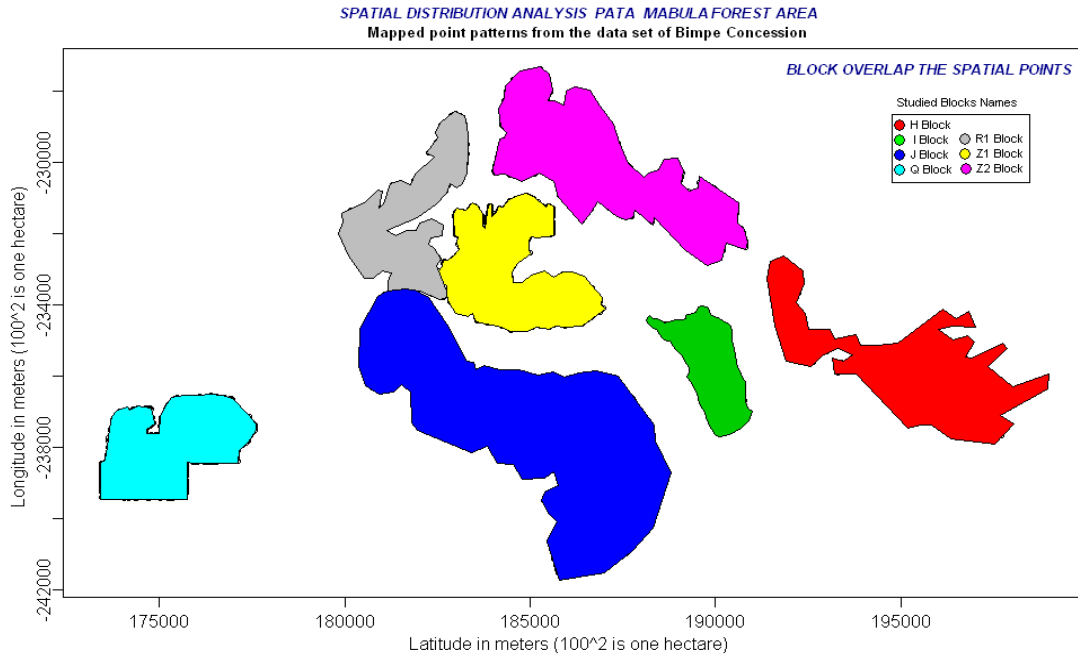
Figure 1: Presentation of the blocks/forest management units of the Bimpe Concession (Pata-Mabula forest area) that contain the mapped point patterns used in this study. Each block's closed polygon area is distinguished by a specific colour. Block H is red, Block I is green, Block J is blue, Block Q is light blue, Block R1 is light purple, Block Z1 is yellow and Block Z2 is gray.

4.2   Investigating the Intensity

This analysis started by doing an exploratory analysis of the 'intensity' (i.e. the number of points per unit of area, such as points per hectare). 2 'intensity' alternatives were investigated: uniform (homogenous) or non-uniform (non-homogenous). In order to investigate the nature of the intensity of the points in the data set, we used the following three methods from the packages spatstat and stat of the R Tool Software: **1 -** the summary of point patterns '`summary(P)`'; **2 -** Kernel Density Estimation with sigma 95 '`density(P,sigma=95,edg=TRUE)`' with the Quadrat Counting for a point patterns methods '`quadratcount(P,nx=dx,ny=dy)`'; and **3 -** the Chi-squared test of CSR '`quadrat.test(P,nx=dx,ny=dy)`'.

Table 1 Points Pattern Summary for the Bimpe Forest Concession Area

	Planar point patterns	Number of vertices	Polygon Area	Average intensity	Coordinate down-left		Coordinate up-rigth	
Blocks	# of points	Polygon Area	in hectares	per hectare	x (meters)	y (meters)	x (meters)	y (meters)
H	944	48	1371.08	0.69	191370	-237900	198990	-232620
I	158	77	480.38	0.33	188130	-237700	190980	-234020
J	3017	47	2964.12	1.02	180380	-241730	188790	-233410
Q	2651	33	864.38	3.07	173410	-239450	177620	-236500
R1	2260	56	1535.73	1.47	183990	-232880	190880	-227300
Z1	2728	68	1013.64	2.69	182530	-234750	187040	-230870
Z2	891	55	823.61	1.08	179840	-233860	183340	-228560

Description: It presents a point patterns summary of the Bimpe Concession in the Pata-Mabula area per block, where the last 2 columns shows the coordinates of the smallest rectangle of each block given by the function summary().

The results of this exploratory analysis provided some insight as to the nature of the spatial distribution of points. In the next step, we carried out a formal analysis of the nature of the spatial distribution was carried out. This started with the application of the Complete Spatial Randomness (CSR) test. In this test, *the null hypothesis is that the spatial point process distribution follows a uniform Poisson point process (CSR).* For this kind of data set, it would not be surprising to have a mix of spatial point process distributions.

The following list presents the observations and preliminary conclusions of the results of the initial spatial analysis from the three statistic methods listed above. These results are presented in various figures and tables (within the document itself or as appendices to the document) specified in the list according to its necessity.

- Block I has the smallest number of the points. H and Z2 Blocks have similar number of points (approximately 900) and blocks J, Q, R1 and Z1 have a relative high number of points (between 2200 to 3020 - see Table 1).

- The high number of vertices within the polygon areas indicated that the approximated boundaries are very irregular (see Table 1). It was noted that the polygon areas defined using the R package followed closely the physical boundary of the real blocks of the Bimpe Forest Concession Map (see Appendix A). Very irregular boundaries are an indication of the presence of natural phenomena such as adjacent swamp forest, while less irregular boundaries indicate that the adjacent areas are external anthropogenic interventions such as logging roads, constructed or planned, and/or property/adjacent concession boundaries. (personal communication with Bwangoy Bankanza, the forest scientist who designed the timber sampling and block layout plan)

- Concerning polygon areas, the Block Q is of particular interest due to a relatively high number

of points (third highest total; 2651) for a relatively small area (third smallest total area). With an average intensity of 3.07 point per hectare, it is reasonable to speculate the presence of a clustered distribution and the absence of historical logging of commercially valuable species. In contrast, Block I, has the smallest area (376.82 hectares) and also has the smallest average point intensity (less than 0.42 points per hectare), which may be an indication of a high dispersion probability or 'thinning' (Baddeley A. et al., 2004).

- Block J is the biggest block in number of points (3017) and in area (2964.12 hectares). Consequently, the isotropic and translation edge corrections were not computed, following a Baddeley et al recommendation that a border method is much faster and statistically efficient for large numbers of points in a data set.

- Blocks Z1 and Q have the highest average intensity of all the blocks (see Table 1), however they potentially have both localized clustered distribution, as well as regular distribution (see Appendix C). Hence, it may be advisable to divide the block according to this suspected pattern and to analyze the two sub areas independently.

- The p-values from the Chi-Squared test of CSR '`quadrat.test(P,nx=dx,ny=dy)`' using Quadrat Counts Method rejected the null hypothesis for all blocks with exception of Block I where the p-value was greater than 5%. Ignoring the warning message from the Chi-squared test (principally for the irregularly boundary of the blocks), the cuts were made for 1 hectare-sized quadrats. This size was used because it is a common measure in forestry (see Table 2). Intuitively, it was assumed that the spatial points of Block I should follow a uniform Poisson distribution (CSR). Additionally, the Chi-Squared test was applied with 4 and 4.5 hectare sized quadrats to verify the independence of the p-value with respect to the size of the quadrat. The results showed a p-value in Block I which indicated particularly high variance. As such, Block I needed to be analyzed in more detail.

Table 2 Chi-squared test of CSR using quadrat counts, until one quadrat is equal one hectare

Blocks	# of hectares	# of cuts		X-squared	df	I	p_value	Ho
		dx	dy					
H	1371.08	76	52	1918.66	1522	1.26	1.40E-11	rejected
I	480.38	28	36	475.58	527	0.90	0.9472	not rejected
J	2964.12	84	83	4061.49	3121	1.30	2.20E-16	rejected
Q	864.38	42	29	1947.74	914	2.13	2.20E-16	rejected
R1	1535.73	68	55	4047.73	1645	2.46	2.20E-16	rejected
Z1	1013.64	45	38	7102.75	1088	6.53	2.20E-16	rejected
Z2	823.61	35	53	1138.34	947	1.20	1.69E-05	rejected

Description: It presents the result of the Chi-quadrat test for each block. The first column has the name of each block. The second column is the number of hectares (area) per block. The third column and fourth column is the number of horizontal and vertical cuts respectively. The fifth column shows the statistic x^2 value per block. The sixth column shows the degrees of free. The seventh column shows the x^2 Index of Dispersion, the eighth column show the p-value of the Chi-Quadrat test and the nine column shows whether the null hypothesis was rejected or not

With these observations and preliminary conclusions, it was possible to develop a path to study the Spatial Point Patterns of the Bimpe Concession data set. The study of Spatial Point Patterns focused on finding what kind of point patterns (random, cluster and regular) characterized the data set. This represents an analysis of the space, distance and interaction between points. The methods that were used were: the F function - 'Estimate of the Empty Space' (Fest is the name of the function in R), the G function - 'Nearest Neighbour Distance' (Gest is the name of the function in R) and the K function - 'Estimate of Ripley's reduced Second Moment' (Kest is the name of the function in R). To validate these function results, it was necessary to use the 'simulation envelope' method (envelope is the name of a function in R), which is similar to the one used in the Monte Carlos test (Schabenberger et al 2005) for obtaining simulation envelopes. Also, it is necessary to mention the sensitivity of this test in relation to the error-I type. The theoretical distribution function is simulated under CSR distribution, but in cases it may be necessary to follow other hypotheses as in the case of 'Q' Block and Block Z1 (See Appendix C). Hence, the null hypothesis 'H_0' is: *The point patterns of the sample follow a CSR distribution* and the alternative hypothesis 'H_a' is: *The point patterns of the sample do not follow a CSR distribution*, in each case.

Accordingly for a preliminary analysis of the blocks, they were divided in three groups. These groups are:

1. Block I

2. Blocks H, R1 and Z2

3. Blocks J, Q and Z1

Block I was analyzed as a model to show the step by step results of the spatial methods selected to find the spatial distribution. In the second and third groups the estimates of function J and Pair

Correlation Function g were the only tests applied which summarize the results of the F, G and K functions. Three groups were used to analyze the assumption of aggregation described in Section 1.

4.3   Spatial Point Patterns for Block I

Block I shows high dispersion (See Appendix C) as the 158 spatial points are spread out over 480.38 hectares. As mentioned previously, the Chi-Quadrat test could not reject the null hypothesis when the number of quadrants was big. This suggests that the point processes follow a CSR. Even with this evidence, the error-II type of the null hypothesis was possible to occur when the warning message is ignored in the Chi-quadrat test. However, given that the boundary of all the blocks is very irregular, it can be expected that quadrats on the edges will have no points.

4.3.1   Estimation of the empty space function F

Figure 2 shows the curves of the Empty Space Function F using the `Fest` function from the `R spatstat` package. When `Fest` function is plotted, the x axis represents the radius of the estimate of $F(r)$ (whose recommended range is $[0, 170.62]$) and the y axis represents the $F(r)$ function in the two panels . The `Fest` function gives more than one estimate of $F(r)$. The Kaplan-Meier estimate curve (black), the Border Corrected Estimate curve (red) and the theoretical Poisson curve (green - based on the estimated intensity) is shown in panel 'a'. The Simulation Envelopes of the Summary Function was applied to the Kaplan-Meier Estimate (See panel 'b'), with 99 simulations (5% level) under CSR and a 0.01 significance level of a point wise Monte Carlo Test.

Figure 2: Spatial Distribution Analysis of Block I from the Pata-Mabula area in the Congo Basic rainforest F function contains 2 panels: a) The Estimate the empty space function F 'Fest' which panels the border corrected estimate 'Fbord(r)', the Kaplan-Meier estimate 'Fkm(r)' curve and the Theoretical Poisson of k(r) curve 'Fpois(r)' and b) The Simulation Envelope F function which presents the cumulative empirical distribution of the observations 'obs(r)', the cumulative theoretical distribution under CSR, the highest simulation envelope under CSR 'hi(r)' and the lower simulation envelope under CSR 'lo(r)'. Both functions are from the spatstat package developed by Baddeley in the statistic R Tool.

When the radius is 7.11 meters, the first contact distance appears, that means the cumulative empirical distribution have value greater than 0 (information having by the summary of function Fest). When the radius is greater than 70 meters; it is possible to observe the cumulative empirical distribution approaching close to the lower simulation envelope, according to the deviation between the Kaplan-Meier estimate function and the theoretical Poisson function. Additionally, looking at the results of the 'a' panel (of Figure 2), the observations curve 'Km(r)' is always below the theoretical curve with the only exception being when the radius is between 14.22 and 42.65 meters. The tendency of the observation curve to be below the lower simulation curve when the radiuses are ascending, may suggest the presence of clusters. However, it can be concluded from the results that: the null hypothesis cannot be rejected because the cumulative empirical distribution of the observations in 'b' panel (distances or radiuses) falls within the CSR simulation envelope.

4.3.2 Estimation of the nearest neighbour distance function G

Whether the null hypothesis from the empty space function (or contact distance function) was not rejected, then there is a high possibility that the Nearest Neighbour Function G will produce a similar result. Once again, the Figure 3 presents two panels: Nearest Neighbour Distance function G and Simulation Envelopes to G function under the same conditions, but the recommended range of the radius is between 0 to 131.69 meters.
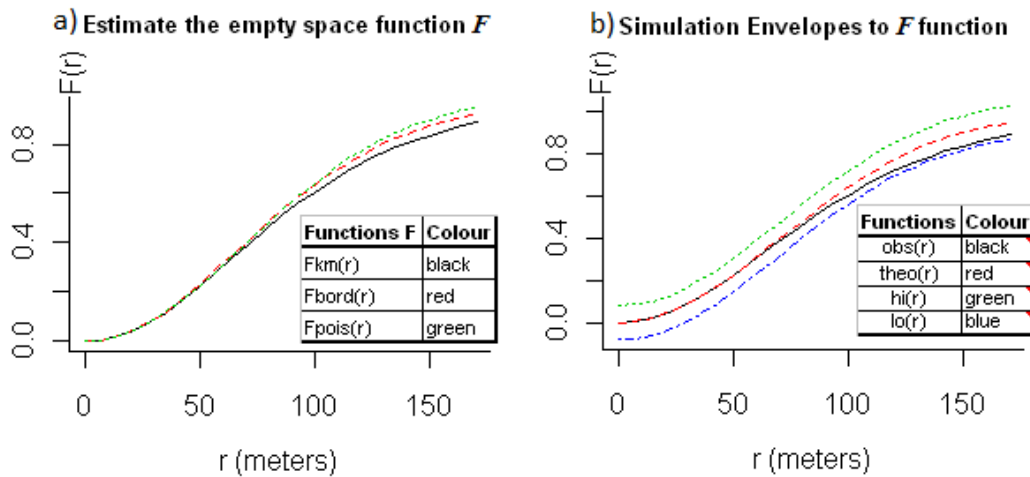
Figure 3: Spatial Distribution Analysis of Block I from the Pata-Mabula area in the Congo Basic rainforest G function contains 2 panels: a) The Estimate the empty space function G 'Gest' which panels the border corrected estimate 'Gbord(r)', the Kaplan-Meier estimate 'Gkm(r)' curve and the Theoretical Poisson of g(r) curve 'Gpois(r)' and b) The Simulation Envelope G function which presents the cumulative empirical distribution of the observations 'obs(r)', the cumulative theoretical distribution under CSR, the highest simulation envelope under CSR 'hi(r)' and the lower simulation envelope under CSR. Both functions are from the spatstat package developed by Baddeley in the `statistic R Tool`.

Regarding the results in Figure 3, the null hypothesis is again not rejected because the cumulative empirical distribution of the observations is falling within the simulation envelope under CSR. In the panel 'a', *the nearest neighbour did not appear until after a radius of 24.77 meters*, showing that there could be an inhibition between points. However, this needs to be proved. Panel 'a' shows that, between the radiuses of 0 to 43.68 meters, the cumulative empirical distribution (black or red) is almost completely below the cumulative theoretical distribution (green). The biggest distance between both curves is 0.07, which occurred when the radius was about 34 meters. This indicated the presence of clusters (according to the Simulation Envelope definition). However, in panel 'b', the curve of the observations continues inside the boundary of the simulations envelope under CSR (hi and lo). This indicates that the null hypothesis cannot be rejected and the presence of clusters is discarded.

4.3.3   Estimation of the reduced second moment K function

The next function that needed to be applied was the K function 'Kest'. It was used for analyzing the interaction between points and their impacts in the patterns of Block I. The K function is the Ripley's Reduced Second Moment Function which analyses the interaction between a given 'x' point of the process and all the points inside a circle with a radius 'r'.

The theoretical function $K(r)$ begins with a radius of 1.4 meters, which is considered to be the first radius for the circle of interaction. However, the first interaction in the observations (or empirical function) does not appear until around r = 25 meters which is 18 times greater than the first radius

(sub-section 4.3.2).

The presence of deviations between all the estimates and the `Kpois(r)` curves in 'a' panel suggests spatial clustering because the various estimate curves tend to be greater than the theoretical curve (Cressie, 1993). In the 'b' panel, the observations curve is under the CSR simulations envelope, but when the radius is equal to and greater than 297 meters, the observations curve is greater than `hi(r)` curve, *hence the null hypothesis cannot be accepted and it is possible to have spatial clustering with radiuses equal to and greater than 297 meters* (see definition of Simulation Envelopes). Another possible conclusion could be that there are dependent interactions between the points inside the ball when radiuses are equal to or greater than 297 meters.
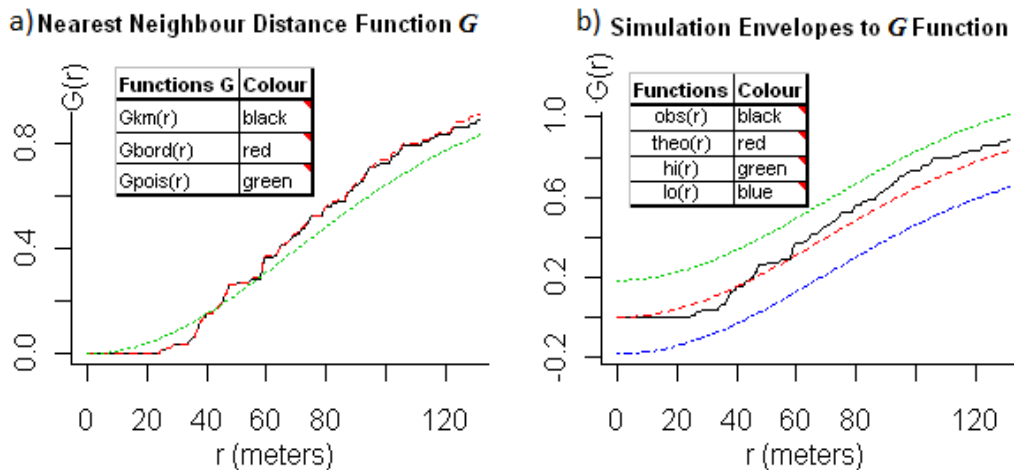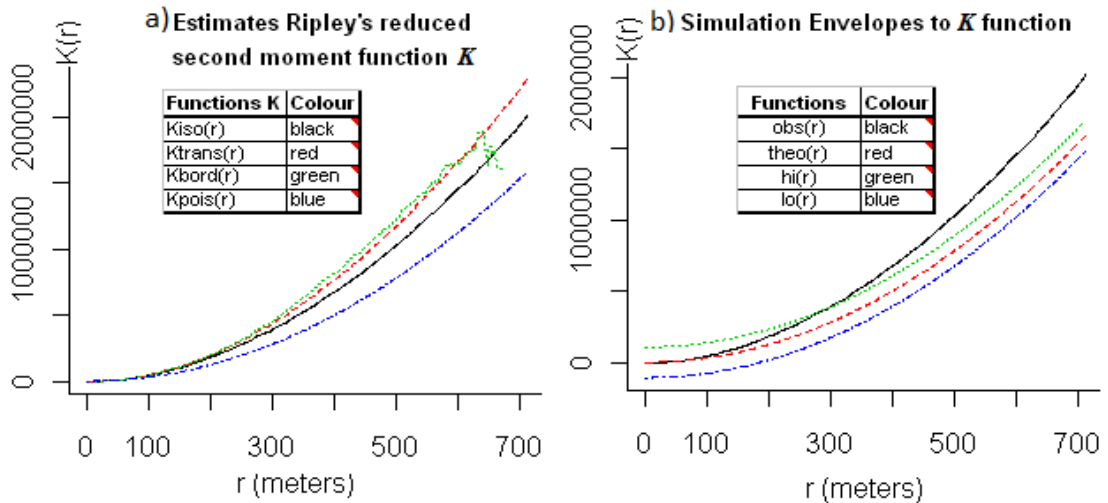

Figure 4: Spatial Distribution Analysis of Block I from the Pata-Mabula area in the Congo Basic rainforest K function shown in 2 panels: a) The 'Kest' function panels the 'Kiso(r)' which is the Reply isotropic correction estimate of K(r), 'Ktrans(r)' which is the translation-corrected estimate of K(r), 'Kbord(r)' which is the border-corrected estimate of K(r) and 'Kpois(r)' which is the theoretical value for a stationary Poisson process and b) The Simulation Envelope to K function which presents the cumulative empirical distribution from the observations 'obs(r)', the cumulative theoretical distribution under CSR, the highest simulation envelope under CSR 'hi(r)' and the lower simulation envelope 'lo(r)' under CSR. Both functions are from the spatstat package developed by Baddeley in R statistic Tool.

4.3.4 First and second moment characteristic analysis for I

To support or to discard the results from the Estimates of the Empty Space Function F and of the Nearest Neighbour Distance Function G (First Moment Characteristic), it is recommended to apply the Estimate the Function J in the spatial analysis of the data set. With the same criterion, it is also recommended to use the Pair Correlation Function$g$ to support or to discard the results of the Estimate of the Function K (Second Moment Characteristic).

4.3.4.1 Estimate the function J   According to section '3.4.3.' and formula (19), the $J(r)$ function of a stationary point process is defined as: $J(r) = \frac{1-G(r)}{1-F(r)}$ where $G(r)$ is the nearest neighbour distance distribution function of the point process and $F(r)$ is its Empty Space Function. For a completely

random (uniform Poisson) point process, the $J(r)$ Function is identically equal to 1. Deviations when $J(r)$ is greater than 1 suggests spatial regularity (inhibition) and when $J(r)$ is less than 1, it suggests spatial clustering (aggregation).

Figure 5 shows the J function on panel (a) and the Simulate Envelopes to J function on panel (b) . For both panels, the x axis represents the radius. The recommended range of the radius r is between 0 to172.0001 meters for this observation in particular. The y axis represents the $J(r)$ function in both panels. The panel on panel (a) presents the following three estimates of the J function: the Kaplan-Meier Estimate (black), the Border Corrected Estimate (red), and the Uncorrected Estimate (green), as well as the theoretical Poisson curve J (blue). The Simulation Envelopes of the Summary Function was applied to the Kaplan-Meier Estimate, panel (b), with 99 simulations (5% confidence level) under CSR with a 0.01 significance level of point wise Monte Carlo Test.
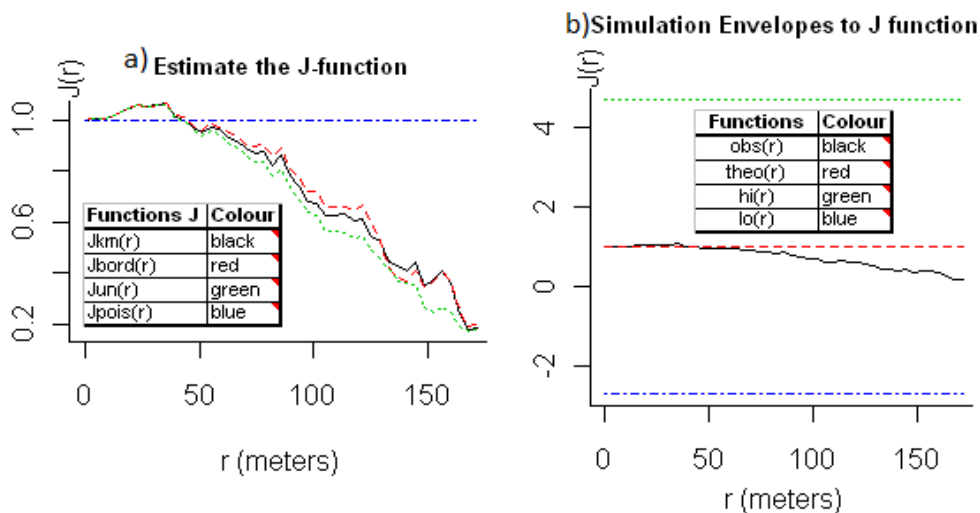
Figure 5: Spatial Distribution Analysis of Block I from the Pata-Mabula area of the Bimpe Forest Concession. J function contains 2 panels: **a)** The J-Function 'Jest' which panels The Kaplan-Meier estimate of J 'Jkm(r)' curve, the reduced-sample or border corrected estimate of J 'Jbord (r)' curve, the uncorrected estimate of J "Jun(r)" and the Theoretical Poisson of J curve 'Jpois(r)' and **b)** The Simulation Envelope J function which presents the cumulative empirical distribution of the observations 'obs(r)', the cumulative theoretical distribution under CSR, the lower critical boundary under CSR 'hi(r)' and the lower critical boundary under CSR. Both functions are from the spatstat package developed by Baddeley in the statistic R Tool.

In analyzing the results from the 'a' panel, it was observed that the first radius greater than 0 is 3.91 meters for all the J functions and this value is used to generate the next radiuses. Between the radiuses of 3.91 and 39.09 (10 times the first radius) meters, the empirical estimate curves track over the theoretical curve and it is possible to see two peaks in this range of radiuses and the greater of these two peaks is 0.06 units (when the radius is 31.27 meters). This is a small value, which suggests that the presence of inhibition should be discarded. For the rest of the observations, (i.e. between 43.00 to 172.00 meters, the maximum recommended radius), the empirical estimate curves are under the theoretical Poisson curve, suggesting the presence of clustering. The lowest point (0.17) from the

Kaplan-Meier estimate has a distance of 0.83. As such, the presence of clustering could be rejected. In conclusion, considering the simulations envelope to J function panel 'b' (Figure 5), the deviations in the observations are small and accordingly all observations fall within the CSR region. Hence, it is possible to discard the presence of inhibition and clustering in Block I. One should remember that the recommended range of radiuses in the 'a' panel is used for the Simulation Envelope Methods to J function, confirming that Block I has a CSR distribution. The only remaining analysis for this block is the Pair Correlation Function Method.

4.3.4.2 Pair correlation function g under CSR   According to sub-section 3.5.2, for a stationary Poisson process, the pair correlation function is identically equal to 1. When $g(r)$ shows values less than 1, it suggests inhibition between points; but when $g(r)$ values are greater than 1, it suggests clustering. The Simulation Envelopes were applied to the '`pcf`' functions, as well as the F, G and K functions.
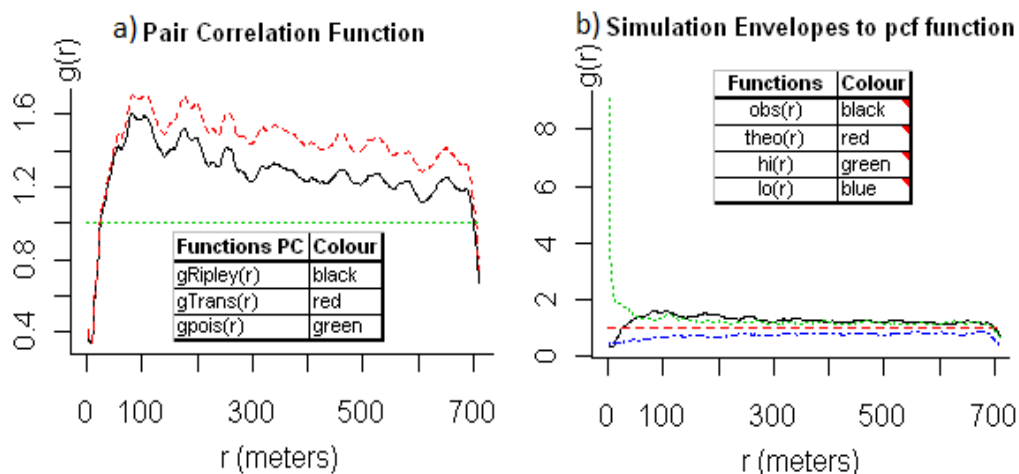
Figure 6: Spatial Distribution Analysis of Block I from the Pata-Mabula area. Pair Correlation function contains 2 panels: a) The Pair Correlation Function 'pcf' which panels The Ripley-corrected estimate of g '`gRipley(r)`' curve, the Translation-corrected estimate of g '`gTrans(r)`' curve and the Theoretical Poisson of g(r) curve '`gpois(r)`' and b) The Simulation Envelope g function which presents the cumulative empirical distribution of the observations '`obs(r)`', the cumulative theoretical distribution under CSR, the lower critical boundary under CSR '`hi(r)`' and the lower critical boundary under CSR. Both functions are from the spatstat package developed by Baddeley in the statistic R Tool.

Both panels 'a' and 'b' of Figure 6 show a recommended range of 'r' radius between 0 and 710.93 meters and the first radius for the circle of interaction is 1.39 meters. These values of curves and the first radius for the circle of interaction were in the same range obtained with the K function which shows the relationship between functions. Taking into account the two estimated empirical values and the theoretical Poisson value in the 'a' panel, when the radius is between 26.38 and 699.82 meters, the $g(r)$ values are greater than 1, which suggests the presence of clustering between points. However, the calculated maximum difference is 0.61 (when the radius is 81.92 meters or 59 times the

first radius), suggesting that the presence of clusters still needs to be proven. Regarding the results of Simulation Envelope of the Function K (Figure 4, 'b' panel) for radiuses less than 297.14 meters, the observations curve is within the CSR region which could suggest that the presence of clustering could occur between 297.14 and 710.93 meters of radius.

In the 'b' panel, the empirical curves are under the theoretical curve in the left extreme g(r) when the radius of the circle interaction is between 1.29 and 12.50 meters which implies the presence of spatial point patterns regularity. However, since the values in the Estimate of the Function K do not appear until 24.99 meters of radius, then the assumed regularity is discarded.
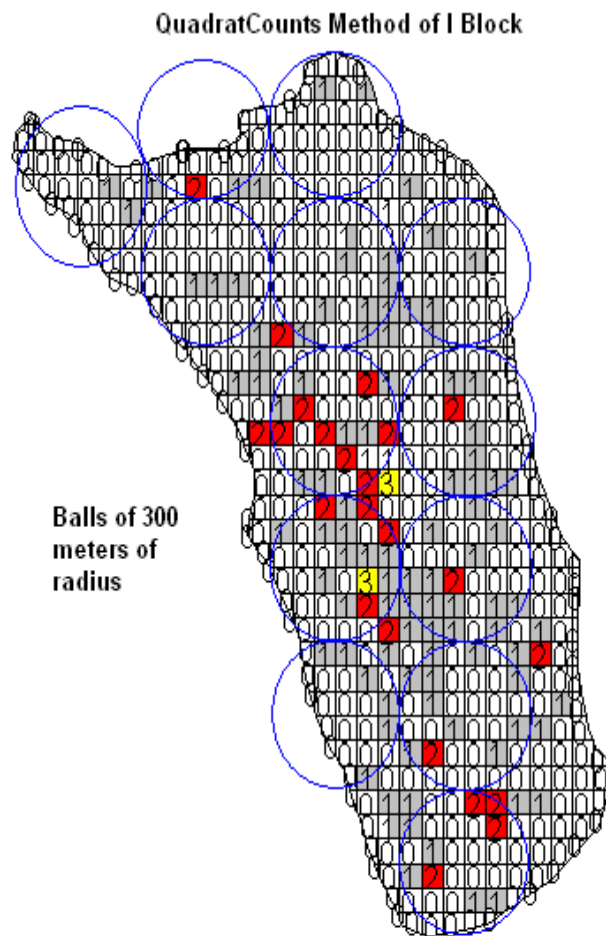
Figure 7: QuadratCount Method from `I' block from the Pata-Mabula area. This figure shows a division of 28 (in x axis) per 36 (in y axis) hectares with the number of points in each quadrat (1 hectare).

Hence the presence of clustering is apparently significant as shown in Figure 7, where the circles have radiuses greater than 297.14. Looking closely at Figure 7, where the circles of interaction all have a radius of 300 meters , it is possible to observe one circle with the most interactive points (19 points) at the center of the Block I and circles with lesser number of points (more than 2 points) throughout the block. Near the northern edge there is one circle with just 2 interactive points. In conclusion,

the null hypothesis cannot be completely accepted for the Second Moment Characteristics and it is necessary therefore to explore Fit Models for Block I.

4.3.5 Fit models for block I

According to Cressie, Baddeley et al, Illian et al and Schabenberger et al, if the null hypothesis of the CSR test is rejected, then the spatial point patterns should have a clustering or regularly tendency or both. Different authors have developed different models to fit the distribution of the data set being studied. Some of these models are Inhomogeneous Points Process, Cox Process (cluster), Poisson Cluster Process and Simple Inhibition Process (regular).

Spatial points from Block I have a CSR distribution under the First Moment Characteristics (intensity), but the average intensity is 0.33 points per hectare, and accordingly the interaction between points should not exist. However, when you apply the methods (Function K) under the Second Moment Characteristics, the null hypothesis is rejected because there is a clustering tendency when the interaction circle radiuses increase. This 'controversial' behaviour of Block I suggests, that one should ignore the potential new models and find the factors that are affecting the uniform Poisson distribution of these spatial points. In another words, in this case, the Fit-Model was using methods where if the null hypothesis is not rejected, the observations show a CSR distribution. Looking at Block I, (see Figure 7) it is possible to make the following observations:

- In the center-west part of the block there is the biggest concentration of the quadrats (one hectare) with points.

- In the northern part of the block there are more quadrats without points and there are two points with distances of contact to other points, greater than 300 meters.

- In the southeastern part, around the edge, there are no quadrats with points.

- In the northwestern corner, there are quadrats without points that extend 400 meters from the last point to northwestern tip of the block.

- Around the eastern edge there are no quadrats with points.

With these observations and knowledge of the study area forests, such as the distribution patterns of swamp and 'terra firma' forests and the tendency for commercial value trees not to be found right on the edge where swamp forests meet terra firma forests, it is possible to suggest that the empty space around the border can be a factor as to why the null hypothesis (spatial points are under CSR distribution) was rejected when Second Moment Characteristic was used. To demonstrate this assumption two Fit methods were used: The Cuts Method and The Exclusion-Edge Methods.

4.3.5.1 Results of the cuts method and the exclusion-edge method for block I The Cuts Method: Block I empty edge was progressively cut until the observation curve tended to fall under the CSR region without rejecting any points for the Second Moment Characteristics. Regarding the cuts, if the observation curve (empirical curve) maintained its original track or increased its slope in relation to the CSR region (using the K Function in the Simulation Envelope Method), the cut was discarded. In this case, there were four potential cuts (see 'a' panel in Figure 8), but only two showed influence without changing the recommended range of Function K. Hence, the observations curve fell within the CSR region up to a radius of 337.41 meters, whereas without the cut, the curve stayed inside the CSR region up to a radius of 270.41 meters. However, the null hypothesis was rejected because there were observations that fell outside the CSR region as shown in the 'a' panel in Figure 9.

The Exclusion-Edge Method: Reviewing again the observations and results for Block I, there was one cut missing in the northern part, but there was significant empty space between two remote spatial points and the rest of the spatial points. Additionally, the map from the Bimpe Concession shows a human intervention (road) bordering the northern edge of Block I. Taking this into consideration, this method consisted in using the remote spatial points as the limit and, in doing so, drawing back the edge of Block I by excluding significant empty spaces on the edge. In the case of the northern edge, this also involved the exclusion of two isolated remote points.
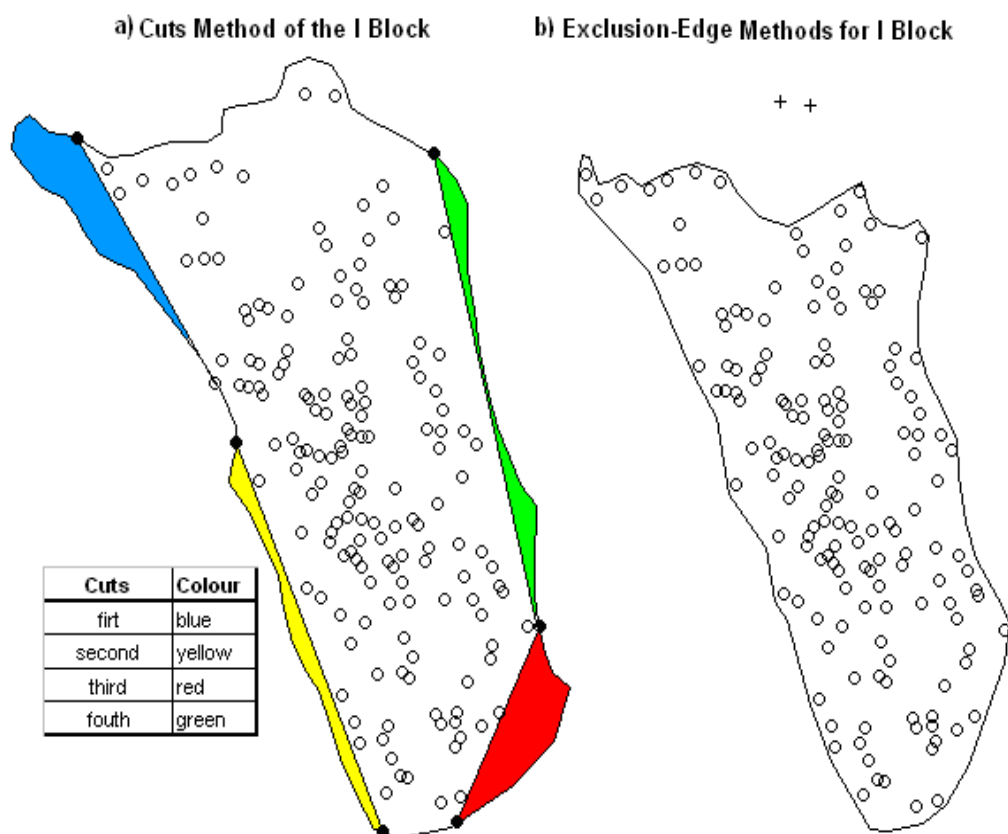
Figure 8: Cuts Method and Exclusion-Edge Method of Block I. In the `a' panel shows the cuts applied to Block I and the `b' panel shows the new border using the more remote points as boundary with the exclusion of the two isolated points in the northern tip of the block.

This new area, with the two isolated points removed (see 'b' panel from Figure 9), now has 156 points with intensity of 0.47 points per hectare within an area of 333.13 hectares. The results of the J-Function (First Moment Characteristics) are under CSR distribution as expected, and for the Second Moment Characteristic, the Reduced Second Moment Function K in the Simulation Envelope Methods was applied. In conclusion, the null hypothesis could not be rejected without the 2 points, as observed in the 'b' panel (Figure 9).

a) Simulation Envelopes to K function with second and fouth cuts

Functions	Colour
obs(r)	black
theo(r)	red
hi(r)	green
lo(r)	blue

b) Simulation Envelopes to K function with Exclusion-Edge Method

Functions	Colour
obs(r)	black
theo(r)	red
hi(r)	green
lo(r)	blue
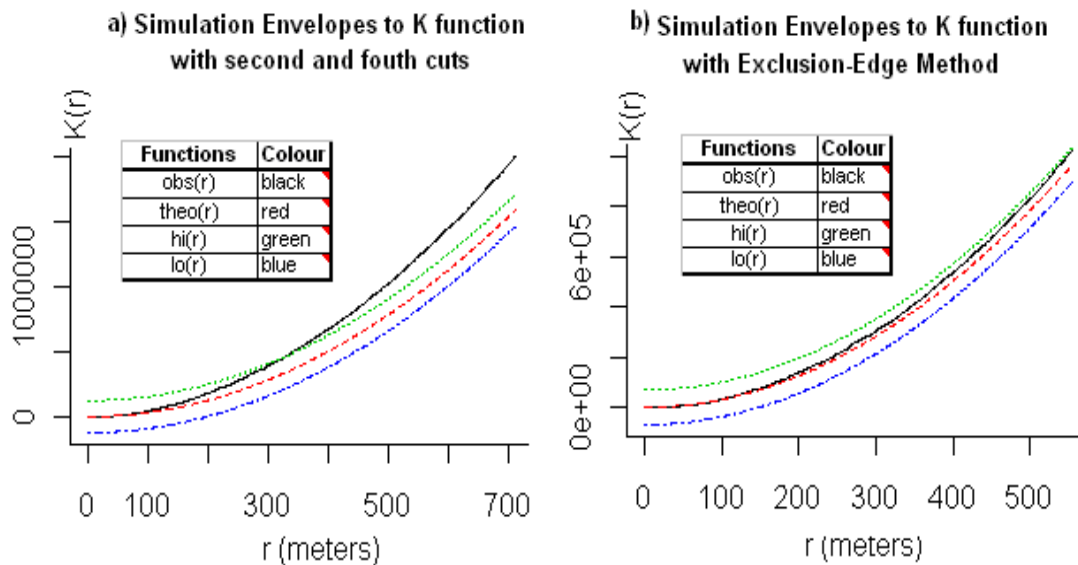
Figure 9: Simulation Envelopes Method applied K function with Cuts Method and Exclusion-Edge Method of Block I. It shows 2 panels with the following curves: the cumulative empirical distribution from the observations 'obs(r)', the cumulative theoretical distribution under CSR, the highest simulation envelope under CSR 'hi(r)' and the lower simulation envelope 'lo(r)' under CSR. Both functions are from the spatstat package developed by Baddeley in R statistic Tool.

The remaining step was the analysis of the results which would make the Pair correlation Function support or discard the results from the K Function for Block I, considering the new border/area with the two isolated points excluded. Between the first radius (1 meter) and up to a radius of 20.5 meters, the observation curve is lower than the lower simulation envelope curve, hence suggesting the presence of inhibition. However, given that the interaction between points did not appear until the radius of 24.8 meters, (from the results returned of the function Gest, see sub-section 4.3.2). Hence the presence of inhibition is discarded again. The rest of the observations are completely under the CSR (uniform Poisson distribution) region as shown in Figure 10. In conclusion, the null hypothesis could not be rejected.
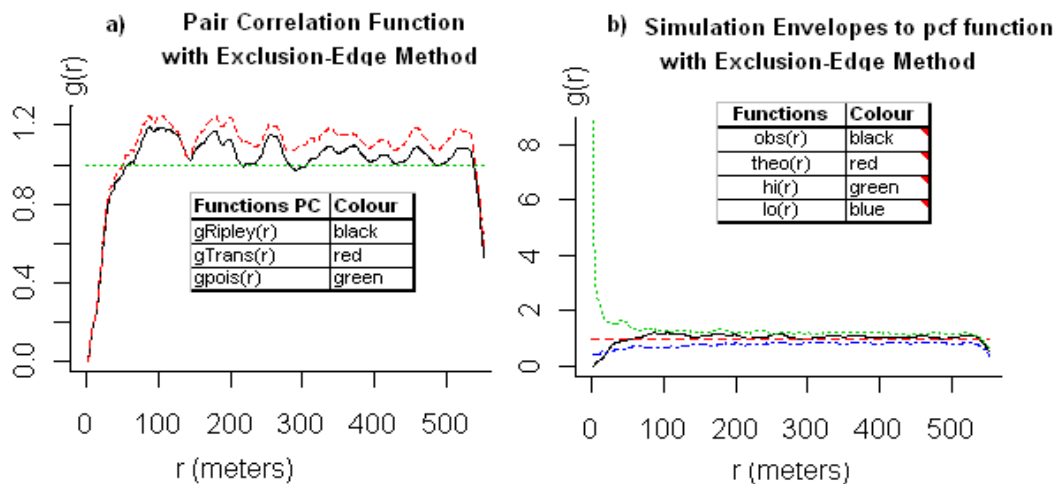
Figure 10: Spatial Distribution Analysis from the Pata-Mabula area in the Congo Basic rainforest Pair Correlation function applied to new border of Block I contains 2 panels: a) The Pair Correlation Function 'pcf' which panels The Ripley-corrected estimate of *g* 'gRipley(r)' curve, the Translation-corrected estimate of *g* 'gTrans(r)' curve and the Theoretical Poisson of g(r) curve 'gpois(r)' and b) The Simulation Envelope *g* function which presents the cumulative empirical distribution of the observations 'obs(r)', the cumulative theoretical distribution under CSR, the highest simulation envelope under CSR 'hi(r)' and the lower simulation envelope 'lo(r)' under CSR. Both functions are from the spatstat package developed by Baddeley in R statistic Tool.

The results of the Exclusion-Edge Method concluding that the greater part of the new area (region with red colour, panel 'a' of Figure 11) of Block I, which excluded the two northernmost isolated points, has a CSR distribution, with an intensity of 0.47 points per hectare. The question remains. What happened within the northern area of exclusion that included the two removed isolated points? The absence of points in this excluded area may suggest the influence of and/or partial presence of swamp forests (i.e. water levels fluctuate between the dry and rainy season). To answer this question it is necessary to obtain knowledge of this area as well as possible human intervention that can potentially be deduced by study of the Bimpe Forest Concession map. The following are two possible scenarios and the resulting impacts on the spatial point distribution analysis:

1. If the empty area between the two rejected points and the CSR area of Block 'I' was swamp forest, then the Exclusion-Edge Method explains the distribution of Block I, provided that supporting forestry evidence could be obtained that would help explain the existence of the two rejected points.

2. If the empty area between the two rejected points and the CSR area of Block I was logging (considering the closed human intervention as the road) or incurred natural tree death, this potential scenario can be solved via a purely mathematical perspective.

4.3.5.2 Simulation of the excluded area with the two isolated points for block I Using the postulates of the homogenous Poisson process distribution (Diggle 1993, 50), it is possible to find

the number of points missing in the rejected area of Block I by simulation.

Denote by A the rejected area (green on Figure 11) under CSR and B the CSR area (red on Figure 11) of Block I.

According to the data set and construction of region in R, A and B are disjoint. According to PP3, $N(A)$ and $N(B)$ are independent.

According to PP2 $N(A) = n$ the number of points which A has a homogenous Poisson process distribution and $N(B) = 156$ from the Block I analysis

According to the Block I analysis, the intensity of $B$is .47 points per hectare, hence using PP1 and knowing the A area (57.67 hectares), it possible to know that the number of points that could have been in the excluded area is 27. As there are already two points in the area, then there are 25 points missing.

Using the function rpoispp from the spatstat package developed by Baddeley in R, it is possible to simulate the 27 points of this block. The simulation could not allocate the exact existing location for the two existing points but using more than 100 iterations it was possible to arrive at the best hypothetical representation of the missing spatial points (see panel (b) on Figure 11).
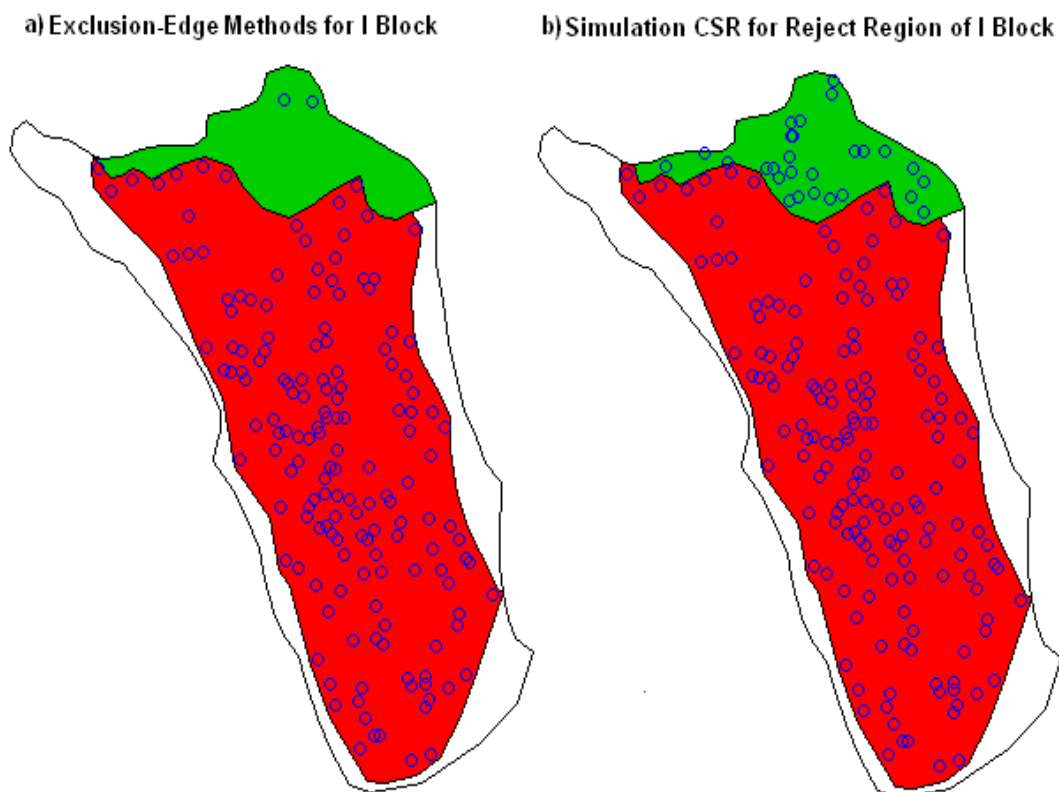
Figure 11: Exclusion-Edge Method and Simulation CRS for Reject Region of Block I from the Pata-Mabula area. The 'a' panel shows the boundary more close to the physic edge of Block I, the new border using the more remote points as boundary in colour red and the reject region or area where are the two rejected points in green colour. The 'b' panel presents the same areas as the 'a' panel, but it included the 27 points obtained by the simulation under CSR.

$N\left(A\bigcup B\right) = 183$ points with average intensity of 0.47 points per hectares over an area of 390.46 hectares within a polygonal boundary with 57 vertices (see 'b' panel Figure 11). The Figure 12 shows the J Function and K Function results for the new area 'C' where the null hypothesis could not be rejected. These results were expected because the two areas have a CSR distribution.
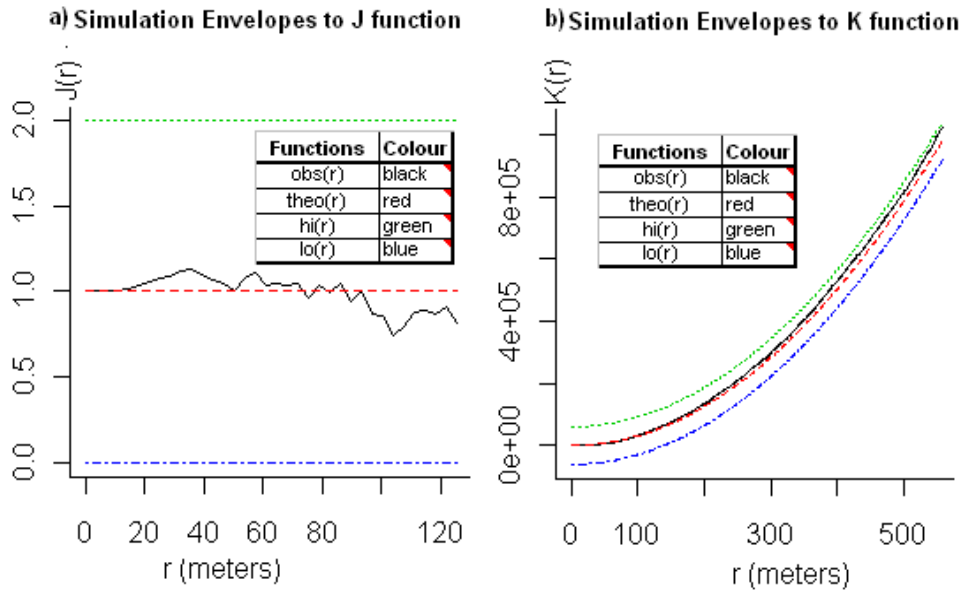
Figure 12: Simulation Envelopes Method for the J-function and K-function applies to the CSR simulation for the rejected area of Block I from the Bimpe Concession data set. Using 2 panels the following curves shown are: the cumulative empirical distribution from the observations 'obs(r)' of the best estimate from each function; the cumulative theoretical distribution under CSR; the highest simulation envelope under CSR 'hi(r)'; and the lower simulation envelope 'lo(r)' under CSR. Both functions are from the spatstat package developed by Baddeley in R statistic Tool.

4.4 First and Second Moment Characteristic Analysis for Blocks H, R1 and Z2

According to the Block I analysis, the F function and G function which analyze the First Moment Characteristics for spatial points are both related used in the calculation of the J Function. Additionally, the application of Simulation Envelope Method within these distance methods brings the best results which can describe the spatial distribution of these point patterns. Thus, blocks H, R1 and Z2 were analyzed only using the function J. The results of the analysis are given below:

Studying the First Moment (intensity) of the blocks H, R1 and Z2, they followed the CSR distribution because the whole observation curve (the best estimator of J-function) was between the lower and higher Simulation Envelope curves under CSR, *hence the null hypothesis cannot be rejected.* However, when the radiuses increase there is a tendency towards clustering (See Appendix E, panels of J Function).

Studying the Second Moment Characteristics (dependence and interaction between points), blocks H

and Z2 followed CSR when the radiuses were small, but as these radiuses grow, clustering is observed (when the observation curve is greater than the higher simulation envelope curve under CSR). However, Block R1 shows that when the radiuses are greater than 700 meters, regularity patterns were observed (when the observation curve is less than the lower simulation envelope curve under CSR - see Appendix E, K Function panels).

4.5   First and Second Moment Characteristic Analysis for J, Q and Z1

Studying the First Moment Characteristic (intensity), J and Q blocks show clustering for radiuses greater than 60 and 45 meters respectively when the observation curve (the best estimator of J-function) is less than the lower simulation envelope curve under CSR, hence the null hypothesis is rejected for both blocks. However, the observation curve of Block Z1 falls completely within the CSR region. Hence the spatial point patterns follow the uniform Poisson distribution, which means the null hypothesis could not be rejected (see Appendix F, J-function panels).

Studying the Second Moment Characteristics (interaction between balls, two intensities) the J, Q and Z1 blocks present peculiar observation curves when the K Function and the Simulation Envelope to function K are applied. Accordingly, the results are described block by block. The point patterns of Block J follow CSR distribution because the observations curve (Block J has more than 3000 points, and accordingly, the estimator border is significant) is between the lower and higher simulation envelopes which are under CSR, hence the null hypothesis cannot be rejected. Block Q shows strong clustering distribution because the observations curve is over the higher simulation envelope under CSR when the radiuses are greater than 112 meters, and the slope of the observation curve is steeper than the slope of the theoretical curve when the radiuses increase. Finally, Block Z1 shows both strong regularity and clustering patterns (i.e. the observations curve crossed the CSR region more than one time). Between 165 and 572 meters of radius, the observation curve is greater than the higher simulation envelope indicating a presence of clustering. Between 590 and 883 meters of radius (the last radius for the observations curve), the observation curve is less than the lower simulation envelope, hence there is a presence of regular point patterns (See Appendix F, K -Function panels).

4.6   Exploring Aggregation

4.6.1   Aggregation analysis for the more dense areas of Bimpe Concession

Previous studies have explored possible aggregation. This study seeks to shed light on this form of possible tree grouping (type of clustering). The distance methods helped us find whether or not the distribution of the trees exhibited a clustering tendency as shown in the results here-above. A literature review directed at finding information on crown diameters of these very large trees with

commercial value shows that the approximate range would be 6 to 30 meters (Asner et al, 2002, 486). Personal communication with the foresters involved in the inventory work revealed that an approximate average crown diameter may be 15 meters. If the average was 20 meters and the distance between two points is less than 20 meters, then these trees are interacting between themselves. Using this information, it is possible to detect which blocks have the higher aggregation levels. To explore aggregation, the blocks were divided into sub blocks that exhibited a presence of clustering (first or second moment characteristics). These sub-blocks have the characteristics of having the highest point densities of the block. These selected sub-blocks were divided into 100 X 100 meter (1 hectare) units. The numbers of points contained within these hectares is shown. The size of these sub-blocks was variable (See Appendix G). For each sub-bock, at least three hectares which had a high number of spatial points (trees) were selected and these were then divided into quadrats of 10 x 10 meters to find the crossing crown points (assuming an average crown diameter of 20 meters) between them. When there are more crossing crown points in a closed area, then the relative level of aggregation is higher. It can be assumed that when there are a relatively high number of points within a hectare there is a higher probability to have more crossing crown points. An analysis of aggregation confirmed this pattern (See table 3 and Appendix H).

Table 3 Analysis of Tree´s Aggregación, assuming an average crown diamater of 20 meters

Result: There is a certain aggreagtion level between trees at 10 meter radius

Sub-Blocks	Number of quadrats with points > 4	Average Intensity (points per hectare)	Max number of points per quadrat	Number quadrats selected	Number of crossing crown points	Max # of grups with crossing crown	Max number of crossing crown
H	6	0.926	8	4	12	2	9
J	8	1.373	7	3	5	2	3
Q1	72	2.239	13	3	15	8	4
Q2	126	5.241	15	3	20	7	5
R1	75	3.163	15	3	33	11	7
Z1	192	9.023	25	3	34	9	9
Z2	20	1.386	8	3	6	2	4
I	0	0.3289	3	3	0	0	0

Description: It presents the results of the aggregation analysis based on an average crown diameter of 20 meters per each sub-block studied. The first column is the name of each sub-block, the second column is the number of quadrats within more than 4 points per sub-block, the third column is the average intensity of points per hectare, fourth column is the maximum number of points per quadrat in each sub-block, the fifth column is the selected number of quadrats used in the aggregation analysis (between points in each sub-block), the sixth column shows the number of crossing crown points per sub-block, the seventh column shows the maximum number of groups with crossing crown points found in the selected number of quadrats and the eighth column shows the maximum number of crossing crown points per group in each sub-block .

To know whether or not the clustering tendency (principally via the Second Moment Characteristics) has changed from the original block, distance methods J Function and K Function were applied to each of the sub-blocks studied. The results showed that, under the First Moment Characteristics (intensity), all the sub-blocks followed the uniform Poisson distribution (see table 4 and Appendix

I). This implied a clustering tendency change vis-à-vis J and Q sub-blocks, because originally these blocks showed a presence of clustering (see Appendix F).

Table 4 Simulation Envelopes to J Function (First Moment)

		Simulation Envelopes			Estimator K-M curve		
Sub-Blocks	Average Intensity (points per hectare)	lower value	higher value	Range	Min Peak value	Max Peak value	Ho
H	0.93	-1.96	3.96	5.91	0.20	1.03	not rejected
J	1.37	0.07	1.93	2.00	0.40	1.07	not rejected
Q1	2.24	0.12	1.88	2.00	0.32	1.10	not rejected
Q2	5.24	0.40	1.60	2.00	0.54	1.26	not rejected
R1	3.16	-0.54	2.54	3.08	0.23	1.07	not rejected
Z1	9.02	0.38	1.62	2.00	0.47	1.27	not rejected
Z2	1.39	-0.28	2.28	2.56	0.40	1.11	not rejected

Description: It presents the results of the Simulation Envelope of J function for all the sub-blocks to which an aggregation analysis was applied (for a radius of 20 meters).

However, using the Second Moment Characteristics (dependence and interaction between points), the results showed all the sub-blocks have a clustering tendency (see table 5) with varying intensities. For example, the observations curves of H, J and Z2 blocks were still very close to the higher simulation envelope, but in the case of Q, R1 and Z1 blocks, the deviations between both the empirical and theoretical curves was still there (see Appendix I). The empirical curves of sub-blocks Z1 and J have appreciable changes from the original blocks (see Appendix E).

In doing this, we observed that both the area and the form of the selected boundary have a tremendous influence over how the spatial point patterns are distributed. In sub-block R1, it was thereby clearly possible to detect the presence of clusters, which were not visible in block R1 when we studied it using the same distance methods.

Table 5 Simulation Envelopes to K Function (Second Moment)

Sub-Blocks	Average Intensity (points per hectare)	First value radius	First Interaction radius	Min radius	Max radius	Ho
	Result: All the sub-blocks have clustering tendency					
				Clustering range		
H	0.93	0.94	6.59	332.22	481.86	rejected
J	1.37	1.37	10.93	411.39	699.78	rejected
Q1	2.24	0.95	9.54	110.67	488.46	rejected
Q2	5.24	0.50	6.51	76.59	256.29	rejected
R1	3.16	0.79	7.10	60.78	404.12	rejected
Z1	9.02	0.59	4.73	225.68	302.48	rejected
Z2	1.39	0.78	12.45	252.20	398.54	rejected

Description: Table 5 presents the results of the Simulation Envelope to K Function for all the sub-blocks to which an aggregation analysis was applied (for a radius of 20 meters). The first column is the names of each sub-block analyzed, the second column is the average intensity (points per hectare), the third column shows the first radius value greater than zero for the theoretical functions, the fourth column shows the first radius value where the function estimates have values greater than zero, the fifth and sixth columns show the range of radius values observed and the last column shows if the Ho (hypothesis) was rejected or accepted.

This semi-manual aggregation analysis focused in the more dense parts of each block. The objective was to explore the level of tree aggregation. Using the geometrical knowledge specifically regarding the distance of two circles with one common point (See Figure 13), and the Nearest Neighbour Distances (`nndis`) and the Nearest Neighbour (`nnwhich`) functions from the spatstat package in R Tool, we developed two functions which are described below.

The first function '`numneigdistlessmaxcrownd`' gives the nearest neighbours and their correspondent distances per each spatial point. These distances need to be less than a given distance (per example: maximum crown distance).

The second function '`aggregationlevel`' gives the same information as the first function, but in this case, it lists only the points with have a least one nearest neighbor. Additionally, it shows the number of nearest neighbours per spatial point which will be used.
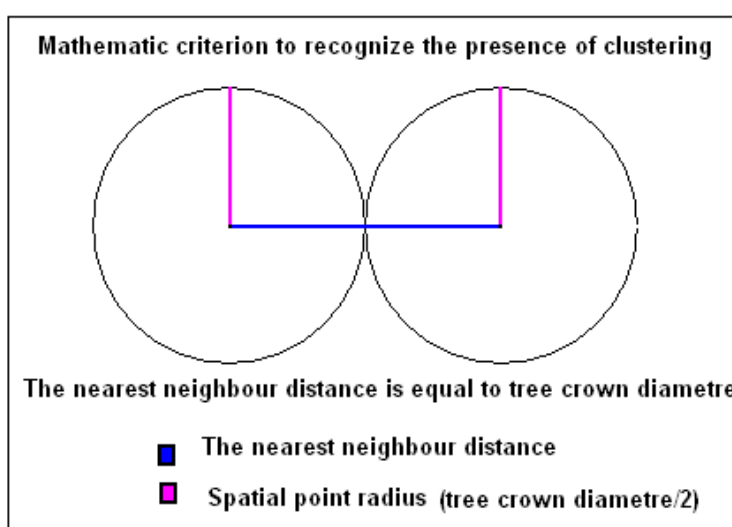
Figure 13: Mathematic criterion to recognize the presence of clustering.

Finally, these functions provided a relative measure of aggregation levels and together they formed an Algorithm of Aggregation called '*First Wave of Clustering*'. The levels measured will be have same name of this Algorithm of Aggregation in this thesis. The reason to call 'First Wave of Clustering' is because it only allows us to know the point-to-point connections (tree-to-tree distances). It does not provide knowledge of more than two points in the same direction or the total mathematic tree. The Figure 14 panels show two clustering groups in Block H within a circle with a set radius (R1=15 meters) with the one center being tree 573 (A group), and the other center being tree 572 (B group). These circles are considered to be the first wave level of clustering.

To show the limitation of this algorithm, there is a circle which represents a hypothetical second wave level of clustering with a radius (R2) of 20 meters and center at tree 573. This circle contains three hypothetical trees which are a first level of clustering (trees 572, 571, 575) as the nearest neighbours of tree (573). Figure 14 has five groups: two empirical and three hypothetical (see Table A). If the hypothetical second wave level of clustering would be computed, with the connection of more than two spatial points following in the same direction, it would be possible to have the connections:573 → 572 → X1, 573 → 571 → X2 or 573 → 575 → X3. Additionally, in Table B of Figure 14, one can see the nearest neighbour distances of the empirical trees of Block H. From this exercise, one can assume that the nearest neighbour hypothetical distances are less than 15 meters for the first wave (nearest neighbours of a center being tree 573) and less than 20 meters for the second wave (again nearest neighbours of a center being tree 573).

Example of First and Second Waves of Clustering for Spatial Points

R2=20 meters

575

X3

574

572

X1

5̶7̶3̶

R1=15 meters

571

X2
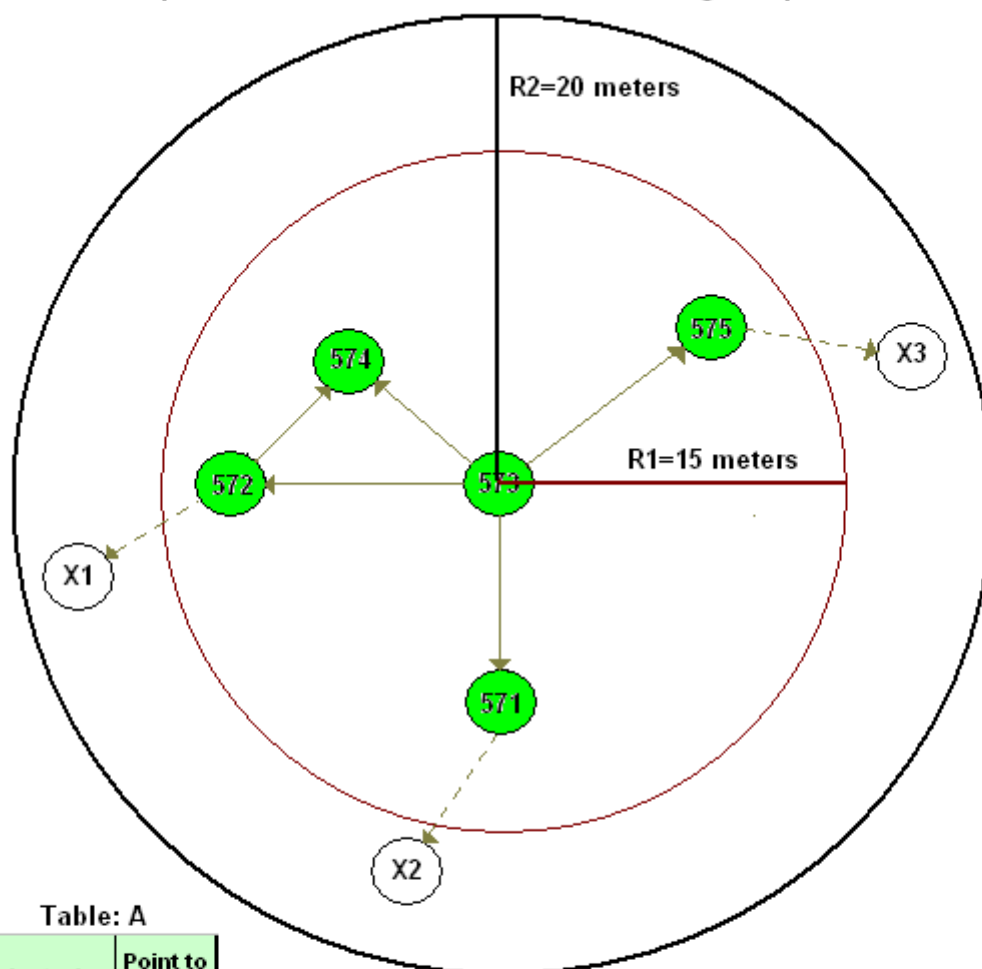
Table: A

clostering	Point to Point
A-573	573-574
	573-571
	573-572
	573-575
B-572	572-574
C-572	572-X1
D-571	571-X2
E-575	575-X3

Table: B

Point to Point	Distance (meters)
573-574	6.74
573-571	10.02
573-572	10.66
573-575	12.01
572-574	6.00

Figure 14: Example showing the aggregation level using the First Wave of Clustering Algorithm.

4.6.2   First wave of clustering algorithm for the all blocks

Using the First Wave of Clustering Algorithm, it was possible to know and to analyze the First Level of Aggregation for all the blocks which form the whole data set of Bimpe Concession Area. The average crown diameter of the trees used was 15 meters (as described in sub-section 4.6.1).

Table 6 First Wave of Clustering Summary based on an Average Crown Diameter of 15 meters for all the trees sample

Name Block	Number of Points	Average Intensity (per hectare)	# of points with a least one nn	% of Clusters	# of points with one nn	# of points with two nn	# of points with three nn	# of points with four nn	# of points with five nn	Total # of nn per # points	# nn points related with more than one point
H	944	0.69	27	2.86%	11	1	1			28	1
I	158	0.33	0	0.00%						0	0
J	3017	1.02	77	2.55%	37	4				83	6
Q	2651	3.07	286	10.79%	130	18	4	1		309	23
R1	2260	1.47	217	9.60%	89	23	11	6	1	264	47
Z1	2728	2.69	795	29.14%	281	118				799	4
Z2	891	1.08	14	1.57%	7					14	0

Description: It shows a summary of the First Wave of Clustering Algorithm assuming an average crown diameter of 15 meters (principal input parameter) as applied to all the trees of the Bimpe Concession Area. 'nn' in this table refers to 'nearest neighbour'.

Table 6 presents a summary of the results of the First Wave of Clustering Algorithm. With it, following observations can be made:

- It appears that the # of points with a least one nearest neighbour is directly proportional to the average intensity per block however, in the case of blocks Q and Z1, this does not happen.

- The maximum number of the nearest neighbours to a spatial point is 5 when the nearest neighbour distance is 15 meters, but this was observed only once (in block R1)

- Block R1 has the most number of nearest neighbours per spatial points.

- Block Z1 shows the most aggregated spatial points and, when one takes into account the % of clustering (column 5), this block also has the highest level of aggregation.

- Block I does not have aggregation with a nearest neighbour distance of 15 meters. As mentioned in sub-section (4.3.3), there is interaction between the points only for radiuses greater than 25 meters.

- The total number of nearest neighbour per number of points (column 11) is not always equal to the number of points with a less one nearest neighbour because there are some trees which share one or more nearest neighbour trees (ex. see point-to-point: 573-574 and 573-574 of Figure 14) as is the case in block H.

- When the number of nearest neighbour points increases, the level of aggregation increases accordingly, because there are more trees which are interactive within a distance less than 15 meters.

- Regarding the % of clustering for all the blocks (column 5), the level of aggregation per block never exceeds 30% of the total corresponding points (trees).

In addition (as per Table 6 results), the First Wave Level of Clustering Algorithm provides the nearest neighbour distance for each spatial point (trees) for aggregations based on a 15 meter radius. For all blocks, this information was summarized along with the distance $(D(K())>)$ at which the interactive radius is greatest and the following observations was obtained (See Table 7).

Table 7 First Wave of Clustering Distances Summary based on an Average Crown Diameter of 15 meters for all the trees sample

# of nn	H Block		J Block		Q Block		R1 Block		Z1 Block		Z2 Block	
	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)
1	14.96	6.00	14.98	10.38	15.00	6.18	15.00	6.63	14.98	4.69	14.61	10.44
2	14.99	10.02	14.82	13.43	14.66	10.47	14.94	9.72	14.98	7.46		
3	10.66	10.66					14.96	12.25	14.98	9.85		
4	12.01	12.01					13.86	13.86	14.83	11.27		
5									14.60	14.60		
D(K())>		5.156		10.34		5.156		5.45		3.792		10.26

Description: This summary shows the maximum and minimum nearest neighbour distance per block and the number of nearest neighbours. Additionally in the yellow part, there is the interaction distance value over which the K Function showed presence of interaction or/and dependence between spatial point patterns.

- The nearest neighbour distance is more frequently between 14 and 15 meters, and the minimum nearest neighbour distance is 4.69 meters and the maximum is 15 meters.

- Regarding the minimum nearest neighbour distance per block, it is always greater than the distance value $(D(K())>)$ at which the interactive radiuses are greater (values from the Estimate of Function K per block). This result shows the accuracy of the spatial methods used.

4.6.3 First wave of clustering algorithm for Wenge and Bosse-Clair

This primary focus of this thesis was to provide answers from a mathematic perspective to the critical hypothesis of the foresters who have been analyzing the forest area from which the study's data set was collected. The foresters suggested that we should focus on Wenge and Bosse-Clair trees, coincidently these are the two species which are densest in the data set. Then, with an average crown diameter of 15 meters, the First Wave of Clustering Algorithm was applied only with the spatial points consisting only of Wenge and Bosse-Clair trees. The results are shown in Table 8, and from those results, one can make the following observations:

- Block J shows more than 50% of its trees aggregated and accordingly it is possible to suggest that these trees have a relatively high clustering pattern. It is useful to remember that Block J has not been subjected to known human intervention (See Appendix A).

- Comparing the number of points with a least one nearest neighbour with the column 5 of the table 6, between 36% and 77% of the total trees aggregated are Wenge or Bosse-Clair, with the

exception of Block Z1 (12%).

- Block R1 shows a relatively strong aggregation level as it has the highest number of nearest neighbour points with more than one point (23 points in total, see last column).

- Regarding the percentage of clustering (column 5) for all tree species, it is possible to suggest a lower level of aggregation for Wenge and Bosse-Clair at an average crown diameter of 15 meters. This conclusion does not take into consideration forestry or ecological factors.

Table 8 First Wave of Clustering Summary based on an Average Crown Diameter of 15 meters for Wenge and Bosse-Clair Trees

Block	Number of points	Average Intensity (per hectare)	# of points with a least one nn	% of Clustering	# of points with one nn	# of points with two nn	# of points with three nn	Total # of nn per # points	# nn points related with more than one point
H	839	0.61	21	2.50%	10	1		23	2
I	1783	0.60	0	0.00%				0	0
J	50	0.10	28	56.00%	18			36	8
Q	1733	2.00	128	7.39%	58	7		131	3
R1	1194	0.78	87	7.29%	43	7	4	110	23
Z1	1067	1.05	98	9.18%	48	7	1	114	16
Z2	519	0.63	8	1.54%	4			8	0

Description: It show a summary of the First Wave of Clustering Algorithm with crown diameter of 15 meters (input parameter) as applied only to trees of Wenge and Bosse-Clair species of the Bimpe Concession Area. 'nn' means 'nearest neighbour' for each time it appears in the table.

The First Wave of Clustering Summary of the distances is presented in Table 9. According to these results, the following observations can be made:

- The nearest neighbour distance is more frequently between 12 and 15 meters, and the minimum nearest neighbour distance is 4.94 meters and the maximum nearest neighbour is 15 meters.

- Regarding the minimum nearest neighbour distance per block, it is always greater than the distance value $(D(K())>)$, which is always less than any interactive radiuses of the estimate of Function $K($ see sub-section (4.6.2)).

Table 9 First Wave of Clustering Distances Summary based on an Average Crown Diameter of 15 meters of Distances for the Wenge and Bosse-Clair

# of nn	H Block		J Block		Q Block		R1 Block		Z1 Block		Z2 Block	
	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)	Max(d)	Min(d)
1	14.96	6.00	14.39	10.59	15.00	6.42	15.00	7.08	14.94	4.94	12.59	12.53
2	14.99	14.99			14.66	11.24	14.70	10.56	14.83	10.86		
3							14.96	12.71	14.60	14.60		
D(K())>		5.16		8.12		5.76		5.45		3.79		11.97

Description: This summary shows the maximum and minimum nearest neighbour distance per block and number of nearest neighbours of only the Wenge and Bosse-Clair species of the Bimpe Concession Area. Additional in the yellow part, there is the interaction distance value over which the Function showed presence of interaction or/and dependence between spatial point patterns.

5   SECTION

5.1   Conclusions and Recommendations

5.1.1   Conclusions

- According to sub-section 4.2 (Investigating the intensity) , the blocks of the Bimpe Forest Concession have very irregular boundaries which are an indication of the presence of natural phenomena such as adjacent swamp forest. Less irregular boundaries indicate that the adjacent areas are external anthropogenic interventions such as logging roads, constructed or planned, and/or adjacent property. Taking these factors into account for Block I, results were obtained after the 'Cuts Methods' was applied (sub-section 4.3.5.1) with the two net areas of influence being less than that of the entire block.

- The spatial distribution analysis of the Bimpe Concession data set was done by grouping the seven blocks according to their intensity characteristics. Block I was treated as an individual group because it presented a particular distribution with a strong tendency to CSR in contrast to the others Blocks. The second grouping consisted of the blocks H, R1 and Z2 given that their Kernel Smoothed Intensities are similar with some small dense areas around the dispersed areas (see Appendix C). The third group included the remaining blocks J, Q and Z1, which have the densest areas. However this group was selected due to the presence of some *unexpected behaviour* such as: Block J has more than 3000 points with an apparently regular distribution in an area of relatively low density, and yet it does not present external anthropogenic interventions (See Appendix A). Block Q has two sub-areas with strong but different patterns, one showing apparent clustering and the other showing regular distribution. It has the highest average intensity (3.07 points per hectare) and it shows external anthropogenic interventions along its borders. Block Z1 presents two sub-areas strongly marked similar to Block Q, but the difference with Block Q is that the panel of JFunction graph/panel showed a strong distribution under CSR, while the panel of K Function completely rejected the null Hypothesis. Hence, for each of these three groups it is possible to conclude:

  - According to the First Moment Characteristic (F, G and J sub-section 4.3), the spatial point patterns of block I is homogeneous. However, the Second Moment Characteristic (J and g sub-section 4.3) cannot accept the hypothesis and the point patterns showed a clustering tendency when the radius was increased. Applying Fit Models (see sub-section 4.3.5.), block I has a CSR distribution with the assumption that the northern area (See Figure 11) was logged (considering the closed human intervention of the adjacent road) or

incurred natural tree death.

– H, R1 and Z2 blocks followed the CSR distribution when analyzed using the First Moment Characteristics (sub-section 4.4.), implying that these blocks have homogenously distributed point patterns. Otherwise, when analyzed with the Second Moment Characteristics, the blocks H and Z2 followed CSR distribution at very smallest radiuses, but when the radiuses grew, clustering was observed. Block R1 followed CSR distribution however when the radiuses were greater than 700 meters, regularity patterns were observed. Hence, blocks H and Z2 have a CSR distribution with interactions and/or dependences between the points when the radiuses are greater than 5.16 meters. Block R1 showed a CSR distribution, which potentially resulted from external anthropogenic and/or natural events (See Appendix A). In the distribution of Block R1, it is not possible to discard a pre-existing clustering pattern, which could have been altered by these events.

– According to the analysis of the First and Second Moment Characteristics, Block J is non-homogenous, but does not have interactions and/or dependences between its point patterns. However, the aggregation analysis (sub-section 4.6.) discarded the absence of interaction between the trees. Block Q shows strong clustering distribution, however its two sub-areas suggest high variance in the level of clustering completely within this block. Finally Block Z1 is homogenous but with interactions and/or dependences between its point patterns.

• Taking into account all the above conclusions, it can be summarized that the spatial point patterns of the Bimpe Forest Concession study area have a mixed distribution. This is not surprising. The presence of swamp forest around each block and anthropogenic and/or naturals events have had a significant impact on spatial distribution of trees reflected in results obtained using spatial analysis methods. Hence, a detailed study of these areas using forest management and forest ecology knowledge is necessary to obtain a more complete understanding of the spatial distribution of trees within these blocks.

• The results of the spatial analysis of block I informs both forest ecology and management in that it presents and explains the importance of key factors to be considered in the spatial distribution of the points (trees) such as the geographic landforms, human interventions such as roads and natural events such as tree death interventions.

• The aggregation analysis in the more dense areas (called sub-areas) of the blocks showed that almost all the blocks are homogeneous with interactions between their points (clustering patterns) with the exception of Block I when a crown diameter of 20 meters is assumed. Accordingly, it is possible to conclude that Block I is under CSR distribution and the presence of a regular

pattern in Block R1 can be completely discarded.

- Using an average crown diameter of 15 meters and applying the First Wave Level of Clustering Algorithm for all the spatial points, the level of aggregation per block never exceeds 30% of the total spatial points (trees). Hence, the overall aggregation level of the Bimpe Forest Concession Area is not strong at the time the data set was collected and compiled. The external anthropogenic and/or natural events could be contributing factors. It is also important to note that aggregation is strongly dependent on the average crown diameter chosen.

- With respect to the main proposal of this thesis, the aggregation analysis applied to the trees of Wenge and Bosse-Clair species of trees, it is possible conclude that all of this subset presented aggregation, with the exception being Block I. The level of aggregation was not dense (less than 10% of points have overlap between its crowns). However the Block J showed that 56% of the Wenge and Bosse-Clair trees have aggregation patterns (interaction and/or dependence). It was noted that Block J has only 5 species, however this block did not show external anthropogenic interventions. Accordingly, it is possible to assume that natural events such as normal tree mortality due to fluctuating water levels given the presence of adjacent swamp forest or generally unsuitable soil nutrients as possible causes.

- When one compared the minimum nearest neighbour distance of all the trees (all species) and the trees of Wenge and Bosse-Clair species in each block (Table 8 and Table 10), with the interaction distance over which the Function K showed presence of interaction or/and dependence between spatial point patterns, it could be determined that the average crown diameter chosen was proper.

- This spatial analysis consists of recently applied theory that attempts to provide some missing answers to the forest managers of the area where the data set came from. In this specific case, these results could help foresters decide the level and spatial characteristics of tree removal/harvesting without altering the fundamental forest structure and composition (sub-section1.2).

5.1.2 Recommendations

- The use of spatial analysis methods, for the spatial analysis of data set of Bimpe Forest Concession Area data set, should be accompanied by background knowledge of the study data (past forest management and forest ecology in this case) so to provide more accurate results, taking into account the large size and complexity of the data set.

- For future applications, it is best to first use, the Estimate of Function J with the Simulate Envelope Method, when one is analyzing the First Moment Characteristics. However, for the

Second Moment Characteristics (interaction and/or dependence) it is recommended to first use, the Estimate of Function K and, depending on the results obtained, then applying Simulate Envelope Method.

- The First Wave of Clustering Algorithm needs to be scaled more to provide the total level of aggregation, and it be would be useful to explore the development of new algorithms for a more complete and accurate, mathematic-based, theory of spatial distribution (particularly for forest data set).

- The results of sub-section 4.6, the existence and location of aggregation (clustering) in any forest are very important to be study because it will be a key factor in determining sustainable extraction rates that will not significantly alter forest structure composition and ecological functions. The goal is to cut just the 'interest' or 'benefit' without going beyond such a level of extraction that would diminishing a threshold forest stock level that is often referred to as the 'principal'.

5.2   Appendix

5.2.1   Appendix A: Map of the Bimpe Concession
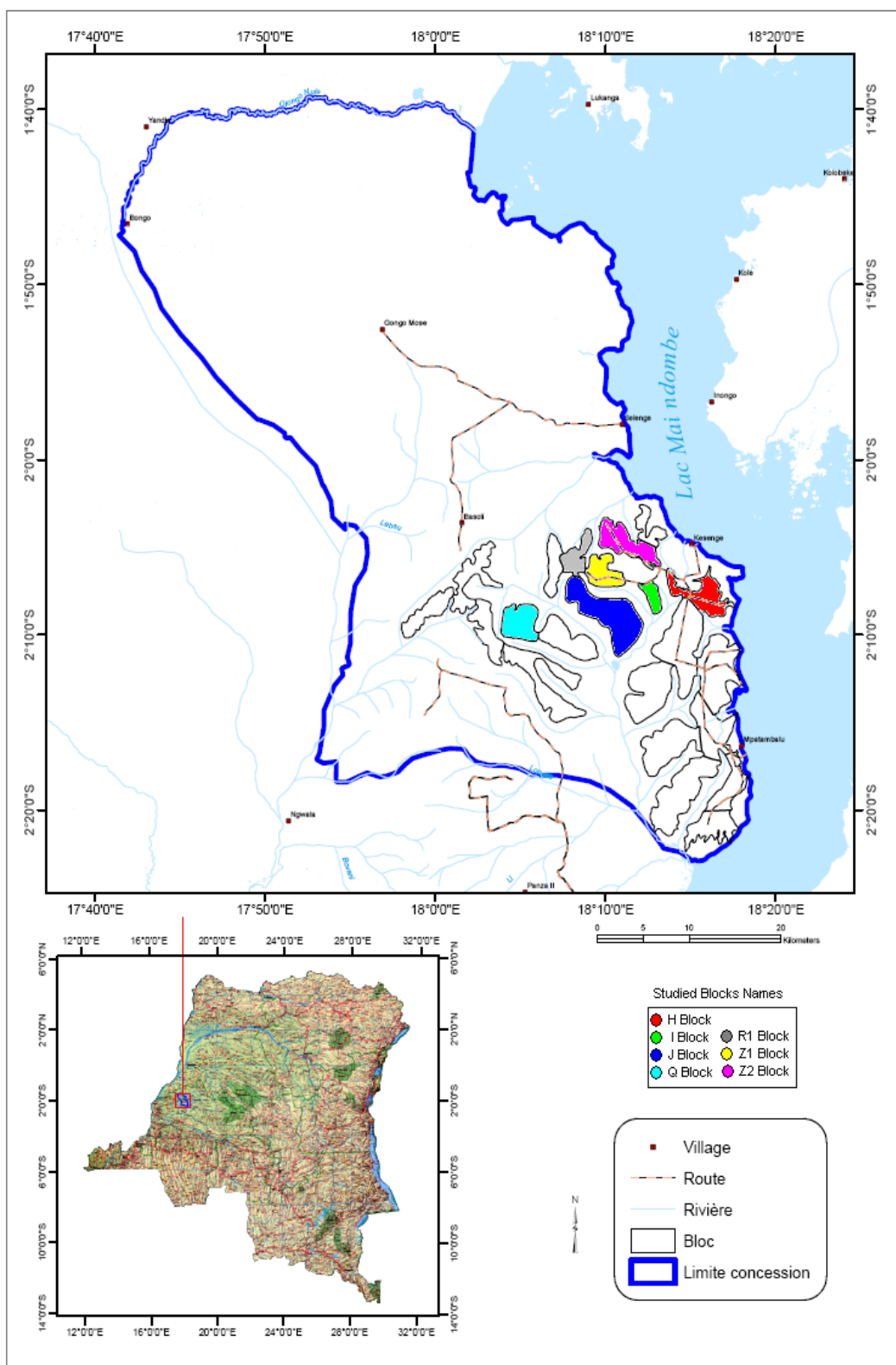
5.2.2 Appendix B: Species of trees of Bimpe Concession Area

Number of Trees per Specie Local Name			
Specie Tree's Local Name	Number of Trees	Specie Tree's Name	Number of Trees
WENGE	5745	RIKIO	17
BOSSE-CLAIR	1440	NKOTE	17
BOMANGA	886	INCONNU	14
TIAMA	785	LONGHI BLANC	9
SAPELLI	620	MUKULUNGU	8
AFINA	469	LONGHI ROUGE	8
WAMBA	318	LONGHI-ROUGE	8
BILINGA	294	EVEUSS	8
NO NAME	265	EBOM	8
ILOMBA	249	DIBETOU	8
KOSIPO	145	BOOTO	8
AIELE	137	BOOSO	6
DABEMA	121	OBOTO	5
ANGUEUK	109	SIPO	4
YUNGU	91	EMIEN	4
ETIMOE	80	IROKO	2
KANDA-BRUN	72	LONGHI	2
TALI	71	BOMBO BOLU	2
PADOUK	69	AXONONG	2
LATI	68	OSSOL	1
MEPEPE	50	OLONVOGO	1
FARO	50	WANDO	1
ESSESSANG	48	MUSANGA	1
MOBALA	44	LIMBALI	1
BODIOA	42	ENGAME	1
ESILI	37	DRAGONIER	1
DIANIA	36	BOSELE	1
BOSSE-FONCE	30	BOSEKI	1
NIOVE	28	BOSANGE	1
ESSIA	27	BONTENDO	1
AKO	25	BOBELE	1
KOTO	22	AVODIRE	1
IATANDZA	22	AFANE	1

5.2.3 Appencix C: Block's density of Bimpe Forest Concession Area

Block	# of points	# of hectares	Average Intensity	# of vertice	p_value
H	944	1371	0.69	48	1.40E-11
I	158	480	0.33	77	0.9472
J	3017	2964	1.02	47	2.20E-16
Q	2651	864	3.07	33	2.20E-16
R1	2260	1536	1.47	56	2.20E-16
Z1	2728	1014	2.69	68	2.20E-16
Z2	891	824	1.08	55	1.69E-05
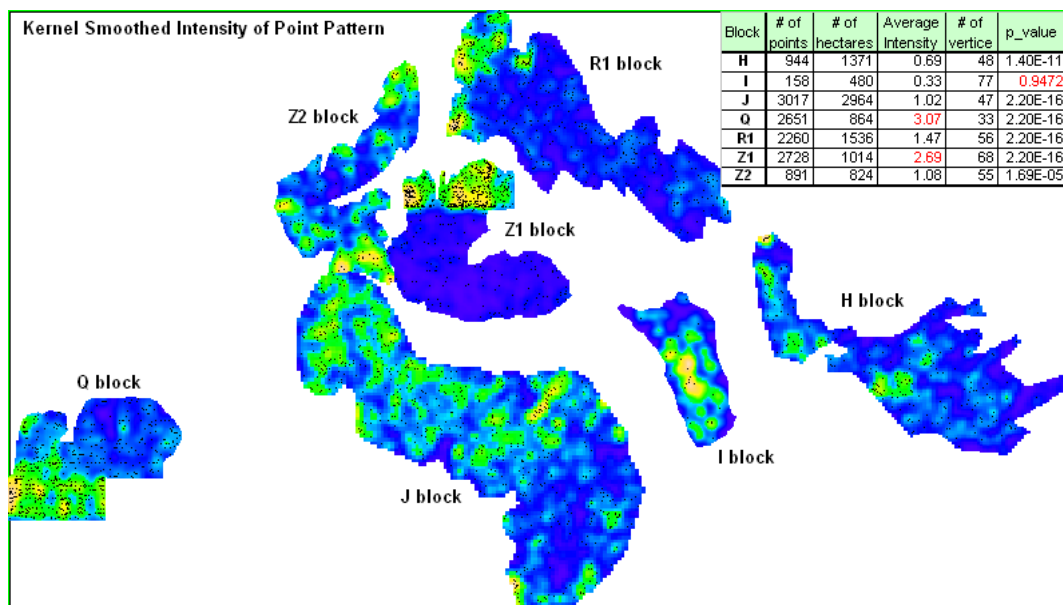
5.2.4 Appendix D: Mapped point patterns from Special Wenge and Bosse-Clair of Trees in Bimpe Forest Concession Area

SPATIAL DISTRIBUTION ANALYSIS

Mapped point patterns from Wenge and Basse-clair Trees in Bimpe data set

The vertices must be listed anticlockwise and no vertex should be repeated (i.e. do not repeat the first vertex).
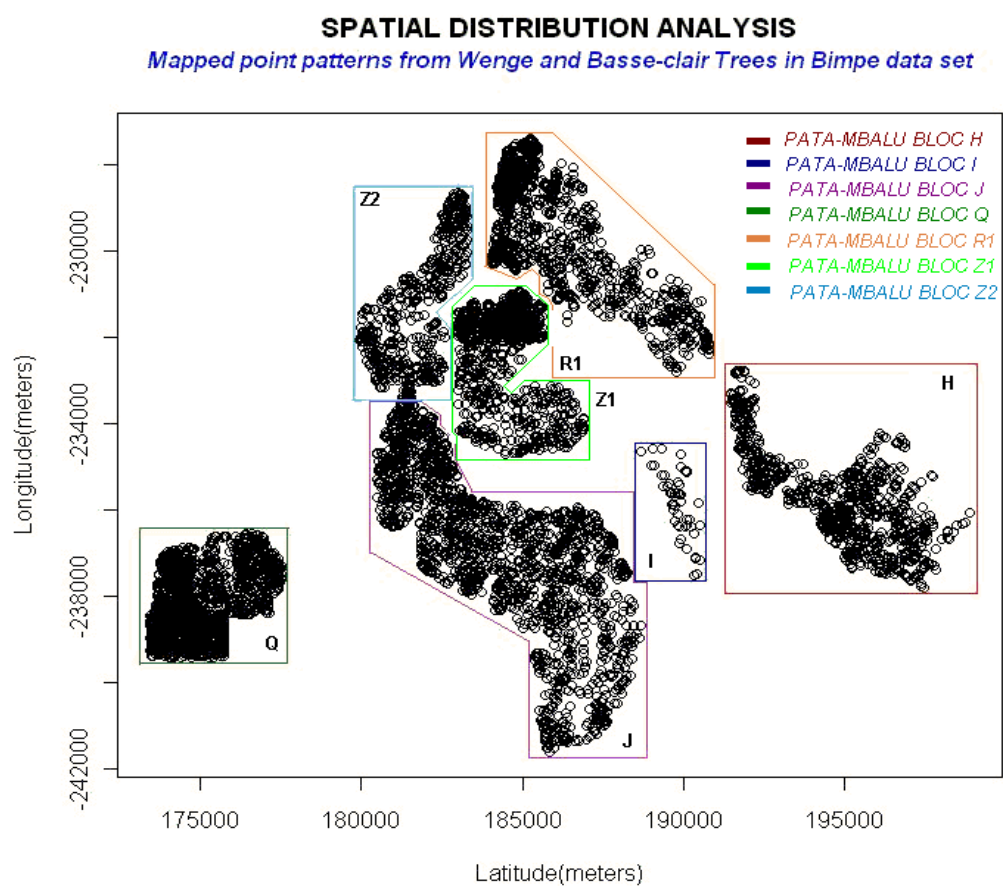
5.2.5 Appendix E: J-function and K-function panels for blocks H, R1 and Z2

H Block , 944 point with intensitsy=0.689 points per hectare
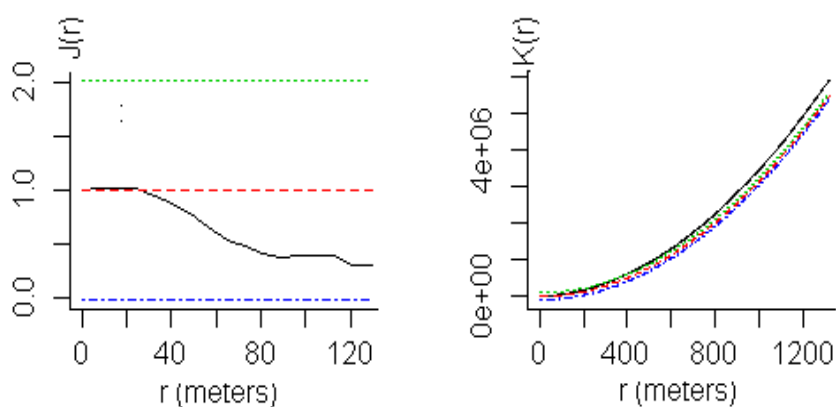
R1 Block, 2260 points with intensity 1.47 points per hectare
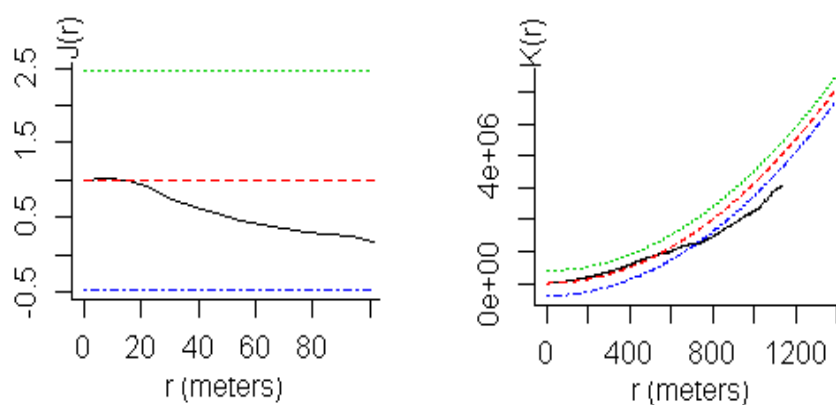
Z2 Block, 891 points with intensity 1.08 points per hectare
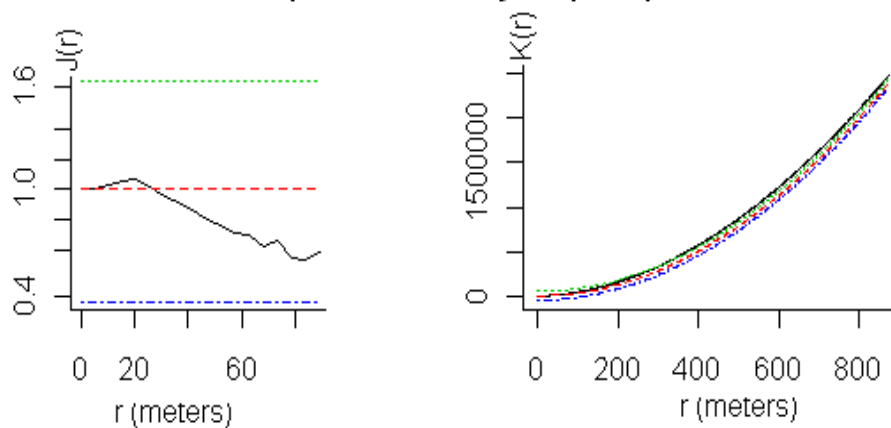
5.2.6 Appendix F: J-function and K-function panels for blocks J, Q and Z1

J Block, 3017 points with intensity 1.02 points per hectare

Q Block, 2651 points with intensity 2.07 points per hectare

Z1 Block, 2728 points with intensity 2.69 points per hectare
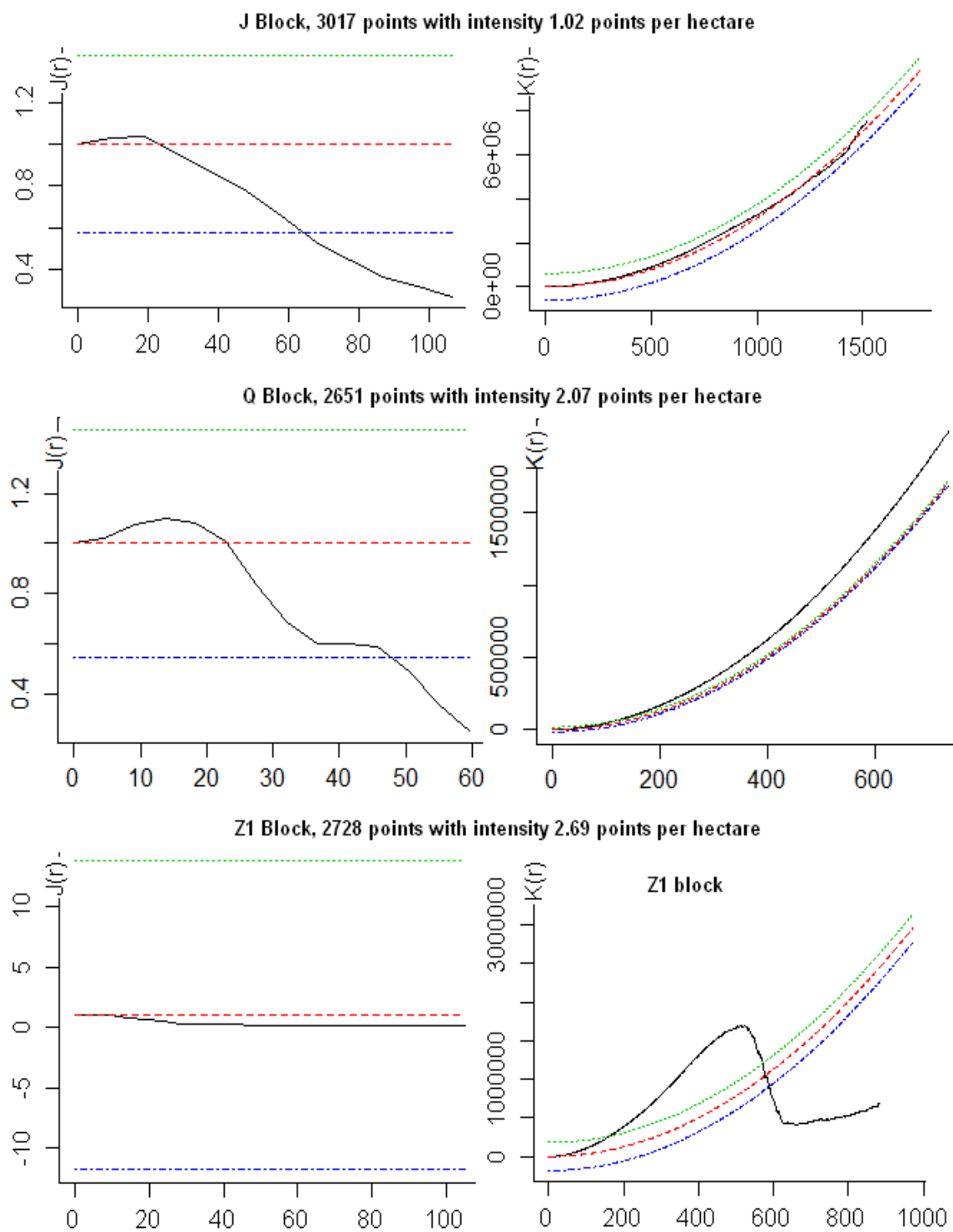
5.2.7 Appendix G: Exploration of aggregation

PATA_MABULA H SUB-BLOCK AGGREGATION EXPLORATION

a) Mapped point patterns
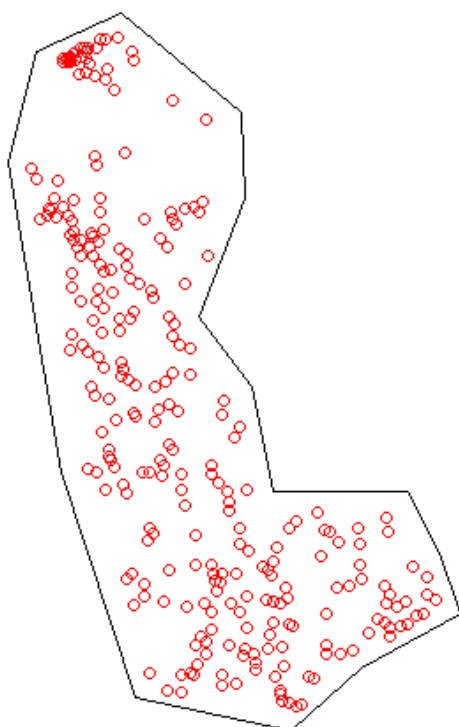
b) QuadratCounts Method
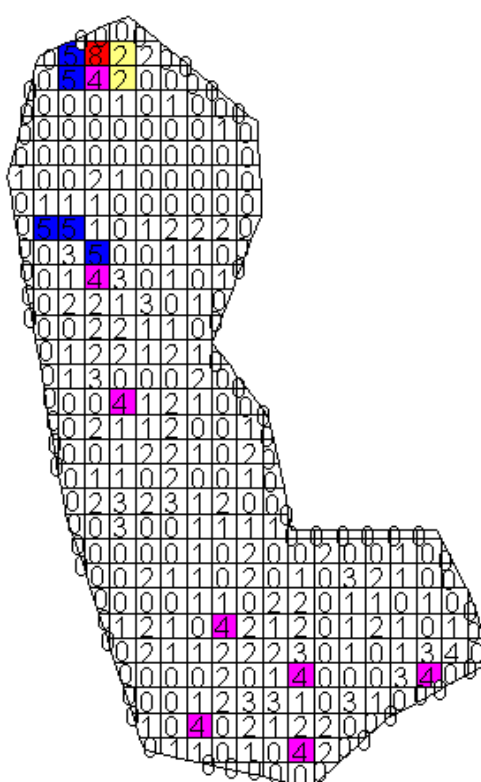
271 points around 293 hectares, with a average intensity 0.9251 points per square hectare

PATA_MABULA J SUB-BLOCK AGGREGATION EXPLORATION

a) Mapped point patters
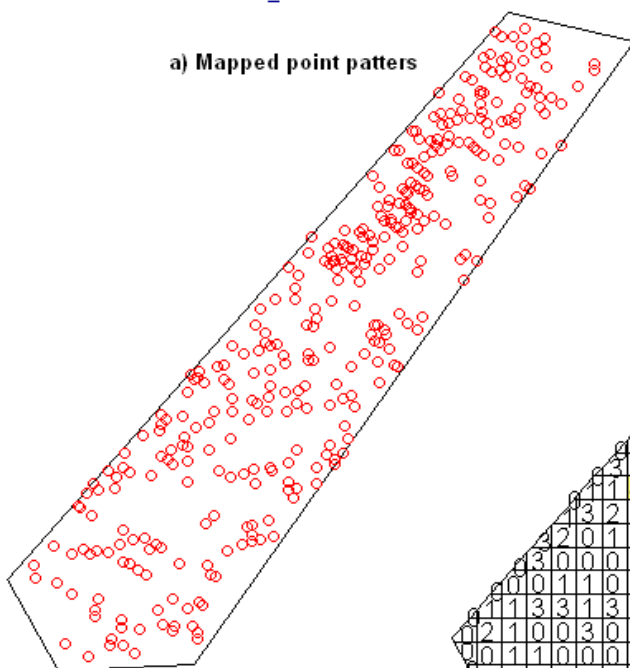
b) QuadratCounts Method
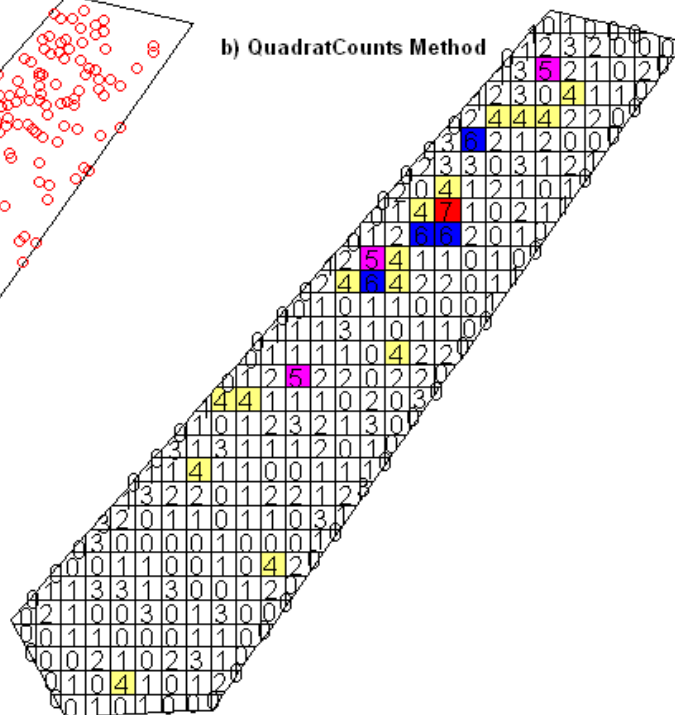
356 points around 259 hectares, with a average intensity 1.373 points per square hectare

PATA MABULA FIRST PART-Q BLOCK AGGREGATION EXPLORATION

a) Mapped point patterns
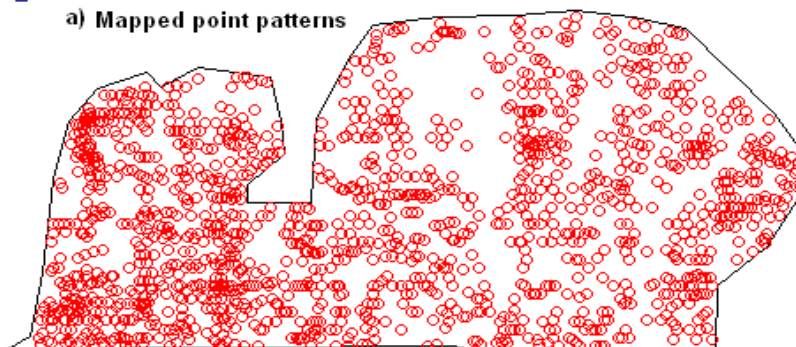
b) QuadratCounts Method
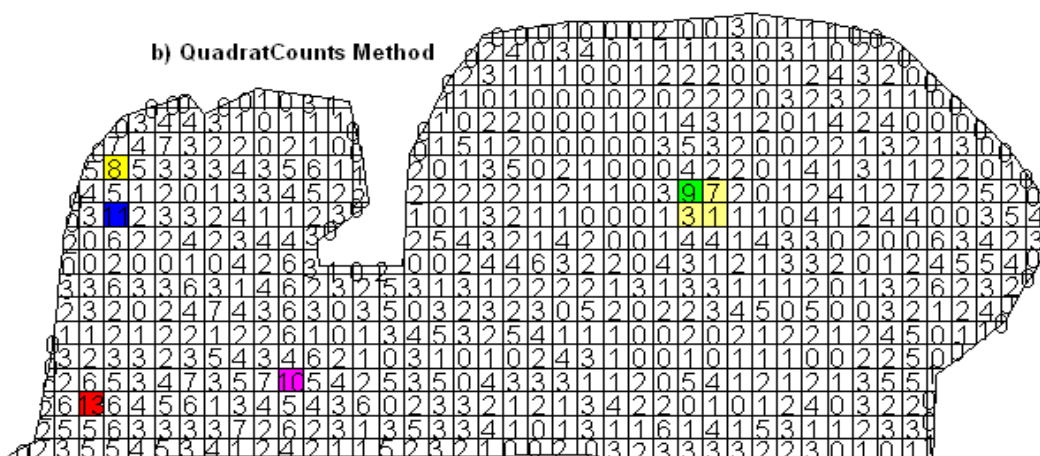
1402 points around 626 hectares, with a average intensity 2.239 points per square hectare

PATA MABULA SECOND PART-Q BLOCK AGGREGATION EXPLORATION

a) Mapped point patterns

b) QuadratCounts Method

10	12	7	6	3	5	6	7	0	4	5	10	3	2	5	5	4	8	4	4	4	9	11
9	6	11	7	5	10	6	2	6	6	5	6	7	6	6	12	6	7	3	4	4	7	7
9	12	6	7	4	6	10	13	7	4	4	2	8	10	5	7	11	6	5	4	8	4	8
12	15	4	8	5	6	5	10	8	0	2	4	7	1	2	1	8	7	4	4	4	8	6
9	13	8	11	5	4	4	12	7	4	2	4	7	6	4	5	2	5	4	4	3	4	4
11	12	3	2	4	2	4	5	2	2	1	4	11	3	6	0	2	6	3	2	4	6	8
7	10	3	3	2	6	8	7	4	4	4	2	7	12	6	1	5	8	0	3	1	8	10
11	9	5	5	9	6	3	5	1	6	6	6	10	4	9	2	8	12	2	5	3	3	5
5	5	5	9	7	3	7	10	7	2	7	3	12	10	4	3	1	6	3	4	3	9	7
3	1	2	1	3	2	4	3	1	3	2	2	3	4	3	1	5	1	4	2	1	1	2
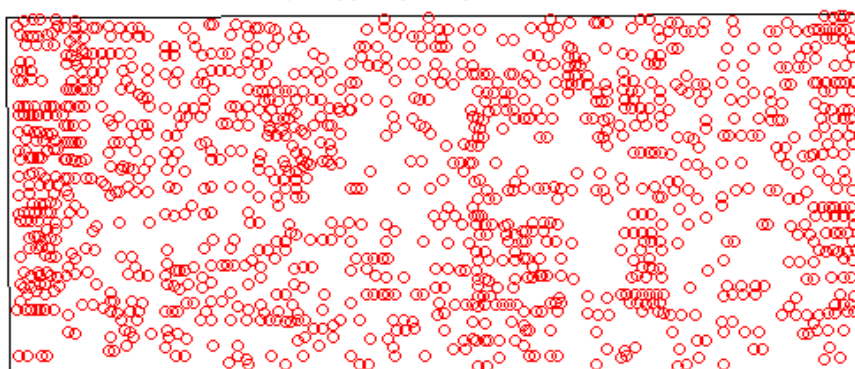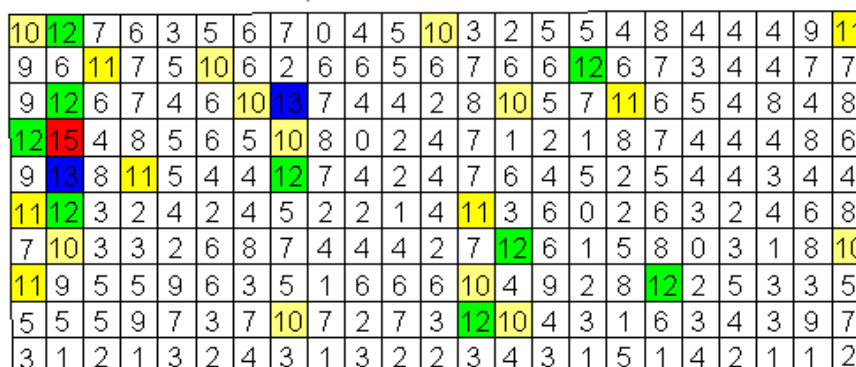
1249 points around 238 hectares, with a average intensity 5.241 points per square hectare

PATA_MABULA R1 SUB-BLOCK AGGREGATION EXPLORATION

a) Mapped point patterns
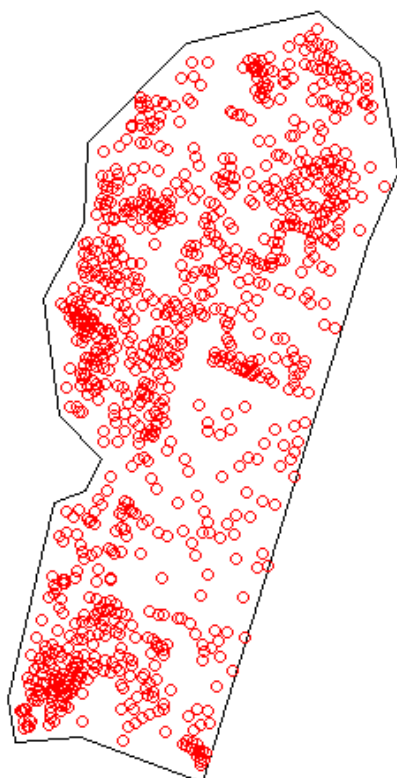
b) QuadratCounts Method
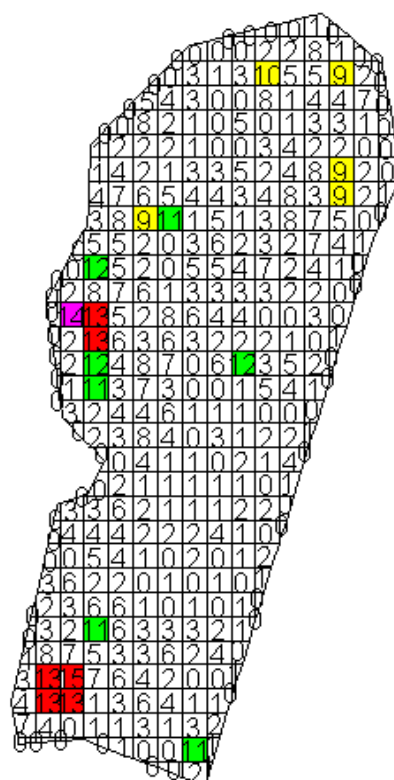
992 points around 314 hectares, with a average intensity 3.163 points per square hectare

PATA MABULA Z1 SUB-BLOCK AGGREGATION EXPLORATION

a) Mapped point patterns
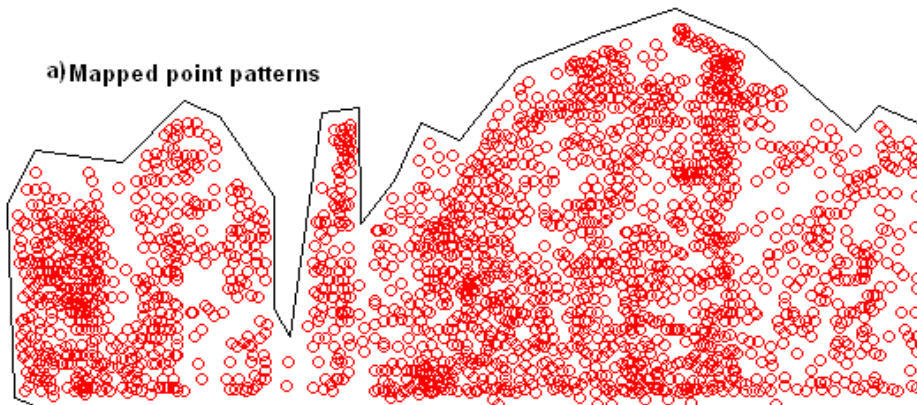
b) QuadratCounts Method
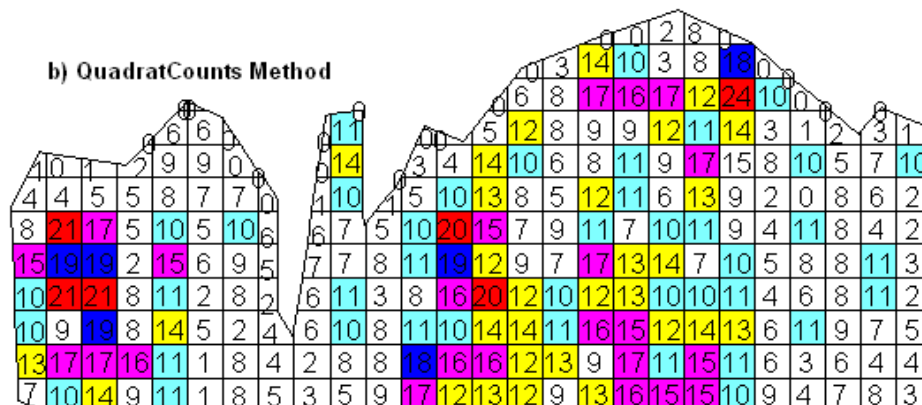
2158 points around 239 hectares, with a average intensity 9.023 points per square hectare

PATA MABULA Z2 SUB-BLOCK AGGREGATION EXPLORATION
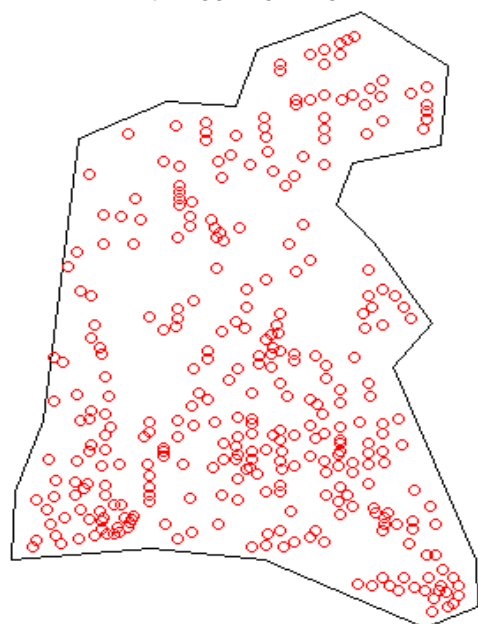
a) Mapped point patterns   b) QuadratCounts Method
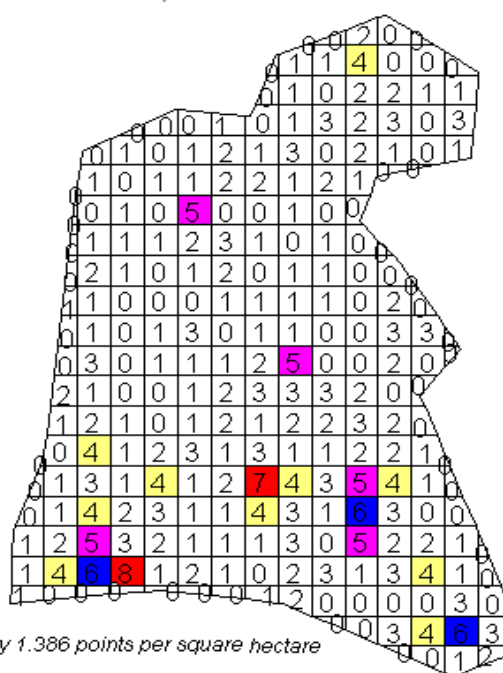
324 points around 234 hectares, with a average intensity 1.386 points per square hectare

5.2.8 Appendix H: The three more dense hectares in exploration of aggregation

SUB-BLOCK H

hectare 21 with 5 points , -1 on edge

hectare 22 with 8 points

hectare 23 with 2 points

hectare 40 with 5 points   hectare 41 with 5 points   hecctare 42 with 2 points
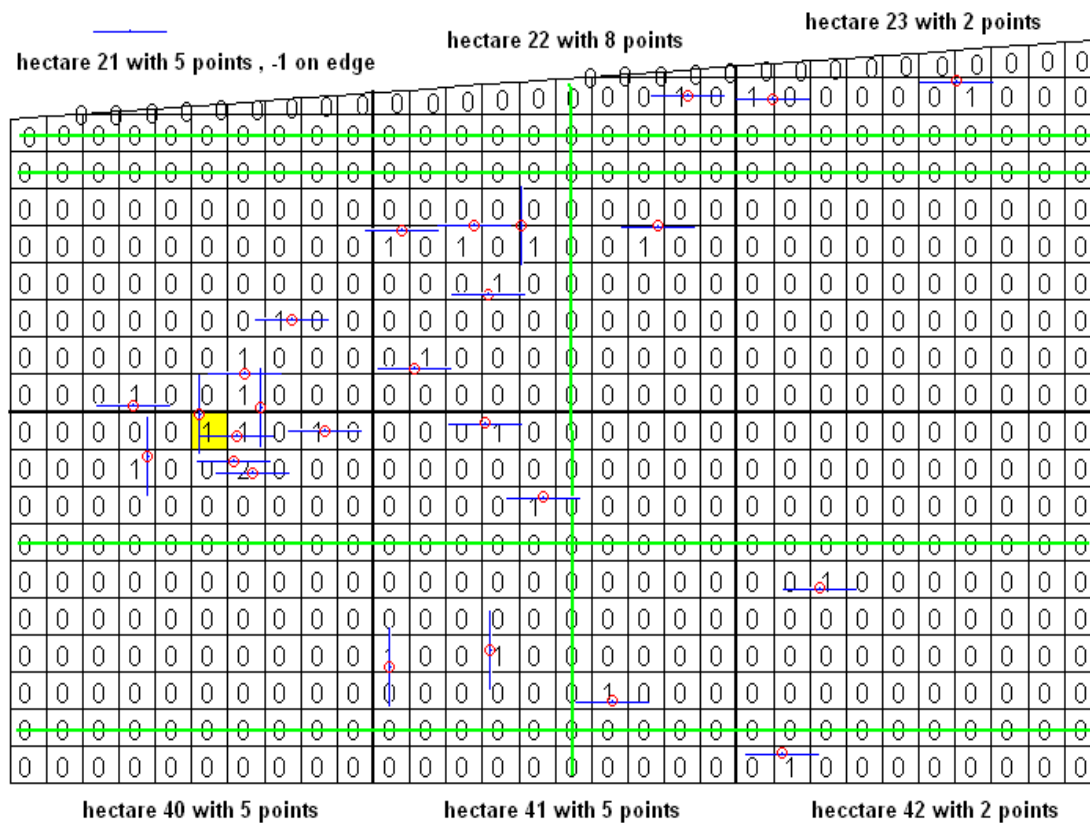
SUB-BLOCK J

hectare 153 with 4 points, -1 on edge   hectare 154 with 7 points   hectare 155 with 1 point

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

hectare 260 with 6 points   hectare 261 with 6 points, -1 on edge   hectare 262 with 2 points

SUB-BLOCK Q-PART 1

hectare 676 with 13 points

0	0	0	1	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0
0	1	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1
0	0	1	1	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0

hectare 341 with 11 points

0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0

hectare 642 with 10 points, +1 on edge

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
1	0	0	1	0	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1
0	1	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

hectare: 322, 323, 364, 365 with 20 points

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

SUB-BLOCK Q-PART 2

hectare 72 with 15 points

0	0	0	0	0	1	1	0	1	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

hectare 94 with 13 points

0	0	0	0	0	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	1	0	1	0
0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0	1	0	0	0	0	0	0

hectare 54 with 13 points

0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	1	0	0	1
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	1	0	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	0

hectare 71 with 12 points   hectare 72 with 15 points   hectare 73 with 4 points

hectare 71 with 12 points

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	0

hectare 72 with 15 points

0	0	0	0	0	1	1	0	1	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

hectare 73 with 4 points

1	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0

SUB-BLOCK R1

hectare 433 with 13 points   hectare 434 with 15 points, -1 on edge

0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
1	1	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
0	0	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	1	2	0	1	0	0	0	0
0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

hectare 449 with 13 points   hectare 450 with 13 points, 1 on Edge

SUB-BLOCK Z1

hectare 73 with 25 points   hectare 158 with 21 points   hectare 168 with 10 points + 1 on edge

SUB-BLOCK Z2

hectare 274 with 8 points , +1 on edge   hectare 233 with 7 points   hectare 251 with 6 points , -2 on edge

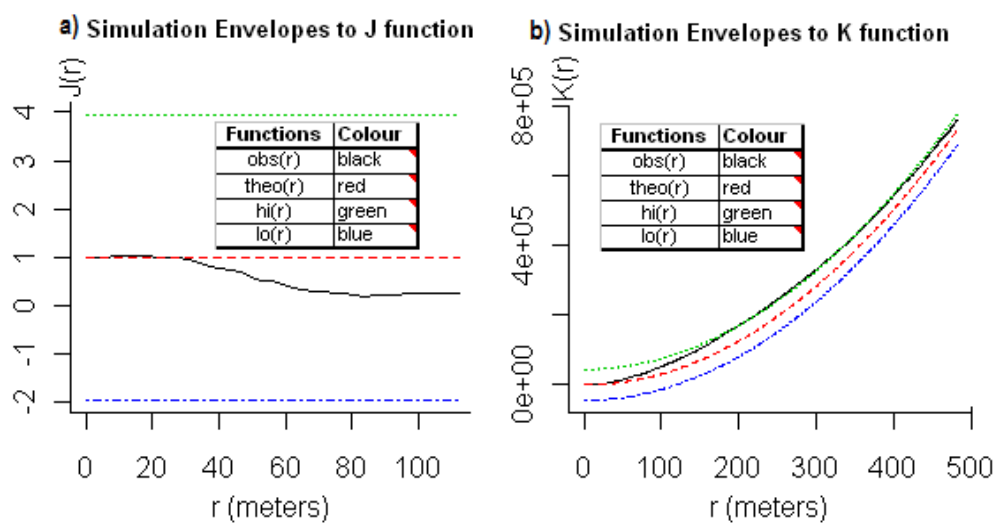hectare 272 with 4 points , -1 on edge   hectare 273 with 6 points   hectare 274 with 8 points , +1 on edge
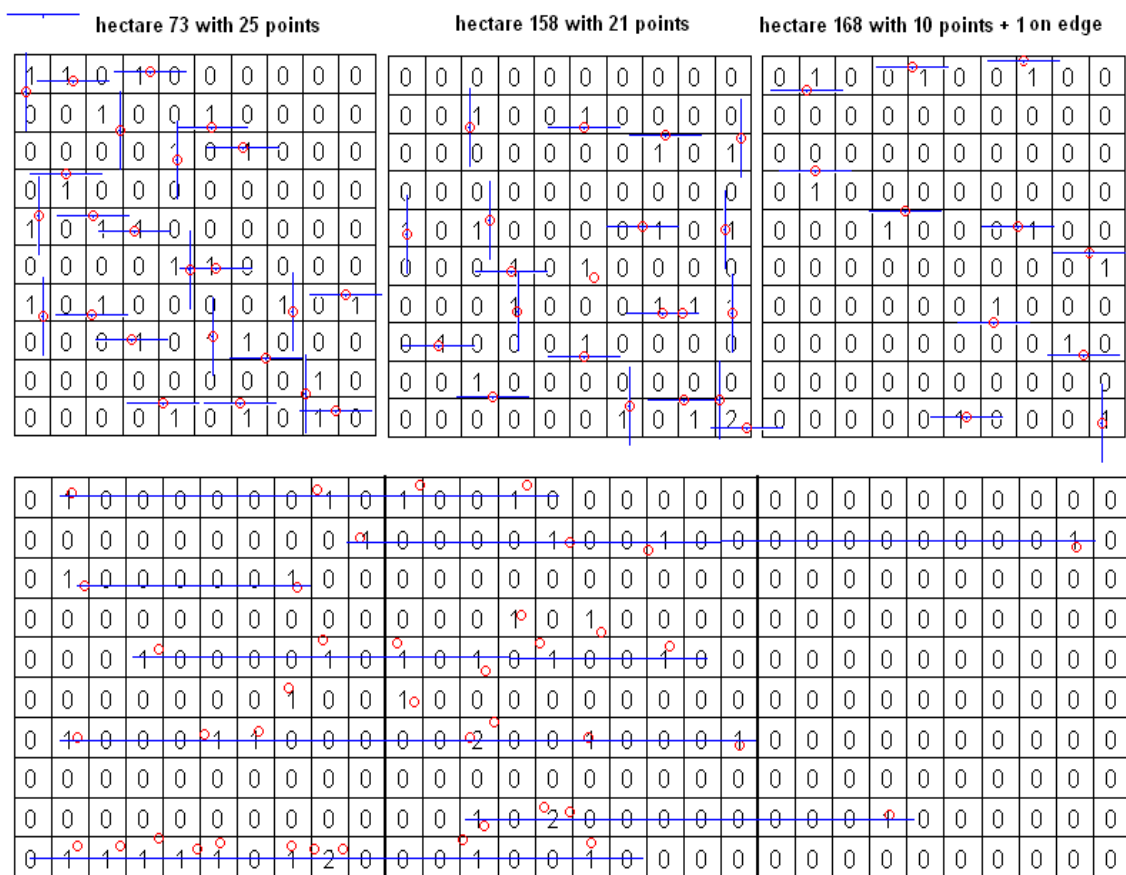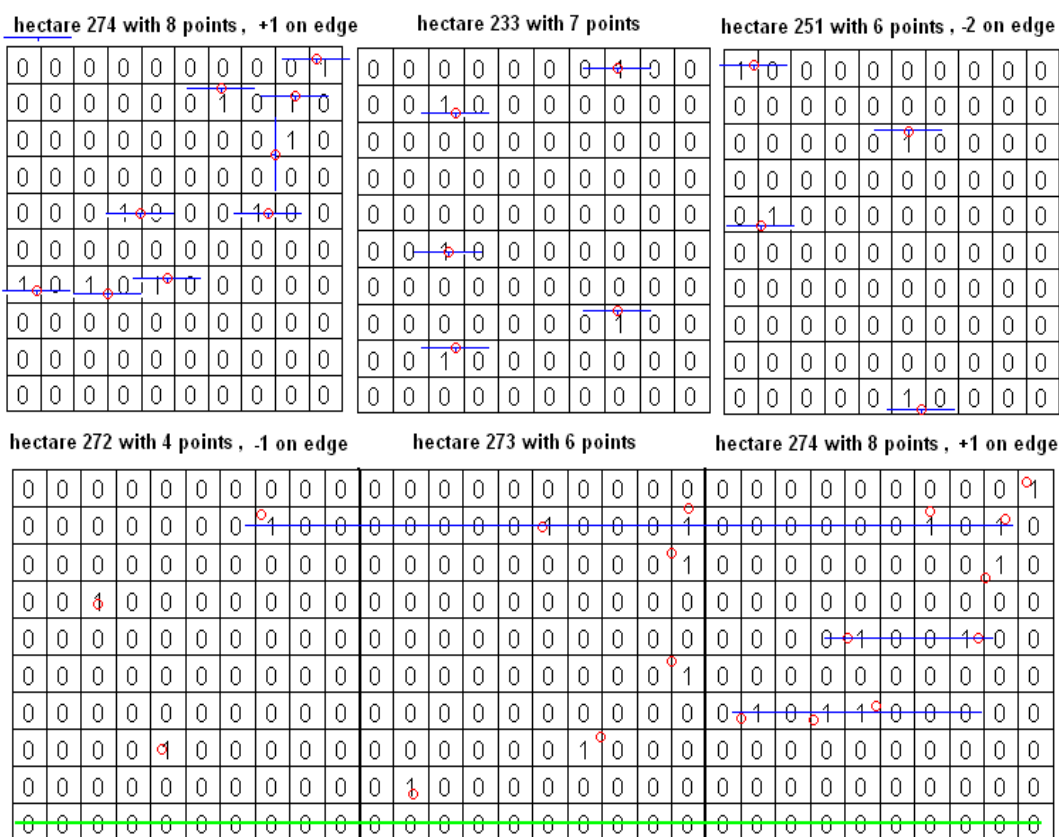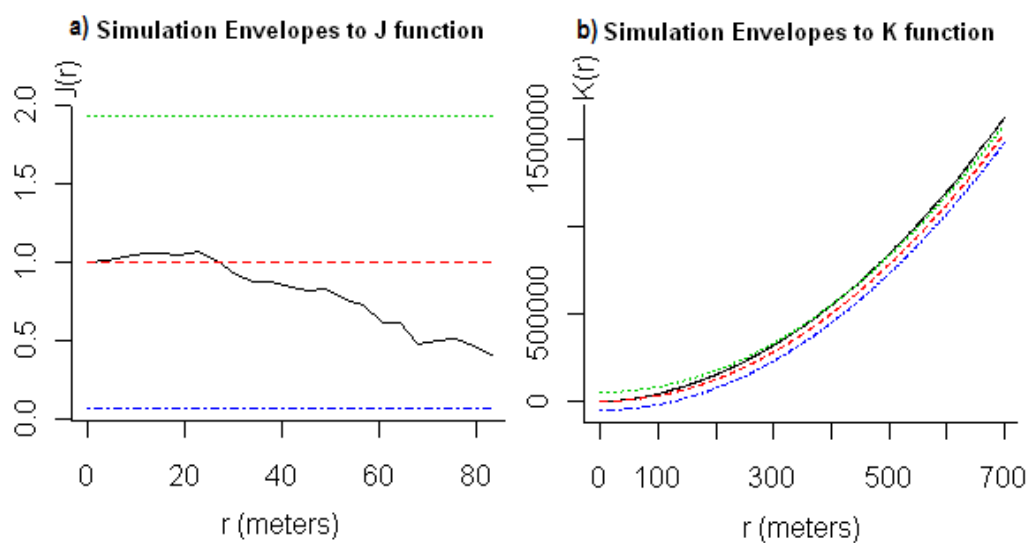
5.2.9   Appendix I: *J*-function and *K*-function panels from the sub-blocks

SUB-BLOCK H

SUB-BLOCK J

SUB-BLOCK Q1

a) Simulation Envelopes to J function

b) Simulation Envelopes to K function

SUB-BLOCK Q2

a) Simulation Envelopes to J function

b) Simulation Envelopes to K function

SUB-BLOCK R1

a) Simulation Envelopes to J function

b) Simulation Envelopes to K function

SUB-BLOCK Z1

a) Simulation Envelopes to J function

b) Simulation Envelopes to K function

Functions	Colour
obs(r)	black
theo(r)	red
hi(r)	green
lo(r)	blue

Functions	Colour
obs(r)	black
theo(r)	red
hi(r)	green
lo(r)	blue

SUB-BLOCK Z2

a) Simulation Envelopes to J function

b) Simulation Envelopes to K function

Functions	Colour
obs(r)	black
theo(r)	red
hi(r)	green
lo(r)	blue

Functions	Colour
obs(r)	black
theo(r)	red
hi(r)	green
lo(r)	blue

5.3   List of Tables

5.4   List of Figures

5.5   Notation Index

5.5.1   Abbrevations

a.s.   almost sure

c.d.f.   cumulative distribution function

i.i.d.   independent and identically distributed

p.d.f   probability density function

r.v.   random variable

5.5.2 Nomenclature

R^d	The d-dimensional Euclidean Space (set of real numbers)
d	dimensions
R	Statistic Software Tool
$X(t)$	Index random variables (p. 7)
B or A	a compact region in R^d
$V_{B,k}$	Basic event N which has k points in the compact region B (p. 7)
N	Point process
$N(.)$	Number of spatial point in a compact region as $N(A)$ (p. 9)
υ	First Moment of point processes
λ	Intensity
$\lambda(.)$	Intensity Function
$E[.]$	Mean value
X	Point Process
υ_2	Second Moment of point processes
K	Ripley's Function
^{0}C	Centigrade degrees
r	Number of rows for of compact region A (p. 13)
c	Number of columns for of compact region A (p. 13)
n	Number of realization of point processes model (p. 13-14)
$\bar{n}$	Expected number of events in any quadrat under CSR (p. 13)
Z	An observed pattern in S
α	Significance level of a test (p. 14)
$I(.)$	Indicator function which returns 1, if (.) is true or false else.
G or $G(.)$	Nearest Neighbour Distance Function
$\hat{G}(.)$	Estimate the Nearest Neighbour Distance Function G (p. 15)
F or $F(r)$	Contact Distance or Empty Space Function
r or s	Measure of a radius of some ball b
$\hat{F}(r)$	Estimate the Empty Space Function F
J or $J(r)$	The Function J
J	A Block from the Bimpe Forest Concession Area
g or $g(r)$	Pair Correlation Function
(x, y)	The coordinates of a spatial point (tree)

5.6 Bibliography

5.6.1 Forestry literature review

Asner, Gregory P., et al. "Estimating Canopy Structure in an Amazon Forest from Laser Range Finder and IKONOS Satellite Observations". BIOTROPICA 34(4) 2002: 483-492. Web. 28 Apr. 2011.

Balakrishman N. *Mehtods and Applications of Statistics in the life and Health Sciences*. USA: Wiley, 2010. Print.

Bwangoy, Jean-Robert B. "Wetland mapping in the Congo Basin optical and radar remotely sensed data and derived topopanelical indices". Remote Sensing of Environment 114 2010: 73-89. Print.

Linares-Palomino R. "Spatial distribution patterns of trees in a seasonally dry forest in the Cerros de Amotape National Park, Northwestern Perú". Bosques relictos del NO de Perú y SO de Ecuador 12(2) 2005: 317-326. Web. 27 Jul. 2010.

Ministry of Forests and Range from British Columbia. *Glossary of Forestry Terms in British Columbia Online*. 2008. Web. 27 Jan. 2011.

Neeff, Till et al. "Markov point process for modeling of spatial forest patterns in Amazonia derived from interferometric height". USA: Remote Sensing of Environment (2005). Wed. 06 Nov. 2010.

Salovaara, Kati J., et al. "Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery". Remote Sensing of Environment 97 (2005): 39-51. Web. 06 Nov. 2010.

Stephenson, Nathan L. "Forum: Use of Tree Aggregations in Forest Ecology and Management". Environmental Management 11.1 (1987): 1-5. Wed. 03 Apr. 2011.

Tarr, Matt and Whitetail-Stewards Inc. *Glossary Online*: Co-dominant tree. 2009. Web. 02 Nov. 2011.

Bwangoy, Jean-Robert B. Personal Communications. From: Jul-2010 to April-2011.

5.6.2 Mathematic literature review

Baddeley, Adrian, et al. *Analysing spatial point patterns in R*. Australia: CSIRO, 2008. Print.

Baddeley, Adrian and Richard Gill D. "Kaplan-Meier Estimators of Distance Distributions for Spatial Point Processes". The Annals of Statistics 25.1 (1997): 263-292. Print.

Baddeley, Adrian, et al. *Stochastic Geometry*. Berlin: Springer, 2007. Print.

Baddeley, Adrian and Rolf Turner. *Help pages for package 'spatstat' version 1.14-8*. Page: Estimate the empty space function F, 2005. Print.

Bivand, Roger S., Edzer J. Pebesma and Virgilio Gomez-Rubio. *Applied Spatial Data Analysis with R*. New Year: Springer, 2008. Print.

Cressie, Noel. *Statistics for Spatial Data*. USA: John Wiley & Sons Inc, 1993. Print.

Diggle, Peter A. C. *Statistical Analysis of Spatial Point Patterns*. London: Academic Press Inc. Ltd, 1983. Print.

Illian, Janine, et al. *Statistical Analysis and Modeling of Spatial Point Patterns*. USA: John Wiley & Sons, Ltd, 2008. Print.

Møller, Jesper and Rasmus Waagepetersen P. *Modern statistics for spatial point processes*. Nordic Conference 11-15 Jun. 2006. Web. 01 Sep. 2010.

Ripley, Brian D. *Spatial Statistics*. USA: John Wiley & Sons Inc, 1981. Print.

Ripley, Brian D. *Statistical Inference for Spatial Processes*. Cambridge: Press Syndicate of the University of Cambridge, 1988. Print.

Schabenberger, Oliver and Carol A. Gotway. *Statistical Methods for Spatial Data Analysis*. USA: Champ & Hall/CRC Press, 2005. Print.