Análisis de la estructura espacial de los árboles del bosque seco tropical mediante teoría de procesos puntuales.

Luis Fernando Parra Camacho

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
March 2019
Análisis de la estructura espacial de los árboles del bosque seco tropical mediante teoría de procesos puntuales.

Luis Fernando Parra Camacho

Dissertation submitted in partial fulfilment for the degree of
Magister in Science - Statistics

Advisor
Martha Patricia Bohorquez Castañeda, Ph.D.
Doctor in Science - Statistics

Research line
Spatial Statistics (Point Processes)

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
March 2019
Title in English
Analysis of the spatial structure of the trees of a tropical dry forest through point process theory.

Título en español
Análisis de la estructura espacial de los árboles del bosque seco tropical mediante teoría de procesos puntuales.

Abstract: A marked point pattern of a tropical dry forest, obtained from the locations of each plant in a one-hectare plot located in the national natural park El Tuparro in the Colombian Orinoquia was analysed. We used two qualitative marks species and size, and a quantitative mark diameter at breast height-dbh. The objective of the present study was to evaluated at an ecological community level, the spatial distribution of the plants with the marked point pattern, to understand the processes of facilitation and competition that have an influence on the generated point pattern. We propose a methodology based on different tools of point process theory: first we estimated the intensities using a kernel and the incorporation of a covariate (distance to rocks); then we used different second order tools to analyse the inter and intra dependences of each of the marks; and finally we used three classification methods to group the species by their spatial behaviour, one based on dispersal indices and the other two on second order functions. The results show an influence of the rocky soil over the spatial distribution of the plants inside the study area; the second order functions indicate an inhibition pattern for all of the species, and for some a cluster pattern at close distances was identify; the size of the plants also influence the spatial distribution, large and medium trees tend to created small clusters pattern with a radius of 1.3 meters, while small plants aggregated more around big plants than around medium plants; the three types of information used for the classification analysis show similarities in the groups of species form, however, the second order functions displays better classification results.

Resumen: Se analiza un patrón puntual marcado de un bosque seco tropical, el cual es generado a partir de la localización de los árboles dentro de una parcela de una hectárea ubicada en el parque nacional natural El Tuparro en la Orinoquia Colombiana. Utilizamos dos marcas cualitativas la especie y tamaño, y una marca cuantitativa el diámetro a la altura del pecho - dap. El objetivo del estudio fue analizar a un nivel de comunidad ecológica, la distribución espacial de las plantas mediante el patrón puntual marcado, para entender los procesos de competencia y facilitación que tienen una influencia en el patrón generado. Proponemos una metodología basada en diferentes herramientas provenientes de la teoría de procesos puntuales: primero estimamos la intensidad utilizando un kernel e incorporando la covariable distancia a las rocas; después utilizamos diferentes herramientas de segundo orden para analizar las intra e inter dependencias de cada una de las marcas; y finalmente usamos tres métodos para clasificar a las especies por su comportamiento espacial, uno basado en índices de dispersión y los otros dos en funciones de segundo orden. Los resultados muestran una influencia del suelo rocoso sobre la distribución espacial de las plantas dentro del área de estudio; las funciones de segundo orden indican un patrón de inhibición para todas las especies, y para algunas se identificó un patrón de agrupación a distancias cercanas; el tamaño de las plantas también influye en la distribución espacial, los árboles grandes y medianos tienden a formar pequeños
patrones agregados con radios de 1.3 metros, mientras que las plantas pequeñas se agrupan más alrededor de plantas grandes que medianas; los tres tipos de información utilizados para el análisis de clasificación muestran similitudes en los grupos de especies formados, sin embargo, las funciones de segundo orden muestran mejores resultados de clasificación.

**Keywords:** Point process, first and second order characteristics, Functional data analysis, spatial distribution, Tropical dry Forest

**Palabras clave:** Procesos puntuales, Características de primer y segundo orden, análisis funcional, distribución espacial, Bosque seco tropical.
Acceptation Note

Thesis Work
Aprobado

“Meritoria o Laureada mention”

________________________________________________________
Jury
Por definir

________________________________________________________
Jury
Por definir

________________________________________________________
Advisor
Martha Patricia Bohorquez Castañeda

Bogotá, D.C., March 15 de 2019
Dedicated to

To my parents and Jenn.
Acknowledgements

This work would not have been possible without the constant support of my advisor Martha Patricia Bohorquez. We also like to thank the Institute of Research of Biological Resources Alexander von Humboldt, for allowing us to use the data collected on the National Natural Park El Tuparro.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>I</td>
</tr>
<tr>
<td>List of Tables</td>
<td>III</td>
</tr>
<tr>
<td>List of Figures</td>
<td>IV</td>
</tr>
<tr>
<td>Introduction</td>
<td>VIII</td>
</tr>
<tr>
<td>1. Theoretical framework</td>
<td>1</td>
</tr>
<tr>
<td>1.1  Point Process</td>
<td></td>
</tr>
<tr>
<td>1.1.1 Stationary and isotropy</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2 Edge effects</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3 Complete spatial randomness CSR</td>
<td>3</td>
</tr>
<tr>
<td>1.1.4 First order characteristics</td>
<td>3</td>
</tr>
<tr>
<td>1.1.5 Second order characteristics</td>
<td>7</td>
</tr>
<tr>
<td>1.1.6 Indices</td>
<td>15</td>
</tr>
<tr>
<td>1.2  Functional data</td>
<td>16</td>
</tr>
<tr>
<td>1.2.1 Functions systems</td>
<td>16</td>
</tr>
<tr>
<td>1.2.2 Coefficients and Smoothing</td>
<td>17</td>
</tr>
<tr>
<td>1.2.3 Functional principal components analysis</td>
<td>18</td>
</tr>
<tr>
<td>1.3  Ecological Concepts</td>
<td>19</td>
</tr>
<tr>
<td>2. Methods</td>
<td></td>
</tr>
<tr>
<td>2.1  Study site and variables</td>
<td>21</td>
</tr>
<tr>
<td>2.2  Statistical methodology</td>
<td>26</td>
</tr>
<tr>
<td>2.2.1 Point Pattern and Marks</td>
<td>28</td>
</tr>
<tr>
<td>2.2.1.1 Marks</td>
<td>28</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Values of the three methods used to estimated the bandwidth for the unmarked point pattern and the sub patterns of the qualitative marks species and size, and selection of the method use. .................................................. 36

3.2 Values of the parameters of the intensity model for the unmarked point pattern and the for some sub patterns of the qualitative marks species and size. ................................................................. 37

3.3 Groups defined in the three methods, and the number of species in common .................................................. 82

4 Information about the parameters of the models .................................................. 102
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Distribution of trees in a 1 ha plot in the National Natural Park El Tuparro.</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Upper left: Location of the National Natural Park El Tuparro in Colombia,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bottom left: Location of the plot in the National Natural Park El Tuparro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(red point), left: Distribution of trees inside the plot</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Distribution of the plants (green circles) and rocks (gray polygons) inside</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>the study area</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Number of plants in each of the 32 species with more than ten plants inside</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>the study area</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Left: Box plot of the values of diameter at breast heigh dbh, Right: Number</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>of plants in each of the categories of the mark size</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>Box plot of the value of diameter at breast heigh dbh for the six species</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>with more plants</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Pie chart with the percentage of each class of habit</td>
<td>25</td>
</tr>
<tr>
<td>2.7</td>
<td>Propose methodology scheme for the analysis of the point pattern generated</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>by the plants of a TDF</td>
<td></td>
</tr>
<tr>
<td>2.8</td>
<td>Top: sector K-functions, Bottom: Rose diagram of estimated probability density</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>for the point pair orientation distribution of the point pattern</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Left: box-plot of the values of intensity of each species, Right:box-plot of</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>the values of proportion of each species</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>For the Unmarked point pattern. Left: kernel, Center: intensity of the</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>parametric model, Right:intensity of the model, rocks and points.</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>For the species Bactris bidentula. Left: kernel, Center: intensity of the</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>parametric model, Right:intensity of the model, rocks and points of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>species.</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>For the species Eschweilera tenuifolia. Left: kernel, Center: intensity of the</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>parametric model, Right:intensity of the model, rocks and points of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>species.</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>For the species Inga gracilifolia. Left: kernel, Center: intensity of the</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>parametric model, Right:intensity of the model, rocks and points of the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>species.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IV</td>
</tr>
</tbody>
</table>
3.6 For the species Morf sp4. Left: kernel, Center: intensity of the parametric model, Right:intensity of the model, rocks and points of the species. 43
3.7 For the species Pachira nukakica. Left: kernel, Center: intensity of the parametric model, Right:intensity of the model, rocks and points of the species. 44
3.8 Box plot of the value of diameter at breast heigh dbh of the plants over rocky soil, shown by species 45
3.9 For the category Large of the size mark. Left: kernel, Center: intensity of the parametric model, Right:intensity of the model, rocks and points of the category. 46
3.10 For the category Medium of the size mark. Left: kernel, Center: intensity of the parametric model, Right:intensity of the model, rocks and points of the category. 47
3.11 For the category Small of the size mark. Left: kernel, Center: intensity of the parametric model, Right:intensity of the model, rocks and points of the category. 48
3.12 Left: inhomogeneous $k$ function, Right: inhomogeneous pair correlation function. For the unmarked point pattern 50
3.13 Probabilities for 32 species 51
3.14 Spatial distribution with the value of diameter at breast heigh dbh for each of the 11 species selected 53
3.15 inhomogeneous $k$ function and inhomogeneous pair correlation function for the 11 species selected 54
3.16 condensed $k$ function and condensed pair correlation function for the 11 species selected 58
3.17 Left: Distribution of the tree in each class of size, Right: Distribution of the trees in different ranges of dbh 61
3.18 Top: inhomogeneous $k$ function, Bottom: inhomogeneous pair correlation function. For the categories of the size mark 63
3.19 Inter-type $k$ function and the cross pair correlation function for the different categories of the mark size. 65
3.20 Top: r- mark correlation function, Bottom: mark variogram. Of the quantitative mark diameter at breast heigh-dbh. 67
3.21 Top: box-plot of the values clark-evans index and Hopkins-Skellam test, Bottom: box-plot of the values the mean distance to the nearest neighbour of the same species. 68
3.22 Cluster analysis (Wards method) of the 32 species using the values of the three indices 70
3.23 Boxplot of the values of dbh for each species in the three groups form.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.24</td>
<td>Smoothed K functions of 32 species used for Functional Principal Components Analysis. Top: Smooth K functions, Bottom: Smooth Pair correlation functions (close up to distance 0 to 10 meters and values 0 to 10).</td>
</tr>
<tr>
<td>3.25</td>
<td>Diagram of the cluster analysis (Wards method) of the 32 species using the scores of the first functional principal component (k functions)</td>
</tr>
<tr>
<td>3.26</td>
<td>k functions of the 32 species separated by the groups form in the cluster analysis</td>
</tr>
<tr>
<td>3.27</td>
<td>Boxplot of the values of dbh for each species in the three groups form. Method k functions</td>
</tr>
<tr>
<td>3.28</td>
<td>Cluster analysis (Wards method) of the 32 species using the scores of the two first functional principal components (Pair correlation functions)</td>
</tr>
<tr>
<td>3.29</td>
<td>k functions of the 32 species separated by the groups form in the cluster analysis</td>
</tr>
<tr>
<td>3.30</td>
<td>Boxplot of the values of dbh for each species in the three groups form. Method pcf</td>
</tr>
<tr>
<td>3.31</td>
<td>Similarities in the groups of the three different methods of cluster analysis of the 32 species selected</td>
</tr>
<tr>
<td>3.32</td>
<td>Two options to construct the mark diameter breast height dbh for the trees with more than one stems. Option 1: sum of dbh of the different stems, Option 2: use of basal area formula.</td>
</tr>
<tr>
<td>3.33</td>
<td>Box plot of the value of diameter at breast heigh dbh for 32 species</td>
</tr>
<tr>
<td>3.34</td>
<td>Kernel and estimated intensity for the species Arrabidaea sp</td>
</tr>
<tr>
<td>3.35</td>
<td>Kernel and estimated intensity for the species Astronium graveolens</td>
</tr>
<tr>
<td>3.36</td>
<td>Kernel and estimated intensity for the species Attalea microcarpa</td>
</tr>
<tr>
<td>3.37</td>
<td>Kernel and estimated intensity for the species Bactris bidentula</td>
</tr>
<tr>
<td>3.38</td>
<td>Kernel and estimated intensity for the species Bursera simaruba</td>
</tr>
<tr>
<td>3.39</td>
<td>Kernel and estimated intensity for the species Clathrotropis macrocarpa</td>
</tr>
<tr>
<td>3.40</td>
<td>Kernel and estimated intensity for the species Erythroxylum macrophyllum</td>
</tr>
<tr>
<td>3.41</td>
<td>Kernel and estimated intensity for the species Eschweilera tenuifolia</td>
</tr>
<tr>
<td>3.42</td>
<td>Kernel and estimated intensity for the species Ficus americana</td>
</tr>
<tr>
<td>3.43</td>
<td>Kernel and estimated intensity for the species Fridericia pubescens</td>
</tr>
<tr>
<td>3.44</td>
<td>Kernel and estimated intensity for the species Guarea glabra</td>
</tr>
<tr>
<td>3.45</td>
<td>Kernel and estimated intensity for the species Gustavia augusta</td>
</tr>
<tr>
<td>3.46</td>
<td>Kernel and estimated intensity for the species Heisteria acuminata</td>
</tr>
<tr>
<td>3.47</td>
<td>Kernel and estimated intensity for the species Himatanthus articulatus</td>
</tr>
<tr>
<td>3.48</td>
<td>Kernel and estimated intensity for the species Inga gracilifolia</td>
</tr>
<tr>
<td>3.49</td>
<td>Kernel and estimated intensity for the species Inga sp</td>
</tr>
</tbody>
</table>
50 Kernel and estimated intensity for the species Licania micrantha . . . . . . . . 112
51 Kernel and estimated intensity for the species Matayba sp . . . . . . . . . . . . 112
52 Kernel and estimated intensity for the species Morf sp2 . . . . . . . . . . . . . . 113
53 Kernel and estimated intensity for the species Morf sp4 . . . . . . . . . . . . . . 113
54 Kernel and estimated intensity for the species Ocotea schomburgkiana . . . . 114
55 Kernel and estimated intensity for the species Pachira nukakica . . . . . . . . 114
56 Kernel and estimated intensity for the species Petrea sp . . . . . . . . . . . . . 115
57 Kernel and estimated intensity for the species Phanera guianensis . . . . . . . 115
58 Kernel and estimated intensity for the species Pouteria plicata . . . . . . . . . 116
59 Kernel and estimated intensity for the species Pouteria sp4 . . . . . . . . . . . 116
60 Kernel and estimated intensity for the species Protium guianense . . . . . . . . 117
61 Kernel and estimated intensity for the species Pseudolmedia sp1 . . . . . . . . 117
62 Kernel and estimated intensity for the species Rudgea crassiloba . . . . . . . . 118
63 Kernel and estimated intensity for the species Sapium glandulosum . . . . . . 118
64 Kernel and estimated intensity for the species Senegalia macbridei . . . . . . . 119
65 Kernel and estimated intensity for the species Siparuna guianensis . . . . . . . 119
Introduction

The tropical dry forest (TDF) is distributed all around the tropical areas of the world. Most of it is located in America: 54.2% in South America and 12.5% in Central America. Although the TDF is considered one of the most endangered ecosystems in the tropical zones of the world, it is one of the least studied [22]. The main threats to this type of forest present are: deforestation due to change to other land uses like agriculture and livestock mostly cattle raising, degradation because of intentional and accidental fires, logging for home use and logging of species of high wood value [47], [22].

In South America the most extensive and continuous areas of the TDF are located in the north east of Brazil, south east of Bolivia and Paraguay, and North east of Argentina. While in Colombia and Peru the areas of the tropical forest are less extensive and not well protected [22].

In Colombia originally there were 8.882.854 ha of TDF, now there’s only a 22.5% of this area with forest. Most of the original area of this type of ecosystem has been converted into grass lands to raise cattle (33.2%) and agriculture lands (28.2%), the rest have been turned into roads, cities or other type of infrastructure [56].

The reduction and degradation of this type of forest is a great environmental problem. This ecosystems are of great importance due to their specific characteristics. The climatic conditions are very particular, the temperature is over 25°C, precipitation between 700 to 2200 mm yearly, and it has three or more months of drought in a year (precipitation under 100 mm a month). Due to this characteristics the TDF shelter many endemic species of fauna and flora, that can adapt to this extreme conditions [56].

Besides the biological importance of the TDF, this ecosystem present many benefits to the to the environment: food and wood supply for the nearby communities, environmental services like soil stabilization, recycle of nutrients and climate regulation [56].

The constant threat on this ecosystem, was a motive to created the Latin American Seasonally Dry Tropical Forest FLoristic Network - DryFlor. They encourage communication and cooperative work between the organizations that are doing research on this type of forest and the governmental and no governmental institutions that are in charge of protecting, preserving, restoring and applying a sustainable use of the TDF. The following institutions are link to this network in Colombia: Fundacion ecosistemas secos de Colombia, Jardin Botanico de Medellin, Instituto de investigaciones biologicas Alexander von Humboldt, Fundacion Orinoquia Biodiversa, Universidad de Pamplona, Asociacion de Becarios del Casanare, Fundacion Yoluka, Fundacion Natura y fundacion ecosistemas secos de Colombia (dryflor.info)
In Colombia the department of environment and sustainable development is in charge of giving an orientation and promoting the different actions of research that is being done over the natural resources. Because of the importance of the TDF in Colombia, the department have declared this ecosystem as a strategic natural resource to conserve. In consequence, the institute of biological research Alexander von Humboldt, entity attach to the department, has being studying this type of ecosystem so they can gather the knowledge to gift advise on different strategies of conservation, restoration and sustainable use of this type of forest (humboldt.org.co).

The focus of the research that gather this kind of knowledge are base on the ecology of the forest, where the ecological processes that have a spatial nature are studied [75]. Different ecological processes like: mortality, establishment, regeneration, growth, facilitation and competition, affect the spatial structure of the trees in a forest [15], so they are of interest in this kind of research.

Forest have complex structures due to the different processes that affects them, this leads to different spatial patterns [14]. This is why the pattern or spacial distribution of the plants has been acknowledge like a foot print of the biological process that determinate the dynamics of the forest [55].

Initially ecological studies did not took into account the spatial distribution of the trees, in some cases, methods of experimental designs were used to correct the spatial effect in their analysis. In the 1980s the spatial distribution started to take an important role in ecological research, this was driven by different factors: computer science developments that provided tools capable of modelling and analysing complex spatial relationships; new available information from aerial photographs and satellites images, providing an overview of ecological patterns; and, the need in ecological studies to incorporated a spatial thinking [75].

With this developments point pattern analysis has become an important tool during the last years. This kind of analysis is based on a variety of tools develop from point processes theory, whichs helps understand the underlying process of the population which influence the spatial distribution of the plants in a forest [36]. A point process is consider a random mechanism where one realization gives a point pattern; the point pattern is a group of points or events with coordinates in a D-dimensional space $\mathbb{R}^d$. This events may have different characteristics that describe them, which can be call marks [8][39].

In the context of the present study the point pattern is generated by the location of the trees, while the characteristics of each tree (species and diameter at breast heigh-dbh) are consider the marks of the point pattern.

Different studies have review the applications that point process theory have on ecological and forest research. Stoyan and pettinen (2000) [68], describe different analysis focus on: the estimate of the intensity of trees in a particular area, the use of correlation functions between trees to improve aggregation and segregation indices, and the implementation of point process models in exploratory and confirmatory analysis of forest. Comas and Mateu [14] updated the study of Stoyan and Pettinen, they show not only the statistical tools of point process but they focus on the applicability to analyse and model forest dynamics.

One way to analyse a point pattern is by the first moments of the random variable number of events or points in a area $B$, this is defined as $N(X \cap B)$ [10]. The intensity of points is the first order characteristic, while the functions $K$ of Ripley, $L$ of Besag and
the pair correlation are considered characteristics of second order [39]. This functions are based on the second moment of $N(X \cap B)$. [8].

In 2012, a research in a tropical rainforest used a pair correlation function restricted to a radius of 10 meters (scale at which many ecological important processes are expected in a rain forests), to show that growing population should present a more cluster type of pattern than shrinking populations of the same current local abundance [28].

There is also other tools that can give a more detail analysis of point patterns, the implementation of different kinds of point process models can help to understand the spatial structure of the trees in a forest. This kind of models plays an important part in the analysis of forest point pattern, they are used as exploratory and confirmatory methods. Usually the analysis starts with a null model as a base for understanding the point pattern, the most used null model is the homogeneous Poisson process with intensity $\lambda$, this model allow to test the hypothesis of complete spatial randomness CSR [68].

The implementation of this kind of models was used to evaluate seed dispersal and heterogeneity of soil properties in a tropical dry forest. The study conclude that the point pattern study of the TDF does not present CSR due to the complex processes that are behind the point pattern [42].

There are other types of point processes models that have applicability in ecological or forest context. Cox processes and Gibbs processes are the most important ones. Picard et al (2009) highlight the following recommendations when modelling a point pattern from a tropical forest: when the interest is to model the species in a separate way the point pattern usually is aggregated, for this kink of analysis Cox process models area suggested. If the species is omitted and the objective is to work with all trees of the forest, a regular point pattern is usually observe for the trees with greater size, here a Strauss (Gibbs) models are the most useful to implement [55].

This kind of analysis can be implemented for the unmark or mark point pattern. The marks can be qualitative or quantitative characteristic, usually the species and some numerical characteristic of each tree is used.

To aboard forest mark point patterns where the species are the categories, is a challenging task because the correlations functions that describe the spatial variability are based on pairwise relationships, for a forest with n species their is $n \times (n - 1)/2$ pairwise relations to analyse. Usually in tropical forest there is a large number of species in a small area, making the analysis of tropical forest point pattern restricted to some dominant species [14].

One way to tread this problem is to use a simpler analysis where there is one interaction between trees of the same species, and another interaction between points of different species, this will give $2 \times n$ relations [8]. An other way is to used functional principal components on the functions of second order of the different subpoint patterns generated by the species. The objective is to detect the different structures of point pattern to classify the species base on their spatial distribution [38].

An other important characteristic to understand the spatial structure of the TDF is the size of the plants. In dense forest the competition between trees is strongly related with the size of the trees, influencing the spatial distribution of the plants in the forest [55].
Different studies have focus on modelling the spatial structure of the forest by point process with a mark that characterize the size of the trees. Grabarnik and Sarkka (2009) used a Gibbs process hierarchical model with multi-type marks, to estimated the potential force between and within the three classes of size: small, medium and large trees. They found that the smaller classes does not influence the spatial distribution of the bigger classes, but they are influence by trees of bigger classes and of their own class.

The objective of the following study is to analyse the spatial structure of a mark point pattern of a TDF located in the National Park Tuparro in the Colombian Orinoquia. To achieve this goal, we propose a methodology based on tools from point process theory, we used the marks: species (qualitative mark), three categories of size based on the diameter at breast heightdbh (qualitative mark) and the values of dbh (quantitative mark).

First, we used a kernel and integrated the relationship with the rocky soil of the study area to estimated the intensity for the unmarked point pattern and for the sub patterns generated by the two qualitative marks. Based on this intensities we used different inhomogeneous second order functions to evaluated the inter and intra dependence inside the categories of the marks species and size, and to complement the analysis on the size mark we used other second order tools to examined the quantitative mark dbh. Finally we used three classification methods, to group the different species of the TDF based on their spatial behaviour, one based on dispersion indices and other two based on second order characteristics.

Based on the estimated intensities we found that the rocky soil in this type of forest have an influence in the locations of the trees; the intensities also allow us to implement a methodology to determined the more dominant species inside the study area. The second order tools allows us to understand the spatial dependence of the sub pattern of the qualitative marks species and size, in general we see that at very close distances an aggregation pattern is form indicating a facilitation process, and as the distance increases a more regular pattern is observed, which shows an influence of a competition process between the plants of the study area. The three classification methods show similarities in the species of each of the groups form, however the dispersion indices method indicate that most of the species have a cluster pattern which is contradictory to the actual spatial behaviour; on the other hand the groups form by the inhomogeneous pair correlation function, represent better the different aggregation patterns at close distances and the strength of the regular pattern at greater distances.

The present document starts with the theoretical framework in which describe the main theory of point processes and the tools that derive from this area of spatial statistics and explain some ecological concepts. Then we present a concise methodology, explaining: the data, and a diagram with a flow structure that shows the point pattern, the marks used and the different tools used in the present research. The third chapter contains all the results: first showing a descriptive analysis of the data, then the first order analysis, follow by the second order analysis and finally presenting the classification analysis for the species. The final chapters present the discussions and conclusions, and future work from a interdisciplinary perspective between statistics and ecology.
CHAPTER 1

Theoretical framework

1.1 Point Process

A point process $X$ is a random mechanism where a realization is a point pattern. Any natural or synthetic phenomenon which results in a spatial point pattern can be viewed as a point process [8].

A point pattern denoted by $x$ is a set of points (events) $x_i$ in a d-dimensional space in $\mathbb{R}^d$.

$$x = \{x_1, x_2, ..., x_n\}$$

The number of points $n(x)$ in the pattern is not fixed and may be any finite non negative number including zero. Even if the data is collected in some order $x_1, ..., x_n$; this ordering is artificial, and we treat the pattern $x$ as an unordered set of points [8].

Usually the data collection of the point pattern is done within a specific subset of the space of interest, this set is called the observation window $W$[39].

So if $x$ is our point pattern and $W$ is our window, we can write $x \cap W$ for the subset of $x$ consisting of the points in the pattern that fall in $W$ and $n(x \cap W)$ will be the number of points falling in $W$ which is a well defined random variable [8].

The objects or events in a point process may be characterised not only by their position but also by mark variables, i.e. other information on each event, that may be quantitative or qualitative. This is call a marked point process $M$. The resulting point pattern of this process can be express as:

$$m = \{(m(x_1)), (m(x_2)), ..., (m(x_n))\}$$

where $x_n$ are the locations and $m(x_n)$ is the mark of the point at location $x_n$ [39]. For every event there can be more than one mark.

When working with a qualitative mark the categories $o = 1, \cdots O$ can be considered as sub point process, this are generated by the points with category $o$ of the qualitative mark.
Chapter 1. Theoretical Framework

1.1.1 Stationary and isotropy

In the analysis of a point pattern, assumptions on the point process are required for making inference about data at locations that are not sampled; a great variety of point process tools used in the analysis of point patterns rely on these assumptions [8], [15].

Assuming a point process as stationary and isotropic, gives a guideline for the analysis of the point pattern. A point process is called stationary and isotropic if its statistical properties are unaffected by shifting or rotating the point process; if the process is view in a window $W$ the observable statistical properties do not depend on the location or orientation of the window [8].

To clarify, if $x$ is a point pattern and a translation vector $v$ is applied to each point of the pattern we have $x + v = \{x_i + v : i = 1, ..., n\}$. So formally a point process $X$ is called a strict or strong stationary field if the statistical properties of $X$ and $X + v$ are identical, for any translation vector $v$ [8] [65].

Most spatial data analysis are satisfied with stationary conditions based on the moments of the spatial distribution. This leads us to the term weak stationary field, which implies that the mean of the random field is constant and the covariance is only a function of their spatial separation [65].

Some statistical test can evaluate some aspects of stationarity, but it is impossible to prove rigorously that a point pattern is a sample from a stationary point process [39].

A isotropic point process $X$ is a concept that is analogous to stationarity, but instead of translation by vectors, the rotation in the origin is considered [39]. If any statistical property of the process change when the process is rotated, the process is considered anisotropic [8].

There are different tools to measure anisotropy in a point pattern: the sector $k$ function, pair orientation distribution and the pair correlation function. These tools are describe later in section 1.1.5.

1.1.2 Edge effects

Edge correction methods compensate for the unrecorded points outside the window $W$. The window is usually a part of the space where the point process occurs, so the points close to the border inside the window usually have nearest neighbours outside the window [75].

Most of the time, the choice of edge correction does not seem to be very important as long as some method is applied. But understanding and implementing edge correction is helpful for discovering potential weakness in the analysis. When discrepancies are found between the results with different edge correction methods, this may suggest unusual features in the data and violation of the stationary assumption [8].

For the statistical tools use in point process analysis, different methods of edge correction are proposed. The choice mostly depend on the size and shape of the window $W$ and in the number of points inside the window.

The edge correction methods are describe in more detail for each of the point process statistical tools describe in the following sections.
1.1.3 Complete spatial randomness CSR

A very important point process, in which many of the analysis are built on, is the homogeneous Poisson point process also known as Complete spatial randomness - CSR. This process is characterised by two key properties: (i) the number of events in any planar region $B$ with area or volume $|B|$ follows a Poisson distribution with mean $\lambda|B|$; (ii) given $n$ events $x_i$ in a region $B$, the $x_i$ are an independent random sample from the uniform distribution on $B$ [21].

In (i), the constant $\lambda$ is the intensity or mean number of events per unit area. Based on (i), CSR implies homogeneity which means that the points have no preference for any spatial location, and the intensity of the events does not vary over the plane. Based on (ii), CSR also implies independence, where information about the outcome in one region of space has no influence on the outcome in other regions [21] [8].

Most of the analysis done on point patterns, uses the theoretical values of a CSR process to analysed the behaviour of the point process distribution. If a pattern follows a CSR behaviour there is not much other formal statistical analysis to be done, however if the test rejects CSR, this could be use as a dividing hypothesis to distinguish between patterns classifiable as "regular" or "aggregated" [21].

1.1.4 First order characteristics

Homogeneous intensity

The intensity $\lambda$ is a first order characteristic for stationary point processes. The analysis of a point pattern usually starts by assuming an homogeneous intensity. A point process $X$ is defined to have homogeneous intensity, if it satisfies that the mean number of points of $X$ in any set $B$ is equal to $\lambda$ multiplied by the area of $B$ [39] [8]:

$$E(n(X \cap B)) = \lambda|B|$$ (1.1)

An unbiased, ergodic and consistent estimator for this constant intensity is

$$\hat{\lambda} = \frac{n(X \cap W)}{|W|}$$ (1.2)

When estimating the intensity in a mark point pattern with a qualitative mark, the intensity $\lambda_o$ of the sub-point process generated with the points of the category $o$ of the mark, is the focus of the analysis. This can be estimated by counting the points with mark $o$ in $W$ and dividing by the area or volume of $W$ [39]

$$\hat{\lambda}_o = \frac{\sum_{x \in W} 1(m(x) = o)}{|W|}$$ (1.3)

In the analysis of a point pattern with quantitative marks, the mark sum intensity $\lambda_S$ is a first order characteristic which takes into account the quantitative mark. The standard estimator is based on the sum of the mark of the points inside the window $W$ and divided by the area or volume of the window [39].
\[ \hat{\lambda}_S = \sum_{x \in W} m(x) / |W| \] (1.4)

**Inhomogeneous intensity**

The last definitions are suitable when the point pattern is homogeneous (homogeneous intensity) but in some cases the point pattern is inhomogeneous, so a more general definition needs to be made. To define this, we can imagine the intensity \(\lambda\) as a spatially varying value. So at any spatial location \(u\), the intensity is \(\lambda(u)\), where \(\lambda(u)\) is a function of the location [8].

To be more precise, in a small neighbourhood of the location \(u\) with small area \(a\), the expected number of points is equal to \(\lambda(u)a\). So for any given region \(B\) the integral of the intensity function should be the expected number of points in region \(B\)

\[ E[n(x \in W)] = \int_B \lambda(u)du \] (1.5)

The estimation of inhomogeneous intensity in a point pattern, can be done by non-parametric techniques such as: quadrat counting, kernel estimation and spatially adaptive smoothing.

Quadrat counting, is based on the number of points in regions of equal area. To estimate the intensity function the observation window \(W\) is divided into subregions \(B_1, ..., B_m\) that are define as quadrats. Inside each quadrat the number of points is counted and we obtain the average intensity in each quadrat, and this is a estimator of the intensity function [8].

The basic idea in kernel estimation is that the intensity function \(\lambda(u)\) at a given location \(u\) is the mean number of points within a small disk centered at \(u\) and divided by the area of the disk [75].

A commonly and simple approach used to estimate the intensity function \(\lambda(u)\) proposed by Diggle for a planar case is

\[ \hat{\lambda}(u) = \frac{n(b(u, r))}{\pi h^2} \] (1.6)

where \(n(b(u, r))\) is the number of points inside a circle \(b\) with center at location \(u\) and radius \(r\), where \(h\) represents the bandwidth of the kernel which determines the degree of smoothing [39]. So this approach estimates the intensity function by the points density in a circle with radius \(h\) centered at \(u\).

The specific kernel function implemented in (1.6) is \(k(u) = 1_{b(0, h)}/|b(0, h)|\), where \(1_{b(0, h)}\) is the indicator which equals one if a point is found inside the circle \(b\) centered at 0 with a radius \(h\) and \(|b(0, h)|\) is the volume or area of that circle \(b\).

This estimator can be generalised to be based on a general kernel function \(\kappa(u)\). The kernel \(\kappa\) must be a probability density, which means that, \(\kappa(u) \geq 0\) for all locations \(u\) and \(\int_{\mathbb{R}^2} \kappa(u)du = 1\). So the kernel estimator without any edge correction can be express as
\[ \hat{\lambda}^{(0)}(u) = \sum_{i=1}^{n} \kappa(u - x_i) \] (1.7)

The estimator [1.7] defined as the uncorrected estimator can be modified by a choice of edge correction. The estimators more commonly used are the uniformly corrected (1.8) and Diggle’s edge correction estimator (1.9) [8].

\[ \hat{\lambda}^{(U)}(u) = \frac{1}{e(u)} \sum_{i=1}^{n} \kappa(u - x_i) \] (1.8)

\[ \hat{\lambda}^{(D)}(u) = \sum_{i=1}^{n} \frac{1}{e(x_i)} \kappa(u - x_i) \] (1.9)

where

\[ e(u) = \int_{W} \kappa(u - v)dv \] (1.10)

is a correction for bias due to edge effect. Diggle’s corrected estimator has better performance overall, with a smaller mean square error [8].

An important aspect of any of these three estimators is the bandwidth \( h \). As mentioned before this determines the degree of smoothing. So this brings a difficulty in the choice of an appropriate bandwidth, small values of \( h \) result in estimated intensity functions too spiky and large values of \( h \) gives smoother intensity functions that may ignore local features.

There is no simple recipe for the selection of the bandwidth. The choice should be made in the context of the specific analysis of the point pattern. Nevertheless there are data base methods that can facilitate the analysis and choosing of the bandwidth for the kernel estimation.

The algorithms developed are based on minimising a measure of error, some of this methods are the Diggle and Berman’s method, mean square error cross validation method and the likelihood cross validation method. To understand more about this methods see [8], [20] and [11].

For each specific context in the analysis of point patterns the kernel estimator can be modified by the correction of the edge effect and with the choice of the bandwidth. Additionally, when estimating the intensity function for a point pattern with a quantitative mark, the kernel estimation can also be modified to include the numerical values of the mark.

This is done by multiplying the value of the mark (weight), this estimator is call the weighted kernel estimator [8]. (1.11) shows the weighted estimator without edge correction.

\[ \hat{\lambda}^{(0,w)}(u) = \sum_{i=1}^{n} w_i \kappa(u - x_i) \] (1.11)
To give to each point, a weight that is determined by the value of quantitative mark is very useful, especially when the mark represents multiplicity (number of cases found in one point of the pattern) or physical mass (the diameter or height of the tree).

The kernel estimators described above used the same kernel and bandwidth to compute estimates at different spatial locations. In some cases this could be a weakness, a fixed kernel and bandwidth is unsatisfactory if the intensity fluctuates greatly across the area of interest. With kernel estimation sharp boundaries between areas of high and low intensity could be smoothed out.

A method to avoid this problem is spatially adaptive smoothing. This adaptive estimators of intensity can be based on Dirichlet-Voronoi tessellations (for more information see [8]). In statistical seismology this methods have performed well, since the point pattern observed in this area of study have abrupt changes in intensity.

**Dependence between intensity and a covariate**

In some applications it is important to understand how the intensity of points depends on the values of a covariate. A covariate is any data treated as explanatory, rather than as part of the response. There are different kinds of covariates: a spatial function \(Z(u)\) defined in all spatial locations \(u\), in a forest analysis could be the terrain, hydrology, soil characteristics, and others; a spatial pattern such as an other point pattern, a line segment pattern, or a collection of spatial features could also be a covariate [8].

There are different ways to determine if the intensity depends on the values of the covariate. One is using quadrats counting method. The subregions \(B_1, \cdots, B_m\) do not have to be rectangles of equal area, these subregions could be of any shape. Using the values of the covariate \(Z(u)\) the window \(W\) could be divided into subregions in which each one has similar values of the covariate. This new tessellation of the window is used to study the dependence of the point pattern with the covariate, if the intensities of the subregions are not similar, this will indicate that the covariate has an influence over the intensity of the point pattern [8].

Another way to study this dependence is by using parametric models of intensity. The intensity function \(\lambda(u)\) can be express in any form depending on the context, as long as the intensity values are non negative and the integral of the intensity over the observation window is finite. Using an exponential function of the covariate to estimate the intensity \(\lambda(u)\) is a good way to test the influence of the covariate. An important model that uses this strategy is:

\[
\lambda(u) = \exp(\alpha + \beta Z(u))
\]

with this model, to the test for a covariate effect is equivalent to testing whether \(\beta = 0\) [8].

**Probabilities**

All these methods for estimating intensity can be used for the sub point pattern form by the categories of a qualitative mark. The intensities of these sub patterns are the main input in estimating the probability distribution of these types (categories). The probability that a point belongs to type \(o\) is
\[ p_o = \frac{\lambda_o}{\lambda_\bullet} \quad (1.13) \]

where \( \lambda_\bullet \) is the intensity of the unmarked point pattern and \( \lambda_o \) is the intensity of the specific type or category of the qualitative mark [8].

For an inhomogeneous mark point pattern, the probability of types is also spatially varying. With the intensity \( \lambda_o(u) \), the conditional probability, given that there is a point at \( u \) and is of type \( o \), is

\[ p(o|u) = \frac{\lambda_o(u)}{\lambda_\bullet(u)} \quad (1.14) \]

### 1.1.5 Second order characteristics

When describing the properties of a point pattern, different kinds of summary characteristics have been propose. The most versatile are the summary characteristics which are based on a function of distance that describe the point pattern. The most outstanding are the \( k \)-function and the pair correlation function-pcf; they study the occurrence of points for a range of distance, by considering the average number of points found within an area (\( k \)-function) or at a distance \( r \) from a point in the pattern (pair correlation function) [10].

This summary characteristics are often call second order characteristics, and are related to counting pairs of points or adding up contributions from each pair of point in the process [40] and [8]. In a point process \( X \), the \( k \)-function could be defined as the expected number of \( r \)-neighbours of a point of \( X \), divided by the intensity \( \lambda \) (1.15) [8]

\[ k(r) = \frac{1}{\lambda} \mathbb{E}[\text{number of } r - \text{neighbours of } u|X \text{ has a point at location } u] \quad (1.15) \]

An alternative expression for (1.15), that does not involve conditional expectation is (1.16), for any bounded region \( B \) with area \( |B| > 0 \).

\[ k(r) = \frac{\mathbb{E}[\sum_{x \in X \cap B} t(x, r, X)]}{\lambda \mathbb{E}[n(X \cap B)]} \quad (1.16) \]

where \( t(x, r, X) = \sum_{j=1}^{n(x)} 1\{0 < ||x - x_j|| \leq r\} \) is the number of \( r \)-neighbours (points between a distance of 0 and \( r \)) of the location \( x \). So \( \lambda k(r) \) would be the expected number of \( r \)-close pairs of points in which the first point falls in \( B \), divided by the expected number of points falling in \( B \) [8].
The $k$-function presents a simple form in the Complete spatial randomness - CSR case $k(r) = \pi r^2$, for a two dimensional space.

Based on this, the shape of $k(r)$ compare to the Poisson process provides valuable information on the point process distribution. If $k(r) > \pi r^2$ we could say that the points are part of a cluster pattern and in result each point has some very near neighbours. While, if $k(r) < \pi r^2$ the points are isolated, and in result there exist a certain distance between each point and his nearest neighbours.

An estimator for the $k$-function for a point pattern with window $W$, without edge correction would be the uncorrected empirical function (1.17)

$$\hat{k}_{un}(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} 1\{d_{ij} \leq r\}$$ (1.17)

However this estimator is severely biased, so an edge correction method is needed. There are several methods to implement in the estimation of the $k$-function, the particular choice is usually not critical as long as some kind of technique is performed. Nevertheless there are some recommendations when selecting the edge correction method.

Depending on the size of the dataset, there are some options of edge correction that are more suitable than others. For small data sets with 100 or less points Horvitz-Thompson type of weighted edge corrections is better because of their greater statistical performance, some of these methods are: translation, isotropic or rigid motion correction [8].

The border method is a too imprecise tool in small datasets because it discard substantial amounts of data, but it performs satisfactorily on moderate large data sets containing 1000 to 10000 points. For very large data sets there is no need for edge corrections and it is computational efficient to avoid this procedure [8].

The counterpart for (1.17) with an edge correction method, would be the empirical $k$-function (1.18), where $e_{ij}(r)$ is the edge correction weights (see section 1.1.2).

$$\hat{k}(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} 1\{d_{ij} \leq r\} e_{ij}(r)$$ (1.18)

One of the most common edge correction methods use in second order characteristics is the Ripley’s isotropic correction. This method regard the edge effect as a form of sampling bias.

To understand this method, consider a pair of points $x_i, x_j$ and the distance between them $d_{ij} = ||x_i - x_j||$, so the second point must be somewhere on the circle $b(x_i, d_{ij})$ of radius $d_{ij}$ center at $x_i$. The probability that $x_j$ falls inside $W$ is the fraction of length $l$ of the circle lying inside $W$ [63].

$$p(x_i, d_{ij}) = \frac{l(W \cap b(x_i, d_{ij}))}{2\pi d_{ij}}$$ (1.19)

So, the edge correction weights $e_{ij}(r)$ in (1.18) would be the reciprocal of the probability $p(x_i, d_{ij})$ for each pair of points $x_i, x_j$ [8].
If the point pattern is spatially inhomogeneous, there are tools that can take into account this inhomogeneity. The counterpart of the $k$ function is the inhomogeneous $k$ function. Here the strategy is to weight each point $x_i$ by $w_i = \frac{1}{\lambda(x_i)}$, the reciprocal of the intensity at $x_i$, and each pair of points $x_i, x_j$ will be weighted by $w_{ij} = w_i w_j = \frac{1}{\lambda(x_i)\lambda(x_j)}$. With this idea, the inhomogeneous $k$ function can be defined as

$$ k_{inhom}(r) = E \left[ \sum_{x_j \in X} \frac{1}{\lambda(x_j)} 1 \{ 0 < ||u - x_j|| \leq r \} \mid u \in X \right] $$

(1.20)

So, the inhomogeneous $k$ is the expected total weight of all points within a distance $r$ of the point $u$, where the weight of a point $x_i$ is $1/\lambda(x_i)$. If the process is stationary, then it will mean that $\lambda(u)$ is constant and the inhomogeneous $k$ reduces to the $k$ function [8].

The estimator (1.18) can be extended to the inhomogeneous $k$ function, as

$$ \hat{k}_{inhom}(r) = \frac{1}{D^p |W|} \sum_i \sum_{j \neq i} 1\{d_{ij} \leq r\} \frac{1}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)} e_{ij}(r) $$

(1.21)

where $\hat{\lambda}(u)$ is an estimate of the intensity $\lambda(u)$ and the constant $D^p$ is the $p$th power of

$$ D = \frac{1}{|W|} \sum_i \frac{1}{\hat{\lambda}(x_i)} $$

(1.22)

which has expected value 1 if the intensity is estimated without error. If the intensity function is known or estimated with high precision, then $\hat{k}_{inhom}(r)$ is an unbiased estimator of $k_{inhom}(r)$ when $p = 0$, and is approximately unbiased when $p = 1$ or 2. The intensity function usually is estimated from the data, so the estimator could be biased.

Exactly as with the $k$ function the inhomogeneous $k$ function can be compared to an inhomogeneous Poisson process with intensity $\lambda(u)$

$$ k_{inhom}(r) = \pi r^2 $$

(1.23)

**Pair correlation function $g(r)$**

While the $k$-function contains contributions for all interpoint distances less than or equal to $r$, the pair correlation function $g(r)$ contains contributions only from interpoint distances equal to $r$. This can be defined by (1.24).

$$ g(r) = \frac{k'(r)}{2\pi r} $$

(1.24)

where $k'(r)$ is the derivative of the $k$-function with respect to $r$ [8]. The $g(r)$ contains the same statistical information as the $k$- function but it presents the information in an easier way to understand. The relationship between this tools is similar of that of a probability density function $f(x)$ to its corresponding distribution function $F(x)$, where $f(x)$ is the derivative of $F(x)$ and in this case $g(r)$ is proportional to the derivative of $k(r)$. 
For the Poisson process or CSR case the pair correlation function is equal to one, reflecting that the location of any point is independent of the location of any other point. When the behaviour correspond to a cluster process, $g(r) \geq 1$ and it can take large values especially when $r$ is small and decreases as $r$ increases. When the behaviour correspond to a regular process, $g(r) \leq 1$ for small $r$.

One of the most used estimators for the pair correlation function is based on kernel smoothing. If we take the edge correction estimator of the $K$-function (1.11) and replace the indicator $1\{d_{ij} \leq r\}$ by a kernel term $\kappa(d_{ij} - r)$ to obtain a smooth estimate of $K'(r)$, then the estimator for $g(r)$ is (1.25)

$$\hat{g}(r) = \frac{|W|}{2\pi r n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \kappa_h(d_{ij} - r) e_{ij}(r)$$

where $d_{ij}$ is interpoint distance between the $i$th and $j$th points $||x_i - x_j||$, $e_{ij}(r)$ is an edge correction weight, and $\kappa_h$ is the smoothing kernel with smoothing bandwidth $h > 0$.

As for the $k$ function, pcf also has its counterpart when working with a point pattern inhomogeneous. The inhomogeneous pair correlation function can be defined as

$$g_{inhom}(r) = \frac{k_{inhom}'(r)}{2\pi r}$$

**Tools for anisotropy based on second order functions**

The sector $k$ function is based on modifications done over $k$-function to measure anisotropy. The changes are made over the geometrical shape used to count the points that fall into a certain distance $r$. The generic $k$-function uses a circle of radius $r$, instead the sector $k$-function uses a part of the disc of radius $r$ that lies between two lines with fixed orientations angles $\alpha$ and $\beta$. The theoretical value for a Poisson process, is given by (1.27) (with $\alpha$ and $\beta$ in radians)

$$k_{sector.pois}(r) = (\beta - \alpha)r^2/2$$

To test for anisotropy, different orientation angles are used and the resulting sector $k$-functions are compared. Anisotropy is suggested if the functions appeared to be unequal.

An other way to measured anisotropy is by the pair orientation function. Consider two fixed distance $r_1$ and $r_2$ with $r_1 < r_2$ and all the pairs of points that lie pass $r_1$ and before $r_2$. For these pairs of points, the direction of the line that joint them is measured as an angle in degrees anticlockwise from the $x$-axis. The pair orientation distribution is the probability distribution of these angles.

The pair orientation distribution can be defined as:

$$O_{r_1, r_2}(\psi) = \frac{E[\text{number of points in } O(u, r_1, r_2, \psi) | X \text{ has a point at location } u]}{E[\text{number of points in } O(u, r_1, r_2, 360) | X \text{ has a point at location } u]}$$
where \( O(u, r_1, r_2, \psi) \) is the bounded region with the radius \( r_1 \) and \( r_2 \) center at \( u \) and with lines at orientation 0 and \( \psi \) through \( u \). These can be estimated just like the \( k \)-function\[8\].

**Mark point processes**

When working with a mark point process \( M \), the second order characteristics aim to characterise not only the variability of the point distribution but also the variability of the marks. This second order characteristics also help to describe the correlation among marks and points of the process \[39\].

To explain the most useful characteristics used in the analysis of mark point patterns, the type of mark has to be identify. There are different tools used for qualitative marks as for quantitative marks. First, we introduce to those characteristics used in processes with qualitative marks or as some authors called multivariate or multitype processes, then we show the tools for quantitative marks \[39\][21][8].

**Qualitative marks**

The analysis of the spatial correlation in processes with qualitative marks, focuses on the aggregation and repulsion between the different subprocesses \[39\].

There are two main strategies when analysing a qualitative mark point pattern. One is the multitype viewpoint where \( X = (x_1, ..., x_n) \) denotes the locations of the points, and \( M = (m_1, ..., m_n) \) represents the marks or types of these points. For this approach there are three possible schemes to analyse the point pattern \[8\]:

- The joint distribution \( P(X, M) \), this scheme focus on regarding the locations and types as been generated at the same time.
- The conditional distribution \( P(M|X) \), this scheme focus on regarding the location of \( X \) as fixed and the study focus only on the types \( M \) attached to these locations.
- The marginal and conditional distribution \( P(X)P(M|X) \), this scheme focus on regarding the location \( X \) as having been generated first and then annotated with types \( M \).

The other strategy is the multivariate viewpoint. If we have only two types of points, \( o \) and \( p \), the point pattern is \( Y = (O, P) \) where \( O \) and \( P \) area the pattern of the two types of points. For this viewpoint there are also three schemes to analyse the point pattern:

- The joint distribution \( P(O, P) \), for this scheme the data is regard as a pair of spatial point patterns.
- The conditional distribution \( P(P|O) \), this scheme regards the location of type \( o \) as fixed and study the points of type \( p \).
- The marginal and conditional distribution \( P(O)P(P|A) \), this scheme regards the data as it was built by first determining the pattern \( O \) and then the pattern \( P \) depending on the pattern of \( O \).

The choice of analysis depends on the scientific context and questions or objectives that the study wants to resolve.
One of the most important second order characteristics that can be extended to a multivariate or multitype point process is the \( k \)-function. This can be generalised in a straightforward way, the multitype \( k \) function \( k_{op} \) is defined as the mean number of points of type \( p \) in a disc or radius \( r \) centred at the points of type \( o \), and divided by the intensity \( \lambda_p \) of the points of type \( p \)

\[
k_{op} = \frac{E_{0o}(n_p(b(0,r))))}{\lambda_p}
\]

where \( E_{0o} \) denotes the conditional mean given that the typical point, at location 0 is of type \( o \) \[39\].

Another useful tool is the condensed \( k \) functions, express by (1.30) and (1.31). The first one can be interpreted as the mean number of points (irrespective of their marks) in a disc of radius \( r \) centered at a point of type \( o \), divided by the intensity of the point process \( \lambda \). The second one is similar, is the mean number of points of type \( p \) in a disc of radius \( r \) centered at a point (irrespective of its mark), divided by the intensity \( \lambda_p \) of the type of points \( p \).

\[
k_o = \frac{E_{0o}(n(b(0,r) \setminus \{0\}))}{\lambda}
\]

\[
k_p = \frac{E_{0o}(n_p(b(0,r) \setminus \{0\}))}{\lambda_p}
\]

Both the multitype and condensed \( k \) functions, have a similar interpretation as the \( k \) function. In these cases the regular or aggregation pattern is between the points of two categories of the mark (multitype \( k \) function), or between the points of a category and the rest of the points in the point pattern (condensed \( k \) function).

This tools based on the multivariate \( k \)-function, are useful in the detection of correlation between the marks of the point pattern. However, the cross-pair correlation functions \( g_{op}(r) \) and condensed pair correlation function are more suitable in the detection of this correlations in an exploratory analysis.

The cross-pair correlation functions \( g_{op}(r) \), considers the co-occurrence of individuals of different types of points \[40\]. In a similar way as the pair correlation function it can be defined as the derivative of the \( k_{op} \)-function.

\[
g_{op}(r) = \frac{k'_{op}(r)}{2\pi r}
\]

The interpretation of the cross-pair correlation function is similar as the pair correlation function. The tool allows to explore if there is repulsion \( (g_{op}(r) < 1) \) or attraction \( (g_{op}(r) > 1) \) between two types of marks in the point pattern.

Just like the \( k \)-function, the pcf also can be used as a condensed function. This function can also be defined using the standard relationship

\[
g_o(r) = \frac{k'_o}{2\pi r}
\]
The interpretation is the same but in this case, the dependence is between the points of a category of the mark with the rest of the points of the point pattern.

The latter functions describe for the analysis of the sub patterns of a qualitative marks, also have an inhomogeneous version. The inhomogeneous analogues of the functions \( k_{op} \) and \( k_{o\cdot} \) are obtained by weighting each point by the reciprocal of the intensity function, and by weighting the contribution from each pair of points by the product of these weights. The estimator for this functions are:

\[
\hat{k}_{inhom}^{op}(r) = \frac{1}{|W|} \sum_{x_i \in X^{(o)}} \sum_{x_j \in X^{(p)}} \frac{e(x_i, x_j)1\{d_{ij} \leq r\}}{\hat{\lambda}_o(x_i)\hat{\lambda}_p(x_j)}
\]

\[
(1.34)
\]

\[
\hat{k}_{inhom}^{o\cdot}(r) = \frac{1}{|W|} \sum_{x_i \in X^{(o)}} \sum_{x_j \in X^{(\cdot)}} \frac{e(x_i, x_j)1\{d_{ij} \leq r\}}{\hat{\lambda}_o(x_i)\hat{\lambda}_\cdot(x_j)}
\]

\[
(1.35)
\]

The cross pair correlation function and the condensed pair correlation function can be defined using the standard relationship \( g(r) = \frac{k'_{op}}{2\pi r} \).

**Quantitative marks**

When working with point patterns with quantitative marks, the analysis is focused on the numerical difference among the marks depending on the distance of the corresponding points. Based on this type of analysis the numerical similarity of the marks of neighbouring points can be studied [39].

There are different second order characteristics that have been adapted to the analysis of point pattern with quantitative marks. In this section the mark correlation function, the multiplicatively weighted pair correlation function and the mark weighted \( k \)-function are described.

The mark correlation function \( k_t(r) \) is a quantity that measures the dependence between the marks of two points of the point process at a distance \( r \). This summary characteristics are constructed based on a different test functions \( t(m_1, m_2) \), which depends on the two marks \( m_1 \) and \( m_2 \), where \( m_1 \) is at location 0 (\( m_1 = m(0) \)) and \( m_2 \) is at distance \( r \) from location 0 (\( m_2 = m(r) \)) [39] [8].

For instance, consider the conditional mean of the product of the marks of a pair points in the point process (here the condition is that there is a point in 0 and in \( r \)) [1.36].

\[
C_{mm}(r) = E_{0r}(m(0) \cdot m(r)) \text{ for } r > 0.
\]

(1.36)

The test function used in [1.36] is \( t(m_1, m_2) = m_1 \cdot m_2 \). Different test functions can be used to construct different summary characteristics, some of the most used are:

\[
t_1(m_1, m_2) = m_1 \cdot m_2
\]

\[
t_2(m_1, m_2) = m_1
\]

\[
t_3(m_1, m_2) = m_2
\]

\[
t_4(m_1, m_2) = \frac{1}{2}(m_1 - m_2)^2
\]

\[
t_5(m_1, m_2) = (m_1 - \mu)(m_2 - \mu)
\]
The general expression for the summary characteristics $C_t(r)$ based on any test function $t(m_1,m_2)$ is presented in (1.37). This summary functions are called non-normalised mark correlation functions [39].

$$C_t(r) = E_0(t(m(0)m(r))) \text{ for } r > 0. \quad (1.37)$$

To facilitate the interpretation of this summary characteristics, it is very useful to normalise the function $C_t(r)$ by dividing it by $C_t(\infty) = C_t$, which is the value the function takes at very long distance, where the marks are independent. For the test functions described above, this are the normalizing factors [39]:

- $t_1 : \mu^2$
- $t_2 : \mu$
- $t_3 : \mu$
- $t_4$ and $t_5 : \sigma^2_{\mu}$
- $t_6 : \frac{\pi}{4}$ for $d = 2$

Some important normalize mark correlation functions are constructed by the test functions $t_1(m_1,m_2)$, $t_2(m_1,m_2)$ and $t_4(m_1,m_2)$ and their corresponding normalizing factors. These functions are: the mark correlation function $k_{mm}(r)$, the r-mark correlation function $k_{m\cdot}(r)$, and the mark variogram $\gamma(r)$.

If the marks in the point process are independent, then $k_{mm} = k_m \equiv 1$ for $r \geq 0$ and $\gamma(r) \equiv \sigma^2_{\mu}$. When there is some kind of dependence among the marks, this tools can help identify it. If $k_{mm} < 1$ and $k_m < 1$ for small $r$, shows a dependence in the marks called inhibition, this is when the points compete and have smaller than average marks when they are close together. When $k_{mm} > 1$ and $k_m > 1$ for small $r$, this shows mutual stimulation, which is consider when the points benefit from being close together and thus have on average larger marks than $\mu$ [39].

The multiplicatively weighted pair correlation function $g_{mm}(r)$ can be interpreted as the second order characteristic of the mark sum intensity $\lambda_S$. This summary function describes the spatial distribution of the mark mass instead of the point distribution [39].

The latter second order characteristics functions gather information from pairs of points lying exactly at a distance $r$. The mark weighted K-functions $k_{mm}(r)$ is a cumulative function that is preferred for hypothesis testing purpose.

The construction and interpretation of $k_{mm}(r)$ is analogous to $k$ and $k_{ij}$ functions; and the form is similar to the $k$-function but it takes into account the quantitative mark of the point process

$$k_{mm}(r) = \frac{E_d(\sum_{x:m(x)} m(o)m(x)1_{b(o,r)}(x))}{\lambda \mu^2} \quad (1.38)$$
so the \( k_{mm}(r) \) is the mean of the sum of the products formed by the mark of a point and the mark of all points in the disc or radius \( r \), centered at that point and divided by \( \lambda \mu^2 \). The division is done to normalised the \( k_{mm}(r) \) \[39\].

### 1.1.6 Indices

Indices are numerical summary characteristics that describe specific aspects of the distribution of a point process. One important characteristic of this indices is that they are easy to determine and easy to understand, making them a very important tool in describing point patterns in ecology and forestry where mapped data is sometimes difficult to collect \[39\].

There are two main types of indices: location-related and point-related indices. The first type are determined by deterministic test locations or sampling points, which are selected independently of the points of the point process. The latter type is relate to points process and yields information on the typical point \[39\].

In this section we focus on a few point-related indices. Formally this types of indices are built by taking each point \( x \) and constructing a mark \( m(x) \) for each one of them, then the mean mark of the typical point \( E_o(m(o)) \) is considered. The index could be this value or some quantity containing it \[39\].

Some of this indices are based on the nearest-neighbour distances, like the aggregation or clark-Evans index and the Hopkins-Skellam test.

The Clark-Evans index could be regarded as the average of the nearest neighbour distance \( d_i \) for \( m \) randomly sampled points or all data points in the point pattern, and divided by the expected value \( E[D] \) for a Poisson process with the same intensity \( \lambda \). For a Poisson process of intensity \( \lambda \) the expected distance to the nearest neighbour is given by \( E[D] = \frac{1}{2\sqrt{\lambda}} \). So the Clark Evans index it could be estimated by (1.39).

\[
R = \frac{d}{E[D]} = \frac{2\sqrt{\lambda}}{m} \sum_{i=1}^{m} m_i
\]

(1.39)

Values greater than one for the Clark-Evans index \( R > 1 \) indicate that the pattern tend to regularity, while \( R < 1 \) suggest clustering.

The Hopkins-Skellam test is used to as a simple summary of the spatial pattern. The test is built on the nearest-neighbour distance of all points in the point pattern, compare to the nearest neighbour distance of a completely random pattern with the same number of points \( \lambda \).

\[
A = \frac{\sum_{i=1}^{m} P_i^2}{\sum_{i=1}^{m} l_i^2}
\]

(1.40)

where \( P_i \) is the nearest neighbour distance of the point \( i \) in the point pattern and \( l_i \) is the nearest neighbour distance simulated in a completely random pattern. The interpretation is similar to the Clark-Evans index, for values \( A > 1 \) indicates spatial regularity, for values \( A < 1 \) indicates spatial clustering and \( A = 1 \) complete spatial randomness \[37\].
Other numerical information can be used to describe specific aspects of the distribution of the point process. The mean or median distance to the nearest neighbour can be used in point process with a qualitative marks, to classify the spatial behaviour of each sub-process. Using the pair correlation function, the minimum distance where the value of the function is 1 is also helpful to explain the spatial behaviour of the point process.

### 1.2 Functional data

To implement a functional principal components analysis, first some key term in functional data analysis have to be review. In this sections we present the most important concepts for functional data analysis, and then focus on the analysis of functional principal components.

In a functional data analysis, the main idea is that the observations are functions and these are interpreted as single entities instead of a consecutive measurements [38].

The functional observation $z(t)$ is build on $n$ pairs $(t_p, y_p)$, where $y_p$ is an observation of $z(t)$ at time or location $t_p$. Usually the observations of the function are taken at a finite number of times $p = 1 \cdots P$, so smoothing techniques have to be implemented [38].

The function is built in a two stages process: first a set of functional building blocks called basis functions $\phi_q$ have to be define; then a vector, matrix or array of coefficients has to be set up to define the function as a linear combination of the basis functions [59].

So the function $z(t)$ is defined by (1.41) which is called the basis function expansion. Here the parameters $c_1, c_2, ..., c_Q$ are the coefficients of the expansion; and in (1.41) $c$ stands for the vector of $Q$ coefficients and $\phi$ denote a vector of length $Q$ containing the basis functions.

$$ z(t) = \sum_{q=1}^{Q} c_q \phi_q(t) = c' \phi(t) \quad (1.41) $$

When a sample of $N$ functions are consider then $z_l(t) = \sum_{q=1}^{Q} c_{ql} \phi_q(t)$ with $l = 1, ..., N$ and the matrix notation becomes $z(t) = C' \phi(t)$, where $z(t)$ is the vector of length $N$ containing the functions $z_l(t)$, and the coefficients are store in the matrix $C$ of $N$ rows by $Q$ columns [59].

#### 1.2.1 Functions systems

Usually the data we want to model falls into two categories: periodic and non-periodic. There are different types of basis systems and depends on the situation, but commonly the Fourier basis system is used for periodic functions and the spline basis system tends to be the choice for non-periodic functions [59].

The Fourier series is used when we need the function to repeat itself over a certain period $T$, this happens often when expressing seasonal trends in a long time series. The basis functions of the Fourier series are:

$$ \phi_1(t) = 1 $$
$$ \phi_2(t) = \sin(wt) $$
\[ \phi_3(t) = \cos(wt) \]
\[ \phi_4(t) = \sin(2wt) \]
\[ \phi_5(t) = \cos(2wt) \]

... where the constant \( w = \frac{2\pi}{T} \). So to define the Fourier basis system only the number of basis functions \( Q \) and the period \( T \) is required \[59\].

On the other hand the splines bases for non-periodic data, are more flexible and thus more complicated than the finite Fourier series. There are many different kinds of splines, all of them are define by the validity, the knots and the order \[59\].

To construct a spline, the interval of observation has to be divided into subintervals, with boundaries at points called break points or breaks. Over each subinterval, the spline function is a polynomial with fixed degree or order, and from one subinterval to an other the nature of the polynomial could change \[59\].

The degree refers to the highest power in the polynomial, while the order of the polynomial is one higher than its degree. For example, a order four splines consist of cubic polynomial segments with degree three \[59\].

A spline basis is defined also in terms of a set of knots. This knots are place in every break point, but there may be multiple knots at certain break points. The number of knots positioned at an specific break point, determined the number of derivatives that must match between the neighbouring polynomials at that break point \[59\].

In the majority of applications, only a single knot is place at every break point, except for the boundary values at each end of the range of \( t \). The purpose to assign as many knots as the order of the spline to the end points, is to drop to zero the function value when is outside the interval over which the function was defined \[59\].

Within this framework, there are different basis systems for constructing spline functions, some of them are: B-spline, M-splines, I-splines and truncated power functions. The most commonly used is the B-spline basis system. In this system each basis function aside from the two end basis functions, they begin at zero, then at a certain knot location goes up to a peak before falling back to zero and remaining there until it reaches the right boundary. The two end basis functions rise from the first and last interior knot to a value of one on the right and left boundary, respectively \[59\].

1.2.2 Coefficients and Smoothing

Once the basis system is selected, the next step to built the functional data is to supply the coefficients. The coefficients are stored in a vector, matrix or array depending on: the number of basis functions, the number of functions or functional observations and in the number of dimensions of the functions.

If only one function is defined we used a vector of length \( Q \), where \( Q \) is the number of basis functions. If we are working with a sample of functional observations of size \( N \), we used a matrix of \( Q \) rows by \( N \) columns. And if functions themselves are multivariate of
To compute those coefficients with the basis function system and define the curves or functions, there are two main strategies that have more careful consideration of measurement error. One is the use of regression analysis, where we used the least squares estimation process over function \( z(t) \) defined as a basis function expansion. The second aims to miss nothing of importance of the data by using a powerful basis expansion and avoids over-fitting by using a penalty on the roughness of the function [59].

We focus on the latter strategy that can be called data smoothing with roughness penalties. This approach uses a large number of basis functions, up to one basis function per observation and even beyond; but also applies smoothness by penalizing some measure of the function complexity [59].

The first step to choose a roughness penalty is to define a measure of roughness of the fitted curve. A popular way to quantify the roughness of a function is by using the squared of the second derivative \([D^2 z(t)]^2\) of a function \(z\) at argument value \(t\), which is often called its curvature at \(t\). Keeping in mind that the second derivative of a straight line (which obviously has no curvature) is zero, a measure of a function’s roughness could be given by the integrated squared second derivative or total curvature (1.42) [59].

\[
PEN_2(z) = \int_t [D^2 z(t)]^2 dt \tag{1.42}
\]

Wherever the function is highly variable, the square of the second derivative \([D^2 z(t)]^2\) it would be large, so this penalty term is a good way to provide smoothing.

Once the measure of roughness of the fitted curve is defined, the final step is to minimize a fitting criterion that trades off curve roughness against lack of data fit. To this purpose we add some multiple of the penalty term to the error sum of squares to define the compound fitting criterion. Using \(PEN_2(z)\), the fitting criterion is given by

\[
F(c) = \sum_p [y_p - z(t_p)]^2 + \lambda \int_t [D^2 z(t)]^2 dt \tag{1.43}
\]

where \(z(t) = c'\phi(t)\) is the basis function expansion, and \(\lambda\) is the smoothing parameter.

The smoothing parameter \(\lambda\) penalized the curvature in the second term of (1.43), this relative to the goodness of fit quantified in the sum of squared residuals in the first part of (1.43). If \(\lambda\) is sufficiently large, then \(D^2 z\) will be essentially zero, implying that \(z\) will be essentially a straight line. If \(\lambda\) tends to zero, then the function \(z\) is free to fit the data as closely as possible with the selected basis system; sometimes resulting in wild variations in the approximating function [59].

### 1.2.3 Functional principal components analysis

A principal component analysis - PCA is an approach to explore the variability in a multivariate data. This technique uses an eigenvalue decomposition of the variance matrix of the data to find directions in the observation space in which the data has the highest variability. The resulting analysis yields a loading vector or weights vector for each principal
component, this loading gives the direction of variability corresponding to that component \([60]\).

In functional principal components analysis - FPCA, instead of working with an eigenvector as in multivariate statistics, there is an eigenfunction associated with each eigenvalue. Another difference is that in PCA, the variables in a multivariate observation can vary greatly in location and scale due to the different choices of origin and unit of measurement of the variables, so PCA is usually based on the correlation matrix. In FPCA the values of \(z_l(s)\) and \(z_l(t)\) have the same origin and scale, so the covariance function and the cross-product function tend to be more useful than the correlation function \([59]\). Equations (1.44), (1.45) and (1.46) show the estimators for these functions, respectively.

\[
cov(s, t) = (N - 1)^{-1} \sum_{i=1}^{N} [z_l(s) - \bar{z}(s)][z_l(t) - \bar{z}(t)] \quad (1.44)
\]

\[
cross(s, t) = (N)^{-1} \sum_{i} z_i(s)z_i(t) \quad (1.45)
\]

\[
corr(s, t) = \frac{cov(s, t)}{\sqrt{cov(s, s)cov(t, t)}} \quad (1.46)
\]

One of the main reasons to use FPCA is to search for a score \(\rho_{\xi}\). A score is a tool that reveals the most important type of variation in the data. Consider function values \(z(t)\) and define \(\rho_{\xi}(z_l) = \int \xi(t)z_l(t)dt\), where \(\xi(t)\) is a weight function. If \(\xi\) has been structured so as to be a template for a pattern of variation of \(z\), then the resulting score value \(\rho_{\xi}(z)\) will be substantially far from zero \([59]\).

With the score define the next question to ask is, for what weight function \(\xi\) would the scores have the largest possible variation? To answer this question we have to impose a size restriction on \(\xi\), \(\int \xi(t)dt = 1\). Under this constraint the next step is to maximise \(N^{-1}\sum_{i} \rho_{\xi}^2(z_l)\) and get an eigenequation (1.47) \([38]\).

\[
\int \text{cov}(s, t)\xi_j(t)dt = \mu_j\xi_j(s) \quad (1.47)
\]

where \(\mu = \max_{\xi}\{\sum_{i} \rho_{\xi}^2(z_l)\}\). In terminology \(\mu\) and \(\xi\) are referred to as the largest eigenvalue and eigenfunction respectively, of the estimated variance-covariance function \(v\). The solution to this eigenequation with the largest eigenvalue and eigenfunction \(\xi_1\) solves the maximization problem and yields the first principal component. The second largest eigenvalue with eigenfunction \(\xi_2\) yields the second component and so on \([38]\).

1.3 Ecological Concepts

The study of forest dynamics focus on the changes in forest structure and composition, as well as the response of the forest to anthropogenic and natural disturbance \([57]\). The forest structure is a concept based on a complex system that consider a variety of forest attributes like: aboveground biomass, abundance, basal area, canopy height, plan density \([17]\).
Due to the importance of forest structure in understanding forest dynamics, modern ecological researches are focus on the spatial structure of the trees [57]. There are a variety of biological processes that affect the forest spatial structure: mortality, seed dispersal, seed establishment, regeneration, growth, facilitation and competition [15].

In the present study the focus is on understanding the facilitation and competition processes in a tropical dry forest. Here facilitation is the process by which an organism profits from the presence of another and competition is the process where both organisms compete for resources.

There are different levels of organization in a forest: individuals, populations, communities and landscape. The spatial structure can vary according to this levels [15]. So defining the level of organization can help understand better the ecological processes under study. Processes such as facilitation and competition are at a population level [15], so our analysis will be based on this level.

Working at this level of organization, the species composition which refers to all the organisms that make up the forest, it is very important to discover how the forest works. In this context the species composition is base on all the species of trees found in the TDF.

So, the species composition will help to analyse the facilitation and competition processes from a interspecific (between members of different species) and intraspecific (between members of the same species) viewpoint [56].

The approach of using point processes methods in analysing spatial patterns observed in nature specifically of trees in a forest, have been increasing over the last decade; due to the accessibility of software and an improved dialogue between statisticians and ecologist, interested in their respective fields. However, spatial point processes are still not considered in the statistical methods used by ecologist, this is evidenced in the limited number of ecological publications using this approaches [10].

Using this approach in the analysis of the spatial structure of the TDF, not only will allow to understand the ecological processes of facilitation and competition, but also will grant the possibility to observed the problems in implementing point process methods in ecological analysis.
CHAPTER 2

Methods

This chapter presents the methodology propose to analysed the spatial structure of the tropical dry forest inside the one hectare plot. The study site and the variables used are presented in section 2.1. While on section 2.2 the methodology scheme is presented, using a flow diagram here we describe the three steps that form the methodology: 1) the analysis of the intensities (first order tools), 2) the analysis of interactions (second order tools) and 3) the analysis of the classification methods used to group the species based on their spatial behaviour.

2.1 Study site and variables

The study area is located in the National Natural Park El Tuparro, in the Colombian Orinoquia. The information of the point pattern was collected from a one hectare plot, establish by the Institute of Research of Biological Resources Alexander von Humboldt in a tropical dry forest. The figure 2.1 shows the location of the national park El Tuparro in Colombia, and the spatial distribution of the 1274 plants recorded in the one hectare plot.

The mean temperature is around 25°C and is located at an altitude of 90 meter above sea level. The annual precipitation is around 1559 mm, and it is distributed in a unimodal bistational seasons: a wet season of 7 months from April to October and a dry season from November to March [72].

The soil conditions are similar in all the study area, with high percentage of sand (over 60%), and a moderate acid soil with values of pH around 4.4 (base on soil analysis done in the plot). An important characteristic of the tropical dry forest in this region of Colombia is the presence of rocky soils; this special characteristic creates a limitation on the availability of soil resources, so usually the plants are not located over the rocks, this can be seen in figure 2.2.

The location of each tree in the plot was captured by the following procedure: first a random point was selected and georeferenced (geography coordinates) inside the tropical dry forest, from this point the plot was divided into 100 subplots of 10 by 10 meters, and the location of each tree was register with a measuring tape inside the subplot.
Figure 2.1. Distribution of trees in a 1 ha plot in the National Natural Park El Tuparro. Upper left: Location of the National Natural Park El Tuparro in Colombia, bottom left: Location of the plot in the National Natural Park El Tuparro (red point), left: Distribution of trees inside the plot.

Figure 2.2. Distribution of the plants (green circles) and rocks (gray polygons) inside the study area.

For each tree the following information was captured: taxonomical information (family, genus and species), plants habit (tree, palm, cacti, shrub or vine), diameter at breast height - dbh in centimetres, estimate of height in meters and a estimate of the crown diameter in meters.
Inside the plot a great number of species were found, a total of 90; which are within 39 families and 77 genus. This is a typical structure of a tropical forest, where there is a large number of species coexisting. However the number of trees per species is low, only 32 species had more than 10 plants inside the plot. Figure 2.3 shows the number of plants for the 32 species with more plants.

**Figure 2.3.** Number of plants in each of the 32 species with more than ten plants inside the study area
The distribution of the values of diameter at breast height (dbh) of the trees are presented in figure 2.4. Here we see that 75% of the trees have a diameter below 13.3 cm, and 50% are under 5.7 cm. There are only 118 trees with values greater than 27.7 cm the upper limit of the box plot, most of them around 30 to 50 cm of dbh, the biggest tree found in the plot is a *Ficus americana* with a diameter of 101.6 cm.

The values of dbh can be used to group the plants into size categories, this can be done based on information of the stage of development of the trees of a tropical dry forest. [18] consider three main stages: a sapling stage which considers young plants from germination to 30 cm of height, a pole stage which considers plants with a dbh between 15 and 30 cm, and a timber stage which considers those plants with more than 30 cm of dbh.

To help evaluated the spatial structure of the forest, we group the plants inside the study area into different size categories. Using the values of dbh and the information of the pole and timber stage, three categories were created: Small, Medium and Large. The category for each plant was assigned as follows:

- Small: Plants with $dbh < 15$ cm
- Medium: Plants with $dbh$ between 15 and 30 cm
- Large: Plants with $dbh > 30$ cm

The number of trees in each of the three categories are presented in figure 2.4. This graph shows that 77% of the trees are in the class Small with dbh under 15 cm, 16% in class Medium with dbh between 15 and 30 cm, and only 6% are in class Large with dbh over 30 cm. From this values we can see that most of the trees are in the sapling stage; indicating that the forest must be in a young stage of a secondary succession.

![Figure 2.4. Left: Box plot of the values of diameter at breast height dbh, Right: Number of plants in each of the categories of the mark size](image)

The distribution of the dbh for the six species with more trees inside the plot is shown in the box-plot on the figure 2.5 (for the rest of the species see appendix A).

Most of the plants of this six species are under 20 cm of dbh. For the species *Bactris bidentula* and *Matayba sp* almost 100% of the points have a dbh less than 10 cm.

The species *Attalea microcarpa* and *Protium guianense*, are the ones with more plants in the large category, with 23 and 11 points respectively; and yet the box-plot for the species *Protium guianense* shows that most of the plants have less than 20 cm of dbh.
Figure 2.5. Box plot of the value of diameter at breast height dbh for the six species with more plants.

An other important aspect to complement the spatial structure of the tropical dry forest, is the habit of the plant. Although this information is not used as a mark of the point pattern, it was very helpful in the analysis done in the present study. The figure 2.6 shows the percentage in each type of habit found inside the plot.

Figure 2.6. Pie chart with the percentage of each class of habit.

The most abundant type of habit is trees which represents a 75.4% of the points of the pattern, follow by the category palms which is represented only in two species with a 15.5%: *Attalea microcarpa* and *Bactris bidentula*. The rest is of the points are vines (5.4%), shrub (3.1%) and cacti (0.39%).

This general information about the marks and the habit of the points in the pattern, shows us that this is secondary successional tropical dry forest at a young stage. Most of the plants have low dbh and 15% of the points are palms which can be consider as heliophyte, plants that grows best in direct sunlight and can be associated with early and intermediate stages of the secondary succession of the forest.
2.2 Statistical methodology

To study the structure of the tropical dry forest and understand the biological processes of competition and facilitation that have an effect in the resulting point pattern of this type of forest, a methodology based on different tools of point process theory was propose.

The methodology scheme is presented in figure 2.7, which is composed of three main parts. Each have an specific objective which all add ups to understanding the spatial structure of the trees of the tropical dry forest.

The following sections explains each part of the methodology scheme, starting by identifying the point pattern and the marks used, followed by the first and second order characteristics and ending with the classification methods. In each section we specify the tools used and the objective in the analysis of the spatial structure.

To complement the analysis, additional information of the species was sought. This information was useful in the analysis of some of the results obtained in the methodology used. All the information collected on field was digitized using the Geographic information system ArcGis. The raw data of the variables of each plant was process and registered in a spread sheet in the software excel, and the ecological information of the species was also registered in a spread sheets. The statistical analysis were done in the software R using mainly the library ”spatstat". 
Figure 2.7. Propose methodology scheme for the analysis of the point pattern generated by the plants of a TDF.
2.2.1 Point Pattern and Marks

In forest ecology there are different processes that affect the spatial arrangement of the plants, the idea is to draw conclusions about this ecological processes by analysing the resulting point pattern.

The point pattern analyse in the present study is under a two dimensional space in $R^2$, the events are the plants $x_i$ for which two coordinates gives the location inside our observation window a one hectare plot located at the National Natural Park El Tuparro.

The point pattern could be assumed that it comes from a stationary point process, this is based on the following non statistical arguments: the window was defined in an arbitrary way inside a tropical dry forest, so is surrounded by a larger forest area; the trees close to the edge of the window have the same behaviour as those in the center; the environmental conditions are the same inside and outside the window; and the soil conditions do not vary greatly inside the observational window.

The isotropic assumption of the point pattern, is based on the sector $K$-function and in the pair orientation distribution. For the sector $K$-function three directions were used, the angles are in degrees and are measured anticlockwise:

- Horizontal: Begin angle $\alpha = -15^\circ$ end angle $\beta = 15^\circ$
- Diagonal: Begin angle $30^\circ$ end angle $60^\circ$
- Vertical: Begin angle $75^\circ$ end angle $105^\circ$

In the top panel of figure 2.8 each of the three directions used for the sector $K$-function are divided by the theoretical values and superimposed. From the graph it can be seen that the three sector $K$-functions are similar, suggesting that there is not anisotropy in the point pattern.

We used the pair orientation function to estimate the probability density of the angles with pairs of points between distance $r_1 = 0.1$ and $r_2 = 6$ meters. This information was plotted on the bottom panel of figure 2.8 which shows a rose diagram of the estimated probability density for the point pair orientation distribution. This diagram indicates that at distances between 0.1 to 6 meters, the plants do not have any particular preference in the direction where they are located.

2.2.1.1 Marks

The variables chosen as marks in the present study are the species and the diameter at breast height - dbh. The name of the specie was used as a qualitative mark and the dbh as a quantitative mark. Also the dbh was used to created size categories that were implemented as an other qualitative mark.

Species

The qualitative mark species contains 90 categories represented by the species found in the one hectare plot. For each one, ecological information was search and stored in a database. This information was used to analyse the different results obtained from the methods used.
Figure 2.8. Top: sector K-functions, Bottom: Rose diagram of estimated probability density for the point pair orientation distribution of the point pattern.

For some functional and point process tools not all the species could be used; some of them had few numbers of plants making impossible to implement the methods used in the study. Most of the analysis used the 32 species which had more than 10 plants inside the plot.

In other specific analysis we focus only on the ten species with more plants and in the species *Pachira nukakica* which is consider one of the most important endemic species of the tropical dry forest of the Colombian Orinoquia [56].

**Diameter at breast height - dbh**

The diameter at breast height was used as a quantitative mark, representing the size or area occupied by the tree, which is an important characteristic for evaluating the competition and facilitation processes inside the TDF. This variable was chosen because of the precise measurement (with metric tape) unlike the height and crown diameter that were taken by observational estimates of the person who captured the information.

Some of the plants develop more than one stem, so they have more than one value of dbh. To treat this, two options were propose: the first one, adds all the dbh of the stems;
the second one, turns the dbh into basal area $g$ with a standard formula $g = \pi \times (\text{dbh}/2)^2$, adds all the basal area of the stems and finally using the same formula a global dbh is determined for the plant.

Using a basal area formula to determined the global value of dbh for the plant, is a strategy that gives a more precise value to represents the area occupied by the plant. The values obtain with the first option compared to those of the second option, are greater. Based on this, the present study uses the latter option.

**Size Category**

To evaluated the spatial structure of the forest, an important aspect to analyse is the size of the trees. The size give an idea of the stage of development of the forest and could also be an effect of competition and facilitation processes.

We used the three size categories (small, medium and large) explain previously, to created the qualitative mark size. This mark is based on the values of dbh and the information of the stages of development of the plants, reported for a tropical dry forest.

### 2.2.2 First order characteristics

The first step in the methodology propose, highlighted in green in figure 2.7, focused on estimating the intensities for the unmarked point pattern and the marked point pattern using the qualitative marks species and size. The objective of this first analysis was to have a first insight in the spatial distribution of the trees. Here we focus not only on understanding the spatial structure of all the trees but also on the structure for each species and size, and the relationship that this categories have with the rocky soils of this type of forest. This first analysis was also the basis for the tools used in the following steps of the methodology propose.

First assuming that the point process had homogeneous intensity, the generic intensity was estimated using equation (1.2) for the unmarked point pattern and equation (1.3) for the categories of the two qualitative marks(species and size). To complement this analysis the sum intensity (equation (1.4)) was estimated using the quantitative mark dbh.

The intensity was also consider as a intensity function $\lambda(u)$. Using equation (1.9), we implemented a kernel estimation as a baseline for the intensity of the unmarked and qualitative marked point patterns. The probability density used in the kernel was an isotropic Gaussian (Normal distribution) and the Diggle edge correction was applied.

To control the degree of smoothing in the kernel estimation of the intensity function, the band width was constantly modified by different algorithms that minimize a measure of error. Three algorithms were used: Diggle and Berman’s method, the likelihood cross validation method (ppl) and Scott’s rule of thumbs.

For the selection of the final bandwidth value for the unmarked point pattern and the sub-patterns of the qualitative marks species and size, we propose an analysis base on: literature information about ecological interaction distances, the size of the plants represented in the dbh, and the habit of the species.

The last step in this first part was to incorporated the influence of the rocks inside the plot to the intensities of the point pattern. For this we compute the distance from each point $u$ inside the plot to the nearest rock. This distance was used as a continuous covariate.
The final intensity for the unmarked point pattern and for the sub-patterns of the qualitative marks species and size, was estimated using a parametric model. The model propose uses an interaction between the values of a kernel estimated intensity which is use as a baseline and an exponential function that captures the influence of the covariate distance to the rocks.

2.2.3 Second order characteristics

The study uses some of the most important tools of functional summary characteristics. This tools were used to measure dependence between points in the point pattern, and study numerical similarities of the quantitative mark between neighbouring points. Depending on the type of mark (qualitative or quantitative) specific methods were implemented.

In figure 2.7 this section of the analysis is presented in blue. Here we used the intensities estimated in the previous part of the methodology scheme.

Unmarked point pattern

First the inhomogeneous k-function (see equation (1.21)) and inhomogeneous pair correlation function (see equations (1.25) and (1.26)) were estimated using the intensity estimated for the unmarked point pattern in the previous section.

Qualitative mark Species

This section is based on pairwise interactions, because of this we focus the analysis only on the dominant species and the endemic species Pachira nukakica. To define the dominant species inside the plot, we used a methodology based on the intensities estimated in the first phase of the methodology, here we selected for every space $u$ the max probability value between the species using equation (1.14); those species with more area with the highest probability were selected.

The objective of this section was to analysed the interactions within and between the species, to identify behaviours of aggregation or repulsion that could be a reflection of the competition and facilitation processes.

The tools used to analyse the point pattern with the qualitative mark species are based on the multitype viewpoint. Here we focus on regarding the locations and types (species) as been generated at the same time.

We used the inhomogeneous $k$-function and the inhomogeneous pair correlation function for each of the sub-pattern to analyse the interaction within the species; and the inhomogeneous condensed $k$-function and inhomogeneous pair correlation function to evaluate the interaction of a species with the rest of the species (see equations (1.34) and (1.35)).

For each of this tools the intensities define in the previous section were used, and the Ripley’s isotropic edge correction was implemented.

Qualitative mark Size

For the qualitative mark size the analysis also focused on the interactions within and between the categories of the mark. We used a multitype viewpoint, where it is assume that the locations have been generated first and then the types are annotated for each point depending on their location.
The tools used are the same as those used for the qualitative mark species, and have the same objective of analysis. This information also was used as a way of testing for random labelling of the categories of the mark.

Quantitative mark dbh

To complement the analysis done for the qualitative mark size, the quantitative mark dbh was evaluated with different tools. The objective was to study the numerical difference and similarities among the mark depending on the distance of neighbouring points.

Because the diameter at breast height could be seen as a numerical value of the dominance of the tree inside the forest; this analysis was helpful in understanding how the competition and facilitation process express themselves in term of the diameter.

The tools used were based on the mark correlation function, here we used two test functions: the r-mark correlation function and the mark variogram (see Quantitative marks in section 1.1.5).

2.2.4 Classification methods

In this section the primary interest was to classify the species by their spatial behaviour rather than considering the inter-specific interactions. Here we worked with the 32 sub-patterns of the species that had more than 10 plants inside plot.

For this purpose three classification methods were applied and compared. The first one is base on applying a hierarchical cluster using Ward’s minimum variance method to a set of ecological indices that where estimated for each of the 32 species.

The second and third approach uses the estimated inhomogeneous $k$-function and pair correlation function of each species. From this data we implemented a functional principal component analysis one for the set of 32 inhomogeneous $k$-functions and one for the 32 inhomogeneous pcf (see section 1.2.3). With the scores of this two FPC we used a hierarchical cluster using Ward’s minimum variance method to group the 32 species.

The resulting aggregation from the three approaches, was analysed and compared. The database with the ecological information was the support to validate the resulting classifications. This part of the scheme is presented in orange in the figure 2.7.
In this chapter we present the results obtained from the analysis of the mark point pattern. First we present the results and different analysis of the intensities of the unmarked and marked point pattern. The Second section shows the analysis based on the second order characteristics used on the qualitative and quantitative marks. And in the final section we present the results for the different classification methods and a comparison between them.

3.1 First order characteristics

This section focus on the analysis of the intensity of the unmarked point pattern and the different categories of the two qualitative marks: species and size. We first show the results assuming homogeneous intensities. Then we present the results of the parametric model propose to estimate the intensities, this model is composed of a baseline given by a kernel and an exponential part which includes the dependence of the point pattern with the rocks of the study area.

3.1.1 Generic intensity

The unmarked point pattern had an intensity $\hat{\lambda}$ of 0.127 points in a squared meter; this could be interpreted as if in average one plant needed a space of 7.8 squared meters to survived. This value is similar to the intensity reported for a secondary lowland rainforest in New Guinea $\lambda = 0.120$ [25].

Having in mind that 75% of the trees have a diameter smaller than 15 cm, the intensity value give an insight of a strong competition process between the trees inside the plot. However this is a general value that does not acknowledge the rocky soil, the type of species and the size of each tree, which are important information in ecological processes like competition and facilitation.

The rocks are an important aspect in analysing the spatial structure of the point pattern under study. The areas with rocks are a space with limited soil resources, so usually the plants are not located in this spots. If the rocks tend to repel the establishment
of plants, this would reduce the space available and have an effect on the intensity of the point pattern.

The species composition refers to the contribution of each species to the overall vegetation of the forest. This is important when trying to discover how the forest works and how important are the different species to the spatial structure in the forest.

The intensities \( \hat{\lambda}_i \) of each species is a way to analyse this composition. This values are shown for the 90 species in the left side of figure 3.1, while on the right side a box-plot with the values of the proportions of the number of trees of each species in reference to the total of trees in the plot, is presented.

As it is common in a tropical forest, a high richness with 90 different species is found in the study area. Nevertheless, the species evenness or the equal distribution of the number of plants in each species is not that high. As it can be seen in the right side of figure 3.1, 75% of the species are represented with less than 2% of the plants inside the study area; this is reflected in low values of intensities on many of the species.

From this values of intensity and proportion, it could be concluded that inside the plot, inter-specific interaction are more common and intra-specific interaction will be less common. However, this are general values that does not let us see the spatial distribution of each species. For example, if one specie have a low intensity value the first thing in mind is that this species is surrounded by other species and will have high inter-specific interactions, but if the low number of trees of this species presents a cluster pattern the interactions of the species will be mostly intra-specific. Even do the intensities values give us an idea of the species composition, the analysis of inter and intra specific interactions have to be based on other point pattern tools.

![Figure 3.1](image)

**Figure 3.1.** Left: box-plot of the values of intensity of each species, Right: box-plot of the values of proportion of each species

The species composition are not only based on the number of species and its density, a more precise description can be obtained when other attributes such as biomass or an area cover variable is determine, in this case we used the diameter at breast height of each tree. The sum intensity \( \hat{\lambda}_S \) of this quantitative mark is 1.33 cm of dbh per squared meter;
so in average we can find one or several trees in a squared meter, that their dbh sum is 1.3 cm.

This value gives more insight to the structure of the tropical dry forest under study. For the mean average value of dbh which is 10.4 cm, the area needed for a tree with this diameter has to be of 7.8 squared meters. This complements the analysis mentioned previously with the intensity $\hat{\lambda}$, where it show that a plant needed this space to survived.

A plant with a diameter of 10.4 cm is inside the category small of the qualitative mark size, which has an intensity $\hat{\lambda}_i$ of 0.0984. The other two categories had an intensity of 0.0204 and 0.0086 for medium and large respectively.

This latter intensities, shows us that the structure of the forest inside the area of study is composed mostly by small plants with a low number of plants inside the other two categories. If in average the plants inside the category small needed a space of 7.8 squared meters to survived this could suggest that the structure of the forest could be strongly influenced by a competition process due to the scarce among of resource.

### 3.1.2 spatially varying intensity

The latter values of intensity are assuming that the point pattern is homogeneous, now we focus on intensities that are a spatially varying value. This is done by a parametric model which is compose of two parts: a kernel estimation which will be the base of the intensity, and an exponential part which incorporates the effect of the covariate distance to the rocks. For the mark species we only use the 32 species that presented at least 10 plants inside the study area.

Because the kernel estimation uses a bandwidth $h$, our first analysis targeted the different values obtain for the bandwidth of the sub-patterns of each species, size categories, and the unmarked point pattern. Table 3.1 presents the bandwidth values found by the three methods used: Diggle and Berman’s method, likelihood cross validation method (ppl), Scott and Stoyan rules of thumbs; the final column of the table shows the method used in the kernel estimation.

The selection of the final bandwidth for the unmarked point pattern was based on distance reported for the interactions between trees inside a tropical forest. [25] reported that biotic interactions, operate mostly in radius of 10 to 20 meters inside a tropical rain forest.

We did not take exactly this values because this was reported to a rain forest and not to a dry tropical forest; instead we used the scott method which reported bandwidth values a little bit lower.

For the bandwidth of the sub-patterns size, we based our decision on the fact that bigger trees need more space than smaller trees, so the bandwidth of a large tree should be greater.

The values of the bandwidth for the categories of the mark species, were selected based on three factors: the habit of the species, the distribution of dbh for the species, and the spatial distribution of the plants of the species. When the habit of the species was a tree, palm, shrub and vine, we assumed that the bandwidth should be bigger in that order; with the dbh we follow the same argument as the one used for the mark size; and the
<table>
<thead>
<tr>
<th>Point Pattern</th>
<th>Category</th>
<th>diggle</th>
<th>ppl</th>
<th>scott W-E</th>
<th>scott N-S</th>
<th>Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>unmark</td>
<td>-</td>
<td>3.8</td>
<td>3.31</td>
<td>8.69</td>
<td>9.05</td>
<td>scott</td>
</tr>
<tr>
<td>mark size</td>
<td>Large</td>
<td>8.45</td>
<td>41.93</td>
<td>13.72</td>
<td>14.85</td>
<td>scott</td>
</tr>
<tr>
<td>mark size</td>
<td>Medium</td>
<td>7.95</td>
<td>21.86</td>
<td>12.72</td>
<td>11.74</td>
<td>diggle</td>
</tr>
<tr>
<td>mark size</td>
<td>Small</td>
<td>3.81</td>
<td>3.75</td>
<td>8.91</td>
<td>9.45</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Arrabidaea sp.</td>
<td>6.45</td>
<td>21.45</td>
<td>17.35</td>
<td>20.18</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Astronium g.</td>
<td>2.47</td>
<td>8.93</td>
<td>15.92</td>
<td>16.34</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Attalea m.</td>
<td>7.12</td>
<td>14.67</td>
<td>12.66</td>
<td>16.85</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Bactris b.</td>
<td>6.82</td>
<td>6.76</td>
<td>10.62</td>
<td>12.37</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Bursera s.</td>
<td>7.82</td>
<td>15.55</td>
<td>23.3</td>
<td>15.73</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Clathrotropis m.</td>
<td>6.52</td>
<td>16.64</td>
<td>10.11</td>
<td>10.69</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Erythroxylum m.</td>
<td>8.09</td>
<td>48.47</td>
<td>19.95</td>
<td>19.27</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Eschweilera t.</td>
<td>8.52</td>
<td>16.48</td>
<td>12.46</td>
<td>12.79</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Ficus a.</td>
<td>2.41</td>
<td>6.02</td>
<td>17.6</td>
<td>10.4</td>
<td>ppl</td>
</tr>
<tr>
<td>mark species</td>
<td>Fridericia p.</td>
<td>6.58</td>
<td>11.45</td>
<td>16.16</td>
<td>13.1</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Guarea g.</td>
<td>5.81</td>
<td>13.85</td>
<td>18.41</td>
<td>16.38</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Gustavia a.</td>
<td>7.65</td>
<td>30.56</td>
<td>16.05</td>
<td>14.49</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Heisteria a.</td>
<td>8.42</td>
<td>38.66</td>
<td>16.15</td>
<td>17.93</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Himatanthus a.</td>
<td>8.35</td>
<td>23.92</td>
<td>16.83</td>
<td>12.09</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Inga g.</td>
<td>4.31</td>
<td>23.66</td>
<td>14.53</td>
<td>13.14</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Inga sp.</td>
<td>8.49</td>
<td>39.5</td>
<td>14.91</td>
<td>20.86</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Licania m.</td>
<td>5.58</td>
<td>13.92</td>
<td>10.84</td>
<td>14.3</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Matayba sp.</td>
<td>8.49</td>
<td>10.38</td>
<td>13.7</td>
<td>19.78</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Morf sp2</td>
<td>4.71</td>
<td>12.65</td>
<td>22.07</td>
<td>20.11</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Morf sp4</td>
<td>8.45</td>
<td>17.93</td>
<td>16.83</td>
<td>16.08</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Ocotea s.</td>
<td>7.42</td>
<td>36.15</td>
<td>19.03</td>
<td>17.61</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Pachira n.</td>
<td>8.35</td>
<td>38.79</td>
<td>16.8</td>
<td>21.4</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Petrea sp.</td>
<td>8.09</td>
<td>8.13</td>
<td>13.71</td>
<td>13.98</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Phanera g.</td>
<td>7.75</td>
<td>48.17</td>
<td>18.28</td>
<td>17.19</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Pouteria p.</td>
<td>5.91</td>
<td>46.23</td>
<td>15.8</td>
<td>19.63</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Pouteria sp4</td>
<td>3.58</td>
<td>17.79</td>
<td>14.59</td>
<td>12.54</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Protium g.</td>
<td>8.42</td>
<td>25.68</td>
<td>14.47</td>
<td>11.48</td>
<td>scott</td>
</tr>
<tr>
<td>mark species</td>
<td>Pseudolmedia sp1</td>
<td>6.18</td>
<td>41.75</td>
<td>22.37</td>
<td>17.09</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Rudgea c.</td>
<td>8.42</td>
<td>24.89</td>
<td>17.35</td>
<td>12.59</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Sapium g.</td>
<td>8.49</td>
<td>38.99</td>
<td>15.16</td>
<td>18.28</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Senegalia m.</td>
<td>6.72</td>
<td>45.3</td>
<td>21.38</td>
<td>16.33</td>
<td>diggle</td>
</tr>
<tr>
<td>mark species</td>
<td>Siparuma g.</td>
<td>5.11</td>
<td>11.25</td>
<td>18.67</td>
<td>17.52</td>
<td>diggle</td>
</tr>
</tbody>
</table>

**Table 3.1.** Values of the three methods used to estimated the bandwidth for the unmarked point pattern and the sub patterns of the qualitative marks species and size, and selection of the method use.
spatial distribution helped us to determined visually an aggregation or repulsion of the points.

The values selected were a good first insight on the spatial behaviour of the species and the intra-specific interactions. For most of them the diggle algorithm was chosen indicating a more cluster or positively correlated behaviour. While for the seven species that have the scott method, the behaviour could be associated with a more regular or negatively intra-specific behaviour.

Now we present the intensity for the unmarked point pattern and the sub-patterns of the categories of the mark size and species. The model used to estimate this intensities is presented in equation (3.1), here the baseline \( b(u) \) is the kernel estimated using the selected bandwidths and \( Z(u) \) is the covariate distance to the rocks.

\[
\lambda(u) = b(u) \exp^{(\alpha + \beta Z(u))}
\]  

(3.1)

The table 3.2 shows the value of \( \alpha \) and \( \beta \), while the figures 3.2 to 3.11 presents the kernel estimation, the intensities constructed by the model, and an overlap of the intensities of the model with the locations of the rocks and the plants. All of this is presented for: the unmarked point pattern, the sub-patterns of the three categories of the mark size and the sub-patterns of the species where the parameters of the model were significant and for the species *Pachira nukakica* (see appendix for the coefficients of all species and the confidence interval of all models).

<table>
<thead>
<tr>
<th>Category</th>
<th>P.P.</th>
<th>alpha</th>
<th>beta</th>
<th>sig alpha</th>
<th>sig beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmarked</td>
<td>-0.13</td>
<td>0.16</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Large</td>
<td>-0.08</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>-0.19</td>
<td>0.21</td>
<td>*</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>-0.13</td>
<td>0.15</td>
<td>**</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Attalea m.</td>
<td>-0.04</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bactris b.</td>
<td>-0.38</td>
<td>0.35</td>
<td>**</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Eschweilera t.</td>
<td>-0.25</td>
<td>0.28</td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Gustavia a.</td>
<td>-0.11</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matayba sp</td>
<td>-0.05</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protium g.</td>
<td>-0.11</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pachira n.</td>
<td>0.29</td>
<td>-0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inga g.</td>
<td>-0.28</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morf sp4</td>
<td>-0.42</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2. Values of the parameters of the intensity model for the unmarked point pattern and the for some sub patterns of the qualitative marks species and size.

For the unmarked point pattern the Z-test shows that the two parameters of this model are significant, indicating that the distance to the rocks have an effect on the locations of the plants.

A location where there is a rock (distance = 0) the model will reduce the baseline, multiplying by \( \exp^{-0.13} = 0.87 \). If the distance to the rock is one meter the value that modifies the kernel would increase by a factor of \( \exp^{-0.13 + 0.16} = 1.03 \); the larger value of distance to the rock is 4 meters so this value would increase by a factor of \( \exp^{-0.13 + (4 \times 0.16)} = 1.66 \).
Figure 3.2 shows how the kernel is modify by the covariate distance to the rock. Here we see that the locations where there are rocks, reduces the intensity of the kernel, an important fact is that in this locations the intensity shown by the model is not reduce to zero, which means that plants can establish over the rocks which resembles to reality. On the other hand locations further from the rocks the intensity almost doubles the kernel estimation.
Figure 3.2. For the Unmarked point pattern. Left: kernel, Center: intensity of the parametric model, Right: intensity of the model, rocks and points.

Only for a few sub-pattern of the mark species the parameters of the model were significant. The number of plants in each species could be the reason for this result. The two species with more plants inside the study area *Bactris bidentula* and *Eschweilera tenuifolia* their parameters were significant (table 3.2).

The figures 3.3 and 3.4 shows how the kernel is modify by the influence of the location of the rocks for these species. The distance where the exponential part equals zero, determines where the kernel starts to increase and decrease. From the rocks to a distance of \(0.38/0.35 = 1.08\) meters for *Bactris bidentula* and \(0.25/0.28 = 0.89\) meters for *Eschweilera tenuifolia* the kernel is reduce, from this distance the kernel increases.

The exponential part of the model which determines the value of modification of the kernel, is stronger in these two species compared to the one reported for the unmarked point pattern.
The other two species with the $\beta$ parameter significant were *Inga gracilifolia* and the undetermined species called Morf sp4. The first one had 45 plants and 3 of them were over a rock, the latter one had 29 plants and only 1 was over a rock.

We see that the relationship of this species with the rocks is very strong, especially for Morf sp4. It occurs the same as for the model of the unmarked point pattern, but the values of the exponential part have a greater influence on reducing and incrementing the baseline kernel; this could be seen on the values of the coefficients of the model and in figures 3.5 and 3.6.

For example, with the model of the Morf sp4 if the location is over a rock the exponential value is $\exp -0.42 = 0.65$, less than the reported for the unmarked model (0.87), indicating a stronger reduction over the kernel value. And the same happens when the location gets further apart from the rock, if we take a distance of 4 meters from the rocks the value would increase by a factor of $\exp -0.42 + (4 \times 0.44) = 3.81$, a little more than twice the value reported for the unmarked model (1.66).
CHAPTER 3. RESULTS

Figure 3.4. For the species Eschweilera tenuifolia. Left: kernel, Center: intensity of the parametric model, Right: intensity of the model, rocks and points of the species.

Although the parameter were not significant for many of the species, the values obtain could help explain the relationship of each species with the rocks. For most of them the behaviour of the exponential part is the same, at the locations of the rocks this value reduces the baseline and at locations further away the value increases the baseline.

However for some of the species, the values estimated for the parameters of the model show a preference for the locations with rocks; this happens when the $\alpha$ is positive and $\beta$ negative. This is the case of the species *Pachira nukakica* which has two big trees over rocks, one with a dbh of 29.7 cm and the other with 71.6 cm, which indicates that this species have the capacity to establish in these locations.

In figure [3.7] we see that the locations with rocks have incremented the value of the kernel by a factor of $exp^{0.29} = 1.3$ for the species *Pachira nukakica*. Although this model is not significant, the new intensities are taking into account the rocks of the study area, while the original kernel does not.
This analysis are based on parameters that are not significant, they only give us an idea of this relationships. There are different ways to improve this intensities models, one way could be using the covariate as a factor, where one category indicates the presence of a rock and the other the absence.

An other way to improve this models of intensity for the species is by using the dbh. The figure 3.8 shows the distribution of the dbh of the plants that are over a rock for each of the species. We see that a very low number of plants have diameters over 20 cm, mainly the species *Pachira nukakica*, this could indicate that plants can establish over the rocks but because of the lack of soil resource the probability of survival for most of the species is very low as the plants growth.
For the sub-pattern of the mark size, the Medium and Small categories the parameters were significant, but the category Large did not report the same. This could be attributed to the small number of plants inside the large category, an effect that we also saw on the models of the species.

The figure 3.9 shows that the intensities for the category Large are the same in almost all locations of the study area but with low values; this is due to the small number of plants in this category and the large values of the bandwidth selected.

The kernel estimation is a good representation of reality, where big trees have large area of influence inside a forest; because this is a secondary succeesional forest this big trees probably had an effect on the establishment of new plants and in the development of younger trees. The modification of the kernel with the covariate distance to the rocks give a more realistic view of the area of influence of the larger trees inside the plot.
CHAPTER 3. RESULTS

Figure 3.7. For the species Pachira nukakica. Left: kernel, Center: intensity of the parametric model, Right: intensity of the model, rocks and points of the species.

The models for the other two categories Medium and Small show that the kernel is reduce at close locations from the rocks, and it starts to increase from a distance of 0.9 and 0.8 meters from the rocks for the Medium and Small categories respectively.

This behaviour is stronger in the Medium model where the estimated values for $\alpha$ and $\beta$ are greater. This could be attributed to the fact that Medium trees are in a stage where they need more resources to develop, so if they are close to rocks where the availability of soil resources are low the chances of survival will be minimum.

The kernel and its modification with the covariate distance to the rock for the Medium category are shown in the figure 3.10.

The values obtain for $\alpha$ and $\beta$ for the model of the sub-pattern of the category Small are very similar to those estimated for the unmarked point pattern. Figure 3.2 with the intensity for the unmarked point pattern and figure 3.11 with the intensity for the sub-pattern Small are very similar. However, if we take a closer look in the latter one the zones of influence of the small trees are very well highlighted.
The model of the intensity for the Small plants, is also very realistic. It shows how the intensity is reduce close to the rocks but is not low, indicating that in this stage of development the plants do not need a great amount of the soil resource to establish. Nevertheless, the probability of survival when they past to the next stage of growth will be reduce and this would generated the point pattern that we see in the Medium category.

The intensities estimated by this models where used in posterior analysis, even if some of the models were not significant we choose to used this spatially varying values instead of the original kernel, because they take into account the influence of the rocks which is an important factor to this type of tropical dry forest.
Figure 3.9. For the category Large of the size mark. Left: kernel, Center: intensity of the parametric model, Right: intensity of the model, rocks and points of the category.
Figure 3.10. For the category Medium of the size mark. Left: kernel, Center: intensity of the parametric model, Right:intensity of the model, rocks and points of the category.
Figure 3.11. For the category Small of the size mark. Left: kernel, Center: intensity of the parametric model, Right: intensity of the model, rocks and points of the category.
CHAPTER 3. RESULTS

3.2 Second order characteristics

The results of this section measure the dependence between the points of the unmarked point pattern and the pattern with each of the marks. First we present the results for the unmarked point pattern, follow by the results of the pattern with the qualitative mark species and finally the results based on the dbh represented in the qualitative mark size and the quantitative mark dbh.

3.2.1 Unmarked point pattern

Using the intensity estimated for the unmarked point pattern, we implemented an inhomogeneous k-function and an inhomogeneous pair correlation function, both with an isotropic edge correction. The results of these functions are shown in figure 3.12 here we use 99 simulated realisations of an inhomogeneous Poisson process with the same number of point to compare with the observed functions and test if the pattern follows an inhomogeneous Poisson process.

In the top of the figure the inhomogeneous $k$-function with the theoretical function of CSR and the maximum and minimum of the simulated inhomogeneous Poisson process (envelope) are presented. In the first $r = 5.6$ meters the observed function is slightly over the theoretical function indicating a cluster pattern, from this distance the observed function starts to go under the CSR.

The difference between the observed and the theoretical functions are not that high. The observed function is mostly under the CSR values; this shows that the unmarked point pattern under study presents an inhibition pattern.

The results are similar with the inhomogeneous pair correlation function, but in this case we can see that in the first 4.2 meters a strong cluster pattern is shown. From 5 to around 12 meters the curve shows an inhibition pattern, and the rest of the curve is inside the envelopes indicating that the points follow a CSR pattern.

3.2.2 Species

Here the interest is on the dependence between the plants of the same species (intra-specific dependence) and between a species with the rest of the species (inter-specific dependence).

Species selected

Due to the large number of species found inside the study area, we focus our analysis on the most likely species found inside the plot. This selection was based on the species with the highest probability at each location of the study area.

Figure 3.13 presents for the 32 species with more plants the locations where that species has the highest probability inside the plot. The figure is organized in four maps, starting from the top left to bottom right we present the species with more to less plants.

We see that most of the area is cover with the 8 species with more plants (top left of the figure); the species Guarea glabra and Astronium graveolens also cover a significant part of the study area (top right of the figure). As the number of plants decreases the areas where that species had the highest probability is reduced.
FIGURE 3.12. Left: inhomogeneous $k$ function, Right: inhomogeneous pair correlation function. For the unmarked point pattern
Figure 3.13. Probabilities for 32 species
However, the highest probability at a location that’s not only depend on the number of plants. This probabilities are based on the intensity values defined in the previous section, so the spatial distribution of the plants and the bandwidth of the kernel have an influence on the values of the probabilities of each species.

For example, the species Petrea sp on the top left of the figure we see that in none of the locations have the highest probability, while for the species called morf sp2 in the bottom left of the figure there are some locations where this species has the highest probability. These two species have 37 and 14 plants respectively, in these case the abundance is not too important but the spatial distribution of the plants is crucial. Petrea sp have more plants but they are more disperse inside the study area, while morf sp2 is less abundant but in the north east of the plot, there is an aggregation, giving this species the highest probability in that location.

This methodology of the highest probabilities based on the intensities was a good approach to define the dominant species inside the plot. Usually in ecology the dominance is examined using different variables like plant density, cover and biomass [34]. Here the intensities are estimated not only by the plant density but it also takes into account the location of each plant and by using the size of the plants to select the bandwidths of the kernel we acknowledge in some way the biomass of the species.

By this method we determined not only the dominant species but also the areas inside the plot where that species is most dominant. Based on this analysis we decided to focus on 11 species instead of the 90 found inside the plot. The species selected where: Bactris bidentula, Eschweilera tenuifolia, Protium guianense, Gustavia augusta, Attlealea microcarpa, Matayba sp, Inga gracifolia, Licania micrantha, Guarea glabra, Astronium graveolens and Pachira nukakica. The latter one is selected because of the ecological importance to this specific tropical dry forest.

In the 11 species selected two of them are palms Bactris bidentula and Attlealea microcarpa, the rest have a habit of tree. An other important characteristic is the light requirement of this plants, eight of them are heliophytes which are plants that adapt to a habitat with a very intensive insolation; while Licania micrantha and Guarea glabra are classified as shade enduring sciophyte i.e plants that are relatively tolerant to high light saturation.

The figure [3.14] shows the spatial distribution and the size of the plants for each of the 11 species selected. This information is very useful to analyse the intra and inter specific dependence of the species.
**Intra specific Dependence**

To evaluated the intra-specific dependence of the 11 species selected, the inhomogeneous $k$- function and inhomogeneous pair correlation function were used. The figure [3.15] shows the results obtained: on the top of the figure we presented the inhomogeneous $k$- function and on the bottom the inhomogeneous pair correlation function for each of the species. The results for the 11 species are separated in two graph for a better visualization.

Figure 3.14. Spatial distribution with the value of diameter at breast heigh dbh for each of the 11 species selected
Figure 3.15. inhomogeneous $k$ function and inhomogeneous pair correlation function for the 11 species selected
To complement the analysis we used the information: number of plants, type of habit, light requirements and the dbh distribution of each of the 11 species.

For all of the species, both functions show a regular pattern after the first 5 meters of distance; and some of the species present a very weak cluster pattern at close distances.

In both functions the species with the stronger regular pattern are: *Astronium g.*, *Inga g.*, *Licania m.*, *Guarea g.*, follow by *Pachira n.* and *Matayba sp.*

All of the species with a strong regular pattern have a habit of tree; we also see a strong relationship of the functions used with the number of plants. Species with less number of plants, their $k$-function tend to be further apart from the CSR curve; which is logical, with a low number of plants is more difficult to find a point of the same species as the distance increases.

It is also important to highlight that the only two sciophytes species of the eleven selected, *Licania m.* and *Guarea g.* show a very strong regular pattern. This species are conditioned to areas where there is a lowered light intensity, this restriction reduces the locations where they can establish.

The values of dbh of *Guarea g.* are low, almost all of the trees of this species have a dbh below 10 cm (see appendix). If we look at the $k$ and pair correlation functions for this species, we see that this is one of the species with a cluster pattern at very close distances. This information allows us to think that in the first stages of growth many trees stabilish in those small areas where the soil and light meet the requirements for the species to growth.

On the other hand the dbh of the trees of *Licania m.* are bigger compared to those of *Guarea g.*. Most of them are below 20 cm, but there are a few large trees with dbh over 40 cm. Unlike *Guarea g.*, the $k$ and pair correlation functions does not show a cluster pattern at close distances.

If we look at the spatial distribution of the trees with their value of dbh of the species *Licania m.* (see 3.14), we see that the large and medium trees in fact show a regular pattern; however the small trees with dbh around 10 cm show an aggregation on the center of the study area. This is not shown in the $k$ and pair correlation functions.

Probably these two species have similar behaviour regarding their spatial distribution, but because of its difference in the dbh of their trees, the cluster pattern at close distance is not captured by the second order functions of the species *Licania m.*

By this, we see that the spatial distribution does not only depend on the species but also on the stage of growth in which each plant is.

The $k$ and pair correlation functions of the species *Astronium g.*, where the ones further apart of the CSR values. Indicating a strong regular pattern, however as we mentioned before the number of trees is directly reflected in the second order functions. Here we are not sure if this is the spatial behaviour of this species or if the results are due to the low number of trees.

An other species with strong regular pattern and low number of trees is *Pachira n.*, however the trees of this species have greater dbh and by looking at figure 3.14 the regular pattern is more obvious compared to the one of the species *Astronium g.*.

We can also see that in the second order functions of this two species, the ones for *Pachira n.* give more information than those for *Astronium g.*, e.i. the pair correlation function of *Pachira n.* shows at distances of 10 to 15 meters peaks that indicate the
distance to the first neighbour, while the function for *Astronium g.* is almost a straight line which is difficult to interpret.

The other two species mentioned as with a strong regular pattern are: *Licania m.* and *Matayba sp.*. This two species have almost the same number of trees inside the study area and are both heliophyte, but the dbh of the trees of the species *Licania m.* are bigger.

If we compared the second order functions of this two species we see that the ones for *Licania m.* are further apart from the CSR values; and other difference is that the functions for *Matayba sp.* show a weak cluster pattern at close distances.

In figure 3.14 we see that for the species *Licania m.* the bigger the dbh of the tree the more disperse is from the other trees of the same dbh; while for *Matayba sp.* most of the trees have low dbh and their spatial distribution is more aggregated.

The other species presented functions closer to the CSR values, this species are: *Bactris b.*, *Eschweilera t.*, *Protium g.*, *Gustavia a.* and *Attalea microcarpa*.

These five species are the most dominant within the study area; they also contain the largest number of plants; all of them are heliophyte; and three have a habit of tree *Eschweilera t.*, *Protium g.*, *Gustavia a.*, and the other two are palms.

The most dominant species is *Bactris b.*, this palms have very small dbh and is spatially distributed in all of the study area. The smaller palms, with dbh lower than 5 cm tend to form small clusters; we can see that the $k$ and the pair correlation functions captured this behaviour, suggesting cluster diameters of 2.8 and 1.3 meters respectively.

The other palm *Attalea microcarpa*, presents dbh above 20 cm and up to 40 cm, indicating that they are in a mature state. In figure 3.14 we can also see that these palms are also spatially distributed in all of the study area, with the exception of the north eastern zone where we can find the largest amount of rocks.

The pair correlation function for this species shows a hard core process with minimum inter-point distance of 0.26 meters, probably caused by the dbh of the palms. At distance of 7 and 13 meters we see two peaks in the curve produced by the first and second neighbours.

The other three species with weaker regular patterns, have a similar distribution of dbh values, it is noted that *Eschweilera t.* and *Protium g.* have some trees with high value reaching 40 cm and 50 cm of dbh respectively.

The spatial distribution of these three species shown in figure 3.14 displays aggregations of points of the same color, which indicates small clusters of trees of the same size. This give an insight on the intra specific dependence, where small trees are not too close to bigger trees of the same species.

Both of the second order functions show that these three species have a cluster pattern at close distances. In the $k$ function the cluster distance are 2.0 4.6 and 1.8, and in the pcf are 1.0 2.4 and 0.6 for the species *Eschweilera t.*, *Protium g.* and *Gustavia a.*.

These values indicate the cluster size found in each species. *Protium g.* is the one that presents the larger cluster size of all of the 11 species. In the pcf of this species we see that the distance between the clusters could be around 7, 15 and 20 meters where we see the highest peaks of the curve.

As for *Eschweilera t.* the cluster size is smaller and based on the pcf the distance between the first and second neighbours of clusters are around 8 and 12 meters.
The initial cluster pattern of the species *Gustavia a.* is strong but with a small cluster diameter; the pcf presents a high peak at around 5 meters probably caused by the first neighbour cluster.

In general this second order functions, led us to conclude that these species of the tropical dry forest tend to have regular patterns probably due to the lack of resources and in some cases the light conditions requirements. However we also saw that some of the species created small clusters, in some cases of plants of the same size in other cases of plants of different sizes.

Although creating new categories that include the size of the plants of a species, will increase the number of categories and reduce the number of points in each category, it could be a good strategy to understand the behaviour of a species in different stages of growth.

**Inter specific Dependence**

In the latter analysis we focus on the intra specific dependence, the relationship between plants of the same species. Now we look at the inter specific dependence, here we do not focus on a one on one dependence between the species but instead we look at the dependence of each of the 11 species with the rest of the plants of the other 89 species found.

The relationship between each of the 11 species selected and the plants of the other species, was analysed using the inhomogeneous condensed $k$ function and the inhomogeneous condensed pair correlation function. The results of both functions for each species is presented in figure /refcondksp. Once again we divided the results of the two tools in two graph for a better visualization.
CHAPTER 3. RESULTS

Figure 3.16. condensed \( k \) function and condensed pair correlation function for the 11 species selected
The condensed $k$ functions for all of the species are mostly below the CSR curve, indicating that around the plants of the 11 species there is not an aggregation of other plants. The condensed pair correlation function shows a similar behaviour, for all of the functions after 1.5 meters of distance the values are below 1.

From this we can deduce that the lack of area with good soil resources is creating a strong competition process, which is reflected in the intra specific dependence (as we saw in last section) and in the inter specific dependence.

Some species tend to repel more the plants of other species. This is the case of *Inga g.*, *Guarea g.*, *Astronium g.* and *Licania m.*. The condensed $k$ functions of these species are far apart from the functions of the other species, and in the condensed pcf the values are below 0.5.

It is necessary to emphasize that the number of plants in these four species is low, however, if we compare their functions with those of *Pachira n.* which is the one with the lowest number of trees of all the 11 species selected, we see that these species are further apart from the CSR values.

The trees of all of these four species have low values of dbh, specially *Guarea g.* and *Astronium g.*; *Licania m.* have some big trees but mostly the trees have values under 20 cm.

The condensed pcf for the species *Astronium g.*, shows values in all distance under 0.1 which indicates that the dependence with other species is of a mutual repulsion. The function does not present any peaks, meaning that there are not many trees from other species around the plants of *Astronium g.*

As we mentioned, most of the trees of this species have very low dbh, this could make us think that the species chooses areas with low density of other plants to stablish their seedlings.

The condensed pcf of *Inga g.* also shows a strong repulsion dependence with the plants of other species; we can also see little peaks, probably representing the small numbers of neighbours of other plants. Unlike *Astronium g.*, the dbh of the trees of this species is more distributed: half is under 10 cm and the other half is between 10 and 30 cm; the other difference is that the trees of this species are more disperse in all the area of study.

The other two with a strong repulsion dependence with the plants of other species are: *Licania m.* and *Guarea g.*. These two species are sciophytes so they need low light exposition at early growth stages. Because of this condition we expected to have a strong aggregation behaviour with other species at closed distances, however the two condensed functions did not show this result.

This can be explain by the area of influence of the crowns of big trees, e.i. in the figure 3.14 we can see that the small trees of *Licania m.* are aggregated in the middle of the study area, if we compared this to the map with the spatial distribution of dbh (figure 3.17), we see that in that same area the biggest tree is found; so the shade provided by a single big tree can generated the light conditions for the plants of these sciophytes species.

The other seven species, show a weaker repulsion dependence with the plants of the other species. Inclusive, they show an aggregation dependence at very close distance under 1.5 meters and two of them at around 5 meters.
If we focus on the condensed pcf of these species, we can see many peaks; having in mind that these species presents the most number of plants in the study area, we can identify relationships between species looking at the peaks of these functions.

In each condensed pcf, the peaks represents aggregation of plants of other species at that distance. Identifying peaks at same distances between different condensed pcf, gives insight to an aggregation dependence between the species and may indicate a facilitation process.

For instance at a very close distance around 1 meter the condensed pcf of \textit{Pachira n.} presents a high peak indicating an aggregation of other species, at this same distance the pcf of \textit{Attalea m.} presents a peak, this could indicate that most of the plants that are at a 1 meter distance from the plant of \textit{Pachira n.} are probably of \textit{Attalea m.}.

We can also see this relationship at a distance of around 3 meters, here the species \textit{Pachira n.} shows an other high peak indicating aggregation of other species, but in this case three species also show a peak in their condensed pcf: \textit{Gustavia a.}, \textit{Attalea m.} and \textit{Matayba sp.}. This functions are showing us a positive dependence between all of these four species.

It seems that \textit{Protium g.} and \textit{Eschweilera t.} also have a positive relationship. This can be seen in their condensed pcf where at very close distances they show a strong aggregation with plants of other species, and at around 4 meters, both functions show a high peak, indicating that the neighbours of \textit{Protium g.} at this distances are probably from \textit{Eschweilera t.}.

At greater distances, is more difficult to analysed the relationship between the species, the fluctuations in the functions probably are caused by a greater mixtures of species.

This tools have helps us generated an idea of the dependence of the species with the rest of the plants. To analysed specific inter dependence between a pair of species, tools like the inter type $k$ function or the cross pair correlation function can be used.

However, because we are using the 11 species with the most number of trees inside the study area, the condensed functions specially the pcf allow us to identify some specific inter dependences between specific species.

These intra and inter specific analysis are complementary, they allow us to understand the different types of dependences between the plants of the species and can help us created an holistic view of the spatial distribution of the species of the tropical dry forest.

### 3.2.3 Diameter at breast height - dbh and size categories

As we saw in the last section the size of the plants have an important role in the spatial distribution of each species. In this section we focus our analysis only on the size of the plants; this analysis is based on the point pattern with the qualitative mark size and the point pattern with the quantitative mark dbh.

The figure 3.17 shows the spatial distribution of the 1274 plants: on the left side of the figure the three categories of size are presented and on the right side the location of the plants based on different ranges of the value dbh are shown.
Figure 3.17. Left: Distribution of the tree in each class of size, Right: Distribution of the trees in different ranges of dbh

Both maps give a representation of the spatial distribution based on the size of the plants, however the map of the categories is an aggregation of the other based on the dbh ranges for the sapling, pole and timber stages reported for the tropical dry forest.

From the maps we see that the category large include too many plants compared to the map on the right where we see that there are only a few very large plants with more than 60 cm of dbh. This aggregation may present difficulties if we are trying to analyse the importance of these few very large trees, but the objective is to understand the relationship of the three categories based on the dbh ranges of the different stages of the tropical dry forest.

In both maps we see that plants with bigger dbh tend to be more apart from each other generating a more regular pattern, while plants with smaller dbh are closer indicating a more cluster pattern. To verify this we applied different tools, first we show the results for the qualitative mark size and then for the quantitative mark dbh.

**Size categories**

To understand the competition and facilitation processes, we analyse the resulting point pattern with the qualitative mark size which was created with the ranges of dbh of the sampling, pole and timber stages reported for the tropical dry forest.

First we present the inhomogeneous $k$ function and the inhomogeneous pair correlation function for each of the three categories, and used it to analyse the behaviour inside each category. Then we show results of a cross pair analysis between the categories to understand the dependence between them.

Figure [3.18] shows the inhomogeneous $k$ function and the inhomogeneous pair correlation function for each of the three categories. The inhomogeneous $k$ function does not give much information, it only indicates that each of the subpatterns presents a regular pattern after 5 meters of distance.

The pair correlation functions are more useful and permits a better analysis of the sub-patterns. Here we see a similar behaviour in all of the three patterns.

First there is a very strong cluster pattern with a diameter of less than 1 meter; then the pair correlation functions are below 1 indicating a more regular pattern; and the range
of correlation seems to be around 25 meters for the large category and around 13 meters for the other two categories.

Part of this behaviour may be attributed to the rocks inside the study area. In the places absent of rocks the plants aggregate forming the first part of the curves, while at greater distance the rocks appear causing a more regular pattern in the trees of the forest.

Although the pair correlation function is similar for the three categories, we can differentiate behaviours between them.

The category Small presents the larger cluster diameter which is at 1 meter. This shows that plants with smaller dbh are less competitive, which allows that a greater number of plants germinate and grow in areas where there are resources available.

After a distance of 1 meter the pair correlation function shows some kind of repulsion generating a weak regular pattern. The peaks of the curve on distance of 2 and 3.5 meters could indicate distance between the small clusters.

At a distance of around 13 meters the curve is pretty close to one which may imply the maximum distance where we can see a dependence between the plants of this category.

For the category Medium the behaviour is similar to the category Small, but here the cluster diameter is smaller 0.7 meters; and the distance between this clusters is around 5.2 meters. This distances show that the competition process is stronger generating smaller clusters and longer distances between this clusters.

In the Large category the competition and the range of correlation is stronger. The first part of the curve shows a very weak or insignificant cluster pattern, probably due to a few plants of this category been close together; the next part of the curve shows a regular pattern until the range of correlation which is about at 25 meters, indicating a strong competition between the plants of this category.

The higher peaks of the function at 2.6, 11.4 and 16 meters are probably caused by the first, second and third near neighbours.

The correlation between the three categories is analyse with the inter-type $k$ function and the cross pair correlation function. This results are presented in figure[3.19] in each graph we show the interaction between two categories in both ways, so we have six functions for each of the two methods used.

The inter-type $k$ functions are similar, around 5 meters of distance they show some kind of repulsion between the plants of the three categories; this repulsion becomes stronger between some of the categories after 16 meters.

From this inter-type functions we see that the interaction between the categories Large and Medium is weaker (less repulsive), contradictory to what we expected (stronger competition); but this could be explain given the low amount of area with good soil resources due to the rocks.

Although the plants from this two categories do not form a cluster pattern they are found closer, in areas further away from rocks; while the plants from the category Small can also be found in areas near the rocks.

If we compare the interactions functions between the categories in both ways, we see that the inter-type $k$ function of the class with greater dbh is less repulsive e.i. we found more Medium plants around a Large plant that the other way around.
This behaviour is typical in a forest and is reflected in the number of plants in each of the growth stages; where we found a greater number of plants with low dbh and as the diameter increases the number of plants decreases, primarily because of the competition for resources.

On the cross pair correlation functions, the behaviour is similar when comparing two categories in both ways, however here the difference is not that clear.

The cross pair correlation function between the three categories, shows that in the first meters of distance there is an aggregation or cluster pattern between the categories, after this first meters the functions are mostly under 1 indicating repulsion between the three different categories.

Although, in general the behaviour of these functions is similar, we can highlight some slight difference.
In the first meters of $r$ where they form a cluster pattern we see that it is stronger for the cross function of the categories Large-Medium, but the cluster diameter is greater for the function Large-Small.

This analysis show us once again that the plants of the categories Large and Medium are close together in small areas where most of the soil resources are available. In the function we see a peak in around 10 and 17 meters of distance, this may indicate the distance to the nearest first and second small cluster, form by Medium and Large plants.

The relationship between the Large and Small categories indicates that small plants are probably forming clusters around big plants. The peaks at 3, 4 and 5 meters may be caused by other clusters of small plants that are not around a big plant.
Figure 3.19. Inter-type $k$ function and the cross pair correlation function for the different categories of the mark size.
To test for random labelling for future models, we can compare the pair correlation function or the $k$ function of the unmarked point pattern with the functions of each of the categories. We see in the figures 3.18 and 3.19 that they are not equal indicating that there is not random labelling in this categories.

**Diameter at breast height - dbh**

The purpose of this section is to evaluated the numerical difference among the mark dbh dependent on the distances between the plants in the study area. This will complement the analyse of the size of the plants and will give more insight in the dynamics of competition and facilitation of the plants.

The figure 3.20 shows the $r$ mark correlation function and the mark variogram not normalized for the point pattern with the quantitative mark dbh. The dotted line in the $r$ - mark and mark variogram graph is the mean and the variance of the mark respectively, which is the value the function takes at very long distance where the mark is independent.

The $r$ mark correlation function reveals that the mark dbh is spatially correlated. There are different types of dependence among the mark depending on the distance.

At very close distance $r < 0.4$ meters the $k_f$ function shows a mutual stimulation between the marks i.e the points benefit from being close together and thus have on average larger marks than the mean dbh. But we can not be sure if there is really a mutual stimulation between the plants (facilitation process), or if this effect is caused by the limited areas with good soil resources in which we see an aggregation of plants with low and high dbh.

After this distance the $r$ mark function is mostly under the mean value of the mark 10.4 cm, showing an inhibition dependence which indicates that the points compete and have smaller than the average mark.

In the context of the present study, this shows that plants with larger dbh create a more regular pattern because of the competition for resources and therefore the only plants we start finding in distance between 0.4 and around 7 meters are plants with low dbh.

In the $r$ mark function we highlight two of the peaks, one at 9 meters and the other at 16 meters. This peaks indicate the distance where we start finding plants with bigger dbh. The first peak probably is including plants with medium size dbh while the second and higher peak is counting the bigger plants. Comparing with the pair correlation functions of the categories Medium and Large of the last section, we see that the highest peaks where at 5.2 and 16 meters for the Medium and Large categories respectively.

The mark variogram complements and support this latter analysis. At very close distances we see the highest peak of the function which indicates an aggregation of plants of all sizes; then the function starts to decrease showing the count only of small plants; at around 5 meters we see an increase in the function probably capturing better than the $r$ mark function the incorporation of the medium plants; and finally we see a rise in the values of the variogram at a distance of 16 meters due to the values of dbh of the bigger plants.

Both functions show a spatial correlation for the mark dbh. The range of correlation is around 28 meters of distance and it can be seen better in the $r$ mark correlation function.
CHAPTER 3. RESULTS

67

Figure 3.20. Top: mark correlation function, Bottom: mark variogram. Of the quantitative mark diameter at breast heigh-dbh.

The latter analysis on the qualitative and quantitative marks which are based on the dbh that represents the size of the plants, have give us an idea of the spatial structure of the plants and there dependence based on the size.

3.3 Classification methods

In this section we present three different ways to classify the species by their spatial behaviour, here we look only at the spatial distribution of each species and not at their inter specific dependence. First we present a method based on indices of dispersion; then two methods using the scores of two functional principal components analysis which are based on the inhomogeneous $k$ function and the inhomogeneous pair correlation function.
For the three approaches we used only the 32 species which presented a minimum of 10 plants inside the study area. We used the dbh and number of plants per species to analyse the groups form, the other ecological information like habit and light requirement was not sufficient to determined a good classification of the methods.

### 3.3.1 Method based on Indices

The values estimated for each of the three indices used on the 32 species are presented in the top box-plots in figure 3.21. Because the numerical characteristics presented different ranges, each value was scale by subtracting the mean and dividing by the standard deviation; this values are shown on the bottom box-plots of figure 3.21.

![Box-plots of indices values](image)

**Figure 3.21.** Top: box-plot of the values clark-evans index and Hopkins-Skellam test, Bottom: box-plot of the values the mean distance to the nearest neighbour of the same species

The clark evans index and the Hopkins-Skellam test suggest that more than 75% of the species tend to have a cluster pattern.

The box-plot of the mean distance to the nearest neighbour of the same species, shows values from 3.4 to 16.5 meters; there is a slight concentration of this distance around 9.5 and 11.6 meters in the third quartile, but in general the box plot seems to be even in all of the four quartiles.

From this distance we can not define if the pattern is cluster or regular, however low values could give an idea of a strong cluster patterns while high values could refer to a weak cluster pattern or a regular pattern of the species.
The cluster analysis was proposed with these three indices so that the first two would help us separate clusters from regular patterns, while the third was used as a tool for determining the strength of the pattern.

The figure 3.22 shows the results from the hierarchical cluster. Three main groups were formed, based on the indices used, the one on the left could be assigned as the species with regular pattern (group I2); the other two main groups show species with a cluster pattern (group I1 on the right and group I3 on the middle).
Figure 3.22. Cluster analysis (Wards method) of the 32 species using the values of the three indices.
Figure 3.23 shows for each of the three groups form the species with their values of the diameter at breast height -dbh.

**Figure 3.23.** Boxplot of the values of dbh for each species in the three groups form. Method indices
The group assign as regular, group I2, contains six species which all have a tree habit; five of them are heliophytes and *Pouteria plicata* is considered schiophyte.

The values of dbh varies between the species of this group. Some present very low dbh, under 10 cm, while others have trees with more than 20 cm of dbh.

The other two groups, as we mentioned, the indices indicate that these species have a cluster pattern. The main differences that we found between them is the number of plants per species and the size express in their dbh.

The right side of the cluster diagram, groups a total of 12 species (group I1). The distribution of the values of dbh for all 12 species are very low, most of them do not have plants with a dbh over 10 cm.

Group I3 located in the middle of the cluster diagram, consist of 14 species. This group contains the species with more number of plants inside the study area, in average 52 plants per species.

The values of dbh for these species, is more diverse. Four species have very low values mostly under 10 cm of dbh; for seven of them, most of their plants have values between 5 and 15 cm; *Sapium g.* have trees between 10 and 35 cm of dbh; the species defined as Morf sp4 have a wide range of values, form 2 to 70 cm; and for *Attalea m.* the values are over 20 cm and under 40 cm of dbh.

### 3.3.2 Methods based on the scores of two FPCA

The inhomogeneous $k$-function and the inhomogeneous pair correlation function with a Ripley’s isotropic edge correction were estimated for each of the 32 species. Because the library spatstat in R, that generates this functions only gives the values for different distances, we needed to create our own functions to implement the functional component analysis and the cluster analysis.

To recreated the two functions for each species, a b-splines basis system of order 6 with 512 breaks and 516 basis functions was used. Then a data smoothing with a roughness penalty was implemented; the measure of roughness used was the integrated squared second derivative from which the fitting criterion was established. Different smoothness parameter $\lambda$ of the fitting criterion were used; the final value for this parameter was $\lambda = 0.000001$, leaving the functions free to fit as closely as possible with the selected basis system.

The recreated $k$-functions and pcf are presented in the figure 3.24. On the top of the figure is the $k$-functions, the solid line represents $\pi * r^2$ the values for CSR; here we see that for most of the species after 5 meters of distance the functions is below the values for CSR, indicating a regular pattern, stronger for some species than others.

For the pair correlation functions on the bottom of figure 3.24 we see a similar behaviour; after 5 meters of distance, all the species are under the value of 1 showing a regular pattern. However with the pcf at distance under 5 meters, some of the species show a clear cluster pattern.

The FPCA with the $k$-functions show that the first principal component explained a proportion of 97 % of the variance. For the FPCA of the pair correlations functions the first and second principal components explain 57 and 28 % of the variance respectively.
Figure 3.24. Smoothed K functions of 32 species used for Functional Principal Components Analysis. Top: Smooth K functions, Bottom: Smooth Pair correlation functions (close up to distance 0 to 10 meters and values 0 to 10).

The cluster analysis for the \( k \)-functions was build only with the scores of the first principal component of FPCA, while the cluster for the pcf was build using the scores of the two principal components.

Figure 3.25 shows the results of the cluster analysis for the method with the \( k \) function. Using the scores of the FPCA was very helpful to separated the species by the behaviour of their second order function. In the case of the cluster analysis for the \( k \) functions, we see four groups that are separated based on the strength of the inhibition pattern or the closeness to the values of CSR.
hclust(*, "ward.D2")

Figure 3.25. Diagram of the cluster analysis (Wards method) of the 32 species using the scores of the first functional principal component (k functions)
In figure 3.26 we present the $k$ functions of the species, distinguish in different colors by the groups form with the cluster analysis. Here we can see more clearly how the scores of the first principal component separates the species by their strength of the inhibition pattern.

The four groups form in the cluster analysis for the $k$ functions, can also be separated into two. The ones on the left of the diagram, contains those species with functions that indicate a stronger inhibition or regular pattern; while the ones on the right their functions are closer to the values of CSR.

We defined as group K1 the species with a stronger inhibition pattern. In the figure 3.25 this is the second group from left to right. In contains 8 species. The dbh of the plants of this species are mostly below 20 cm (figure 3.27).

The species with the second stronger regular pattern are contain in group K4, this species are located on the left side of the cluster diagram. This group is form by 10 species. Their dbh are similar to group K1, most of the plants of these species are under 20 cm of dbh (figure 3.27).

The only difference between this first two groups based on the information that we have, is the mean number of trees in each group. Group K1 presents 17.5 plants per species while group K4 have 26.8 plants per species. One more time the number of plants may be influencing the strength of the inhibition pattern.
Figure 3.27. Boxplot of the values of dbh for each species in the three groups form. Method k functions.
Group K3 contains the species with their $k$ function a little bit closer to the values of CSR. Here we see a total of 8 species. The dbh of the trees of this species varies: we see three species with dbh mostly under 10 cm, other three with dbh between 2 to 40 cm, and *Pachira n.* with diameters from 2 to 71 cm (figure 3.27).

The six species with their $k$ function closer to the values of CSR are in group K2. The values of dbh also varies like in group K3: one of the palms *Attalea m.* presents dbh between 20 and 37 cm, the other palm *Bactris b.* have small values of dbh between 2 and 8 cm, the species *Gustavia a.*, *Eschweilera t.* and *Protium g.* present dbh between 1.5 and 37 cm, and the one defined as Morf sp4 presents dbh that goes up to around 60 cm.

One more time the difference between group K3 and K2 is the mean number of plants per species. Here the group closer to the values of CSR have more plants, 79 per species, while group K3 only presents 25.2 plants per species.

We see that the number of plants have an important influence on the $k$ function, the more number of plants the weaker the inhibition pattern. However, if we compared groups K3 and K4, we see that the latter one present a greater mean number of plants per species and have a stronger inhibition pattern.

The group K3 has three species with higher values of dbh compared to the species of the K4 group; but four of the species in group K3 have very low values like those in group K4. For these last species we did not find many differences with the species in the group K4. If the number of plants and the size of the plants do not cause the difference in their spatial behaviour, there must be an intrinsic ecological characteristic that explains these differences; in the present study we can not defined which are those characteristics, but the methodology used here can be implemented to study which ecological characteristics influence more the spatial distribution of the trees.

As mentioned, for the cluster analysis with the pcf we used the scores of the first and second principal components. The figure 3.28 shows the results of the cluster analysis. Here we can identify five groups which are separated by two factors: (1) the strength of the inhibition pattern; and (2) the a cluster pattern identify at close distances.

In figure 3.28 the five groups are obtain by cutting at a heigh of 4. From left to right the groups contains species with stronger to weaker inhibition patterns and with reduced to bigger aggregation patterns at very close distance.
Figure 3.28. Cluster analysis (Wards method) of the 32 species using the scores of the two first functional principal components (Pair correlation functions)
In figure 3.29 we present the pcf of the 32 species, here we use one color per group to visualise the difference between the five groups.

![Figure 3.29. k functions of the 32 species separated by the groups form in the cluster analysis](image)

We start by describing the three groups on the left. The first one on the far left of the cluster structure is defined as group P3. This group contains eight species, their pcf are the ones further apart from the value 1 and only *Pachira n.* and *Petrea sp.* show an aggregation pattern at very close distances. The mean number of plants per species is of 25.5, and the dbh of the eight species varies: three have very low values under 10 cm, other three have values from 2.5 cm to around 30 or 40 or even 70 cm, and the values of dbh of the palm *Attalea m.* are between 20 to 37 cm (figure 3.30).

Twelve of the species are in group P2, this group is the second set of species from left to right. Here the inhibition pattern is strong, but most of them have an aggregation pattern at very close distances. The values of dbh for almost all plants of these species are under 20 cm and for seven of them their dbh is even smaller than 10 cm (figure 3.30).

The group next to P2 is group P1. Here we see a total of 5 species. All of the pcf of these species show a cluster pattern with a cluster diameter a little bit larger than those on group P2. Nearly all plants of these species have values of dbh under 10 cm (figure 3.30).

As in the groups based on the *k* functions, here we see that the number of plants per species also is important. The main difference between P1 and P2, it is also the number of plants per species, P1 has a mean number of 31 while P2 has 23 plants per species.

The other two groups P5 and P4, are on the right side of the cluster diagram. Group P5 only presents two species: *Bursera s.* and *Ficus a.*. The values of the pcf of these species are very close to 1 and show a cluster pattern with the larger cluster diameters, specially *Bursera s.* which is around 5 meters. The trees in both species have dbh between 2.5 and around 30 cm (figure 3.30).
The last five species are contained in group P4. The pair correlation function of these species is similar to those of group P5, but their cluster diameter is smaller and the values of the function are closer to 1. The dbh of the palm are very small under 10 cm, three of the species have diameters between 1.2 and around 30 cm, and the other species have dbh up to 60 cm.
CHAPTER 3. RESULTS

Figure 3.30. Boxplot of the values of dbh for each species in the three groups form Method pcf
3.3.3 Comparison between methods

We compared the species in each of the three, four and five groups defined for the three methods implemented. The methods have similarities in the groups form, in table 3.3 we present the groups for each of the methods and the number of species that those groups have in common, and on figure 3.31 we show a structure based on the number of species in common between the groups of the different methods, here the geometric figures represent the affinity between the groups of the three methods.

<table>
<thead>
<tr>
<th>G. Indices</th>
<th>G. inhomK</th>
<th>G. inhompcf</th>
<th>N. Species in common</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1 K1 P1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I1 K1 P2</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>I1 K1 P5</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I1 K3 P1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I1 K3 P3</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I1 K4 P1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I1 K4 P2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>I1 K4 P5</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I2 K1 P2</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I2 K3 P3</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>I2 K4 P2</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I2 K4 P3</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I3 K1 P2</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I3 k2 P3</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I3 k2 P4</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>I3 K3 P1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I3 K3 P3</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>I3 K4 P1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>I3 K4 P2</td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3.3. Groups defined in the three methods, and the number of species in common

We defined three main groups represented by the geometric figures circle, square and diamond. 18 of the 32 species share the same geometric figure in the three methods, 13 species had the same figure in two of the methods, and only Matayba sp. had a different geometric figure in each of the methods.
The circle was assigned to group I1 of the indices method, groups K1 and K4 for the \( k \) functions method, and groups P1, P2 and P5 for the pcf method. In total, 10 species had the geometric figure circle in all of the three methods. Based on the second order functions these species show a strong inhibition pattern with a small aggregation pattern at close distances; however the values of the indices indicate a cluster pattern.

Because the indices are single values that express the spatial behaviour of the species, we think that for these species the indices are capturing only the aggregation pattern at close distances and its true regular pattern, which it is obvious in the second order functions, is not captured.

Some characteristics of the species with a circle figure in all methods are: mean number of plants per species 18; and almost all of the plants of these species have dbh below 20 cm.

The diamond figure contains group I2, group K3 and group P3. Only three species had in common this geometric figure in all methods, however, between the groups of the \( k \) function and pcf methods we find that this group has in common a total of six species. The second order functions show that these species also have a strong inhibition pattern but they do not show an aggregation pattern at close distances; in this case the values of the indices also show a regular pattern. The mean number of plants per species is 15, and the dbh of the plants of these species are between 2 cm to 40 cm in two of the species, and up to 70 in the other species.
The square figure was assigned to group I3, group K2, and group P4. Six species had in common this geometric figure in all three methods. These six species show a weaker inhibition pattern, with stronger cluster patterns at very close distances. The species with the square figure present the larger number of plants inside the plot, and the dbh of the plants of the species varies: one species presents dbh under 10 cm, three have values from 1.2 cm up to around 40 cm, and the other species have values up to 70 cm.

Although there are similarities between the groups of species formed in the three methods, the methods based on the second order functions are more alike. Ten species besides the 18 mentioned, share the same figure in both of the second order methods. In total 28 of the 32 species have a common figure between both of the second order methods.

If we focus only on this two methods we see that the sciophytes species are included in the species with the circular figure, which indicates as mentioned, species with strong inhibition patterns but with aggregation patterns at close distances.

The use of these methods, specially the ones based on the second order functions, can help us create groups of species based on their spatial distribution. Relating this groups to different ecological and growth information can be used as a strategy to understand the main characteristics that influence the spatial behavior in the species of the tropical dry forest.

In the present research we only focus on a few characteristics, size based on the dbh, and some ecological information: light requirements and habit. We see a great influence on the strength of the pattern due to the number of plants per species and the values of dbh; species with less number of plants and species with larger values of dbh show a stronger inhibition pattern. For the ecological information we can only say that the four sciophytes species have a stronger inhibition pattern.

Understanding the main characteristics that influence the spatial behavior of the species, may lead us to determine better categories that could be analyzed using different tools for qualitative mark point patterns.
Discussions and Conclusions

In this study, we proposed a methodology to evaluate the spatial structure of the plants of a tropical dry forest, based on different tools of point process theory. The three-phase methodology allows us to evaluate the resulting mark point pattern, which led us to comprehend the facilitation and competition processes that influence the spatial distribution of the plants.

Each phase of the methodology permits us to understand some aspect of the spatial distribution. We used the first phase to estimate the intensities, taking into account the effect of the rocky soil. These intensities were the basis for the latter analysis, including the methodology to define the dominant species and the use of inhomogeneous second-order functions. The second-order functions show us the inter and intra-specific dependences of the qualitative mark point patterns, allowing us to understand competition and facilitation processes between the species and the size of the plants. With the final phase proposed, we were able to group the species by their spatial behavior.

Although the application studies that use point processes theory follow a similar methodology, here we proposed a unique model to capture the effects of the rocks in the estimated intensities, reduced the categories of the mark species using a different methodology from those proposed in ecology to identify the dominant species, and evaluated three methods to classify the species by their spatial behavior.

This methodology can be used in other applications where the mark point pattern under study presents qualitative marks with a great number of categories and where the influence of a covariate is determined in the spatial distribution of the points.

Now, we present a set of conclusions ordered in the three phases of the methodology proposed.

First order characteristics

The model proposed to estimate the intensities of the unmarked point pattern and the sub-patterns of the categories of the qualitative marks species and size, have not yet been implemented to evaluate the effect of the rocky soils in a tropical dry forest.

The structure of the model uses a kernel as the base value of intensity and with the covariate distance to the rocks, this intensity is modified. We used three methods for the bandwidth estimation of the kernel, our methodology also uses growth and ecology information to define the bandwidth use. For most of the sub-patterns the diggle algorithm was defined.
The use of an exponential part to captured the relationship between the point of the pattern with the covariate, allow us to measured at which distances the intensity increases and decreases due to the effect of the rocks.

For the unmarked point pattern we found that at a distance of 0.8 meters the exponential part of the model equals one, so the kernel is not modified; at greater distances the intensity increases up to a factor of 1.66 this happens at four meters of distance from the rocks; and at shorter distances the intensity decreases to a factor of 0.8, which occurs over the rocks.

Most of the species and the three categories of the mark size presented the same behaviour but with different values of the factors and distances that modify the kernel. Based on this models we were able to identify which species have develop a better adaptability around rocky soils. We also identify differences between the categories of the mark size, specially in the Medium and Small classes; the model indicates that it is easier for Small plants to stablish around rocks than it is for Medium plants to survive around rocks.

The results of these models are consistent with reality, we expect limited soil resources around rocky soils, which increases the competition between the plants and this is reflected in the reduction of the intensity. This competition is stronger in Medium plants because they required a greater among of soil resources to survived, while on small plant they can establish with a fewer soil resources.

It is necessary to highlight that the parameters of the model were not significant for some of the species and for the category Large of the mark size. We conclude that this happens because of the small among of points in each of these categories. To improve the model and captured the relationship of these categories with the rocks, the covariate could be modify as a factor with two categories: presence and absence of rocks.

The model can be used to test other covariates that may influence the spatial behaviour of the plants in this type of tropical forest or in any other kind of forest.

**Second order characteristics**

The second phase of the methodology used, have presented the criteria to analyse the spatial structure of the this tropical dry forest. This methodology is based not only on the unmarked point pattern but its focus is primarily on the marks of the point pattern. This approach has giving us the tools to understand spatial dependence between species and the size of the plants, which influenced on our objective of understanding the competition and facilitation processes of this type of forest.

While the second order functions of the unmarked point pattern show a general spatial behaviour of the point pattern, indicating an aggregation pattern at distances lower than 5.6 and 4.2 meters (inhomogeneous $k$ function and inhomogeneous pcf respectively) and from this distance a regular pattern, the second order functions used on the marks gave us a more information about the specific dependence between the plants of the forest.

An important aspect in the analysis of the species mark, was the forest structure found in the study area, which presented a high species richness with low species evenness, something common in many tropical ecosystems. This brings statistical problems when using mark point process techniques in the mark species. The first big problem is that we have to analyse all the inter specific dependences, so for a high number of species this pairwise analysis becomes computational impossible. The second big problem is that if a
DISCUSSION AND CONCLUSIONS

species have a very low number of trees, some point process techniques are unsuitable to implemented.

To deal with this problem, we used only the most dominant species inside the study area. To select these species we propose a methodology based on the intensities estimated in the previous models; for each location $u$ the species with the greater probability was selected. Our methodology for selecting the most dominant species, uses not only the plant density but it also takes into account the location of each plant and through the bandwidths of the kernel the size of the plants. The locations of the plants is not use in most of the dominance indices, usually in ecology the dominance is examined using different variables like plant density, cover and biomass [34].

The advantage of this methodology is that not only gives the dominant species but it also indicates the location where that species is the most dominant, unlike most use dominance indices which give only a value of dominance for each species. The disadvantages is that this methodology needs the location and dbh of each plant inside the observational window, while the dominance indices only required a sample data of the study area and the location of the plants is not necessary.

By using the inhomogeneous $k$ and pcf of each of the 11 species selected we were able to identify the spatial behaviour of the plants of each species. This analysis was very helpful in understanding the intra specific dependences. For most of the species, small aggregation patterns were form at close distances, the species *Protium g.* was the one with the larger cluster pattern up to 2.4 meters of distance. At larger distances the species show a regular pattern, for some species this pattern was stronger.

The inhomogeneous condensed functions were used to evaluated the inter specific dependence. The used of these functions are useful when analysing qualitative marks with a great number of categories, this strategy presents different advantages: first it reduces the number of functions to analyse compared to the strategy of using multi type or cross pair correlation functions; second it allow us to see the dependence of a species with the rest of the species and not only the other 10 species selected; and third, we saw in the analysis that with this strategy we were able to identify specific dependences between species, this can be done by comparing the condensed pcf of the species, identifying peaks at a certain distances in different species might indicate a inter specific dependence between those species.

To complement the analysis of the inter and intra specific dependences of the species, is needed to evaluated each stage of growth in each species separately. That is created new categories based on the size of each species. But this presents difficulties, because it reduces the number of plants in these new categories, we believe that using techniques from spatio-temporal point patterns might solve this problem. To be able to used this type of analysis the information of the point pattern of the forest has to be measured over a period of time, which is a expensive task, however the technologies on satellite images, radar information and aerial photographies are improving, this reduces the costs of collecting this type of information.

The other advantage of the methodology used is that by using the other two marks generated by the dbh, we were able to analyse the dependences in relation to the size of the plants. Here we saw that Small plants tend to form aggregation patterns of around 1 meter of radius, while medium presented a radius of 0.7 meters and Large did not show a real aggregation pattern. The pcf allows us to identify that the distances between clusters
of Small plants are around 2 and 3.5 meters, between clusters of Medium plants around 5.2 meters. The max correlation distance indicate that Small and Medium plants have a lower range of influence compared to the 25 meters shown for the Large category.

The dependence between the different categories, show that the Large and Medium plants, form very strong aggregation patterns with small radius, while the relationship Large Small indicates that small plants are establishing around big plants but with less density and in a bigger area.

An other important aspect of our methodology is that by analysing the quantitative mark dbh, we can support the analysis done on the qualitative mark size. The results show how at very close distances up to around 0.4 meters there is a mutual stimulation that generates values greater than the mean value of dbh and high values of variance, this indicates an aggregation pattern of plants of all sizes. After this distances, between 0.4 and 7 meters the value of variance and dbh drops, which shows that we mostly find plants with small dbh. The peaks at 5.2 and 16 meters in the curves of the r-mark correlation function and in the mark variogram, show how at this distances the Medium and Large plants are incorporated respectively, this matches the high peaks on the inhomogeneous pcf of the categories Medium and Large.

The result from this phase of the methodology which incorporates the intensities estimated in the first phase , can be used to define inter and intra dependence distances for the categories of the qualitative marks used. This information is very useful for future point pattern models, defining this models can help simulate the spatial structure of the forest and study the effect of the absence of a category to the hole spatial pattern.

**Classification methods**

The three method presented similarities in the species of the groups form. However the indices used in the first method, show that most of the species had an aggregation pattern which is contradictory to what the second order functions showed

This first method which uses indices that are single values that represent the spatial distribution of the species, can be used as a classification strategy but does not represent the actual spatial distribution of the species.

On the other hand, the scores used from both of the functional component analysis allow us to disaggregate better the species based on their spatial distribution. Here the key element was the FPCA, this technique allow us to express the spatial behaviour of the species in one value (score of the first component) in the case of the $k$ functions and in two values (scores of the first two components) in the case of the pcf. In some way we could say that this methodology creates an index based on the second order functions.

The main objective of these classification methodologies was to group the species by their spatial behaviour. Here we conclude that the method with the pcf has a better performance in achieving this goal. The scores of the FPCA done with the pcf, allows us to separated the functions based on the characteristics of the cluster found at very close distances and the strength of the regular pattern at greater distances. This conclusion agrees with the results found by [38] where via simulation they determine that the pcf is better for this type of analysis.

Although, the result of the methods with the $k$ functions are very similar, the classification is based more on the strength of the regular pattern. This leaves out a special
characteristics in the spatial structure of the plants of this tropical dry forest, which is the very strong cluster pattern that is formed at very close distances.

This strategy can also be used to identify special ecological characteristics that influence the spatial behavior of the species. This can be done comparing species from different groups that have similar dbh values and number of plants but different ecological characteristics.

Comparing the results of these classification methods with ecological or growth information can also help to identify variables to be used as qualitative marks that can explain better the spatial behavior of the plants of the tropical dry forest.

This methodology, especially the two methods based on the second order functions, can be used in any kind of forest, the use of additional information about the species like seed dispersal strategies, is a good way to probe that the classification methods are working. In the present study the lack of this kind of information did not allow us to define which of the two second order methods is a better choice based on ecological characteristics.
Future work

• Updating the intensity model proposed can be done in two different ways. One is modifying the kernel used, specially in selecting the bandwidth, other approaches than using growth and ecological information can be propose. Second the covariate related to the rocks can be modify, one way is to used a factor variable of presence and absence of rocky soil for each location $u$.

• The analysis of the qualitative mark species for any tropical forest still has some challenges, not only because of the great number of species and the pairwise interaction analysis that this implies, but because of the different growth stages that each species has in a specific period of time. To analyse each growth stage for each species would mean reducing the number of points in each of these categories. Probably the way to tread this problem is with spatial temporal patterns, thanks to the constant advances in technology the point pattern of a tropical forest will be easier to captured over a period of time.

• Point patterns in a three dimensional space, is a strategy that has not been used in the analysis of the spatial structure of any kind of forest. The used of a three dimensional space, using the location and height of the trees could give a better understanding of the dependence between trees of different sizes.

• Although the categories of the mark size were defined by information of the stages of growth of Colombian tropical dry forest, we think that a new classification would work better for this specific data. This new ranges could be based on the groups form in the classification methods used in the present study.

• The results of the two first phases of the methodology can be used to propose models that would help us simulate the point pattern. Constructing this type of models can help to understand the importance of an specific species to the overall spatial distribution of the trees inside the forest.
Notation Index and Glossary

Notation Index

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Point process</td>
</tr>
<tr>
<td>x</td>
<td>Point pattern</td>
</tr>
<tr>
<td>W</td>
<td>Observation window</td>
</tr>
<tr>
<td>M</td>
<td>Mark point window</td>
</tr>
<tr>
<td>m</td>
<td>Mark point pattern</td>
</tr>
<tr>
<td>λ</td>
<td>Intensity</td>
</tr>
<tr>
<td>λ_o</td>
<td>Intensity subpattern with category σ_i(r)</td>
</tr>
<tr>
<td>λ_S</td>
<td>Sum intensity</td>
</tr>
<tr>
<td>B</td>
<td>Region</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>Spatial location</td>
</tr>
<tr>
<td>Z(u)</td>
<td>Spatial covariate</td>
</tr>
<tr>
<td>r</td>
<td>Radius</td>
</tr>
<tr>
<td>h</td>
<td>Bandwidth</td>
</tr>
<tr>
<td>b</td>
<td>Circle</td>
</tr>
<tr>
<td>κ</td>
<td>Kernel</td>
</tr>
<tr>
<td>e</td>
<td>Edge correction</td>
</tr>
<tr>
<td>D</td>
<td>Diggle’s edge correction</td>
</tr>
<tr>
<td>U</td>
<td>No edge correction</td>
</tr>
<tr>
<td>w</td>
<td>weights</td>
</tr>
<tr>
<td>d_ij</td>
<td>Distance between x_i en x_j</td>
</tr>
<tr>
<td>k</td>
<td>k-function</td>
</tr>
<tr>
<td>g</td>
<td>Pair correlation function</td>
</tr>
<tr>
<td>k_m</td>
<td>Mark correlation function</td>
</tr>
<tr>
<td>k_m</td>
<td>r-mark correlation function</td>
</tr>
<tr>
<td>σ_i(r)</td>
<td>Mark variogram</td>
</tr>
<tr>
<td>R</td>
<td>Clark-Evans index</td>
</tr>
<tr>
<td>A</td>
<td>Hopkins-Skellam</td>
</tr>
<tr>
<td>P_i</td>
<td>Nearest neighbour distance to point i</td>
</tr>
<tr>
<td>L_i</td>
<td>Nearest simulated neighbour distance to point i</td>
</tr>
<tr>
<td>O</td>
<td>Pair orientation distribution</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>y_p</td>
<td>Observation of z</td>
</tr>
<tr>
<td>t_p</td>
<td>Time or location in z</td>
</tr>
<tr>
<td>φ_q</td>
<td>Basis functions</td>
</tr>
<tr>
<td>c_q</td>
<td>coefficients of the expansion</td>
</tr>
<tr>
<td>Q</td>
<td>Vector of coefficients functional data</td>
</tr>
</tbody>
</table>

Glossary

**Aboveground biomass:** This term encompasses all living biomass which is located above the ground. Generally speaking, it includes the stem, stump, branches, bark, seeds and foliage.

**Aboveground biomass:** This term encompasses all living biomass which is located above the ground. Generally speaking, it includes the stem, stump, branches, bark, seeds and foliage.

**Abundance:** Number of individuals of each species in relation to the total that make up the community or subcommunity.

**Basal area:** Basal area is the area of a given section of land that is occupied by the cross-section of tree trunks and stems at the base. The term is used in forest management and forest ecology.
**Canopy height:** Forest canopy height \( h \) (m), the height of the highest vegetation components above ground level.

**Forest dynamics:** Forest dynamics describes the underlying physical and biological forces that shape and change a forest ecosystem.

**Family, genus, species:** In biological classification, taxonomic rank is the relative level of a group of organisms (a taxon) in a taxonomic hierarchy. Examples of taxonomic ranks are species, genus, family.

**Secondary succession:** Occurs in areas where a community that previously existed has been removed; it is typified by smaller-scale disturbances that do not eliminate all life and nutrients from the environment.

**Species composition:** Species composition refers to the contribution of each plant species to the vegetation.

**Species richness:** Number of species, can be measured with a species accumulation curve.

**Species evenness:** Number of individuals of each species, relative abundance of each species.

**Species diversity:** A measure combining richness and evenness

**Types of light requirement species:** Heliophytes are species that are adapted to a habitat with a very intensive insolation; sciophytes are species that endures or thrives best at lowered light intensity.
Bibliography


[43] Robert John, James W Dalling, Kyle E Harms, Joseph B Yavitt, Robert F Stal- 
lard, Matthew Mirabello, Stephen P Hubbell, Renato Valencia, Hugo Navarre-
rete, Martha Vallejo, and Others, Soil nutrients influence spatial distributions of tropi-

[44] Tero Kokkila, A Makela, and Eero Nikinmaa, A method for generating stand struc-


[46] Georges Kunstler, Daniel Falster, David A Coomes, Francis Hui, Robert M Kooy-
man, Daniel C Laughlin, Lourens Poorter, Mark Vanderwel, Ghislain Vieilledent, S Joseph Wright, and Others, Plant functional traits have globally consistent effects on competition, Nature 529 (2016), no. 7585, 204–207.

[47] Guillermo Castro Marín, Robert Nygård, Benigno Gonzales Rivas, and Per Chris-
ter Oden, Stand dynamics and basal area change in a tropical dry forest reserve in Nicaragua, Forest ecology and management 208 (2005), no. 1, 63–75.


[56] Camila Pizano and Hernando García, El bosque seco tropical en colombia, Instituto de Investigación de Recursos Biológicos, 2014.


In this appendix we present the tables and graph that were not put inside the results chapter.

Variables

The figure shows the two options to construct the mark diameter at breast height. The first option overestimates the DBH; the second option due to the use of the basal area formula gives a more precise information about the area occupied by the tree.

The figure shows a boxplot of the values of DBH for each of the 32 species with more than 10 plants inside the study area.
Figure 32. Two options to construct the mark diameter breast height dbh for the trees with more than one stems. Option 1: sum of dbh of the different stems, Option 2: use of basal area formula.
Figure 33. Box plot of the value of diameter at breast height dbh for 32 species
**First order characteristics**

The following table presents the result of the different parametric models used to estimated the intensities of the unmarked point pattern and the categories of the qualitative marks species and size.

<table>
<thead>
<tr>
<th>PP</th>
<th>parameter</th>
<th>Estimate</th>
<th>S.E.</th>
<th>CI95.lo</th>
<th>CI95.hi</th>
<th>Ztest</th>
<th>Zval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmark</td>
<td>Alpha</td>
<td>-0.13</td>
<td>0.04</td>
<td>-0.21</td>
<td>-0.06</td>
<td>***</td>
<td>-3.60</td>
</tr>
<tr>
<td>Unmark</td>
<td>Beta</td>
<td>0.16</td>
<td>0.03</td>
<td>0.11</td>
<td>0.21</td>
<td>***</td>
<td>6.03</td>
</tr>
<tr>
<td>Large</td>
<td>Alpha</td>
<td>-0.08</td>
<td>0.14</td>
<td>-0.36</td>
<td>0.20</td>
<td>**</td>
<td>-0.56</td>
</tr>
<tr>
<td>Large</td>
<td>Beta</td>
<td>0.10</td>
<td>0.11</td>
<td>-0.12</td>
<td>0.31</td>
<td></td>
<td>0.90</td>
</tr>
<tr>
<td>Medium</td>
<td>Alpha</td>
<td>-0.19</td>
<td>0.09</td>
<td>-0.37</td>
<td>-0.00</td>
<td>*</td>
<td>-2.00</td>
</tr>
<tr>
<td>Medium</td>
<td>Beta</td>
<td>0.21</td>
<td>0.06</td>
<td>0.09</td>
<td>0.33</td>
<td>***</td>
<td>3.38</td>
</tr>
<tr>
<td>Small</td>
<td>Alpha</td>
<td>-0.13</td>
<td>0.04</td>
<td>-0.21</td>
<td>-0.04</td>
<td>**</td>
<td>-2.99</td>
</tr>
<tr>
<td>Small</td>
<td>Beta</td>
<td>0.15</td>
<td>0.03</td>
<td>0.09</td>
<td>0.21</td>
<td>***</td>
<td>4.90</td>
</tr>
<tr>
<td>Arrabidae sp</td>
<td>Alpha</td>
<td>-0.24</td>
<td>0.34</td>
<td>-0.91</td>
<td>0.43</td>
<td></td>
<td>-0.70</td>
</tr>
<tr>
<td>Arrabidae sp</td>
<td>Beta</td>
<td>0.24</td>
<td>0.22</td>
<td>-0.20</td>
<td>0.68</td>
<td></td>
<td>1.08</td>
</tr>
<tr>
<td>Astronium g.</td>
<td>Alpha</td>
<td>-0.01</td>
<td>0.24</td>
<td>-0.47</td>
<td>0.45</td>
<td></td>
<td>-0.03</td>
</tr>
<tr>
<td>Astronium g.</td>
<td>Beta</td>
<td>-0.08</td>
<td>0.27</td>
<td>-0.62</td>
<td>0.45</td>
<td></td>
<td>-0.30</td>
</tr>
<tr>
<td>Attalea m.</td>
<td>Alpha</td>
<td>-0.04</td>
<td>0.18</td>
<td>-0.39</td>
<td>0.31</td>
<td>-0.23</td>
<td></td>
</tr>
<tr>
<td>Attalea m.</td>
<td>Beta</td>
<td>0.05</td>
<td>0.14</td>
<td>-0.23</td>
<td>0.33</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>Bactris b.</td>
<td>Alpha</td>
<td>-0.38</td>
<td>0.12</td>
<td>-0.61</td>
<td>-0.15</td>
<td>**</td>
<td>-3.25</td>
</tr>
<tr>
<td>Bactris b.</td>
<td>Beta</td>
<td>0.35</td>
<td>0.06</td>
<td>0.23</td>
<td>0.46</td>
<td>***</td>
<td>5.93</td>
</tr>
<tr>
<td>Bursera s.</td>
<td>Alpha</td>
<td>0.63</td>
<td>0.30</td>
<td>0.03</td>
<td>1.22</td>
<td></td>
<td>2.07</td>
</tr>
<tr>
<td>Bursera s.</td>
<td>Beta</td>
<td>-651.22</td>
<td>9011.98</td>
<td>-18314.37</td>
<td>17011.93</td>
<td></td>
<td>-0.07</td>
</tr>
<tr>
<td>Clathrotropis m.</td>
<td>Alpha</td>
<td>-0.12</td>
<td>0.27</td>
<td>-0.65</td>
<td>0.41</td>
<td>-0.44</td>
<td></td>
</tr>
<tr>
<td>Clathrotropis m.</td>
<td>Beta</td>
<td>0.17</td>
<td>0.24</td>
<td>-0.30</td>
<td>0.64</td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td>Erythroxylum m.</td>
<td>Alpha</td>
<td>-0.16</td>
<td>0.35</td>
<td>-0.84</td>
<td>0.52</td>
<td>-0.46</td>
<td></td>
</tr>
<tr>
<td>Erythroxylum m.</td>
<td>Beta</td>
<td>0.21</td>
<td>0.30</td>
<td>-0.37</td>
<td>0.79</td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>Eschweilera t.</td>
<td>Alpha</td>
<td>-0.25</td>
<td>0.14</td>
<td>-0.53</td>
<td>0.03</td>
<td>-1.77</td>
<td></td>
</tr>
<tr>
<td>Eschweilera t.</td>
<td>Beta</td>
<td>0.28</td>
<td>0.09</td>
<td>0.10</td>
<td>0.45</td>
<td>**</td>
<td>3.11</td>
</tr>
<tr>
<td>Ficus a.</td>
<td>Alpha</td>
<td>-0.39</td>
<td>0.41</td>
<td>-1.19</td>
<td>0.41</td>
<td></td>
<td>-0.96</td>
</tr>
<tr>
<td>Ficus a.</td>
<td>Beta</td>
<td>0.44</td>
<td>0.28</td>
<td>-0.12</td>
<td>1.00</td>
<td></td>
<td>1.54</td>
</tr>
<tr>
<td>Fridericia p.</td>
<td>Alpha</td>
<td>0.09</td>
<td>0.24</td>
<td>-0.38</td>
<td>0.56</td>
<td></td>
<td>0.36</td>
</tr>
<tr>
<td>Fridericia p.</td>
<td>Beta</td>
<td>-0.15</td>
<td>0.24</td>
<td>-0.61</td>
<td>0.32</td>
<td>-0.62</td>
<td></td>
</tr>
<tr>
<td>Guarea g.</td>
<td>Alpha</td>
<td>0.10</td>
<td>0.21</td>
<td>-0.31</td>
<td>0.52</td>
<td></td>
<td>0.48</td>
</tr>
<tr>
<td>Guarea g.</td>
<td>Beta</td>
<td>-0.13</td>
<td>0.19</td>
<td>-0.51</td>
<td>0.25</td>
<td>-0.67</td>
<td></td>
</tr>
<tr>
<td>Gustavia a.</td>
<td>Alpha</td>
<td>-0.11</td>
<td>0.16</td>
<td>-0.42</td>
<td>0.20</td>
<td>-0.69</td>
<td></td>
</tr>
<tr>
<td>Gustavia a.</td>
<td>Beta</td>
<td>0.13</td>
<td>0.12</td>
<td>-0.10</td>
<td>0.36</td>
<td></td>
<td>1.14</td>
</tr>
<tr>
<td>Heisteria a.</td>
<td>Alpha</td>
<td>-0.27</td>
<td>0.28</td>
<td>-0.83</td>
<td>0.28</td>
<td>-0.97</td>
<td></td>
</tr>
<tr>
<td>Heisteria a.</td>
<td>Beta</td>
<td>0.27</td>
<td>0.16</td>
<td>-0.04</td>
<td>0.57</td>
<td></td>
<td>1.71</td>
</tr>
<tr>
<td>Himatanthus a.</td>
<td>Alpha</td>
<td>0.32</td>
<td>0.32</td>
<td>-0.31</td>
<td>0.95</td>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td>Himatanthus a.</td>
<td>Beta</td>
<td>-0.55</td>
<td>0.42</td>
<td>-1.38</td>
<td>0.28</td>
<td>-1.30</td>
<td></td>
</tr>
<tr>
<td>Inga g.</td>
<td>Alpha</td>
<td>-0.28</td>
<td>0.21</td>
<td>-0.70</td>
<td>0.14</td>
<td></td>
<td>-1.30</td>
</tr>
<tr>
<td>Inga g.</td>
<td>Beta</td>
<td>0.26</td>
<td>0.13</td>
<td>0.01</td>
<td>0.51</td>
<td>*</td>
<td>2.05</td>
</tr>
<tr>
<td>Inga sp</td>
<td>Alpha</td>
<td>-0.21</td>
<td>0.36</td>
<td>-0.91</td>
<td>0.49</td>
<td>-0.59</td>
<td></td>
</tr>
<tr>
<td>Inga sp</td>
<td>Beta</td>
<td>0.23</td>
<td>0.23</td>
<td>-0.23</td>
<td>0.68</td>
<td></td>
<td>0.99</td>
</tr>
</tbody>
</table>
### Table 4 – Continued from previous page

<table>
<thead>
<tr>
<th>PP</th>
<th>parameter</th>
<th>Estimate</th>
<th>S.E.</th>
<th>CI95.lo</th>
<th>CI95.hi</th>
<th>Ztest</th>
<th>Zval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Licania m.</td>
<td>Alpha</td>
<td>-0.26</td>
<td>0.22</td>
<td>-0.69</td>
<td>0.17</td>
<td>-1.19</td>
<td></td>
</tr>
<tr>
<td>Licania m.</td>
<td>Beta</td>
<td>0.27</td>
<td>0.14</td>
<td>-0.01</td>
<td>0.55</td>
<td>1.88</td>
<td></td>
</tr>
<tr>
<td>Matayba sp</td>
<td>Alpha</td>
<td>-0.05</td>
<td>0.19</td>
<td>-0.42</td>
<td>0.31</td>
<td>-0.27</td>
<td></td>
</tr>
<tr>
<td>Matayba sp</td>
<td>Beta</td>
<td>0.06</td>
<td>0.16</td>
<td>-0.25</td>
<td>0.37</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Morf sp2</td>
<td>Alpha</td>
<td>0.01</td>
<td>0.35</td>
<td>-0.67</td>
<td>0.70</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Morf sp2</td>
<td>Beta</td>
<td>-0.02</td>
<td>0.30</td>
<td>-0.61</td>
<td>0.57</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td>Morf sp4</td>
<td>Alpha</td>
<td>-0.42</td>
<td>0.26</td>
<td>-0.93</td>
<td>0.09</td>
<td>-1.60</td>
<td></td>
</tr>
<tr>
<td>Morf sp4</td>
<td>Beta</td>
<td>0.44</td>
<td>0.15</td>
<td>0.15</td>
<td>0.74</td>
<td>2.93</td>
<td></td>
</tr>
<tr>
<td>Ocotea s.</td>
<td>Alpha</td>
<td>-0.10</td>
<td>0.27</td>
<td>-0.64</td>
<td>0.43</td>
<td>-0.38</td>
<td></td>
</tr>
<tr>
<td>Ocotea s.</td>
<td>Beta</td>
<td>0.16</td>
<td>0.24</td>
<td>-0.31</td>
<td>0.63</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Pachira n.</td>
<td>Alpha</td>
<td>0.29</td>
<td>0.30</td>
<td>-0.31</td>
<td>0.88</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Pachira n.</td>
<td>Beta</td>
<td>-0.53</td>
<td>0.40</td>
<td>-1.32</td>
<td>0.26</td>
<td>-1.32</td>
<td></td>
</tr>
<tr>
<td>Petrea sp</td>
<td>Alpha</td>
<td>0.11</td>
<td>0.21</td>
<td>-0.30</td>
<td>0.53</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>Petrea sp</td>
<td>Beta</td>
<td>-0.15</td>
<td>0.19</td>
<td>-0.53</td>
<td>0.23</td>
<td>-0.79</td>
<td></td>
</tr>
<tr>
<td>Phanera g.</td>
<td>Alpha</td>
<td>0.00</td>
<td>0.29</td>
<td>-0.56</td>
<td>0.57</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Phanera g.</td>
<td>Beta</td>
<td>-0.00</td>
<td>0.26</td>
<td>-0.51</td>
<td>0.50</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Pouteria p.</td>
<td>Alpha</td>
<td>-0.45</td>
<td>0.42</td>
<td>-1.27</td>
<td>0.37</td>
<td>-1.07</td>
<td></td>
</tr>
<tr>
<td>Pouteria p.</td>
<td>Beta</td>
<td>0.41</td>
<td>0.22</td>
<td>-0.02</td>
<td>0.83</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td>Pouteria sp4</td>
<td>Alpha</td>
<td>-0.14</td>
<td>0.28</td>
<td>-0.68</td>
<td>0.41</td>
<td>-0.48</td>
<td></td>
</tr>
<tr>
<td>Pouteria sp4</td>
<td>Beta</td>
<td>0.13</td>
<td>0.21</td>
<td>-0.28</td>
<td>0.55</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Protium g.</td>
<td>Alpha</td>
<td>-0.11</td>
<td>0.14</td>
<td>-0.38</td>
<td>0.16</td>
<td>-0.80</td>
<td></td>
</tr>
<tr>
<td>Protium g.</td>
<td>Beta</td>
<td>0.14</td>
<td>0.10</td>
<td>-0.07</td>
<td>0.34</td>
<td>1.32</td>
<td></td>
</tr>
<tr>
<td>Pseudolmedia sp1</td>
<td>Alpha</td>
<td>-0.01</td>
<td>0.29</td>
<td>-0.58</td>
<td>0.56</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Pseudolmedia sp1</td>
<td>Beta</td>
<td>0.02</td>
<td>0.23</td>
<td>-0.44</td>
<td>0.48</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Rudgea c.</td>
<td>Alpha</td>
<td>-0.11</td>
<td>0.33</td>
<td>-0.75</td>
<td>0.53</td>
<td>-0.35</td>
<td></td>
</tr>
<tr>
<td>Rudgea c.</td>
<td>Beta</td>
<td>0.12</td>
<td>0.23</td>
<td>-0.33</td>
<td>0.56</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Sapium g.</td>
<td>Alpha</td>
<td>0.01</td>
<td>0.25</td>
<td>-0.47</td>
<td>0.50</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Sapium g.</td>
<td>Beta</td>
<td>-0.03</td>
<td>0.20</td>
<td>-0.42</td>
<td>0.37</td>
<td>-0.13</td>
<td></td>
</tr>
<tr>
<td>Senegalia m.</td>
<td>Alpha</td>
<td>-0.26</td>
<td>0.43</td>
<td>-1.11</td>
<td>0.59</td>
<td>-0.59</td>
<td></td>
</tr>
<tr>
<td>Senegalia m.</td>
<td>Beta</td>
<td>0.30</td>
<td>0.31</td>
<td>-0.31</td>
<td>0.91</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Siparuna g.</td>
<td>Alpha</td>
<td>-0.06</td>
<td>0.27</td>
<td>-0.59</td>
<td>0.47</td>
<td>-0.23</td>
<td></td>
</tr>
<tr>
<td>Siparuna g.</td>
<td>Beta</td>
<td>0.05</td>
<td>0.18</td>
<td>-0.30</td>
<td>0.40</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>

In the following figures we present the kernel and the final intensity estimated with the parametric model, for the 32 species with more than 10 plants inside the study area. These were the intensities used in the second order functions.
Figure 34. Kernel and estimated intensity for the species Arrabidaea sp

Figure 35. Kernel and estimated intensity for the species Astronium graveolens
Figure 36. Kernel and estimated intensity for the species *Attalea microcarpa*

Figure 37. Kernel and estimated intensity for the species *Bactris bidentula*
Figure 38. Kernel and estimated intensity for the species Bursera simaruba

Figure 39. Kernel and estimated intensity for the species Clathrotropis macrocarpa
Figure 40. Kernel and estimated intensity for the species Erythroxylum macrophyllum

Figure 41. Kernel and estimated intensity for the species Eschweilera tenuifolia
**Figure 42.** Kernel and estimated intensity for the species *Ficus americana*

**Figure 43.** Kernel and estimated intensity for the species *Fridericia pubescens*
Figure 44. Kernel and estimated intensity for the species Guarea glabra

Figure 45. Kernel and estimated intensity for the species Gustavia augusta
Figure 46. Kernel and estimated intensity for the species Heisteria acuminata

Figure 47. Kernel and estimated intensity for the species Himantanthus articulatus
Figure 48. Kernel and estimated intensity for the species Inga gracilifolia

Figure 49. Kernel and estimated intensity for the species Inga sp
Figure 50. Kernel and estimated intensity for the species Licania micrantha

Figure 51. Kernel and estimated intensity for the species Matayba sp
Figure 52. Kernel and estimated intensity for the species Morf sp2

Figure 53. Kernel and estimated intensity for the species Morf sp4
Figure 54. Kernel and estimated intensity for the species Ocotea schomburgkiana

Figure 55. Kernel and estimated intensity for the species Pachira nukakica
Figure 56. Kernel and estimated intensity for the species Petrea sp

Figure 57. Kernel and estimated intensity for the species Phanera guianensis
Figure 58. Kernel and estimated intensity for the species Pouteria plicata

Figure 59. Kernel and estimated intensity for the species Pouteria sp4
Figure 60. Kernel and estimated intensity for the species Protium guianense

Figure 61. Kernel and estimated intensity for the species Pseudolmedia sp1
Figure 62. Kernel and estimated intensity for the species Rudgea crassiloba

Figure 63. Kernel and estimated intensity for the species Sapium glandulosum
Figure 64. Kernel and estimated intensity for the species Senegalia macbridei

Figure 65. Kernel and estimated intensity for the species Siparuna guianensis