REPPLICATED SPATIAL POINT PROCESSES IN NEUROSCIENCE

Author: Jesús Pérez Rodríguez
Supervisor: Concha Bielza Lozoya
Supervisor: Pedro Larrañaga Mugica

July 2012
# Contents

1 Introduction ............................................................................. 11
   1.1 Case study basis ................................................................. 13
       1.1.1 Neuronal morphology ................................................ 13
       1.1.2 Confocal laser scanning microscopy .......................... 15

2 Theoretical background ............................................................. 17
   2.1 Definitions and notation ...................................................... 17
   2.2 Fundamental properties ..................................................... 19
   2.3 Distributional characteristics .............................................. 21
       2.3.1 Point intensity .......................................................... 21
       2.3.2 Inter-point interactions .............................................. 22
       2.3.3 Location density function ....................................... 25
   2.4 Point process models ......................................................... 26
       2.4.1 Homogeneous Poisson process ................................... 27
       2.4.2 Inhomogeneous Poisson process .............................. 28
       2.4.3 Gibbs processes ...................................................... 30
   2.5 Historical overview ........................................................... 32

3 Replicated point patterns .......................................................... 35
   3.1 Point count ...................................................................... 36
   3.2 Intensity analysis ............................................................. 37
       3.2.1 Kernel density .......................................................... 37
       3.2.2 Other methods ......................................................... 38
   3.3 Interaction analysis .......................................................... 39
       3.3.1 Aggregation recipes ................................................ 40
       3.3.2 Group differentiation .............................................. 41
   3.4 Model fitting ..................................................................... 43
       3.4.1 Ad-hoc inference ...................................................... 43
       3.4.2 Berman-Turner device ............................................. 44
   3.5 Additional issues ............................................................... 47


Acknowledgments

• To my supervisors, Concha Bielza Lozoya and Pedro Larrañaga, for their sincere trust and valuable advises during the whole execution of this Master’s Thesis.

• To Isabel Fernaud-Espinosa, Ruth Benavides-Piccione and Javier DeFelipe from the Laboratorio Cajal de Circuitos Corticales (CSIC), for all their contributions to this work.

• To Cajal Blue Brain Project Scholarship - C080020-09.

• To my colleagues from CIG and MUIA for their company and guidance.

• To my family and friends everywhere for their unconditional support.
Resumen

Los procesos espaciales puntuales son modelos matemáticos que describen la distribución y naturaleza de objetos que se encuentran localizados aleatoriamente en el espacio, formando patrones. El marco de trabajo es sencillo: unas coordenadas describen su localización espacial y marcas proporcionan información acerca de sus propiedades. En este contexto, es posible caracterizar estos patrones en base a la dependencia de los objetos de su posición, propiedades, posibles variables externas y especialmente, de la interacción con otros individuos. Este enfoque ha sido utilizado repetidamente en diversos campos, e.g. biología, física, astronomía, etc. y en la actualidad está experimentando un crecimiento importante.

Los procesos espaciales puntuales replicados se refieren a una situación particular en este campo en la que diversos patrones son considerados en el análisis como instancias de un mismo proceso. Este escenario ha sido muy poco tratado (apenas una docena de trabajos) y requiere la adaptación y el desarrollo de técnicas específicas cuyo uso está ya asentado en el caso de patrones simples. Motivados por la escasez de trabajos en este área, el objetivo de esta Tesis de Máster es doble: (i) proporcionar una introducción didáctica a las profundas bases teóricas de los procesos espaciales puntuales, junto con un exhaustivo análisis del estado del arte en patrones replicados; (ii) aplicar algunas de estas técnicas a un problema real relacionado con neurociencia: la distribución de espinas dendríticas sobre la superficie de la dendrita.

Las espinas dendríticas son protuberancias que se encuentran distribuidas a lo largo de la membrana de las dendritas pertenecientes a algunos tipos de neuronas. Es sabido que su forma y distribución puede variar con el tiempo y está relacionada con los procesos del aprendizaje y la memoria. De hecho, su pérdida o alteración ha sido descrita en la patogénesis de algunos desórdenes neurológicos como el Alzheimer. Por esta razón, se ha llevado a cabo un análisis de su distribución espacial utilizando patrones puntuales replicados. El principal objetivo es dar soporte formal a alguna de estas hipótesis, así como contribuir al conocimiento y el empleo de las técnicas de estadística espacial que nos ocupan.
Abstract

Spatial point processes could be defined as mathematical models that describe the arrangement of objects that are irregularly or randomly distributed in space forming patterns. The framework is simple: point coordinates describe the spatial location and marks provide additional information about their properties. In this context, it is possible to characterize these patterns in terms of the dependency of the points on their position, marks, covariates and, specially, interaction with other individuals. This approach has been extensively used in several application fields, e.g. biology, physics, astronomy, etc. and it is currently an important line of research.

Replicated spatial point processes refer to a particular situation on this field, in which several patterns are considered in the analysis as independent instances of the same process. This scenario has not been usually considered (about a dozen published works), and requires the adaptation of techniques the use of which is already long established in the single pattern case, as well as the development of novel methods that take advantage of data availability. Motivated by this shortage of works, the main goal of this Master’s Thesis is twofold: (i) provide a didactic introduction to theoretical foundations of spatial point processes, along with a comprehensive review on pattern replication; (ii) apply some of these techniques to a real case study related to neuroscience: the distribution of dendritic spines over dendrite surface.

Dendritic spines are small protrusions situated throughout the membrane of dendrites belonging to some specific kinds of neurons. It is known that these undergo both remodelling and rearrangements through time and that they are related with learning and memory formation. In fact, loss or alteration of these structures has been described in the pathogenesis of major neurological disorders such as Alzheimer’s disease. For this reason, an analysis of their spatial distribution using replicated point patterns has been conducted. The main goal is to provide formal support for these hypothesis, as well as to contribute to the extension and use of the spatial statistics techniques we are dealing with.
Chapter 1

Introduction

Spatial point processes could be defined as mathematical models that describe the arrangement of objects that are irregularly or randomly distributed in space. These elements are naturally represented by point coordinates of an arbitrary dimensionality, describing their spatial location, and marks, providing additional information about their properties (e.g., class). Each whole pattern of objects is then considered a sample of the random phenomenon which is required to model.

Based on the described framework, spatial point process statistics provides the necessary tools to characterize patterns, in terms of the number, distribution and nature of the objects. The main objective is to obtain reliable statistical conclusions about the dependence of these random variables on location, covariates and marks. This is generally performed by analysing two main aspects: average point density and inter-point interactions; in fact, the latter could be considered the main characteristic that makes spatial point processes differ fundamentally from classical statistics. We usually assume that some features, e.g. nearest neighbor distances or marks attached to close points, may be highly correlated; hence the statistical analysis is strongly concerned with detecting and describing these relations.

Although spatial point process statistics is a relatively unknown methodology among data mining community, it is one of the most developed branches of spatial statistics. The simple representation, as well as abovementioned ability to effectively capture complex correlations among elements, have produced an increasing interest on applying these methods in many different fields. Historically, this topic has been more related to applications in which data collection used to be a costly manual process (e.g. forestry), implying that the availability of several independent samples as realisations of the same underlying process was not usually considered. This scenario is known as point patterns replication, and requires the adaptation of techniques the use of which is already long established in the single pattern case, as well as the development of novel methods that take advantage of data availability.
Neuroscience is one of those fields which find in spatial point processes a useful modelling tool. In addition, recent technical advances on applied microscopy allow now the use of large databases regarding the nature and spatial distribution of some of the neuronal elements that make up the brain; consequently, the availability of replicated data is becoming increasingly usual in this particular field. However, there are only a few works in the literature dealing with replicated neurobiological data; furthermore, most of the contributions on replicated point patterns use the same benchmark datasets and are quite dispersed over the literature.

The abovementioned context motivates the development of this work, the aim of which is twofold. First, we give a comprehensive survey on those works within the field of spatial point processes which have recently arisen about replicated spatial point patterns. Second, some of those techniques are then applied to a real case study related to neuroscience, in collaboration with the Laboratorio Cajal de Circuitos Corticales (CSIC). More precisely, the statistical analysis of human dendritic spine distribution over dendrite surface.

Thus, the main contributions of this document are the following. On one hand, although the effectiveness of this methodology has already been proven, we aim to go further into the unexplored scenario of point patterns replication; thus, we provide an exhaustive compilation of techniques and put additional attention to some problems which are yet to be solved (such as the current shortage of validation methods). The long-term goal is to find a way of bringing spatial point processes closer to the field of machine learning, in order to adapt existing algorithms to deal with the spatial distribution of points.

On the other hand, the case study offers a practical justification for the employment of this methodology. Dendritic spines are known to be critical in learning, memory and cognition; moreover, loss or alteration of these structures has been described in major neurological disorders, including schizophrenia and Alzheimer’s disease. For those reasons, a statistical study of their spatial distribution could provide formal support for these hypotheses.

As mentioned above, there are only a few publications in the literature dealing with replicated neurobiological data and spatial point process statistics; those could be seen as an inspiration for this work and will be reviewed in the next chapter. In what concerns particularly to dendritic spines studies using these techniques, a recent preliminary analysis has concluded that, considering isolated sections of each dendrite, spines appear to be independently and uniformly distributed over dendrite surface. However, a simple visual inspection as well as domain specific knowledge suggest that spine density varies along dendrite longitudinal axis; moreover, it seems reasonable to think that spatial distribution could be related in some way to spine interaction. This work aims to extend abovementioned study, considering each whole dendrite as an independent realisation of the same spatial point process.
Finally, all the necessary code has been implemented in R language, as an extension for the extensively used `spatstat` package from [6]. This library provides a powerful framework for performing spatial point pattern analysis but currently lacks of explicit support for replicated samples. Thus, taking into account that this scenario is being increasingly usual, the implementation itself could be considered an additional contribution of this work.

The contents of this thesis are organized as follows: the rest of the present chapter introduces the fundamentals on dendritic spines, which are necessary to understand the work; chapter 2 provides a deep background to the theory of spatial point processes along with a brief historical overview; chapter 3 goes deeper into pattern replication works with a full review; in chapter 4, all methodological issues concerning those techniques applied in the case study are further explained; chapter 5 contains numerical and graphical outcomes resulting from the spine analysis and, finally, in chapter 6, some conclusions are presented, along with a discussion about the contributions of the whole work.

1.1 Case study basis

In the following, we present a brief introduction to the fundamentals of neuron physiology and neuronal dendritic spines, which are necessary to understand the motivation and scope of this work. Related to this topic is that of confocal laser microscopy, which is the sampling technique that is mostly employed in this kind of scenarios. Since it has been used as well for data collection in the case study, basic concepts about this methodology are also provided.

1.1.1 Neuronal morphology

Neurons are electrically excitatory cells that process and transmit information by electrical and chemical signaling; these are the core components of the nervous system, including the brain, the spinal cord and the peripheral ganglia. Several kinds of specialized neurons exist –sensory neurons, motor neurons, interneurons, etc.– though in this case study we are specially interested in the particular class of Purkinje cells, located in the cerebellar cortex.

The cellular organization of neurons is similar to that of other cells; however, they are characterized by their specialization in intercellular communication. This affects their general morphology, the components that made up their membrane –for improving electrical signaling–, and the structural and functional properties of the synaptic contacts among neurons. The typical neuron is composed by the following elements (Figure 1.1, Left): (i) the body or soma, a compact structure that contains the cell nucleus; (ii) the axon, a special cellular extension that arises from the soma
at the axon hillock and extends up to one meter in humans or even more in other species; and (iii) dendrites, thin structures that arise from the soma, often extending for hundreds of micrometres and branching multiple times, giving rise to a complex “dendritic tree”. The elaborate arborization of dendrites is the most obvious sign of neuronal specialization for communication via electrical signaling.

Figure 1.1: Left: Schematic representation of the typical neuron, showing the soma, the axon and the dendritic tree. The figure is not at real scale; the axon often extends much further to benefit intercellular communication. Right: three-dimensional reconstruction of an apical dendritic segment showing three spines. Figure from [3].

In general, dendrites are the primary target for synaptic input from other neurons whereas the axon conducts electrical impulses away from the neuron’s soma. To aid in the task of electrochemical signaling, axon terminal contains synapses: specialized structures where neurotransmitter chemicals are released to communicate with target neurons. In what concerns to dendrites, those belonging to some specific kinds of neurons (e.g., pyramidal cells, the major neuronal type of the cerebral cortex) contain small membranous protrusions, referred to as “spines”, which represent the target of most excitatory synapses of an axon.

**Nuronal dendritic spines**

Dendritic spines are protrusions that act to conduct the stimulation received from other neural cells to the cell body of the neuron from which dendrites project. Dendrites of a single neuron can contain hundreds to thousands of spines; these provide an anatomical substrate for memory storage and synaptic transmission, as well as serve to increase the number of possible contacts between neurons. In general, most spines have a bulbous head and a thin neck that connects the head of the spine to the shaft of the dendrite (Figure 1.1, Right).
1.1. Case study basis

Despite of this characterization, spines are highly motile and can undergo remodeling even in the adult nervous system; morphological rearrangements have been found in vitro and in vivo and appears to be associated in some cases with developmental or behavioral plasticity [3]. Thus, these protrusions are activity-dependent processes and are known to be related to learning and memory formation; in fact, loss or alteration of these structures has been described in the pathogenesis of major neurological disorders, including schizophrenia and Alzheimer’s disease [38]. For those reasons, a statistical study of the number of spines and their spatial distribution could be interesting to provide formal support for these and other hypotheses.

1.1.2 Confocal laser scanning microscopy

Confocal laser scanning microscopy (CLSM) is an extensively used technique for obtaining high-resolution optical images with depth selectivity, i.e. its key feature is the ability to acquire in-focus images from selected depths (optical sectioning). Briefly, depth selection is attained with the variation of microscope pinhole size; this allows to exclusively image a thin optical slice out of a thick specimen (typically, up to 100 nm). For a given depth, each image is acquired in three main steps: (i) point-by-point scanning of the specimen with a focused laser beam, (ii) detection of the scattered and reflected laser light by means of a photomultiplier tube (PMT) and (iii) digitalisation of the object information contained in the electrical signal provided by the PMT. In biological applications especially, the specimen may be injected with some fluorescent material, so that light from each illuminated spot can be more easily collected; this usually implies filtering fluorescent wavelengths while blocking the original excitation wavelengths.

In either case, in addition to the possibility of observing a single slice of a volume in good contrast (Figure 1.2), optical sectioning allows a great number of slices to be cut and recorded at different planes; this is generally done moving the specimen along the optical axis by controlled increments. Depending on the thickness of the slice and the spacing between successive planes, different quality levels may be obtained. The overall result is a three-dimensional dataset which provides information about the spatial structure of the specimen; in this context, various aspects of the object can be manually identified (e.g. dendritic spines) which, in last instance, represent the raw data for our statistical analysis.
Figure 1.2: Confocal microscope image obtained from a basal dendrite segment at a specific depth level. The original image size is $76.8 \times 76.8 \, \mu m$. Figure from [13].
Chapter 2

Theoretical background

This chapter is a complete introduction to the notation and extensive theory of spatial point processes. Its purpose is not to serve as a deep dissertation about the field, but to provide the basic knowledge for the proper understanding of the state of the art and applied techniques. For those readers interested on going further into this topic, the available literature is quite rich. First, a good general introduction to the topic can be found in [47]; this work represents the main source of information that has been used for the development of this chapter. Nevertheless, for the comprehension of more complex theoretical basis we suggest [31, 32, 53, 56], among others. Finally, a more practical-oriented explanation can be found in [8], along with several case studies providing a good reference for the application of these techniques. In addition, the latter work also includes a useful introduction to the use of spatstat.

2.1 Definitions and notation

A point pattern is defined as a collection of points located in some definite area or subset \( W \) of an abstract space \( S \); in this context, each pattern is interpreted as a sample or realisation of a stochastic model which is known as point process. Additional information can be also attached to those points in the form of categorical or numerical properties in a set \( M \), termed marks. This is the special case of marked point processes which will not be deeply explored in this work but represent an even more powerful modelling tool. In either case, point patterns are assumed to fulfill the simplicity property, i.e., there is not more than one point lying in the same location. Figure 2.1 shows a well-known example of a marked point pattern, which has been extensively analysed.

In general, the term “process” is employed in accordance with classical statistics literature; it usually implies that some development over time is considered, but this is not necessarily the case and we will restrict ourselves to scenarios in which
Figure 2.1: The “amacrine cells” pattern, composed of 294 points with a categorical mark attached to them; it represents the location of cells with two different levels of light sensitivity, in planar projection within a 1060 by 662 µm section of the retina.

\[ S \equiv \mathbb{R}^d, d \leq 3. \] Consequently, there is no meaningful order of the points inside the pattern and this should be also taken into account during the statistical analysis; as we will see, this restriction implies that no directional dependence could be considered among points, which turns out to complicate modelling interaction. The nature of the sampling area \( W \) is not such an obvious matter and has a great impact on the statistical analysis; this issue will be discussed in the next section.

For notational clarity, we will represent individual locations, points and marks with lower case letters (like \( x \)); point patterns using lower case bold letters (like \( \mathbf{x} \)); space and mark sets with upper case letters (like \( W \)), and point processes using upper case bold letters (like \( \mathbf{N} \)). In what concerns to notation, it is also necessary to clarify the difference of the terms we will use to refer to pattern objects (commonly, “events” or “points”), from those we will employ for sites in \( \mathbb{R}^n \) which may not necessary belong to the pattern itself (“positions” or “locations”). Within this framework, a point pattern could then be formally defined as a set of points made up of a location and possibly a vector of property values of the form: \( \mathbf{x} \equiv \{(x_i, m_i)\}_{i=1}^n, x_i \in W, m_i \in M. \) In this scenario, two mathematical definitions of a point process are considered. All through the document, we will use indistinguishably both; the correct interpretation should be clear from context.

1. \( \mathbf{N} \) may be interpreted as a random counting measure, \( \mathbf{N}(A) \in \mathbb{N}, \forall A \subseteq W, \) which determines the number of events in any subspace \( A \subseteq W. \) The function is assumed to be locally finite, i.e. \( \mathbf{N}(A) < \infty, \forall A \neq \mathbb{R}^n; \) moreover, regarded as a function, it must have the fundamental property of additivity, i.e. \( \mathbf{N}(B_1 \cup B_2) = \mathbf{N}(B_1) + \mathbf{N}(B_2), \) for all disjoint \( B_1, B_2 \subseteq W. \)
2.2. Fundamental properties

2. \(\mathbf{N}\) may also denote a random set made up of all possible patterns in the configuration space of the process, \(\mathbf{N} \equiv \{\emptyset, \{x_1\}, \{x_1, x_2\}, \ldots\}, x_i \in W\), and so we accept as valid expressions such as \(\mathbf{x} \in \mathbf{N}\). For ease of use, we will employ the set notation \(\mathbf{x} = \{x_1, ..., x_n\}\) to refer to any of these possible configurations; the inclusion of the empty pattern should be considered implicit.

Finally, in this work we will address in depth the scenario of replicated patterns; this implies the existence of several samples as i.i.d. realisations of the same process. Furthermore, in some cases, replicates will be divided into groups which, in some sense, are interesting from a statistical point of view. In those cases, we will use the following notation: \(P\) is the whole set of replicated samples with \(|P| = m\) and a partition of this set into \(g\) disjoint groups \(\{G_i\}_{i=1}^{g}\) is considered, where the number of replicates in the \(i\)th group is \(|G_i| = m_i\). In this context, \(x_{ij}\) will represent the \(j\)th pattern within the \(i\)th group, and its number of points is denoted by \(|x_{ij}| = n_{ij}\), i.e. \(x_{ij} = \{x_{ijk}\}_{k=1}^{n_{ij}}\). Finally, we will consider \(n_i = \sum_{j=1}^{m_i} n_{ij}\) the number of points in the \(i\)th group, and \(n = \sum_{i=1}^{m} n_i\) the total number of points.

2.2 Fundamental properties

We have already introduced that the sampling area \(W\) usually depends on the nature of the problem to solve. Sometimes, \(W\) is naturally imposed by the physical limits of the environment in which the objects exist (e.g. the dendritic spine distribution over dendrite surface); thus, boundaries have some influence on the geometry of the point pattern and may attract or repulse points. In those cases, the process is considered a “finite point process” and the effect of the sampling area, which is known as existence window, should not be eliminated. In the real world, all point patterns should be considered finite, in the sense that both the number of points and the extension of the sampling area are measurable quantities; however, some point patterns may be regarded as part of much larger structures in which points are distributed according to the same laws (e.g. positions of stars in a region of the galaxy). Those processes are considered “infinite point processes” and the sampling area, which is known as observation window, is arbitrarily chosen following some criterion (Figure 2.2).

When dealing with infinite patterns, points close to the edge of the observation window should not be considered to behave differently from those in the centre. However, a lot of measurements which are of common use in point process statistics (e.g. counting the number of points within a distance of a given location) could be severely biased in those regions close to the boundary. To overcome this limitation, edge-correction techniques are usually applied throughout the whole statistical analysis. Briefly, these methods aim to remove the influence of the window using varied mechanisms, e.g. considering only those locations within a distance from the edge,
Figure 2.2: Arbitrary chosen observation window for a given infinite point pattern; only dark points belong to the pattern though light ones could be considered to account for edge-effects. Figure from [4].

giving larger weights to unbiased measurements or enlarging the pattern by periodic continuation (Figure 2.3). Although the case study we present in this work seemed a clear exemplar of a finite process due to dendrite natural boundaries, we would need to use an approximation to the latter method during the analysis; this happens due to the toroidal nature of the sampling window (dendritic spine surface).

Figure 2.3: A point pattern in a rectangular window \( W \) and its periodic continuation (red points). Those measurements performed close to the boundaries consider points in the opposite extreme region of the window.

In addition to finiteness, there are two more properties which are essential: stationarity and isotropy. A process is formally considered stationary if the overall point distribution is invariant to pattern translations; analogously, a process is isotropic if this happens independently of rotations around the origin. In practical terms, these so-called motion-invariance properties imply that the expectation of observing some point configuration is independent of the particular location, in or outside the sampling area. This fact could seem related to finiteness, since we do not expect a pattern which is part of some infinite structure to be non-stationary; however, these are independent properties and should be analysed separately. In general, patterns are expected to exhibit stationary and isotropic behaviors in one or more of its dis-
tributional properties, commonly inter-point interaction. Finally, it is necessary to specify that throughout this work, all point processes are assumed to fulfill a property referred to as *simplicity*, which means that all points are different and do not coincide, i.e. \( x_i \neq x_j \iff i \neq j \). Note that the theory of point processes also considers models with multiple points lying on the same location.

### 2.3 Distributional characteristics

As we mentioned before, point process statistics aims to characterize patterns distribution in terms of its dependence on location, covariates and marks. In practical terms, this usually implies analysing two distributional properties of the process: *point intensity* and *inter-point interactions*. Thus, we mainly aim to discover existing correlations between abovementioned variables and these properties; then, an elegant mechanism is usually employed to summarise and build a model based on this information: the *location density function*.

#### 2.3.1 Point intensity

Point intensity is the simplest distributional property. It is defined as the average number of points per unit area or volume and is usually denoted with the letter \( \lambda \). This characteristic resembles the use of the sample mean in classical statistics, and relates to our random counting measure definition of a spatial point process \( N \) as follows. As we have seen above, the value \( N(A) \in \mathbb{N}, \forall A \subseteq W \) is a random variable. It seems clear that the expected number of points in a particular region, \( E(\mathbb{N}(A)) \), may depend on the specific characteristics of that region; thus we could define a deterministic function operating on sets \( \Lambda(A) = E(\mathbb{N}(A)) \), what is called *intensity measure*—more formally, *first order moment measure*, \( \alpha^1(A) \). Under continuity conditions, which are usually satisfied in practical applications, there exists a function \( \lambda(x), x \in W \) such that:

\[
\Lambda(A) = E(\mathbb{N}(A)) = \int_A \lambda(x)dx, \forall A \subseteq W
\]  

(2.1)

It is clear that \( \lambda(x) \) is proportional to the point density around a location \( x \). In probabilistic terms, this loosely implies that if \( dx \) is the volume of an infinitesimal sphere centred at \( x \), then \( \lambda(x)dx \) is the probability that there is a point in the sphere. This particular function is called *intensity function* and is one of the most studied summary characteristics in the statistical analysis of spatial point patterns. An alternative formulation which is usually required in practice is the *(Papangelou)* *conditional intensity*. It is defined as \( \lambda(u|x), \forall u \in W \), where \( u \) is a deterministic
location and \( x \) is a point pattern. This can be understood in the context of the probabilistic interpretation introduced above: \( \lambda(u|x)du \) is the conditional probability that there is a point in an infinitesimal sphere centred at \( u \), given the realisation of \( x \) outside of the sphere. This is an important tool for some point process models where explicit formulas for \( \lambda(u) \) cannot be given, since \( E(\lambda(u|x)) = \lambda(u), \forall u \in W \), and so it could serve as a proper estimator.

In either case, based on its nature, intensity will be considered “uniform” or “homogeneous”, if \( \lambda(x) \) is constant, or “non-uniform” or “inhomogeneous” if it varies from location to location (Figure 2.4). This variation is sometimes correlated with covariates and marks. While covariates are often functions defined at every possible location in the window (e.g., terrain altitude influencing the abundance of trees), marked point patterns may exhibit some kind of regional segregation among points with different mark values (e.g., competing plant species being separated). In practical terms, modelling these dependences is one of the main goals of point process statistics. However, at this point, the introduced characterization is enough to understand the rest of the document; more details about the practical estimation of the intensity function and the purpose of the Papangelou conditional intensity will be given in the following sections.

Figure 2.4: Comparison between two simulated spatial point patterns, one approximately uniform (left) and the other with an inhomogeneous intensity function (right). Figure from [4].

### 2.3.2 Inter-point interactions

Despite of the importance of average point density, as we have mentioned above, point process statistics focuses on analysing existing correlations among points, that is, inter-point interactions. This stochastic dependence is related to distances be-
tween points and expected to be stronger for those which are close to each other. In that sense, correlation among point locations is the simplest aspect which can be studied (Figure 2.5). On one hand, patterns are considered to exhibit "clustering" (also "aggregation" or "positive interaction"), if points tend to be closer than should be for a given intensity. On the opposite side, "regular" patterns show "repulsion" (also "inhibition" or "negative interaction"), if inter-point distances are larger than expected. Otherwise, locations are considered to be independent; a formal definition of independence will be introduced in the next section.

![Figure 2.5: A comparison between three simulated point patterns with homogeneous overall intensity and different interaction behaviours: independence (left), inhibition (center) and clustering (right). Figure from [4].](image)

In either case, it is not enough to conclude with the existence of interactions, but also necessary to assess the spatial scale at which these operate. This implies that the same process could produce patterns that exhibit both clustering and inhibition properties, but operating at different scales. For instance, it is known that celestial objects are organized into large and dense structures (galaxies, nebulae, etc.), and so may be considered to form clustered patterns; however, these massive objects also repulse each other and so are not so close as should be under independence. As a consequence, if the observation window is large enough, both behaviours will be shown in the same process. Moreover, if long distances are considered, clustering can be easily misunderstood with the existence of high intensity regions. This shortcoming can be avoided assuming that: (i) long-range variations are often associated with the intensity function, while interaction analysis is focused on short-range effects; (ii) the intensity function may be considered constant across samples of the same process, and so inter-point interactions are assumed to be responsible of pattern specific artifacts. This will be studied in the context of replication.

More complex scenarios not only consider abovementioned correlations, but also interactions among marks. For instance, in the case of qualitative marks, this usually implies the existence of some dependence among points of different classes at certain
scale (e.g. cooperating plant species appearing clustered). Nevertheless, this particular situation is not being considered in this work and our analysis will be restricted to the correlation among point locations and how it varies, between and within groups of replicates of the same process.

In general, interactions are not so uniquely formalized as the variation of the average point density with the intensity function. In practical terms, non-parametric analysis are based on several summary characteristics where inter-point distances are considered. Briefly, the following functions are generally considered in practical applications. Here only theoretical definitions are provided, estimation issues will be addressed in chapter 4.

- **Empty-space function**, $F(r)$. Defined for homogeneous processes as the cumulative distribution function of the empty space distance at the typical location $u$, $F(r) = P\{d(u, x) \leq r\}$, with $u \in W$ and $d(u, x) = \min\{||u - x_i|| : x_i \in x\}$. As far as we know, it was first suggested in [67].

- **Nearest-neighbor distance function**, $G(r)$. Defined for homogeneous processes as the cumulative distribution function of the nearest-neighbor distance at the typical point $o$, $G(r) = P\{d(o, x \{o\}) \leq r\}$, with $o \in x$. As far as we know, it was first proposed in [28].

- **J-function**, $J(r)$. Recently defined in [77] as $J(r) = (1 - G(r))/(1 - F(r))$; it is claimed to have better readability and statistical properties than the previous ones, mainly due to the theoretical reference model value, which is 1.

- **Ripley’s K-function**, $K(r)$. This function was introduced in [14] for homogeneous processes and extended to heterogeneous environments in [12]. It is defined as the expected number of other points of the process within a distance $r$ of the typical point $o$, scaled to remove the effect of the overall intensity, $K(r) = \lambda^{-1}E\{N(B(o, r) \setminus o)\}$; here, $b(o, r)$ is the sphere of radius $r$ centered at $o$. Technically, it is defined based on the second order factorial moment measure, $\alpha^2(B_1 \times B_2) = E(N(B_1)N(B_2))$, for disjoint $B_1, B_2 \subseteq W$; thus, it is related to abovementioned intensity measure $\lambda(A) = \alpha^1(A) = E(N(A))$. For more details, we refer to [56] p. 29.

- **Besag’s L-function**, $L(r)$. A transformation of the $K$-function defined in [22] as $L(r) = \sqrt{K(r)/\pi}$, which is easier to interpret because of its theoretical value under the reference model and has better statistical properties (smaller asymptotic variance).

- **Pair-correlation function**, $g(r)$. Another transformation of the $K$-function defined as $g(r) = K(r)/(2\pi r)$, but which appeared previously in [59]. Roughly
speaking, the pair-correlation function is the probability of observing a pair of points separated by a distance $r$, and for this reason, it is easier to interpret.

Models are sometimes characterized so that theoretical values for these characteristics are available, and so departures from those may be properly assessed. Authors do not often consider a unique functional summary characteristic to be superior to the rest, though Ripley’s K-function and its derivations have been extensively used and are usually selected as the preferable choice.

### 2.3.3 Location density function

In general, infinite point processes are defined in a quite abstract way which is similar to the theory of time series. On the other hand, finite point patterns can be also regarded as samples from a given multivariate distribution in which each variable represents a point in the pattern. Classical statisticians are more familiar with this distributional approach and so it is often used in all scenarios; within this framework, a point process in a window $W$ can be defined using two elements:

- A *discrete probability distribution*, $p_n = P(N(W) = n)$, which determines the random number of points in the pattern. In some scenarios, this value may be considered deterministic, though this is not the usual case.

- A family of *multivariate density functions* $f_n(x_1,\ldots,x_n)$ which are symmetric in their arguments; this means that independently of the order of the vector, functions take on the same value (there is no meaningful order).

In this context, $f_n$ is called *location density function* and its probabilistic interpretation is straightforward. Under the constraint that $n = N(W)$, and considering $n$ infinitesimal spheres of volumes $dx_1,\ldots,dx_n$ centred on locations $x_1,\ldots,x_n$, the probability of finding one point in each of them is: $f_n(x_1,\ldots,x_n)dx_1,\ldots,dx_n$. If the point count condition is relaxed, then the probability must be multiplied by $(p_nn!)$ because the number of points is random, and all permutations of the $n$ points yield the same value of the function. Thus, the density function $L(x) = p_n!f_n(x_1,\ldots,x_n)$ is considered the *likelihood* of a process defined this way and plays a major role in the estimation of model parameters.

To avoid fixing the number of points, an alternative formulation of this probability density is sometimes used, $p(x)$ for $x = \{x_1,\ldots,x_n\} \in N$. The function is defined with respect to the distribution of the Poisson process of unit intensity $\Pi$, that is, $\int_N p(x)\Pi(dx) = 1$. If we condition the number of points in the pattern to $N(W) = n$, the relation between $p(x)$ and the location intensity function is:

$$f_n(x_1,\ldots,x_n) = \frac{p(\{x_1,\ldots,x_n\})}{\int_W \ldots \int_W p(\{x_1,\ldots,x_n\})dx_1\ldots dx_n} \quad (2.2)$$
2.4 Point process models

Practical techniques for spatial point patterns will be treated in depth in chapter 4; however, as in all areas of statistics, it is known that the study of a point process is mainly composed of three complementary stages: exploratory analysis, modelling and validation. In particular, models are essential because of their benefits in three directions: (i) understanding empirical patterns and their summary characteristics, (ii) giving information about their possible stochastic fluctuations and (iii) providing unique frameworks for their simulation.

The most important models for point processes are Poisson processes, which are considered reference models for assessing the behaviour of the distributional properties we have introduced above. These processes have been extensively studied and characterized, as they represent the basis for complex models, which are more often constructively defined. In general, this means that one or more of the following formally defined operations are applied to simple models in order to derive more sophisticated ones:

- **Superposition**: A superposition operation superimposes several point processes \( \{N^i\}_{i=1}^m \) onto each other such that the result is made up of their set-theoretic union \( N = \cup_{i=1}^m N^i \). This general operation is applied in several scenarios.

- **Thinning**: A thinning operation uses a specific rule determining which points in a basic process \( N_I \) are retained. The resulting random set \( N_F \) fulfills \( N_F \subseteq N_I \). This operation is usual in the definition of regular processes, e.g. hard-core processes: points within a given distance \( r \) to their nearest neighbor are deleted (Figure 2.6).

- **Clustering**: In a clustering operation, every point \( x \) in a given process \( N_I \) is replaced by a cluster of points \( N^x_F \). The superposition of these clusters yields patterns belonging to the target process \( N_F \). As it is expected, this operation is often used to define aggregated processes (Figure 2.7).

A great amount of processes follow this constructive definition (e.g., Neyman-Scott processes, Hard-core processes, etc.), and most of them have also been formally characterized. That means theoretical forms for their summary characteristics are available to aid in the task of model selection. However, the main drawback of these approaches is that are usually aimed to model a specific behaviour and lack of a proper generalization. For this reason, in this work, we will focus our attention on two general and extensively used families: (i) Poisson processes (both homogeneous and inhomogeneous), which play an essential role in this field as null models, and (ii) Gibbs processes, which are general models able to reflect inter-point interactions in an elegant manner, that is quite familiar to classical statisticians.
2.4. Point process models

Figure 2.6: A thinning operation from a Poisson process in the unit square with intensity $\lambda = 200$ (left), to a hard-core process with minimum distance among points $r_m = 0.025$ (right). Deleted points are shown in red.

Figure 2.7: A clustering operation from a Poisson process in the unit square with intensity $\lambda = 10$ (left), to a clustered process with $n_c = 25$ average points per cluster uniformly distributed in the disc of radius $r_c = 0.1$ (right).

2.4.1 Homogeneous Poisson process

The homogeneous Poisson process is considered the null model in spatial point process statistics; that is because it is formally characterized by two fundamental properties, each of them serving as a reference for one of the distributional characteristics that have been introduced before:

- **P1. Poisson distribution of counts.** Given any bounded region $A \subseteq W$, $N(A)$ follows a Poisson distribution with mean $\lambda v(A)$, for some constant $\lambda$, being $v(A)$ the area or volume of the region.

- **P2. Independent scattering.** For any number of disjoint regions, $\{A_i\}_{i=1}^k$, $A_i \subseteq W$, resulting count values $N(A_1), \ldots, N(A_k)$ are independent random variables.
On one side, \( P1 \) clearly requires that the expected number of points in an arbitrary set \( \Lambda(A), A \subseteq W \) does not depend on the specific region; this implies the intensity function is constant with \( \lambda(x) = \lambda, \forall x \in W \) and the process must be considered “homogeneous”. On the other side, \( P2 \) does not allow point counts in disjoint regions to be correlated in any manner; thus interactions cannot exist and point locations must be considered “independent”. In addition, stationarity and isotropy are further implications of these properties. In practical terms, \( \lambda \) is the unique characteristic parameter of this model and is assumed to be positive and finite. More precisely, \( p_n \) is a discrete Poisson distribution with mean \( \lambda v(W) \) and points are i.i.d. following a uniform distribution throughout the whole existence window; thus, it can be easily shown that the resulting location density function is:

\[
f_n(x_1, \ldots, x_n) = \frac{\lambda^n}{\lambda v(W)} \quad \text{for } x_i \in W, \tag{2.3}
\]

\[
p\{x_1, \ldots, x_n\} = \exp \left( (1 - \lambda)v(W) \right) \lambda^{N(x)} \tag{2.4}
\]

Thus, the likelihood of \( \lambda \) for the homogeneous Poisson process is:

\[
L(\lambda|x) = p_n(x)f_n(x) = \lambda^{N(x)}\exp (-\lambda v(W)) \tag{2.5}
\]

In summary, the homogeneous Poisson process is considered to exhibit what is called Complete Spatial Randomness (CSR), and is used as the reference model in point process statistics. Testing this CSR hypothesis is an essential stage of analysis, since acceptance implies there is no need to consider more complicated models and that it is not possible to find indicators of interaction among points. To aid in this task, the model has been studied in depth, providing theoretical forms for summary characteristics. This information is summarized in Table 2.1.

### 2.4.2 Inhomogeneous Poisson process

The homogeneous Poisson process may be generalised in a straightforward way by introducing inhomogeneity to yield the inhomogeneous Poisson process. That means the value of the intensity function is no longer constant; in terms of the introduced characterization, the generalization affects to \( P1 \) property, which is replaced with:

- \( P3 \). Poisson distribution of counts. Given any bounded region \( A \subseteq W \), \( N(A) \) follows a Poisson distribution with mean \( \Lambda(A) = \int_A \lambda(x)dx \), for some function \( \lambda(x), \forall x \in W \).

Since the expected number of points in any region is not constant, stationarity and isotropy properties are no longer guaranteed. Nevertheless, \( P2 \) remains unchanged
2.4. Point process models

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CSR</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensity, $\lambda$</td>
<td>Constant</td>
<td></td>
</tr>
<tr>
<td>Empty-space function, $F(r)$</td>
<td>$1 - \exp(-\lambda b_d r^d)$</td>
<td>$&lt;$</td>
</tr>
<tr>
<td>Nearest-neighbor distance function, $G(r)$</td>
<td>$1 - \exp(-\lambda b_d r^d)$</td>
<td>$&gt;$</td>
</tr>
<tr>
<td>$J$-function, $J(r)$</td>
<td>$1$</td>
<td>$&lt;$</td>
</tr>
<tr>
<td>Ripley’s $K$-function, $K(r)$</td>
<td>$b_d r^d$</td>
<td>$&gt;$</td>
</tr>
<tr>
<td>Besag’s $L$-function, $L(r)$</td>
<td>$r$</td>
<td>$&gt;$</td>
</tr>
<tr>
<td>Pair-correlation function, $g(r)$</td>
<td>$1$</td>
<td>$&gt;$</td>
</tr>
</tbody>
</table>

Table 2.1: Values of the most commonly used summary statistics for CSR. The third column shows the expected relationship with respect to Poisson theoretical values for patterns exhibiting clustering. The opposite is expected for regular patterns. Here, $b_d$ depends on the dimensionality: $2$ ($d = 1$), $\pi$ ($d = 2$) and $4\pi/3$ ($d = 3$).

and this model is still useful as a reference for assessing the existence of inter-point interactions in those patterns showing inhomogeneity. In fact, this is one of the most common models in applications. In practical terms, analogously to the homogeneous Poisson process, the model is completely defined based on a parametric intensity function $\lambda_\theta(x)$: $p_n$ is a discrete Poisson distribution with mean $\Lambda(W)$ and points are i.i.d. following a location density function which is proportional to $\lambda_\theta(x)$:

$$f_n(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\lambda_\theta(x_i)}{\Lambda(W)} \text{ for } x_i \in W,$$

(2.6)

$$p(\{x_1, \ldots, x_n\}) = \prod_{i=1}^{n} \lambda_\theta(x_i) \exp\left(-\int_W (1 - \lambda_\theta(u)) \, du\right).$$

(2.7)

Thus, the likelihood of the parameter vector $\theta$ is:

$$L(\theta|x) = p_n N(x)! f_n(x) = \prod_{i=1}^{N(x)} \lambda_\theta(x_i) \exp\left(-\int_W \lambda_\theta(u) \, du\right).$$

(2.8)

Although homogeneity restriction has been relaxed, independent scattering is still a necessary condition, i.e. realisations of the inhomogeneous Poisson process are not supposed to exhibit neither aggregation nor repulsion among points. Nevertheless, as we have previously commented, heterogeneity can be easily misunderstood with inter-point interaction and so not all functional summary characteristics can be directly applied. Besides of the inhomogeneous $K$-function suggested in [12], some other methods have been developed in that sense – *locally scaled point processes*—though these are out of the scope of this work.
2.4.3 Gibbs processes

Unlike Poisson processes, these are versatile models which are mainly aimed to represent inter-point interactions. The term Gibbs processes is commonly used in the physics jargon, where these models have been first described; however, statisticians usually refer to them as Markov point processes. In either case, the main characteristic of these models is that interactions are described in the form of “forces” among points, which may be either repulsive or attractive. Forces are summed up together into a “potential” or “energy” $U$, that determines the value of the location density function according to the following general expression:

$$p(x_1, \ldots, x_n) = Z^{-1} \exp \left(-U(x_1, \ldots, x_n)\right) \text{ for } x_i \in W. \quad (2.9)$$

Here, $Z$ is a normalising factor and the exponential form of the location density function is not arbitrary but motivated by the physics concept of maximising the entropy of a system; however its theoretical justification is out of the scope of this work. Anyway, what really determines the behaviour of Gibbs processes is the nature of the potential function $U$. A common generalization of its form is the following:

$$U(x_1, \ldots, x_n) = \sum_{i=1}^{n} V(x_i) + \sum_{i=1}^{n} \sum_{j=i+1}^{n} V(x_i, x_j) + \ldots \quad (2.10)$$

Although this general form theoretically allows interaction terms of any possible degree, those are usually restricted to the first and the second orders, which leads to the specific class of pairwise Gibbs processes. Thus, the general form results:

$$p(x) = Z^{-1} \exp \left(-\sum_{i=1}^{n} \alpha(x_i) - \sum_{i=1}^{n} \sum_{j=i+1}^{n} \phi(||x_i - x_j||)\right). \quad (2.11)$$

The term $\alpha(\cdot) \in \mathbb{R}^+$ is usually called “self-potential” and represents the overall variation of the intensity function to account for inhomogeneous environments. The term $\phi(\cdot) \in \mathbb{R}$ is denominated “pair-potential”, and measures the energy contribution due to interaction among points, as a function of their distance $r$. Sometimes, this density function is expressed in a different form which is easier to operate with:

$$p(x) = Z'^{-1} \prod_{i=1}^{n} \beta(x_i) \prod_{i=1}^{n} \prod_{j=i+1}^{n} h(||x_i - x_j||). \quad (2.12)$$

Here, $\beta(\cdot) = e^{-\alpha(\cdot)} \in \mathbb{R}^+$ represents the spatial variation in the average density, $h(\cdot) = e^{-\phi(\cdot)}$ is formally called “interaction function” and $Z'^{-1}$ is again a normalising
factor. In either case, interaction terms have a straightforward interpretation: (i) if \( \phi(r) > 0 \) \( (h(r) < 1) \), the value of the exponent is reduced and so these distances are less likely to occur, which is reflected as “repulsion”; (ii) if \( \phi(r) < 0 \) \( (h(r) > 1) \), the opposite effect happens and thus this results in “attraction”; (iii) if \( \phi(r) = 0 \) \( (h(r) = 1) \), \( \forall r > 0 \), there is no real interaction among points and the model yields the Poisson process. In either case, note that the model still assumes that interactions are homogeneous and do not depend on location; in that sense, more complex formulations have been proposed still in the context of Gibbs processes.

Modelling correlations using distances allows to define a range \( r_{\text{max}} \) beyond which no interactions are considered, and so exploit Markovian properties to provide an explicit factorisation of the density. Moreover, both distributional characteristics that have been considered are separately represented in the form of \( \alpha(\cdot) \) and \( \phi(\cdot) \) terms, providing a flexible approach to model construction, what has favored the appearance of a great amount of pair-wise models. However, there is a great drawback on these processes: the normalising factor \( Z \) depends on both terms as well as on the existence window \( W \), and is usually analytically intractable. This means that usually there is no closed form for the likelihood, or it is intractable, and valid approximations to \( Z \) must be used, mainly: sparse data \([63, 65]\), and Monte-Carlo simulation, \([42, 69]\). An alternative option which is extensively used lies on performing parameter estimation based on the Papangelou conditional intensity of the process; that is because its formulation does not contain any normalising factor:

\[
\lambda(u|x) = \frac{p(x, \{u\})}{p(x)} = \exp \left( -\alpha(u) - \sum_{j=i+1}^{n} \phi(||x_i - u||) \right) = \prod_{i=1}^{n} \beta(x) \prod_{i=1}^{n} \prod_{j=i+1}^{n} h(||x_i - u||). \tag{2.13}
\]

Based on this relaxed definition, a pseudolikelihood is defined in \([21]\):

\[
\text{PL}(\theta|x) = \prod_{i=1}^{N(x)} \lambda_{\theta}(x_i|x \\setminus \{x_i\})\exp \left( -\int_{W} \lambda_{\theta}(u|x)du \right). \tag{2.15}
\]

Due to their complex definition, these models have not been characterized in depth in the sense that, in general, no theoretical forms for summary characteristics are available. The main advantage of this family is the great expressive power resulting from the flexible definition of the interaction function; this allows to easily define specific models which may take advantage of domain-specific knowledge. Nowadays, a great effort is being put into improving simulation and inference techniques within these processes, e.g. \([5]\).
2.5 Historical overview

Although point processes could be considered a young field within the novel approach of spatial statistics, literature on both methodology and applications is quite extensive. In this chapter, we have introduced only a small part of the whole theoretical basis, which is complex and large but, in general, it might be considered that applications have addressed the growth of spatial point process statistics. The appearance of real scenarios in which point interactions are of importance –forestry, physics, astrostastics, cytology, neuroscience, etc.– has motivated the development of novel techniques to exploit available data. A great number of pure theoretical works has been also presented, specially in the last thirty years; however, there are not so many “classical” works providing a compilation of theory and methods for spatial point processes, e.g. [33]. The last ten years have seen an increasing interest on the topic; this may be motivated by the appearance of new data sources and technical improvements done on practical sampling methods, which allow the use of more precise and abundant datasets, e.g. sensor networks, confocal laser scanning, etc. This is the case of neuroscience applications, which have been specially important for the development of pattern replication techniques, e.g. [78, 79]. In fact, it can be seen in the bibliography that a lot of methodological books have just been published relatively recently –in the last ten years–, e.g. [31, 32, 47, 53, 56].

To our understanding, this may indicate that the field is currently starting to settle down as a unite entity, and the number of applications in which these techniques are applied is rapidly growing up. In the following, we present a historical overview of spatial point process statistics with a single divulgation purpose; its goal is to provide a brief idea of the conceptual evolution of this field, and so only a small part of the whole literature has been compiled. For a detailed historical review, we refer to methodological books such as [31, 47], among others.

Until 1930: Early attempts

Early attempts to the analysis of spatial point patterns were focused on studying the intensity. The term was initially coined by Gottlob König in the book “Forestry Mathematics” (1835). In this work, an estimator is suggested based on the distance among several randomly chosen trees to their nearest neighbor. Some time later, the number of points in a pattern was addressed by Ernst Abbe (1879), who introduced the Poisson distribution as a suitable model for this magnitude, while analysing the amount of blood particles in a sample. These ideas were further developed by [75], who explored not only the number of points but also the distribution of plants in large communities. This work supposed three key methodological advances: (i) the variation of the intensity function was first considered splitting the observation window into several quadrants and comparing counts; (ii) the Poisson distribution was
established as a “reference model” to distinguish among normal, clustered and regular patterns; and (iii) the size of these squares was noted to highly influence on the obtained results, which can be seen as an early concept analogous to the interaction range. The same approach for sampling was employed in the contemporary work of \[39\]; here, the index of dispersion of point counts from different quadrants was first used to assess interactions among bacterias, which is still used nowadays.

1930–1960: First interaction analysis

Since then, more attention was put to correlation analysis and spatial point process statistics started to grow faster driven by application fields, mainly ecology and astrostatistics. On one side, ecology contributions were focused on modelling interactions among individuals; in this context, the same previous idea of over-dispersion was used for the analysis of plant communities in \[27\]; furthermore, first functional summary characteristics were introduced: the nearest-neighbour function in \[28\] and the empty-space function in \[67\]. In addition, some preliminary modelling attempts were conducted, among other works, in \[54\]. On the other side, astrostatistics analyses were aimed to demonstrate that the universe was distributed in clusters on small scales. This situation resulted in the first aggregated models being introduced in \[58, 61\], along with the first use of the pair-correlation function as a method to evaluate inter-point interactions \[59\].


The following years were characterized by the appearance of a great number of models. For instance, well-known Cox processes were proposed in the context of time-series \[29\] (point processes restricted to \(\mathbb{R}^+\)); these would be posteriorly generalized to any possible space \[30\]. Furthermore, important clustering models were developed, such as Neyman-Scott processes \[60\] and Strauss processes \[74\]; moreover, general purpose Gibbs processes, which are of special importance nowadays, were formalized during this period, e.g. \[15, 40\]. Some works also addressed the simulation of these models and provided several techniques, such as the use of thinning for inhomogeneous Poisson processes \[52\] or the simulation of random packing of spheres \[49\]. Contemporary, non-parametric exploratory techniques continued being proposed. For instance, during this age, the extensively used Ripley’s \(K\)-function and its transformation, the Besag’s \(L\)-function, were suggested in \[14, 22\]; in addition, some testing techniques were suggested for assessing spatial randomness \[15, 45\] and inter-point interaction \[70\]. These are only a sample of the numerous methods appeared during this period that are still in use nowadays; consequently, this stage concluded with the first methodological compilations being published, e.g. \[30, 33\], which in some sense unified previous unrelated ideas on the field.
1980–Today: Field consolidation

Since 1980, scientific community has shown an increasing interest on spatial point processes, maybe caused by the improvement of sampling techniques, which facilitated the general availability of data. This has produced a great amount of case studies in several fields which favored the development of new research areas, such as marked point processes [37, 43, 48, 51, 72], pattern replication [11, 25, 36, 78, 79] and spatio-temporal processes [62, 68, 82].

Many methodological contributions continued previous trend on modelling, since there is a great interest on solving parameter estimation related problems, specially in the case of Gibbs processes. Consequently, several notable works dealt with the approximation of the maximum likelihood expression using sparse data [63, 65], or through Monte-Carlo simulation [42, 69]. Other contributions took the pseudo-likelihood definition in [21] and developed novel techniques for generalizing its use [5, 18, 19, 55] or improving its performance [46]. Alternatively to both approaches, model fitting based on minimum contrast of summary characteristics was also suggested [44]. Related to this topic, simulation techniques based on the Markov Chain Monte-Carlo method [20, 41, 69] and validation techniques for fitted models [7] have also been studied. Despite of this interest on modelling, non-parametric techniques were still of importance. For instance, kernel methods were suggested for the estimation of the intensity function [17, 34] and new functional summary characteristics were introduced: the $J$-function [77] and the inhomogeneous $K$-function [12]. Furthermore, several recent works have shown its preference for non-parametric or semi-parametric techniques, e.g. [36, 50], among others.

The latter is only a sample of the great number of works presented in the last years. Consequently, an important number of methodological compilations have been published recently [8, 31, 32, 47, 56], which settled down theoretical foundations of spatial point process statistics. In general, current research lines are very close to those in last paragraph: (i) exploiting available data in the form of models with higher expressive power (e.g. space-time, replication, marks, etc.); (ii) improving existing parameter fitting, validation and simulation techniques and (iii) developing more reliable non-parametric methods for the analysis of spatial point patterns.
Chapter 3

Replicated point patterns

As we have introduced before, in spatial point process statistics, it is not usual that several samples of the same random phenomenon are used in a statistical analysis. This fact might be caused by historical reasons, since in the early years of its development, data collection used to be a costly manual process. This task was specially complicated in certain fields, e.g. forestry or biology, in which technical or logistical limitations do not allow to easily obtain pattern images. Moreover, in some scenarios, theoretical basis usually consider single patterns to enclose enough information about the process to get reliable statistical conclusions, and so this may be considered an additional cause of the situation.

Nevertheless, that is not usually the case of finite processes, for which stationarity and isotropy assumptions do not always hold; that means there might not be enough evidence to evaluate distributional characteristics which vary across patterns or throughout the space. In addition, using replicated samples provides more reliability to the statistical analysis and allows to extract additional conclusions of two main kinds: (i) discriminating between process properties and pattern-specific artifacts and (ii) finding possible statistical differences among groups within the pool of samples. Moreover, advances in applications have made sampling an easier task; in fact, most works on replicated point patterns are related to biological-related topics, such as neuroscience or citology [9, 11, 35, 36, 78, 79]. In these fields, technical improvements on microscopy allow now the use of large databases regarding the properties of the elements that make up living organisms.

In this chapter, we conduct a methodological analysis of existing techniques for replicated point patterns. This will be done in the form of a comprehensive review of the small number of works that have been published up to the moment. Following the characterization of spatial point processes introduced in the last chapter, we will classify these into four areas: (i) point count modelling (ii) non-parametric intensity analysis, (iii) non-parametric interaction analysis and (iv) parametric model fitting.
Some contributions belong to several of these classes and so this division has been chosen only for clarity purposes. Finally, brief conclusions of this field are given, addressing some additional aspects related to the current lacks of this methodology as well as improvements that are yet to do.

### 3.1 Point count

This area focuses on modelling the discrete distribution of the number of points per pattern. As we have introduced, CSR behaviour assumes that $p_n$ is a Poisson distribution with mean $\mu = \lambda v(W)$, and so $E(n) = V(n) = \mu$. Finding departures from this model might be interesting and usually implies using replicated count data to fit a different model in which realisations of $N(W)$ depend on covariates and possibly random-effects. Only a few works related to pattern replication have addressed explicitly point count distributions; that is because it turns out to be kind of straightforward and more related to classical statistics. However, we are going to comment them for a completeness purpose.

The first works \cite{9, 11} analyse the three-dimensional spatial distribution of osteocyte lacunae –osseous cells– in the skull bones of Macaque monkeys. In this work, authors first exploit the fact that the estimated pattern intensity $\hat{\lambda}$ is a ratio-unbiased estimator, to propose an unbiased estimator for the constant intensity across samples based on the ratio of sums:

$$\lambda_i = \frac{\sum_j N(W_{ij})}{\sum_j v(W_{ij})} = \sum_{j=1}^{m_i} \frac{\hat{\lambda}_{ij} v(W_{ij})}{\sum_{j=1}^{m_i} v(W_{ij})}$$

for $i = 1, \ldots, m_i$. (3.1)

Here, samples are divided into $g$ groups with $m_i$ samples for each group; in this context, $n_{ij}$ denotes the number of points of the $j$th replicate within the $i$th group. These point counts are fitted, using ratio regression, to a standard additive-effects model of the form: $n_{ij} = \alpha_i v(W_{ij}) + e_{ij}$. In this expression, $\alpha_i$ terms are considered independent random effects for each group, with mean $\lambda$ and variance equal to the between-group sampling variance of the intensity $\text{var}(\hat{\lambda}_i)$; errors $e_{ij}$ are assumed conditionally independent given $\alpha_i$, with variance proportional to $v(W_{ij})$. The work is complemented performing an analysis of variance within groups compared with the variance between groups using the EM algorithm for parameter estimation.

A more general method was used in \cite{35}. This work deals with the distribution of pyramidal neurons in the cingulate cortex of human subjects in three diagnostic groupings; in this context, the main goal is finding significant statistical differences among these groups. Authors suggest fitting point counts $n_{ij}$ to a quasi-Poisson log-linear model, maybe allowing pattern-specific means $\mu_{ij}$ and a dispersion parameter $\phi$ equal to the variance-to-mean ratio. In this context: $E(n_{ij}) = \mu_{ij}$ and $V(n_{ij}) = \phi \mu_{ij}$. 

If the dispersion magnitude is compared with the theoretical value associated with CSR, $\phi = 1$, it is possible to look for interesting departures from the reference model in a more straightforward manner.

More precisely, for a given fitting, extra-Poisson variation may be indicative of two statistical conclusions: (i) there exists some departure from CSR within each pattern (e.g. point clustering) or (ii) there is variation between samples in the mean point count (e.g. missing covariates, random-effects, etc.). This methodology is convenient since allows the use of standard statistical software to fit arbitrary complex models; in this particular case, the saturated model considers a different mean for each pattern and more parsimonious ones were compared using a classical analysis of deviance based on $\chi^2$ and $F$ statistics. Due to its flexibility, some works have taken the same methodology and extended it to other scenarios. One of these contributions [79] deals with the spatial distribution of vacuoles in slides of brain tissue affected with different kinds of Transmissible Spongiform Encephalopathies (TSE). In this case, a deeper analysis of point counts is performed considering both different possible groupings (e.g. TSE strains, brain regions) and covariates (e.g. age, sex). Due to the existing parallelisms between this and our work, abovementioned method will be further explained in chapter 4.

### 3.2 Intensity analysis

Since a great part in the theory of spatial point process statistics assumes stationarity, inhomogeneous processes are not so usually considered; this situation is specially noticeable in the presence of pattern replication, where only a few works have been published [37, 78, 79]. However, defining a common spatial trend for the intensity $\lambda(\cdot)$ may be useful to distinguish between the overall effect of average point density and pattern-specific variations due to inter-point interactions. This section focuses on using non-parametric procedures for this estimation.

#### 3.2.1 Kernel density

A quite usual approach in the field is to find non-parametric estimations of the intensity using kernel methods. This approach, first suggested in [34] for single patterns, is quite natural: for an arbitrary pattern $x$, the point density in the vicinity of a location $u$ is: $\hat{\lambda}(u) = e(u) \sum_{x_i \in x} \kappa(u|x_i)w_i$. Here, $\kappa(u|x_i)$ is a kernel function which is often considered a uniform disc or an isotropic Gaussian; $e(u)$ is an edge-correction factor and $w_i$ are weights associated to each of the data points. In either case, the main difficulty lies on selecting the parameter of the kernel function (commonly termed bandwidth, $h$), since it determines the smoothness of the generated surface.
Several approaches has been suggested to this task [17]; in the case of replicated spatial point patterns, a single solution has been proposed in [79]. In this work, which we have already introduced above, all patterns are first registered and normalised into a common sampling area $W$; then, replicates within each group are assumed to have proportional intensity functions $\lambda_{ij}(\cdot) = \alpha_{ij} \lambda_i(\cdot)$ with constants $\alpha_{ij} = n_{ij}/n_i$ where $n_i = \sum_j n_{ij}$. For the estimation of each $\lambda_i$, the smoothing parameter (in this case, the deviation $\sigma$ of an isotropic Gaussian function) is selected using leave-one-out cross-validation, i.e. $h$ is chosen to maximise the log-likelihood:

$$L(h) = \sum_{j=1}^{m} \sum_{k=1}^{n_{ij}} \log \hat{\lambda}^{-j}(x_{jk}) \quad \text{for } x_{jk} \in x_{ij}. \quad (3.2)$$

In this expression, $x_{jk}$ denote the location of the $k$th point in the $j$th pattern of the group; $\hat{\lambda}^{-j}(x_{jk})$ is the kernel estimate resulting when the points from all replicates except the $j$th are superimposed. Note that each of the estimations must be scaled to integrate to 1, so that $L(h)$ can be interpreted as a cross-validated log-likelihood. Due to its simplicity and usefulness, the described method will be used in the case study within the context of anisotropic Gaussian kernels.

### 3.2.2 Other methods

Apart from the latter, an alternative approach for intensity estimation has been suggested in [24, 25, 71]. In these works a new method is used for the analysis of the three-dimensional distribution of (i) sacral parasympathetic nucleus neurons and (ii) locus coeruleus neurons. Briefly, the authors suggest considering the intensity function $\lambda(u)$ to be locally constant, what means CSR assumptions can be done within the vicinity of each location. More precisely, if we consider $X_{ik}$ the three-dimensional distance from a point in the common sampling area $u$ to its $k$-nearest neighbor in the $i$th pattern (Figure 3.1), under CSR, the probability density function of the random variable $X_{ik}$ is:

$$f(x_{ik}) = \frac{(4\lambda\pi)^k}{3^{k-1}(k-1)!} x_{ik}^{3k-1} e^{\frac{-4\lambda\pi}{3} x_{ik}^3}. \quad (3.3)$$

As it can be seen, this expression depends on the intensity value $\lambda$ around the location $u$, and the value of $k$ works as a smoothing parameter. When considering all $m$ samples the log-likelihood of the distances is $L(X_{1k}, \ldots, X_{mk}) = \prod_{i=1}^{m} f(x_{ik})$. This likelihood can be maximized analytically with respect to $\lambda$, what yields the unbiased maximum likelihood estimator:

$$\hat{\lambda} = \frac{(nk - 1)}{4\pi \sum_{i=1}^{n} x_{ik}^3}. \quad (3.4)$$
3.3 Interaction analysis

In this context, the existence of an optimal $k$ value, minimizing the root mean square estimation error, was experimentally demonstrated, though no further analysis were performed. In that sense, authors advocate in favor of a both position and intensity adaptive choice of this parameter; this situation is reminiscent of the bandwidth selection problem in kernel estimation, and so adapting selection strategies, such as abovementioned cross-validation, should be a matter of future investigations.

![Figure 3.1](image)

Figure 3.1: (A) Three replications of the same spatial point process. (B) Samples are superimposed within the same sampling area; measure distances $x_{i2}$ from a location $p$ to the second nearest neighbor in the $i$th pattern. (C) Regular lattice of locations where the intensity function is estimated. Figure from [25].

Finally, there have been a few more works dealing with some kind of non-parametric estimation of the intensity [78]; however these are performed in the context of parametric model fitting and so will be reviewed in section 3.4.

3.3 Interaction analysis

This area of interaction analysis has focused in two complementary directions:

- **Aggregating individual estimates of summary characteristics.** There is a great amount of techniques to assess inter-point interactions based on functional summary characteristics, such as the Ripley’s K-function. Replicated patterns may be regarded as samples of the same random phenomenon, and so the aim is to determine the general summary characteristics of the process $K(\cdot)$ by aggregating the statistical results from the single samples $K_i(\cdot)$; thus, more reliable statistical conclusions may be extracted about departures from CSR.
• **Looking for statistical significant differences among groups of patterns.** When several groupings are considered, abovementioned aggregation procedures can be used to obtain estimations of group-specific $\bar{K}_i(\cdot)$ and overall $\bar{K}(\cdot)$ summary characteristics; this may allow to find interesting statistical differences by comparing raw pattern estimations $K_{ij}(\cdot), j = 1, ..., m_i$ with mean values.

### 3.3.1 Aggregation recipes

As it has been noted in [47], if the windows of the $m$ considered patterns are congruent (i.e. have the same shape and size), simply forming arithmetic means may be a good strategy. Nevertheless, this situation is not usual and often other aggregation recipes should be applied, depending on the specific summary characteristic and, in certain cases, on the edge-correction method. In general, given a functional summary statistic $U(r)$, these aggregation recipes are weighted means of the form: $\hat{U}(r) = \sum_{i=1}^{m} \hat{U}_i(r) w_i$; here, $w_i = C_i/C$, $C_i \geq 0$ are values depending on the specific function and $C = \sum_{i=1}^{m} C_i$. More precisely, considering the main functional summary statistics we have:

- **Empty-space function, $F(r)$:** $C_i = v(W_i)$ as suggested in [11].

- **Nearest-neighbor distance function, $G(r)$:** $C_i = N(W_i \ominus r)$ which was also proposed in [11]; here, $W_i \ominus r$ is the result from reducing the $i$th window a distance $r$ to account for edge effects.

- **Ripley’s $K$-function, $K(r)$:** $C_i = N(W_i)$ was first introduced by [35] though a different alternative was proposed in [11]: $C_i = N(W_i)^2$. The former has been more used in the literature.

- **Pair-correlation function, $g(r)$:** $C_i = \gamma_{W_i}(r)$, where $\gamma_{W_i}(r)$ is the isotropised set covariance of $W_i$. This expression has been found in [47], though original contributors are unknown to us.

The first two aggregation formulas come from the fact that summary characteristics are usually ratio-unbiased estimators of the form $\hat{U}(r) = N_i(r)/D_i$; thus, [11] extended the idea previously introduced within the context of intensity and proposed the ratio of sums as the preferable aggregated estimator. A justification for this intuitive approach can be found in [47]. Similar derivations can be done for the pair-correlation function, but the Ripley’s $K$-function weights have been only recommended based on simulations.

These aggregation formulas were practically applied in posterior works, e.g. [79]. However, recently, [64] has compared several approaches, in the context of the $G(r)$
and $K(r)$ functions, and found some of them appropriated. On one hand, alternative weights which were originally proposed in [73] are tested. These are: $C_i = 1$, $C_i = v(W_i)$, $C_i = N(W_i)$ and $C_i = N(W_i)^2$. On the other hand, two estimators based on the ratio of sums, similar to those mentioned above, are developed for the particular case of the Ripley’s $K$-function. Briefly, results show that the use of squared point number weights $C_i = N(W_i)^2$, as originally suggested in [11], provides better results. Moreover, ratios of sums outperformed weighted averages in most cases. Consequently, these estimators should be also considered as a possible alternative in future works.

### 3.3.2 Group differentiation

This research direction has been closely related to that in the previous section, since functional summary characteristics are useful tools for assessing inter-point interactions. More precisely, Ripley’s $K$-function has received special attention due to its expressive power and straightforward generalization to account for heterogeneity.

#### Pure non-parametric methods

The first work suggesting this kind of analysis was [35]. Here, samples are divided into $g$ groups with $m_i$ samples for each group. In this context, the main purpose is to answer the following question: “adjusting for density differences and acknowledging possible departures from purely random patterns, do the observed patterns differ significantly from group to group on the average?”.

Briefly, the authors consider the estimated Ripley’s K-function for all patterns $\hat{K}_{ij}(r)$, and compute group-specific $\bar{K}_i(r)$ mean functions; then, these are also averaged into an overall mean $\bar{K}(r)$, representing the whole pool of samples. In all these cases, the aggregation recipe introduced above is used, i.e. $C_i = N(W_i)$. In this context, the following statistic is proposed to measure differences between groups in the range $r \in [0, r_0]$: 

$$D_g = \sum_{i=1}^{g} \int_{0}^{r_0} \left[ \hat{K}_{i}^{1/2}(r) - \bar{K}^{1/2}(r) \right]^2 dr.$$  \hspace{1cm} (3.5)

Note that, as the sampling variation of $\hat{K}_{ij}(r)$ increases with $r$, a variance-stabilizing square-root transformation is used in the definition. The statistic is loosely analogous to a residual sum of squares in a conventional one-way ANOVA. However, in this case the distribution of $D_g$ is intractable and so inference must be done using a Monte-Carlo significance test of the null hypothesis of no difference between the groups. Briefly, this is done in three steps:
1. Compute a set of residual K-functions of the form: \( \hat{R}_{ij} = \frac{n_{ij}^{1/2}}{2} [\hat{K}_{ij}(r) - \bar{K}_i(r)] \), which are approximately exchangeable quantities under both the null and the alternative hypotheses.

2. Obtain an empirical approximation \( D^* \) to the distribution of \( D_g \) by recomputing its values from bootstrap samples \( \hat{K}_{ij}^* = \hat{K}_{ij}(r) + \frac{1}{n_{ij}} \hat{R}_{ij}^* \). Here, \( \hat{R}_{ij}^* \) are random permutations of the computed set and group sizes \( n_{ij} \) keep fixed.

3. Repeated sampling leads to a set of realisations \( \{D_g, D_1^*, \ldots, D_N^*\} \); if \( D_g \) is the \( r \)th smallest value among them, the attained significance level of a test of the null hypothesis of no group differences is given by \( p = r/(N + 1) \).

Similar approaches have been taken in posterior works. First, in [36], a comparison against parametric methods is presented; this will be reviewed in the following section. Second, in [51], the non-parametric analysis is extended to consider bivariate marked point patterns; more precisely, the work studies the spatial interaction between neurons and astrocytes (brain cells that are functionally related to neurons) in HIV associated dementia. Finally, in [79], the same technique is employed using the non-homogeneous extension of the \( K \)-function [12]. In either case, in what concerns exclusively to the one-way analysis, the only distinction between these works and the original lies on the mechanism which is used to reflect the asymptotic variance of \( \bar{K}(r) \); in these cases, the statistic is defined as:

\[
D_g = \sum_{i=1}^{g} \int_0^{r_0} n_i r^{-2} [\bar{K}_i(r) - \bar{K}(r)]^2 dr.
\] (3.6)

Semi-parametric methods

Some other alternative approaches have been proposed, though these have not been so extensively used and should not be considered pure non-parametric methods. On one hand, we have already introduced that in [11], authors noted that summary characteristics are usually ratio-unbiased estimators; that means that the same additive-effects model previously suggested in the context of point count modelling may be used with functions. In this context, replication is exploited to yield standard errors for the pooled estimates and an analysis of variance is developed. On the other hand, a rather similar approach has been taken in a recent work [50]; there, the same dataset of [51] is used to extend the residual bootstrapping approach. Briefly, the authors propose to assume a linear mixed-effects model for the empirical K-function values at a given distance:

\[
K_i(r) = X_i \beta(r) + Z_i b_i(r) + \epsilon_i(r) \quad \text{for } i = 1, \ldots, g.
\] (3.7)
Here, $K_i(r)$ is an $m_i$-dimensional vector of $K$-function values at a distance $r$, where $m_i$ is the number of replicates in the $i$th group; $X_i$ and $Z_i$ are $m_i \times p$ and $m_i \times q$ known design matrices, to account for both fixed and random effects; $\beta(r)$ is a $p$-dimensional vector of fixed effects; $b_i(r)$ is a $q$-dimensional vector of pattern random effects and $e_i$ is an $m_i$-dimensional vectors of errors. This provides a framework that can account for correlation between patterns from the same group, as well as accommodate any type of between-groups or within-groups covariate. Moreover, covariance matrices used for random-effects $b_i(r)$ and errors $e_i(r)$ may potentially vary with distance. In this context, authors develop a residual bootstrapping procedure similar to that in [35], but adapted for $K$-functions arising from abovementioned model, and suggest an analogous statistical test. In general, though flexible, this method is rather complex and technical details are out of the scope of this general review; thus, we refer to the original work [50] for a deeper explanation. In the case study, for practical reasons, we will take the simpler pure non-parametric approach in [79], which has already been proven useful in several scenarios.

3.4 Model fitting

Although models are essential in statistics, as we have already introduced, those are not so highly developed for spatial point processes. There are quite a lot of possible approaches but not so many commonly established frameworks that could account for any kind of random behaviour. That is because this scenario usually requires to deal with the high-dimensionality of multivariate joint probability densities, where variables are strongly correlated. This situation is also reflected in the specific case of pattern replication where only a few works related to model fitting have been proposed. In general, those attempts are framed in the context of Markov point processes; more precisely all of them use some parameter estimation method based on the pseudolikelihood definition [21]. We refer to the previous chapter for a brief introduction to this topic.

3.4.1 Ad-hoc inference

As far as we know, the first study in this area was presented in [36]; this work compares the non-parametric (design-based) method in [35] with a parametric (model-based) approach, to the analysis of replicated spatial point patterns in several experimental groups. In this case, the overall goal is to assess which approach is better for the task of group differentiation. Authors consider a pairwise Gibbs model with constant intensity parameter $\alpha$ and three previously used definitions of the pair-potential function $\phi(\cdot|\theta)$. Consequently, the conditional intensity results of the following general form:
\[ \lambda(u|x, \theta) = \exp \left( -\sum_{i=1}^{n} \phi(||u - x_i||; \theta) \right) = \alpha \lambda_\theta(u|x). \] (3.8)

As usual, authors consider a set of \( g \) groups, with \( m_i \) samples per group, and allow different values of intensity parameters \( \alpha_{ij} \) across replicates. Since the intensity is considered constant within each pattern, those \( \alpha_{ij} \) may be maximized analytically and so be treated as nuisance parameters. More precisely, differentiation of the log-pseudolikelihood function of each pattern with respect to \( \alpha_{ij} \) yields the relation

\[ n_{ij} = \alpha_{ij} \int_W \lambda_\theta(u|x) du. \]

After elimination of nuisance parameters, the pooled log-pseudolikelihood within the \( i \)th group results

\[ PL_i(\theta_i) = \sum_{j=1}^{m_i} PL_{ij}(\theta_i), \]

where:

\[ PL_{ij}(\theta_i|x_{ij}) = \sum_{k=1}^{n_{ij}} \log \left\{ \lambda_\theta(x_{ijk}|x_{ij}^{-k}) \right\} - n \log \left\{ \int_W \lambda_\theta(u|x_{ij}) du \right\}. \] (3.9)

Here, \( x_{ijk} \) represents the location of the \( k \)th point in the \( j \)th pattern of the \( i \)th group and \( x_{ij}^{-k} \) stands for the \( j \)th pattern in the \( i \)th group, once the \( k \)th point has been removed. Thus, the expression is maximized to find parameter estimates, \( \hat{\theta}_i \), and a simple statistic is used to test the null hypothesis of no differences among groups, \( T = PL^1 - PL^0 \), where:

- \( PL^1 = \sum_{i=1}^{g} PL_i(\hat{\theta}_i) \), that is, the sum of log-pseudolikelihood across groups.
- \( PL^0 = PL(\hat{\theta}) \), considering a common parameter value \( \hat{\theta} \) for all samples.

Following the approach in [35], a Monte-Carlo based significance test is performed. There a large number of values of the statistic \( \{T^*_1, \ldots, T^*_N\} \) are recomputed from simulations of the model; in this context, if the original \( T \) is ranked in the \( r \)th position, the resulting significance level is \( p = r/(N + 1) \). A comparison of \( p \)-values obtained using non-parametric and parametric approaches with synthetic data is performed. The overall results show that, even in idealized circumstances where the model is known, the non-parametric approach introduced in the previous section may be still competitive; moreover, misspecification of the model can lead to invalid tests of differences between groups in the parametric case. Based on this results, in the case study we will show preference for the non-parametric approach when we look for differences among groups.

### 3.4.2 Berman-Turner device

The previous work already succeeded in modelling a spatial point process based on replicated samples. The following contributions mainly focused on finding more general methods that may account for spatial heterogeneity or arbitrary fixed and
mixed-effect models. This was accomplished by adapting the approach in [19] for fitting the spatial trend of (single pattern) Poisson models using standard statistical software, which was subsequently generalized by [5] to non-Poisson spatial interaction models. This methodology, usually termed “Berman-Turner device”, works as follows. First, it assumes the conditional intensity is log-linear, so that it can be written as $\lambda(u|x) = \exp(\theta^T X(u|x))$; here, $X$ is a vector of covariates at location $u$ and $\theta$ is the vector of parameters to be estimated. Then, the window $W$ is split into a quadrature scheme (for instance, a Voronoi tessellation of the pattern, Figure 3.2) with exactly one dummy point being placed within each tile; this allows approximating the integral in the pseudolikelihood by quadrature:

$$\int_W \lambda(u|x) du \approx \sum_{i=1}^{n^*} \lambda(u_i|x)w_i.$$  \hfill (3.10)

Figure 3.2: A Voronoi tessellation of a pattern being used as quadrature scheme for Berman and Turner device. Both data (dark) and dummy (light) points can be seen. Figure from [78].

Here, $u_i, i = 1, \ldots, n^*$ are points in $W$, including both data and dummy points, and $w_i$ are quadrature weights summing up to $v(W)$. In this context, $z_i$ is defined to be an indicator of data (1) or dummy (0) status and we let $y_i = z_i/w_i$, $\lambda_i = \lambda(x_i|x, \theta)$ and $w_i = v_i/n_i$, where $n_i$ is the number of quadrature points in the $i$th tile and $v_i$ its area. Then, the log-pseudolikelihood equation can be rewritten as:
\[
\log PL(\theta_i, x) \approx \sum_{i=1}^{n^*} (y_i \log \lambda_i - \lambda_i) w_i.
\]

The key benefit of this reformulation is that, for fixed \(x\) the right side is formally equivalent to the log-likelihood of independent Poisson variables \(Y_k\) with means \(\lambda_k\) taken with weights \(w_k\) and logarithmic link. Thus, the expression can therefore be maximized using standard software for fitting generalized linear models (GLM); the unique conditions is that the conditional intensity function \(\lambda(\cdot|x, \theta)\), for fixed \(x\), is related to any explanatory variable by: \(\lambda(u|x) = \exp(\theta^T X(u|x))\). A great number of models, including Poisson and Gibbs processes, can be formulated using conditional intensity functions of this form; as a consequence of the ease of use, subsequent works related to pattern replication have extended and modified this framework to account for several scenarios:

1. **First attempt.** The first work adapting Berman-Turner device to this scenario was [55]. The group differentiation analysis conducted in [36] is repeated using the same real data as in [35], in addition to considering the misspecification of the model. Extension to the replication case is straightforward since it is only necessary to join the whole data; i.e. design matrices corresponding to each sample are concatenated, what means quadrature weights \(w_k\) and conditional intensity values for each pattern \(\lambda_k\) are evaluated separately.

2. **Fixed and mixed-effects models.** There have been two main works interested in whether random variation in patterns can be recovered through modeling, and in comparing fixed and mixed models.

   (a) First, [18] introduced a simple mixed-effects model in which abovementioned formulation of the Papangelou conditional intensity is substituted by \(\log \lambda(u|x) = \theta^T X(u|x) + DZ(u|x)\), where \(Z\) is the random-effects design vector at \(u\) and \(D\) is a normally distributed vector of random-effects. In this context, the model may be fitted with standard statistical software for Generalized Linear Mixed Models (GLMM) and both synthetic and real experiments with the extensively used dataset of [35] are performed. In some aspects the mixed model was found to be superior to parametric solutions.

   (b) Second, [78] presented a complex semi-parametric approach to account for mixed-effects and heterogeneous environments; however it does not consider inter-point interactions. Briefly, the authors modify the conditional intensity expression as \(\log \lambda(u|x) = \theta^T X(u|x) + S(u)\), where \(S(u)\) is a smoothing component, more precisely, a linear combination of the form
\[ S(u) = \sum_{t=1}^{K} Z_t(u)D_t. \]

Here, \( D \) is again a normally distributed vector of random-effects but covariates \( Z_t(\cdot) \) are derived using a spatial extension of the penalized spline. Using this framework, the authors perform both synthetic and real experiments, in the context of drug effects on neuronal activation patterns in the brain of rats, and present performance outcomes along with a two-way ANOVA analysis for group differentiation. Results show that, though generalization has been lost because standard statistical software can no longer be used, this method outperforms non-parametric kernel density estimation in several aspects.

3. Marked processes. One additional work has taken this approach and extended it to the analysis of marked point processes in heterogeneous environments [37]. This is a case study dealing with the distribution of two different species of plant roots in soils. The main novelty lies on the definition of the interaction range which varies throughout the space, depending on the degree of inhomogeneity.

In summary, these works provide adaptations of extensively used techniques within the context of single patterns, and the consideration of mixed-effects models seems to be the straightforward extension to account for pattern replication. However, in general, validation techniques have not been explicitly adapted to the replicated scenario. In the case study, we will employ the simple methodology in [55] along with a leave-one-out cross-validation procedure for model selection.

3.5 Additional issues

In general, works on replication have focused on extending classical techniques of spatial point patterns to considerer several samples. A few specific methods have been proposed but, in such cases, these are straightforward generalizations (e.g. aggregation recipes) or are based on the same basis as single pattern versions (e.g. Berman-Turner device). Consequently, this methodology does not seem to be so worried about a usual objective in the context of machine learning as it is generalization; current model fitting techniques are just extensions of the single pattern scenario and so tend to generalize across points instead of across patterns. Mixed-effect models could be considered a way to achieve this objective but, in either case, further selection and validation techniques should be adapted to this scenario. The long-term goal is to develop more general methods, that can account for pattern replication using a different framework – closer to usual machine learning techniques.
Chapter 4

Methodology

In this chapter, we will further explain all methodological issues that have to do with the specific methods and data which have been used in the case study. In general, theoretical foundations have already been introduced in either chapters 2 or 3; however, in this case, a more practical-oriented approach will be taken, giving details about how selected techniques should be implemented and used in practical applications. Consequently, a kind of short reference guide to the statistical analysis of spatial point patterns is provided, though this does not aim to be a deep tutorial on the topic. For more detailed methodological compilations, we refer to [31, 32, 47, 53, 56]; for practical guides to the use of spatstat R package, see [4, 8]. Briefly, the statistical analysis will be divided in the following sections: (i) data collection, (ii) exploratory analysis, (iii) intensity function estimation (iv) inter-point interaction analysis and (v) model selection and validation.

4.1 Data collection

Here, we will briefly describe the sample material under study and the processes by which data for our analysis have been extracted. These include some preprocessing stages that have been necessary to perform to prepare data samples.

4.1.1 Sample material

As we have introduced in the first chapter, this case study deals with the spatial distribution of dendritic spines over dendrite surface, in two different diagnostic groupings; more precisely, samples correspond to apical dendrites, that is, emerging from the apex of pyramidal cells (cerebral cortex). Brain tissue, from which sample data is obtained, was supplied by Dr. I. Ferrer from Instituto de Neuropatología, Servicio de Anatomía Patológica, IDIBELL (Hospital Universitario de Bellvitge, Barcelona,
Spain). The material was extracted at autopsy (2–3 hours post-mortem) from two human males, aged 40 and 85; from this moment, groups will be labeled as $G_{40}$ and $G_{85}$, respectively. The causes of death were traffic accident ($G_{40}$) and pneumonia plus interstitial pneumonitis ($G_{85}$), respectively. Brains were immediately immersed in cold 4% paraformaldehyde in 0.1M phosphate buffer, pH 7.4 (PB), for their proper conservation; then, these were sectioned into 1.5-cm-thick coronal slices. In this case, cells have been obtained from the same cortical area and neuronal cortex: the anterior cingulate gyrus (Brodman area 24). Each pyramid cell was individually injected with a marker – Lucifer Yellow, LY. This was applied by continuous current until the distal tips of each cell fluoresced brightly, indicating that dendrites were completely filled and ensuring that the fluorescence did not diminish at a distance from the soma. The whole process left sample material ready to take captures of the fluorescence at several depth levels using Confocal Laser Scanning Microscopy (see introduction chapter); in the last instance, these images are employed to recover an approximation to the original three-dimensional shape of the whole neuron using specialized software.

![Figure 4.1: Fragment from an apical dendrite shown with Imaris software. A) Original dendrite body. B) Isosurface of the dendrite. C) Semi-automatically segmented dendritic spines. D) Closer view. Original image size: 76.8×76.8 μm; pixel size: 76.8×76.8 nm. Figure from [13].](image)

### 4.1.2 Extracting sample data

Sample data has been manually collected by neuroscience experts at Laboratorio Cajal de Circuitos Corticales (CSIC), Madrid, Spain. A software for image processing and visualization, Imaris (Bitplane Scientific Software), was used to assist in this task. Briefly, this tool takes raw confocal microscope images and process them, using oversampling, filtering and segmentation, to yield the three-dimensional volume of the whole specimen model (Figure 4.1). Then, it is possible to recover the branched distribution of dendrites, as well as characteristics and locations of dendritic spines over their surface (manually selecting and enumerating the objects of
interest); however, due to practical limitations of image sizes, this procedure must be performed considering only a fragment of each whole dendrite at a time. This results in a total amount of 59 samples from 18 different apical dendrites (10 from G85 and 8 from G40); each of them comprises information about a section of dendrite of approximately 100 $\mu m$ length, including:

1. A polyline representing the curve-skeleton –the longitudinal axis– of dendrite sections (approximately 10 points per chunk). Each dendrite is composed of a variable number of these fragments: from 2 to 4.

2. Several polylines representing the shape of each spine, from the most eccentric point in its head to the point where it is inserted into dendrite’s surface. Each spine is approximated by a variable number of points: from 2 to 4.

3. Additional data about each of the spines: three-dimensional model, length, area, volume, angle, etc., which are not explicitly being used in this case study.

A simple visual inspection as well as domain specific knowledge suggest that spine density varies along dendrite longitudinal axis, and so the process should be analysed considering each whole dendrite as a single sample. However, in Imaris, space locations are expressed with respect to a local coordinate system which is translated between images. As a consequence, three additional synthetic points are selected in each overlapped region between two consecutive fragments; thus, these translation vectors can be properly estimated from the coordinates of common points in each system.

### 4.1.3 Preprocessing samples

Raw data extracted from software must go through three main preprocessing stages:

1. *Joining all dendrite fragments.* We estimate translation vectors between coordinate systems which are local to each sample belonging to the same dendrite. Three common points were arbitrarily chosen in the overlapped region between any two consecutive fragments $A$, $B$. These define three translations vectors $\{\vec{v}_{1AB}, \vec{v}_{2AB}, \vec{v}_{3AB}\}$, where $\vec{v}_{iAB} = (x_i^A, x_i^B)$ and $x_i^X$ is the coordinate vector of the $i$th common point in the $X$ fragment. In this context, the overall translation between each pair of coordinate systems is estimated as the average of these vectors. Sections belonging to the same dendrite are iteratively joined (transforming both axis and spine points) to yield 18 samples, each of them representing a whole dendrite of the dataset.
2. **Converting samples into planar representations.** In this particular case study, we are interested in the spatial distribution of spines over dendrite surface, and so we need to convert the three dimensional representation into a planar one. This has been done using stretching and unrolling transforms inspired by [57], in three main stages as follows:

(a) **Smoothing axis.** First, the polyline representing the dendrite axis is interpolated using a cubic spline procedure; this approximates the continuous smoothed form of the real dendrite using a piecewise cubic polynomial. More precisely, the new polyline is built using a number of knot points \( n^* \) which depends on the length of the original axis \( n \), as follows \( n^* = (n - 1)/0.05 \) (Figure 4.2). Smoothing is done to reduce the accumulated error in the movement of points during the stretching stage. For more information about the procedure, we refer to the documentation of the command `spline` from ©MATLAB R2010 scientific software.

![Figure 4.2: Smoothing and stretching operation using cubic spline interpolation and iterative polyline alignment. Comparison between the projection over XY plane of original (red) and after stretching (blue) dendrites. Raw axis data before smoothing is shown with circles.](image)

(b) **Stretching dendrite.** Second, the bended shape of the dendrite is straightened so that it can be easily approximated using a simple cylinder. For
performing this operation, let \((a_1, \ldots, a_{n^*})\) be the points in the smoothed dendrite axis. In this context, each spine polyline is assigned to its nearest point in the axis. Then, for \(i = 3, \ldots, n^*\), we iteratively consider vectors \(\vec{v}_0, \vec{v}_1\) with \(\vec{v}_r = (a_{i-r}, a_{i-r-1})\), compute the angle \(\phi = \hat{\vec{v}_0} \hat{\vec{v}_1}\) and the vectorial product \(\vec{v} = \vec{v}_0 \times \vec{v}_1\). Then, rotating \((a_i, \ldots, a_{n^*})\) along with spine polylines assigned to them– an amount \(\phi\), using \(\vec{v}\) as the rotation axis, yields a straight representation of the dendrite (Figure 4.2). It is clear that inter-point distances are modified during the process and so this is something it is necessary to consider when analysing obtained results.

(c) **Unrolling dendrite.** Third, the dendrite is aligned with \(X\) coordinate axis and spine points are unrolled to yield a planar representation. This is done assuming that the shape of the straightened dendrite is approximately cylindrical, with radius \(r_d\) equal to the average distance from spines’ insertion point to dendrite axis. Thus, we consider a spine point \(p = (x_p, y_p, z_p)\), the normal vector to axis passing through the point \(\vec{n}_p = (0, y_p, z_p)\) and the angle \(\theta = \hat{\vec{n}_p} \hat{\vec{z}}\), where \(\hat{\vec{z}} = (0, 0, 1)\). In this context, the new coordinates of \(p\) are \((x_p, r_d \theta, 0)\) (Figure 4.3). Resulting from this operation, the spatial distribution of spines over dendrite surface is finally represented in a planar pattern.

![Figure 4.3: Unrolling operation using a cylinder approximation. Comparison between the projection over YZ plane of original spines projected into the average cylinder (red) and the dendrite after unrolling (blue).](image)

3. **Registering and scaling patterns into a common area.** Since we will require to perform spatial analysis involving all dendrites simultaneously, it is necessary to register and scale patterns into a common sampling area. Firstly, all resulting patterns are aligned rectangles \(\{W_i\}_{i=1}^{18}\) and so we chose \(X\) coordinate axis as
the common registering point; then, the average rectangle is computed and patterns are scaled into the resulting existence window, $W^*$. It can be seen that this transformation also distorts average inter-point distances, and so this is something we must consider as well, during the statistical analysis.

As a result from this preprocessing stage, samples are ready to be used for their statistical analysis using spatial point processes. The final dataset comprises a total of 18 patterns, 10 from $G_{85}$, defined in a rectangular existence window $W^*_{85}$ of $224.71 \times 4.14 \, \mu m$; and 8 from $G_{40}$, with $W^*_{40}$ of $233.93 \times 4.35 \, \mu m$. When all patterns are registered and scaled into the same group, the existence window results $W^*$ of $228.81 \times 4.23 \, \mu m$ (Figure 4.4).

### 4.2 Exploratory analysis

In general, exploratory analysis comprises a series of non-parametric techniques, such as the computation of summary characteristics, that helps in the identification of the main properties of the process. In this case, exploratory analysis is divided in four main stages. First, we study basic measures of central tendency and dispersion of pattern main characteristics, e.g. point count, window size or average intensity; this provides information about the overall variability of the process. Second, we observe basic plots, e.g. histograms of each Cartesian coordinate at point locations, to informally assess intensity spatial trend. Third, we use specific techniques designed for testing CSR in each pattern. Finally, we look for some evidence of inter-point interaction using estimations of well-known summary characteristics that have already been introduced in previous chapters. In the following, we will further explain the two latter tasks, since the methodologies of the former are trivial.

#### 4.2.1 Testing CSR

Once basic measures and plots have been computed, the first stage of a spatial point process analysis comprises performing general tests to assess departure from the pure homogeneous Poisson process. In general, these are focused on the variation of the intensity spatial trend, and so some additional testing is usually necessary (see next section). In this case study, we consider two well-known techniques: the Pearson $\chi^2$ test of quadrat-counts and the Kolmogorov-Smirnov test of spatial coordinates.

**Quadrat-counts test**

This method is based on the definition of $P2$ in the characterization of the homogeneous Poisson process. This property implies points are uniformly and independently distributed throughout the space; thus, if the existence window $W$ is divided
Figure 4.4: The final dataset of registered and scaled patterns that are considered.
into \( m_q \) subregions of equal area (termed “quadrats”), point counts in each quadrat \( n_i, i = 1, \ldots, m_q \) are supposed to be i.i.d. under the null hypothesis. In this context, both expected and observed values for \( n_i \) are computed and a classical Pearson \( \chi^2 \) goodness-of-fit test can be used. In practical terms, we have employed the default implementation of this technique in R’s `spatstat` package: `quadrat.test` command.

**Experimental set up** – Window is divided into a regular grid of tiles following six different configurations (\( length \times width \)): \((5 \times 1), (5 \times 2), (5 \times 3), (15 \times 1), (15 \times 2)\) and \((15 \times 3)\). In this context, both individual and aggregated (all patterns superimposed) tests are performed. Note that the selected configuration of the tile grid is arbitrary and may influence results because it carries an implicit assumption about the scale.

**Kolmogorov-Smirnov test**

This alternative method also exploits the \( P2 \) property but using a more powerful Kolmogorov-Smirnov test. Briefly, this works comparing expected and observed distributions of some real-valued function \( T(u), \forall u \in W \); if simple Cartesian coordinates are considered, e.g. \( T(u) = x_u \) with \( u = (x_u, y_u) \), it is possible to evaluate this function on pattern points and perform a classical Kolmogorov-Smirnov test under the null hypothesis of uniformity. In practical terms, we have employed the default implementation of this technique in R’s `spatstat` package: `kstest` command.

**Experimental set up** – Two distribution functions are considered: \( T_x(u) = x_u \) and \( T_y(u) = y_u \). This configuration allows to evaluate departures from spatial homogeneity taking into account separately both Cartesian coordinates. In this context, individual and aggregated (all patterns superimposed) tests are performed. Note that there is no need to perform arbitrary selections which may affect the final result; thus, this methodology is often considered more reliable.

### 4.2.2 Exploring interaction

Basic interaction assessment is done by computing estimations from summary characteristics and performing a Monte-Carlo goodness-of-fit test for the null hypothesis of no interaction. Since no estimations of the intensity function are still available, only homogeneous versions of the estimators are considered at this stage; furthermore, a proper edge-correction procedure should be defined.

**Edge-correction technique**

Patterns represent the spatial distribution of spines over dendrite surface, i.e. a continuous medium. Consequently, existence window should be considered periodical
in the $Y$ axis, so that points can interact with each other in any direction. Thus, a custom edge-correction technique is defined for this case study. Let $[x_m, x_M]$ and $[y_m, y_M]$ be the ranges of the length $l_W$ and width $w_W$ of the aggregated window, $W$:

1. For each $x$, we define the periodically extended pattern $x^e = \{x \rightarrow (0, w_W) \cup x \cup x \rightarrow (0, -w_W)\}$, in the extended window $W^e$ with ranges $[x_m, x_M]$ and $[y_m - w_W, y_M + w_W]$. Here, the notation $x \rightarrow v$ stands for the result of translating all the points in $x$ a vector $v$.

2. For each $x$, and given a maximum interaction range $r$ we consider the border-reduced pattern $x^r = \{x \in W^r\}$, with $W^r \subset W$ defined as the reduced window with ranges $[x_m + r, x_M - r]$ and $[y_m, y_M]$.

In this context, given a maximum interaction range $r$, values of summary characteristics are computed for the pattern $x^r$ but considering all possible neighboring points in $x^e$ within a distance $r$. This allows to take into account both periodical continuation along the $Y$ axis and edge-effects produced in the extremes of the dendrite. The main drawback is that this definition restricts the maximum interaction range to $r_{\text{max}} \leq w_d/2$, so that the same point is not considered several times. However, the limitation is already greater than the value which is usually advised for $r_{\text{max}}$ (approximately $r_{\text{max}} \leq w_d/4$). For further clarity, see Figure 4.5.

![Figure 4.5](image)

Figure 4.5: The custom edge-correction applied. Summary characteristics are computed at the solid line pattern (black points), thought all possible neighbors in the dashed line pattern are considered (black or red points). Blue discs indicate some examples of interaction areas.

**Summary characteristics estimators**

In this stage, interaction analysis is restricted to the homogeneous version of $F(\cdot)$, $G(\cdot)$, $K(\cdot)$ and its derivations; considering the employed edge-correction technique, along with the definition of the extended $x^e$ and restricted $x^r$ patterns, the following expressions are used for the estimation of each of these summary characteristics. If
\[ d(u, x) = \min \{ ||u - x_i|| : x_i \in x \} \] and \( u_1, \ldots, u_n \in W^r \) is a constructed grid of locations in the reduced existence window, which can be of an arbitrary size, then:

\[
\hat{F}(r) = \frac{1}{n_l} \sum_{u_i \in W^r} 1\{d(u_i, x^e) \leq r\} \quad (4.1)
\]
\[
\hat{G}(r) = \frac{1}{N(x^r)} \sum_{x_i \in x^r} 1\{d(x_i, x^e \setminus x_i) \leq r\} \quad (4.2)
\]
\[
\hat{K}(r) = \frac{v(W^r)}{\pi(N(x^r) - 1)} \sum_{x_i \in x^r} \sum_{x_j \in x^e \setminus x_i} 1\{(d(x_i, x_j) \leq r)\} \quad (4.3)
\]

Figure 4.6: Left: estimations of \( F(\cdot), G(\cdot), J(\cdot) \) and \( K(\cdot) \) functions compared with theoretical Poisson values (dashed blue). Right: Monte-Carlo envelope produced by \( n_s = 39 \) simulated estimations of the Ripley’s K-function. In either case, the pattern shows a clear trend towards regularity.

The estimators for the rest of the summary characteristics are derived from these, as it is described in the theoretical chapter 2. In addition, aggregated versions are considered to follow the definition we have already introduced and justified in chapter 3; more precisely, we have chosen \( C_i = v(W_{i}^r) \) weights, for \( F(\cdot) \) and \( C_i = N(W^r) \) weights, for both \( G(\cdot) \) and \( K(\cdot) \). Finally, due to the –continuous– nature of our sampling area \( W \), we ignore the usual advise on the maximum allowed interaction range and take the value \( r_{max} \leq w_d/2 \) instead. In practical terms, it has been
necessary to extend default implementations of these estimators in R’s `spatstat` package – `Fest`, `Gest` and `Kest` commands (Figure 4.6) – in order to account for the specific edge-correction procedure. Furthermore, estimators of the aggregated version of each function have also been implemented.

**Monte-Carlo goodness-of-fit testing**

Commonly, during the exploratory analysis, inter-point interactions are evaluated looking for departures from the reference model, i.e. CSR. Techniques introduced in the last section, are more focused on long-range interactions, that is, the intensity spatial trend; this methodology gives us further evidence about the possibility that the analysed process is not even Poisson, by considering short-range interactions in detail. This is often done based on a Monte-Carlo goodness-of-fit test for the null hypothesis of no interaction: given any functional summary characteristic $U(\cdot)$, we compute the empirical estimation based on available data $U(\cdot|D)$ – which may be an aggregated value among several replicated patterns. Then:

1. A number $m_s$ of sample patterns $s_i, i = 1, \ldots, m_s$ are simulated from the null hypothesis process, i.e. homogeneous Poisson (CSR), and their corresponding summary characteristics are calculated $U(\cdot|D), i = 1, \ldots, m_s$.

2. For each value of the distance $r$, the $m_s$ simulated values $U(r|s_i), i = 1, \ldots, m_s$ are sorted; then, the $k$th lowest $D_r^k$ and highest $U_r^k$ values are used to define an “envelope” for the empirical value of the function $U(r|D)$.

In this context, though pointwise “envelopes”, i.e. $I(r) = [D_r^{(k)} U_r^{(k)}]$, cannot be considered confidence bands for the true value of the function, they can be used as the critical points for a Monte-Carlo test. Thus, for each value of the distance $r$, the null hypothesis of no interaction at that range is rejected if the empirical function value $U(r|D) \notin I(r)$. This test has exact significance level $\alpha = 2k/(1 + m_s)$. In practical terms, Monte-Carlo goodness-of-fit testing is already implemented in R’s `spatstat` package – `envelope` command (Figure 4.6) – and so we have extended it to consider custom edge-corrected and aggregated estimators.

**Experimental set up** – Both individual and aggregated estimations of $F(\cdot), G(\cdot), J(\cdot)$ and $L(\cdot)$ are computed for $r \in [0, w_d/2]$. Then, we plot results to visually compare them with theoretical homogeneous Poisson values (Table 2.1), and perform the corresponding goodness-of-fit tests to assess departure from the reference mode for $\alpha = 0.05 (n_s = 39)$ and $\alpha = 0.01 (n_s = 199)$. 
4.3 Non-parametric study

During this stage, more sophisticated non-parametric methods are used to obtain some specific results on the nature of the process properties. More precisely, based on indicators of departure from the homogeneous Poisson model, we conduct two complementary analysis. First, the intensity function of the process is estimated non-parametrically, considering several possible groups of replicated patterns. Then, these are exploited to perform a deeper analysis of inter-point interaction, which includes: (i) estimating summary characteristics defined for heterogeneous environments and (ii) looking for statistically significant differences between groups.

4.3.1 Intensity estimation

As we have already mentioned in previous chapters, the intensity function is estimated following the method suggested in [79], that is, a cross-validated kernel density approximation. As introduced above, for an arbitrary single pattern $x$, the point density in the vicinity of a location $u$ is: $\hat{\lambda}(u) = e(u) \sum_{x_i \in x} w_i \kappa(u|x_i)$; here, $\kappa(u|x_i)$ is some kernel function, $e(u)$ is an edge-correction factor and $w_i$ are weights associated to each of the data points. In practice, this implies splitting the sampling area into a regular grid of regions of equal area, and assuming the intensity in each “quadrat” to be constant and equal to that in its centre.

Kernel density estimation

In the case study, two additional issues have been considered for this estimation:

1. The dendrite surface is a continuous medium and so a similar edge-correction to that in the estimation of summary characteristics should be applied. More precisely, given some pattern $x$, with $N(x) = n$, we define its periodically extended version $x^e$ in $W^e$, as described in the previous section. In this context, the intensity function $\hat{\lambda}(\cdot)$ is estimated at a regular region grid $q_1, \ldots, q_m \in W$, but considering each possible point $x_i \in x^e$. Depending on the used bandwidth, this may cause the same points to be counted several times and so the intensity become overestimated. To avoid this problem, final values are scaled as $\hat{\lambda}_v(\cdot) = n\hat{\lambda}(\cdot)/\sum_{q_i \in W} \hat{\lambda}(q_i)v(q_i)$, so that expected number of points is still $\hat{\Lambda}(W) = n$, as it is considered in the definition of intensity function.

2. The shape of the dendrite is highly irregular, in the sense that it is far more large than wide. In this context, assuming isotropy may cause kernels to be too thick to capture the variation of the intensity along the width of the window, or too fine to correctly generalize its behaviour along the length. For this reason,
we consider anisotropic Gaussian kernels with diagonal covariance matrix $\Sigma = \text{D}(\sigma_x, \sigma_y)$, i.e. the estimator of the intensity value at each location $u$ results:

$$\hat{\lambda}(u|\Sigma) = e(u) \sum_{p_i \in \mathcal{X}} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{(x_u-x_i)^2}{2\sigma_x^2} + \frac{(y_u-y_i)^2}{2\sigma_y^2}\right)}$$

(4.4)

Here, $u = (x_u, y_u)$ and $p_i = (x_i, y_i)$, the bandwidth -covariance matrix- is fixed throughout the sampling area and $e(u) = 1/\gamma(u, \Sigma)$ is an edge-correction factor, where $\gamma(u, \Sigma)$ is the convolution of the Gaussian kernel centered at $u$ with the existence window. The additional edge-correction term has been considered so that estimations at the extremes of the dendrite are also reliable. In this case, there has been no need to use different weights $w_i$ for each of the points in the process.

In this context, the main task is to select the parameters of the kernel function since these determine the smoothness of the generated surface. This is done using a cross-validation procedure, as we have introduced in the previous chapter.

**Cross-validation procedure**

The group of $m$ patterns $\{r_1, \ldots, r_m\}$ are assumed to have proportional intensity functions $\lambda_i(\cdot) = \alpha_i \lambda(\cdot)$ with constants $\alpha_i = n_i/n$ where $n_i = N(r_i)$ and $n = \sum_i n_i$. For the estimation of $\lambda(\cdot)$, smoothing parameters are selected using leave-one-out cross-validation, i.e. $\Sigma$ is chosen to maximise the log-likelihood:

$$L(\Sigma|\{r_1, \ldots, r_m\}) = \sum_{i=1}^m \sum_{k=1}^{n_i} \log \hat{\lambda}^{-i}(p_{ik}|\Sigma) \quad \text{for } p_{ik} \in r_i$$

(4.5)

In this expression, $p_{ik}$ denote the location of the $k$th point in the $i$th pattern; $\hat{\lambda}^{-i}(p_{ik}|\Sigma)$ is the kernel estimate resulting when the points from all replicates except the $i$th are superimposed. Estimations must be scaled to integrate to 1, so that $L(\Sigma)$ can be interpreted as a cross-validated log-likelihood.

In practical terms, non-parametric intensity estimation is implemented in R’s `spatstat` package for both isotropic and anisotropic Gaussian kernels –`density.ppp` command. In that sense, it only has been necessary to extend the procedure to account for the specific edge-correction technique. On the other side, the maximisation of the log-likelihood has been done implementing Equation 4.5, as the objective function for a general-purpose optimization procedure provided by the `optim` command. This function allows to choose among several well-known techniques, and some facts should be considered to perform a proper selection, mainly:
The expression of the function gradient is available \( \nabla L = \left( \frac{\partial L}{\partial \sigma_x}, \frac{\partial L}{\partial \sigma_y} \right) \):

\[
\frac{\partial L}{\partial \sigma_x} = \sum_{i=1}^{m} \sum_{k=1}^{n_i} \sum_{p_{io} \in \mathbf{x}(i)} \left( (x_{ik} - x_{io})^2 + 2 \sigma_x^2 (x_{ik} - x_{io}) e^{-g(p_{ik},p_{io}|\Sigma)} \right) \sigma_x^2 \sum_{p_{io} \in \mathbf{x}(i)} e^{-g(p_{ik},p_{io}|\Sigma)} \tag{4.6}
\]

\[
\frac{\partial L}{\partial \sigma_y} = \sum_{i=1}^{m} \sum_{k=1}^{n_i} \sum_{p_{io} \in \mathbf{x}(i)} \left( (y_{ik} - y_{io})^2 + 2 \sigma_y^2 (y_{ik} - y_{io}) e^{-g(p_{ik},p_{io}|\Sigma)} \right) \sigma_y^2 \sum_{p_{io} \in \mathbf{x}(i)} e^{-g(p_{ik},p_{io}|\Sigma)} \tag{4.7}
\]

\[
g(a, b|\Sigma) = \frac{(x_a - x_b)^2}{\sigma_x^2} + \frac{(y_a - y_b)^2}{\sigma_y^2} \quad \text{for} \quad a = (x_a, y_a) \tag{4.8}
\]

Evaluating the log-likelihood expression is time consuming since it requires to compute \( m \) estimations of the intensity function, and to check its value at each point in the removed pattern. Preliminary tests were performed to estimate this cost, which turns out to be in the order of several seconds per evaluation.

Considering the shape and continuous nature of the existence window, lower and upper bounds for the parameters can be estimated, so that kernels are large enough to overlap and points are not considered several times.

Following these indications, we have selected a gradient hill-climbing optimization technique \[26\], which is a limited-memory and box-constrained modification of the well-known BFGS quasi-Newton method. This choice is justified since more modern global optimization techniques, such as simulated annealing, usually requires a large number of calculations of the objective function; the possibility of getting stuck at a local optima is avoided following a multi-start approach.

**Experimental set up** – Both group-specific and common intensities are estimated. We approximate the smooth surface of the function by a regular \( (64 \times 512) \) grid of regions of equal area. The lower and upper bounds are defined as \( \sigma_x \in [d_m/\sqrt{2\pi}, w_W/(2\sqrt{2\pi})] \) and \( \sigma_y \in [d_m/\sqrt{2\pi}, l_W/(2\sqrt{2\pi})] \); here \( d_m \) is the minimum distance from each point to its nearest-neighbor and \( l_W \) and \( w_W \) are the length and width respectively of the existence window. These values have been chosen so we can informally ensure that the maximum radius of kernels, for each Cartesian coordinate, is approximately in the range \( r_x \in [d_m, l_W/2] \) and \( r_y \in [d_m, w_W/2] \). In this context, the algorithm is carried out 100 times using equally spaced starting points; then, 10 best results are selected and refined by using a smaller gradient step size. The best of the resulting values is selected as the final bandwidth to be used within each group.
4.3.2 Interaction analysis

In this stage, we go further into non-parametric analysis of inter-point interactions:

- On one side, previous results have shown that there is clear evidence of variation in the spatial trend of the intensity, and so specific summary characteristics for heterogeneous environments should be used. This is accomplished computing estimations of the Ripley’s K-function extension to heterogeneous environments [12]. Thus, considering the extended $x^n$ and border-reduced $x^r$ patterns introduced above, the estimator for a given intensity function $\hat{\lambda}(\cdot)$ is:

$$\hat{K}_{inhom}(r) = \frac{1}{\sum_{i=1}^{n} \hat{\lambda}(x_i)} \sum_{x_i \in x^n} \sum_{x_j \in x^n \setminus x_i} \left\{ \frac{1}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)} \right\}$$

(4.9)

- On the other side, we also aim to assess statistical differences among groups by exploring variation of non-parametric measures; in that sense, we are interested in two different statistical questions: (i) “Is there any change in interaction due to normal ageing?” and (ii) “Do interaction varies along the length of the dendrite?”. In either case, this task is performed following exactly the same approach in [35], which has already been explained in detail in chapter 3. In this case, we consider the inhomogeneous L-function as summary characteristic and employ the definition of the statistic in [36, 51, 79] instead of the original.

In practical terms, inhomogeneous L-function estimation is already implemented in R’s spatstat –Linhom command; consequently, in that sense, it has only been necessary to extend it to consider the specific edge-correction technique we have defined above. In what concerns to group differentiation, the whole procedure has been implemented as an extension for package built-in functions.

Experimental set up – Two grouping approaches are considered. First, replicates are divided into two groups –{G40, G85}– to evaluate the influence of human ageing. Second, each pattern is split into pieces along its longitudinal axis –X coordinate– and groups are made up of patterns from the same region; more precisely, two different configurations are tested –{G21, G22} and {G41, G42, G43, G44}– arising from splitting patterns into 2 or 4 chunks respectively. This allow to assess the variation of interaction along the dendrite at two different detail levels. In either case, aggregated summary characteristics are computed and two analysis are conducted: (i) Monte-Carlo goodness-of-fit tests for the inhomogeneous Poisson model (within each group, $n_s \in \{39, 199\}$) and (ii) Monte-Carlo goodness-of-fit tests for assessing statistical differences (across groups, $n_s = 99$).
4.4 Parametric study

Finally, we conduct a parametric study of the whole process trying to find statistical models that explain this random phenomenon; more precisely, we address two separated problems: (i) modelling point counts distribution $p_n$ based on simple covariates and (ii) selecting a specific model for the distribution of these points throughout the space. In either case, some suitable models are considered and model selection and validation is performed using several techniques.

4.4.1 Point count modelling

As we have introduced in the previous chapter, in this case study, point counts are modelled using a Generalized Linear Model (GLM). Briefly, these models describe the dependence of a scalar variable $n_i, i = 1, \ldots, m$ on a vector of regressors $x_i$. The conditional distribution of $n_i$ given $x_i$ is a linear exponential family with probability density function of the form:

$$f(n|\lambda, \phi) = \exp\left(\frac{n\lambda - b(\lambda)}{\phi} + c(n, \phi)\right)$$

(4.10)

Here, $\lambda$ is a canonical parameter that depends on the regressors via a linear predictor and $\phi$ is a dispersion parameter that is often known. The functions $b(\cdot)$ and $c(\cdot)$ are known and determine which member of the family is used, e.g., the normal, binomial or Poisson distribution.

Considering the particular characteristics of the reference model in spatial point processes, the simple Poisson distribution, which is a special case of this framework, is often used in point count modelling. In this case, the canonical parameter is $E(n) = V(n) = \mu$, the dispersion parameter is $\phi = 1$ and the relationship between the mean and the linear predictor is assumed logarithmic. In this context, it is usual to find that data is over-dispersed and so the pure Poisson distribution is not completely adequate. We overcome this limitation by considering a quasi-Poisson model, which keeps the form of the original distribution, but leaves dispersion parameter $\phi$ unrestricted and estimates its value from data (as the variance-to-mean ratio). This strategy leads to the same coefficients as the standard model though inference is adjusted for over-dispersion. Moreover, the estimated value for $\phi$ provides a measure of departure from the reference model.

In this context, the nuisance parameter $\phi$ is estimated and coefficients are fitted using the well-known iteratively reweighted least squares algorithm (IWLS). In practical terms, this is implemented in R by the `glm` command. Then, several models are considered and selection is performed using classical statistics techniques for inference. More precisely, on one hand, regression coefficients in each model are assessed
via partial Wald tests using sandwich standard errors. On the other hand, an analysis of deviance is used to compare nested models, evaluating the reduction in the residual deviance through a classical F test (which is more appropriate for models with an estimated dispersion parameter). For more technical details on how these statistical tests are performed we refer to [81]. In practical terms, all these techniques are implemented in two commands: `anova.glm`, one of R’s built-in functions, and `coeftest.glm` from `lmtest` package.

**Experimental set up** – Four possible explanatory variables are considered in the model: age group, length, radius and surface area. Firstly, the saturated linear model is analysed using Wald tests to assess coefficient importance; then, higher order models, based exclusively on significant explanatory variables, are studied and compared using analysis of deviance.

### 4.4.2 Point process model fitting

In the previous, we have already introduced the Berman-Turner device for model fitting [5, 19] based on the pseudolikelihood definition in [21]. In this case study, we use this flexible framework to fit several variations of two different models: (i) the classical inhomogeneous Poisson process and (ii) the Area-interaction Gibbs process introduced in [10]. In practical terms, model fitting is already implemented in `spatstat` through the `ppm` command; thus, it has been necessary to extend this function to the replicated scenario. Briefly, this is done precalculating design matrices and quadrature weights corresponding to each of the patterns in the dataset and concatenating them into a unique one, which is finally used to fit the Generalized Linear Model. Within these conditions, arbitrary complex models may be considered, including terms to account for spatial trend, inter-point interaction, spatial covariates and other possible explanatory variables. In the following, we describe in detail practical issues concerning the different models that has been used, as well as employed model selection and validation techniques.

**Poisson models**

At this point of the study, previous results have already shown a clear spatial trend in the average point density. Moreover, non-parametric analysis of the inter-point interaction are not enough significant to ensure that patterns manifest either clustering or regularity behaviours. For this reason, the inhomogeneous Poisson process is first considered as a suitable model for this phenomenon.

As we have introduced before, the pseudolikelihood definition is based on the conditional intensity function $\lambda(|x)$; in the case of inhomogeneous Poisson processes, $\lambda(|x) = \lambda(\cdot)$, because no interaction among points is considered. This fact implies
that the expression of the pseudolikelihood is formally equivalent to that of the likelihood, and so Berman-Turner device can be used to obtain the maximum likelihood estimator of the parameters. In the context of the framework introduced in chapter 3, this implies that \( \lambda(u) = \exp(\theta^T X(u)) \), where \( X(u) \) is a vector of covariates at each location \( u \). For instance, for a log-linear model of the intensity function, the parameter vector is \( \theta = (\theta_0, \theta_x, \theta_y, \theta_{xy}) \) and the covariate vector is \( X(u) = (1, u_x, u_y, u_x u_y) \). Thus, for a set of \( m \) patterns with \( n_i^* \), \( i = 1, \ldots, m \), total entries per pattern (including data and dummy points), the resulting design matrix is made up of the simple concatenation of covariate vectors \( X_{ij}(u) \), \( j = 1, \ldots, n_i^* \). In general, we will restrict possible models of the intensity function to log-polynomial expressions depending on both Cartesian coordinates. However, two additional issues has been considered in this case study:

- As in the case of non-parametric estimation, we are interested on letting the intensity function vary proportional across patterns; this is justified since, exploratory analysis reflect a high variation on the number of points, though it seems to exist a common spatial trend across all patterns. In practical terms, this implies that covariate vectors in the previous example are re-defined as \( X_i(u) = (\log(n_i), 1, u_x, u_y, u_x u_y) \), with pattern-specific constants \( n_i = N(x_i) \); and the new covariate is assumed to be just an offset, i.e. \( \theta_i = (1, \theta_0, \theta_x, \theta_y, \theta_{xy}) \). Consequently, the final expression of the intensity results \( \lambda_i(u) = \exp(\log(n_i) + \theta_0 + \theta_x u_x + \theta_y u_y + \theta_{xy} u_x u_y) \) implying \( \lambda_i(u) = n_i \lambda(u) \).

- Since we have already obtained non-parametric approximations of the intensity, we also consider fitting semi-parametric models based on this precomputed kernel density. In practical terms, this is simply done by taking the estimation \( \hat{\lambda}_i(\cdot) \) as a spatial covariate in the model, and so \( X_i(u) = (\log(\hat{\lambda}_i(u)), 1) \) and \( \theta = (1, \theta_0) \). In the simple Poisson case this is trivial and only a single parameter has to be estimated, \( \lambda(u) = \hat{\lambda}(u) e^{\theta_0} \). However this approach will get more sense when considering also inter-point interactions.

**Experimental set up** – Patterns are considered all together and grouped by age in \{G21, G22\}. In each case, both proportional and regular variations of several log-polynomial forms of the intensity function are fitted. More precisely, we consider expressions from degree 0 (null model) to degree 8 (saturated model). For comparative purposes, also the trivial form of the semi-parametric model mentioned above is computed. In either case, the integral in the pseudolikelihood is approximated by quadrature using a Voronoi tesselation. More precisely, each pattern is completed with a regular grid of dummy points (every 0.5 spatial units) which are randomly perturbed to add some variation; this results into a total number of points in each pattern \( n_i^* \sim 4000 \).
Section 4.4: Parametric study

Area-interaction models

Preliminary results have shown that there is some evidence of inter-point interaction. However, patterns seem to exhibit both clustering and regularity behaviours at different scales and these may differ between G40 and G85 groups. For this reason, we aim to further explore spine interaction by fitting more complex models belonging to the Gibbs family. It has been shown that pairwise interaction processes are not suitable for modelling aggregation due to some instability conditions (see [47], pp. 157); thus, we resort to a more complex structure of interactions which is offered by the flexible Area-interaction process [10], also known as Widom-Rowlinson penetrable sphere model [80]. Briefly, this process expresses the interaction in the pattern in terms of the area covered by the union of individual “influence regions” of the points. More precisely, for a given interaction range \( r \), the density function is:

\[
p(x|r) = Z^{-1} \prod_{i=1}^{n} \beta(x_i) \gamma^{-A(x|r)}; \quad A(x|r) = v \left( \bigcup_{i=1}^{n} b(x_i, r) \right) \bigcap W
\]

Here, \( Z^{-1} \) is a normalising constant, \( \beta(\cdot) \in \mathbb{R}^+ \) represents the spatial variation of the intensity as usual and \( b(x_i, r) \) is the disc of radius \( r \) centered at the point \( x_i \). Depending on the value of the parameter \( \gamma \), this process can be used as a model for moderate regularity (\( \gamma < 1 \)) or clustering (\( \gamma > 1 \)). This flexibility turns out to be quite useful in the particular scenario of this case study. In this case, conditional intensity function is not so straightforwardly defined as for pairwise interaction processes:

\[
\lambda(u|x, r) = \beta(u) \gamma^{-B(u, x|r)}; \quad B(u, x|r) = A(x \cup u|r) - A(x|r)
\]

This formula has a direct conceptual interpretation. Points in the pattern represent individuals sharing or competing for resources. We may interpret \( B(u, x|r) \) as the area of the “unclaimed zone” (Figure 4.7) where a new spine would be able to “draw resources” without competition from other individuals. For \( \gamma < 1 \), a random point is less likely to occur when the available region is small (regularity). For \( \gamma > 1 \), the inverse behaviour is shown (clustering). In either case, whether this model results on clustered or regular patterns usually depends on the selected interaction range. Thus, given \( r \), two points interact if the distance between them is less than \( 2r \) and the value \( \gamma = 0 \) corresponds to a process with hard-core distance \( r \). In practical terms, the intensity expression can be recast as:

\[
\log \lambda(u|x, r) = \log(\beta(u)) - \log(\gamma)B(u|x, r)
\]

This follows the general form of the Berman-Turner device introduced above, i.e. for a log-linear spatial trend, \( X_i(u) = (1, u, u_x, u_y, u_xu_y, -B(u|x, r)) \) and \( \theta_i = \)
Chapter 4. Methodology

Figure 4.7: Representation of the “unclaimed region” (red) belonging to a location in the space when Area-interaction process is considered. For a given region, interaction values such that $\gamma > 1$ penalize large areas yielding smaller values, and thus intensity grows when points form clusters. For $\gamma < 1$, the opposite effect happens.

$(\theta_0, \theta_x, \theta_y, \theta_{xy}, \gamma)$. Note that interaction radius $r$ is an irregular parameter and cannot be estimated directly using the Berman-Turner device; in those cases, the general approach suggests fitting the model for different values of $r$ and selecting that which maximizes a given scoring criterion.

Experimental set up – Patterns are considered all together and grouped by age in \{G21, G22\}. The spatial trend is modelled employing three different log-polynomial expressions ($P_2$, $P_4$, $P_6$) as well as the semi-parametric estimation using kernel density (denoted NP). In each case, the model is fitted for a series of interaction ranges $r \in [0.1, 1.5]$, and the best fit is selected considering several scoring criterion. Pseudolikelihood is approximated using the same quadrature weights as in the Poisson scenario.

Model selection procedures

Theoretical studies on model selection have been only related with single point patterns, and so they are not so concerned with exploiting the generalization capabilities of these models. Previous non-parametric results have only shown a slight departure from the Poisson assumption; for this reason, we focus on model selection to avoid overfitting, with the purpose of finding forms of the spatial trend that only capture the variation in the intensity that is enough to let points interact with each other. Thus, selection is performed using several procedures and scoring criterion:

- Log-pseudolikelihood. The log-pseudolikelihood of the fitted GLM.
• Analysis of deviance. Differences of deviances (two times log-pseudolikelihood ratio) between nested models are calculated and a classical two-sided $\chi^2$ test—for the null hypothesis of preference for the simpler model—is performed. There is no statistical theory available to support this for non Poisson models.

• Akaike information criterion (AIC). This is a scoring measure of the relative goodness of fit of a statistical model presented in [1]. It is defined based on the log-pseudolikelihood value as $AIC = 2k - 2\log PL$, where $k$ is the number of parameters of the statistical model. Hence, AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters, to avoid overfitting. In this case, we uses a correction of AIC for finite sample sizes defined as:

$$AIC_c = AIC + \frac{2k(k + 1)}{n - k - 1}$$

(4.14)

Here, $n$ denotes the sample size, and thus AICc is AIC with a greater penalty for extra parameters. This measure has been chosen because it is recommended to be employed regardless of the sample size, since AICc converges to AIC when $n$ gets large.

• Bayesian information criterion (BIC). Another scoring measure defined in [2] which is closely related to Akaike information criterion. However, in this case, the number of parameters are penalized in a stronger manner. More precisely, it is defined as $BIC = -2\log PL + k\log(n)$. This provides an alternative point of view in the measure of performance to account for overfitting.

• Cross-validated prediction error. For a given set of patterns $\{x_1, \ldots, x_m\}$ with $n_i^*$ locations, considering both data and dummy points, and a model $M$, we consider the leave-one-out cross-validated prediction error, defined as:

$$CV(M|\{x_1, \ldots, x_m\}) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n_i^*} \sum_{j=1}^{n_i^*} (y_{ij} - \hat{y}_{ij}^{-i})^2$$

(4.15)

Here, $y_{ij} = z_{ij}/w_{ij}$ is the value of the independent variable which is used to fit the model by a GLM (see Section 3.4.2), corresponding to the $j$th location of the $i$th pattern; correspondingly, $\hat{y}_{ij}^{-i}$ is the predicted value for this variable, by the model that has been fitted considering all but the $i$th pattern. Thus, this is the cross-validated error of the fitted GLM; this is a simple approach but lacks of a proper interpretation directly related with conditional intensity. An alternative that has been also tested is the cross-validated
average of residuals; in such case the expression \((y_{ij} - \hat{y}_{ij})^2\) is replaced by 
\(R^{-i}(x_{ij}) = z_{ij} - w_{ij}\lambda_{ij}\): the raw residuals at the \(j\)th location point of each \(i\)th pattern, for the model that has been fitted considering all but the \(i\)th pattern. This measure of performance, which is here firstly considered, is based on the definition of the “residual” concept. In spatial point processes, this is an informal measure of goodness-of-fit which is defined at each region \(A \subseteq W\) as \(R(A) = N(A) - \int_A \lambda(u|\mathbf{x})du\). These quantities are supposed to be close to and distributed around zero. Outcomes are similar but we have obtained more consistent results and we are going to show preference for our first suggestion.

In general, model selection is usually performed using a classical analysis of deviance, when considering nested Poisson models, or the AIC otherwise. In this case, we are interested on further exploring the generalization capabilities of these models, and so AICc, BIC and CV are also being used and compared. For the purpose of investigating their effects, we give preference to our scoring criterion.

**Model validation procedures**

For model validation, both informal assessment and formal tests are used:

- **Informal validation:**
  
  1. Visual assessment based on simulations from the fitted model.
  2. Analysis of the spatial trend by considering *lurking variables* of residuals: cumulative distributions of Pearson residuals along each of the coordinate axis. Pearson residuals are defined at each point as \(R_p(x_i) = R(x_i)/\lambda_i, x_i \in \mathbf{x}\). These are standardised, in the sense that if the model (true and fitted) is Poisson, then the sum of the Pearson residuals in a spatial region \(A \subseteq W\) has variance equal to the \(v(A)\) (Figure 4.8).

- **Formal validation:**
  
  1. *Kolmogorov-Smirnov goodness-of-fit test.* General reformulation of above-mentioned CSR test to the scenario of inhomogeneous Poisson models. It validates the spatial variation of the intensity function.
  2. *Monte-Carlo goodness-of-fit test.* General reformulation of the previously introduced test of inter-point interaction, based on the simulation from the fitted model and the computation of an arbitrary summary characteristic.
4.4. Parametric study

Figure 4.8: Example of cumulative distributions of Pearson residuals along X and Y axis. For a good fit, lines should not lie outside significance bands, which are constructed based on the asymptotic variance under the model.

Figure 4.9: Example of a Q-Q plot of residuals under the model. In this case, it indicates that selected model is grossly inappropriate for data (outside critical envelope).

3. Q-Q plot of residuals. Plot comparing quantiles of observed residuals to those of the theoretical distribution under the fitted model (estimate by simulation). The resulting line should be within the pointwise 5% critical envelopes that are shown for a good fit (Figure 4.9). This technique validates both the spatial trend of the intensity and inter-point interactions.

Experimental set up – First, a tentative set of Poisson models is selected and validated using both lurking variables and Kolmogorov-Smirnov tests. These models are considered to capture in enough detail the variation of the overall intensity and are used as the spatial trend expression within the more complex Area-interaction process. Then, several of the latter models are fitted and validated by means of a
Q-Q plot of residuals. Note that simulation from fitted models when considering replication has not been implemented yet and so individual tests for each of the patterns are used.
Chapter 5

Results

In this chapter, we present in detail both visual and quantitative results of the considered case study. These results provide evidence for some interesting conclusions, about both the distribution of spines over dendrite surface, and the current lacks on patterns replication methodology. These results have been validated by experts in the field belonging to the Laboratorio Cajal de Circuitos Corticales (CSIC) and a short version of them has been recently presented in a international conference on spatial statistics [66]. Results are divided in non-parametric and parametric analysis.

5.1 Non-parametric analysis

This section includes: (i) preliminary data analysis, (ii) Complete Spatial Randomness testing, (iii) basic intensity and interaction analysis, (iv) group differentiation related results and finally (v) concluding remarks.

5.1.1 Preliminary data analysis

In general, it can be seen that patterns belonging to G40 group have noticeable higher values of basic characteristics, specially in what concerns to the number of points per pattern and the length of the dendrites. In addition, there exists a high variability within samples belonging to each group (Table 5.1), which we can observe is significantly higher in the case of G85. This may indicate that the reduction in the number of spines –consequence of normal aging– does not happen in an uniform manner throughout all the brain, i.e. dendrites in some specific neurons or regions may be more susceptible of suffering changes.

Basic analysis on the average point density shows that there is a clear spatial variation in the intensity function, in both the X coordinate axis and the Y coordinate
Table 5.1: Summary of patterns main characteristics, including average values, standard deviations and coefficients of variation for point count ($n$), dendrite radius ($r$), length ($l$), surface area ($a$), and average intensity ($\lambda$).

<table>
<thead>
<tr>
<th></th>
<th>$n$ (μm)</th>
<th>$r$ (μm)</th>
<th>$l$ (μm)</th>
<th>$a$ (μm$^2$)</th>
<th>$\lambda$ (μm$^{-2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>238.0000</td>
<td>0.6739</td>
<td>228.8071</td>
<td>964.9989</td>
<td>0.2495</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>67.9221</td>
<td>0.0711</td>
<td>42.6774</td>
<td>185.6741</td>
<td>0.0653</td>
</tr>
<tr>
<td>CV</td>
<td>0.2854</td>
<td>0.1056</td>
<td>0.1865</td>
<td>0.1924</td>
<td>0.2617</td>
</tr>
<tr>
<td>G85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>112.4444</td>
<td>0.3663</td>
<td>124.8381</td>
<td>929.7459</td>
<td>0.2233</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>53.3337</td>
<td>0.0349</td>
<td>50.7851</td>
<td>210.3981</td>
<td>0.0629</td>
</tr>
<tr>
<td>CV</td>
<td>0.4743</td>
<td>0.0952</td>
<td>0.4068</td>
<td>0.2263</td>
<td>0.2815</td>
</tr>
<tr>
<td>G40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>282.5000</td>
<td>0.6922</td>
<td>233.9303</td>
<td>1009.0651</td>
<td>0.2823</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>58.9334</td>
<td>0.1002</td>
<td>32.4546</td>
<td>151.0461</td>
<td>0.0554</td>
</tr>
<tr>
<td>CV</td>
<td>0.2086</td>
<td>0.1448</td>
<td>0.1387</td>
<td>0.1497</td>
<td>0.1962</td>
</tr>
</tbody>
</table>

**Figure 5.1:** Histogram showing the distribution of the X and Y coordinates for the points in the G85 group (first row), in the G40 group (second row) or in any pattern (third row). In each case, first column corresponds to a number of bins $n_b = 10$, while in the second column $n_b = 40$. 

axis (Figure 5.1). In either case, there are no clear signs of differences among groups, except from scale (G40 yields higher counts). Two main aspects may be noticed:
5.1. Non-parametric analysis

- It may be surprising, considering the way data is sampled, that the spatial trend along the width of the patterns varies following this -kind of- sinusoidal form; orientation is supposed to be lost during the process and so there is no sense on finding this kind of trend. Our best guess is that this artifacts were produced by the point of view of the camera within the software that is used to manually identify spines, i.e. a bias towards selecting locations closer to the sides of the dendrite.

- The variation along the length of the patterns is suitable and consistent with previous studies in this field. It is clearly noticeable that the closest region to the soma, of about 1/6 of the total length, is often empty. This may indicate no interchange of signals with other neurons happen at this range of distances from the cell nucleus.

<table>
<thead>
<tr>
<th>Dendrite</th>
<th>QC(_{(5 \times 1)})</th>
<th>QC(_{(5 \times 2)})</th>
<th>QC(_{(5 \times 3)})</th>
<th>QC(_{(15 \times 1)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>if6 1 1</td>
<td>*1.043\times10^{-01}</td>
<td>*1.498\times10^{-01}</td>
<td>*2.200\times10^{-01}</td>
<td>*1.957\times10^{-01}</td>
</tr>
<tr>
<td>if6 1 2</td>
<td>1.584\times10^{-10}</td>
<td>2.978\times10^{-09}</td>
<td>7.827\times10^{-12}</td>
<td>4.386\times10^{-07}</td>
</tr>
<tr>
<td>if6 1 3</td>
<td>4.360\times10^{-06}</td>
<td>5.790\times10^{-05}</td>
<td>1.357\times10^{-04}</td>
<td>2.976\times10^{-05}</td>
</tr>
<tr>
<td>if6 1 4</td>
<td>6.663\times10^{-09}</td>
<td>5.808\times10^{-09}</td>
<td>6.488\times10^{-10}</td>
<td>2.846\times10^{-05}</td>
</tr>
<tr>
<td>if6 1 5</td>
<td>1.123\times10^{-05}</td>
<td>1.735\times10^{-06}</td>
<td>4.850\times10^{-06}</td>
<td>9.799\times10^{-06}</td>
</tr>
<tr>
<td>if6 1 6</td>
<td>5.308\times10^{-08}</td>
<td>5.103\times10^{-06}</td>
<td>6.165\times10^{-07}</td>
<td>3.113\times10^{-07}</td>
</tr>
<tr>
<td>if6 2 0</td>
<td>3.024\times10^{-09}</td>
<td>1.168\times10^{-08}</td>
<td>2.021\times10^{-08}</td>
<td>3.158\times10^{-07}</td>
</tr>
<tr>
<td>if6 2 2</td>
<td>†2.080\times10^{-02}</td>
<td>3.991\times10^{-03}</td>
<td>4.728\times10^{-03}</td>
<td>†8.775\times10^{-02}</td>
</tr>
<tr>
<td>if6 2 11</td>
<td>2.469\times10^{-05}</td>
<td>1.019\times10^{-03}</td>
<td>1.145\times10^{-06}</td>
<td>1.663\times10^{-03}</td>
</tr>
<tr>
<td>if6 2 20</td>
<td>1.347\times10^{-06}</td>
<td>8.063\times10^{-07}</td>
<td>3.225\times10^{-05}</td>
<td>1.206\times10^{-04}</td>
</tr>
<tr>
<td>m16 1 1</td>
<td>3.070\times10^{-06}</td>
<td>9.836\times10^{-05}</td>
<td>1.343\times10^{-06}</td>
<td>2.138\times10^{-04}</td>
</tr>
<tr>
<td>m16 1 3</td>
<td>3.738\times10^{-06}</td>
<td>1.311\times10^{-05}</td>
<td>2.818\times10^{-07}</td>
<td>6.973\times10^{-04}</td>
</tr>
<tr>
<td>m16 1 6</td>
<td>2.711\times10^{-08}</td>
<td>3.021\times10^{-06}</td>
<td>2.000\times10^{-06}</td>
<td>1.146\times10^{-06}</td>
</tr>
<tr>
<td>m16 1 8</td>
<td>8.762\times10^{-13}</td>
<td>3.114\times10^{-12}</td>
<td>3.557\times10^{-16}</td>
<td>5.075\times10^{-10}</td>
</tr>
<tr>
<td>m16 1 9</td>
<td>1.140\times10^{-14}</td>
<td>3.226\times10^{-12}</td>
<td>1.363\times10^{-11}</td>
<td>2.498\times10^{-13}</td>
</tr>
<tr>
<td>m16 2 8</td>
<td>1.408\times10^{-07}</td>
<td>4.775\times10^{-08}</td>
<td>4.024\times10^{-13}</td>
<td>4.412\times10^{-07}</td>
</tr>
<tr>
<td>m16 2 10</td>
<td>5.655\times10^{-09}</td>
<td>5.277\times10^{-12}</td>
<td>1.306\times10^{-10}</td>
<td>1.533\times10^{-07}</td>
</tr>
<tr>
<td>m16 1 13</td>
<td>1.501\times10^{-07}</td>
<td>8.838\times10^{-06}</td>
<td>1.593\times10^{-05}</td>
<td>9.161\times10^{-07}</td>
</tr>
</tbody>
</table>

Table 5.2: Resulting p-values for each pattern when several CSR tests based on Quadrat-Counts (QC) are executed -configurations (5 × 1), (5 × 2), (5 × 3) and (15 × 1). Values over 0.1 are marked with *; values over 0.05 are shown with † and those over 0.01 are indicated with †.
5.1.2 Complete Spatial Randomness

In general, both quadrat count and Kolmogorov-Smirnov tests reject the null hypothesis of the homogeneous Poisson process in almost all samples (except from two outliers belonging to G85 group: “if6 1 1” and “if6 2 2”). This can be seen in more detail in Table 5.2 and Table 5.3. There, results corresponding to the aggregated outliers belonging to G85 group: “if6 1 1” and “if6 2 2”). This can be seen in more detail in Table 5.2 and Table 5.3. There, results corresponding to the aggregated test for each group are not shown because attained p-values are zero in all cases. In either case, we can formally ensure that dendritic spine distribution over dendrite cannot be considered purely random (which clarify previous results on the topic); however, it is not clear whether this departure is caused by the clear variation on the average point density or there also exist some kind of interaction among points.

<table>
<thead>
<tr>
<th>Dendrite</th>
<th>QC_{(15 \times 2)}</th>
<th>QC_{(15 \times 3)}</th>
<th>KS_x</th>
<th>KS_y</th>
</tr>
</thead>
<tbody>
<tr>
<td>if6 1 1</td>
<td>*2.425 \times 10^{-01}</td>
<td>9.419 \times 10^{-02}</td>
<td>17.215 \times 10^{-02}</td>
<td>7.215 \times 10^{-02}</td>
</tr>
<tr>
<td>if6 1 2</td>
<td>8.925 \times 10^{-07}</td>
<td>4.726 \times 10^{-09}</td>
<td>2.917 \times 10^{-09}</td>
<td>2.917 \times 10^{-09}</td>
</tr>
<tr>
<td>if6 1 3</td>
<td>1.569 \times 10^{-05}</td>
<td>9.166 \times 10^{-05}</td>
<td>6.161 \times 10^{-05}</td>
<td>6.161 \times 10^{-05}</td>
</tr>
<tr>
<td>if6 1 4</td>
<td>2.192 \times 10^{-05}</td>
<td>4.719 \times 10^{-07}</td>
<td>3.175 \times 10^{-08}</td>
<td>3.175 \times 10^{-08}</td>
</tr>
<tr>
<td>if6 1 5</td>
<td>1.351 \times 10^{-07}</td>
<td>5.195 \times 10^{-08}</td>
<td>5.966 \times 10^{-05}</td>
<td>5.966 \times 10^{-05}</td>
</tr>
<tr>
<td>if6 1 6</td>
<td>1.481 \times 10^{-04}</td>
<td>1.345 \times 10^{-04}</td>
<td>9.200 \times 10^{-07}</td>
<td>9.200 \times 10^{-07}</td>
</tr>
<tr>
<td>if6 2 0</td>
<td>4.078 \times 10^{-07}</td>
<td>1.193 \times 10^{-07}</td>
<td>4.335 \times 10^{-08}</td>
<td>4.335 \times 10^{-08}</td>
</tr>
<tr>
<td>if6 2 2</td>
<td>7.394 \times 10^{-02}</td>
<td>4.975 \times 10^{-03}</td>
<td>1.816 \times 10^{-03}</td>
<td>1.816 \times 10^{-03}</td>
</tr>
<tr>
<td>if6 2 11</td>
<td>2.115 \times 10^{-03}</td>
<td>9.261 \times 10^{-06}</td>
<td>1.449 \times 10^{-04}</td>
<td>1.449 \times 10^{-04}</td>
</tr>
<tr>
<td>if6 2 20</td>
<td>9.662 \times 10^{-04}</td>
<td>1.201 \times 10^{-03}</td>
<td>1.060 \times 10^{-06}</td>
<td>1.060 \times 10^{-06}</td>
</tr>
<tr>
<td>m16 1 1</td>
<td>9.642 \times 10^{-03}</td>
<td>1.294 \times 10^{-04}</td>
<td>6.917 \times 10^{-04}</td>
<td>6.917 \times 10^{-04}</td>
</tr>
<tr>
<td>m16 1 3</td>
<td>1.454 \times 10^{-04}</td>
<td>3.523 \times 10^{-06}</td>
<td>1.635 \times 10^{-05}</td>
<td>1.635 \times 10^{-05}</td>
</tr>
<tr>
<td>m16 1 6</td>
<td>2.262 \times 10^{-05}</td>
<td>3.454 \times 10^{-06}</td>
<td>8.354 \times 10^{-07}</td>
<td>8.354 \times 10^{-07}</td>
</tr>
<tr>
<td>m16 1 8</td>
<td>6.065 \times 10^{-12}</td>
<td>1.650 \times 10^{-17}</td>
<td>1.206 \times 10^{-10}</td>
<td>1.206 \times 10^{-10}</td>
</tr>
<tr>
<td>m16 1 9</td>
<td>5.501 \times 10^{-11}</td>
<td>1.464 \times 10^{-10}</td>
<td>6.933 \times 10^{-13}</td>
<td>6.933 \times 10^{-13}</td>
</tr>
<tr>
<td>m16 2 8</td>
<td>2.265 \times 10^{-11}</td>
<td>1.959 \times 10^{-18}</td>
<td>1.201 \times 10^{-05}</td>
<td>1.201 \times 10^{-05}</td>
</tr>
<tr>
<td>m16 2 10</td>
<td>3.417 \times 10^{-10}</td>
<td>6.899 \times 10^{-09}</td>
<td>7.382 \times 10^{-07}</td>
<td>7.382 \times 10^{-07}</td>
</tr>
<tr>
<td>m16 1 13</td>
<td>6.221 \times 10^{-07}</td>
<td>3.325 \times 10^{-06}</td>
<td>4.364 \times 10^{-04}</td>
<td>4.364 \times 10^{-04}</td>
</tr>
</tbody>
</table>

Table 5.3: Resulting p-values for each pattern when the rest of CSR tests based on Quadrat-Counts (QC) configurations (15 × 2) and (15 × 3) – and Kolmogorov-Smirnov (KS) are executed. Values over 0.1 are marked with *; values over 0.05 are shown with † and those over 0.01 are indicated with ‡.
5.1. Non-parametric analysis

Figure 5.2: Non-parametric estimations of the intensity; along with the colored representation of the function, averaged values along X and Y coordinate axis are shown for more clarity. In either case, results for G85 (O), G40 (Y) and all patterns together (B) are presented.
5.1.3 Intensity and interaction

Results of the non-parametric estimation of the intensity function can be seen in Figure 5.2. Average plots along the X and Y coordinate axis seems to be consistent with histograms in the preliminary analysis. Once more, there are no clear signs of differences in the spatial trend between groups (except from scale).

![Intensity and interaction plots](image)

Figure 5.3: Computation of aggregated summary characteristics: $F(\cdot)$, $G(\cdot)$, $J(\cdot)$ and $L(\cdot)$ in reading order. All patterns are considered together; the aggregated value is shown in black, individual values are shown in gray and the theoretical Poisson value is shown in red. A slight clustering behaviour can be noticed.

In what respect to inter-point interactions, averages of basic summary functions are computed (Figure 5.3); and the corresponding Monte-Carlo envelope tests are performed. Obtained results show there seems to be some differences among groups though, in all cases, only a small departure from Poisson toward clustering is appreciable (Figure 5.4). This trend is clearly higher in the case of G40 group, at which signs of clustering may be seen specially in empty-space, J and Besag’s L functions; when considering G85 group, part of this behaviour is still noticeable, mainly in the empty-space function.

In either case, these estimators are only defined under stationarity conditions and so results may be invalid. When using the inhomogeneous L-function to account for heterogeneity (Figure 5.5), the following conclusions may be extracted:
5.1. Non-parametric analysis

Figure 5.4: Results of CSR Monte-Carlo goodness-of-fit tests for a number of simulations $n_s = 39$ (maximum significance level $\alpha = 0.05$). Four summary functions are considered – $F(\cdot)$, $G(\cdot)$, $J(\cdot)$, $L(\cdot)$ – and results for G85 (first row), G40 (second row) and all patterns (third row) are shown. Clustering behaviour is confirmed.

Figure 5.5: Results of CSR Monte-Carlo goodness-of-fit tests for a number of simulations $n_s = 39$ (maximum significance level $\alpha = 0.05$). In this case, only the $L_{inhom}(\cdot)$ summary function is considered and results for all patterns (left), G40 (centre) and G85 (right) are shown. Note that red dashed line is no more the theoretical value for CSR but the empirical mean computed from simulations.
• Both clustering and regularity behaviours are present.

• Regularity is shown at short ranges \( r \sim [0, 0.3] \) in G40 and G85 groups.

• Clustering is shown at larger ranges \( r \sim [1.3, 1.5] \) only in G40 group.

In either case, only a slight departure from Poisson assumption is detected and this is significantly more noticeable in the case of clustering in G40. Thus, this fact can be considered as the first clear evidence on the existence of differences between pooling groups (except from intensity function scale).

5.1.4 Group differentiation

In this context, group differences may be appreciated in more detail in Figure 5.6. A quick visual inspection reflects that both age groups share a similar behaviour up to \( r \sim 0.5 \), from where G40 group starts to show a more clear trend towards clustering. In what respect to location-related groups, in general, regions further from the soma seem to be a bit over-aggregated. However, differences are not so appreciable as in
5.1. Non-parametric analysis

the case of age groups. Furthermore, Monte-Carlo tests provide some formal results which are summarized in Table 5.4; it can be noticed that only differences related to age yield $p$-values that are close to statistical significance. In this context, we may conclude that there is some evidence on the influence of aging in dendritic spine distribution but it is not so significant, at least under the assumptions we have imposed in this statistical analysis.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Age</th>
<th>Length (2)</th>
<th>Length (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$-value ($\sigma_p$)</td>
<td>0.1 (0.03)</td>
<td>0.69 (0.04)</td>
<td>0.52 (0.04)</td>
</tr>
</tbody>
</table>

Table 5.4: Formal results of Monte-Carlo tests for the null hypothesis of no differences among groups. For each considered configuration, both the mean and the standard deviation of obtained $p$-values are shown.

5.1.5 Concluding remarks

In what respect to non-parametric analysis, we extract the following conclusions:

1. There are clear differences between groups in what respect to both point count ($\bar{n}_{G40} = 282.5$ vs. $\bar{n}_{G85} = 112.4$) and dendrite length ($\bar{l}_{G40} = 233.9$ vs. $\bar{l}_{G85} = 124.8$). Furthermore, in both cases, values seem to be overdispersed, which is specially noticeable in the case of G85 ($CV_{G40}^{n} = 0.209$ vs. $CV_{G85}^{n} = 0.474$); this may be due to a non-uniform reduction in spine number caused by age, i.e. differences depending on the region of the brain.

2. We can ensure that dendritic spine distribution over dendrite surface cannot be considered Complete Spatial Randomness. In general, non-uniformity seems to be a more probable reason than departure from independence assumption.

3. There exists a clear spatial trend in the average point density, whose shape is shown at Figure 5.2. Variation along X axis is suitable, but non-uniform distribution along Y axis might be an artifact consequent of data sampling.

4. There are some differences between age groups in what respect to inter-point interaction. Both groups show a slight Poisson departure towards regularity at $r \sim [0, 0.3]$, while G40 group is noticeably more aggregated in the range $r \sim [1.3, 1.5]$. Formal results state that this difference is only close to significance ($p = 0.1$ with $\sigma_p = 0.03$).
5.2 Parametric analysis

This section includes: (i) point count model selection, (ii) Poisson inhomogeneous fitting, selection and validation (iii) Area-interaction model fitting, selection and validation and finally (iv) concluding remarks.

5.2.1 Point count modelling

Several quasi-Poisson models have been fitted and a summary of results are shown in Table 5.5; in general, it can be seen that only age, length and area influence the number of points in each party. Moreover, there is not significant improvement on considering non-linear models, and so both \( n \sim Age + Length \) and \( n \sim Age + Area \) are suitable models (Table 5.6). In either case, fits are quite overdispersed (\( \phi = 10.323 \) and \( \phi = 10.255 \)), which indicates a clear departure from Poisson assumption or missing covariates. Dispersion may be reduced a bit by considering non-linear models, but the improvement does not seem to be significant.

<table>
<thead>
<tr>
<th>Model</th>
<th>R. Dev.</th>
<th>Wald Sig.</th>
<th>( \phi )</th>
<th>Comparison</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>337.3496</td>
<td>Yes</td>
<td>19.3841</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Width (W)</td>
<td>331.9879</td>
<td>No</td>
<td>20.1966</td>
<td>Null</td>
<td>0.2655</td>
<td>0.6134</td>
</tr>
<tr>
<td>Length (L)</td>
<td>264.6756</td>
<td>Yes</td>
<td>16.1232</td>
<td>Null</td>
<td>4.5074</td>
<td>*0.0497</td>
</tr>
<tr>
<td>Area (A)</td>
<td>247.1302</td>
<td>Yes</td>
<td>14.9228</td>
<td>Null</td>
<td>6.0457</td>
<td>*0.0257</td>
</tr>
<tr>
<td>Age</td>
<td>218.4621</td>
<td>Yes</td>
<td>13.2840</td>
<td>Age</td>
<td>8.9497</td>
<td>*0.0087</td>
</tr>
<tr>
<td>Age+W</td>
<td>218.3954</td>
<td>No</td>
<td>14.1812</td>
<td>Age</td>
<td>0.0447</td>
<td>0.9462</td>
</tr>
<tr>
<td>Age+L</td>
<td>160.5991</td>
<td>Yes</td>
<td>10.3225</td>
<td>Age</td>
<td>5.6055</td>
<td>*0.0318</td>
</tr>
<tr>
<td>Age+A</td>
<td>163.0489</td>
<td>Yes</td>
<td>10.2549</td>
<td>Age</td>
<td>5.4063</td>
<td>*0.0345</td>
</tr>
<tr>
<td>Age×L</td>
<td>159.1214</td>
<td>No</td>
<td>10.9797</td>
<td>Age+A</td>
<td>0.1346</td>
<td>0.7192</td>
</tr>
<tr>
<td>Age×A</td>
<td>162.9901</td>
<td>No</td>
<td>10.9880</td>
<td>Age+A</td>
<td>0.0054</td>
<td>0.9427</td>
</tr>
<tr>
<td>Age+L(^2)</td>
<td>143.7409</td>
<td>No</td>
<td>9.8572</td>
<td>Age+L</td>
<td>1.7102</td>
<td>0.2120</td>
</tr>
<tr>
<td>Age+A(^2)</td>
<td>133.6017</td>
<td>No</td>
<td>9.0803</td>
<td>Age+A</td>
<td>3.2430</td>
<td>0.0933</td>
</tr>
<tr>
<td>Age×L(^2)</td>
<td>117.5806</td>
<td>No</td>
<td>9.3382</td>
<td>Age+L</td>
<td>1.5356</td>
<td>0.2560</td>
</tr>
<tr>
<td>Age×A(^2)</td>
<td>130.9623</td>
<td>No</td>
<td>10.3404</td>
<td>Age+A</td>
<td>1.0343</td>
<td>0.4123</td>
</tr>
</tbody>
</table>

Table 5.5: Results of model selection for point count modelling. Several model fits are shown specifying in this order: (i) residual deviance, (ii) coefficients significance based on Wald tests, (iii) estimated dispersion parameter, \( \phi \) (iv) model to compare with, (v) F-statistic value and (vi) final p-value of the F-test. Significant improvements are marked with *. 
5.2. Parametric analysis

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4,6469</td>
<td>0,2928</td>
<td>15,8684</td>
<td>&lt;2.20·10^{-16}</td>
</tr>
<tr>
<td>Age</td>
<td>0,3125</td>
<td>0,0985</td>
<td>3,1722</td>
<td>0,0015</td>
</tr>
<tr>
<td>Length</td>
<td>0,0029</td>
<td>0,0012</td>
<td>2,3707</td>
<td>0,0177</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4,6844</td>
<td>0,2823</td>
<td>16,5920</td>
<td>&lt;2.20·10^{-16}</td>
</tr>
<tr>
<td>Age</td>
<td>0,2852</td>
<td>0,0998</td>
<td>2,8580</td>
<td>0,0120</td>
</tr>
<tr>
<td>Area</td>
<td>0,0007</td>
<td>0,0003</td>
<td>2,3230</td>
<td>0,0347</td>
</tr>
</tbody>
</table>

Table 5.6: Estimations of selected models for point count modelling –Age+Length (top) and Age+Area (bottom)– along with coefficients standard errors and significances based on Wald tests.

5.2.2 Inhomogeneous Poisson modelling

In general, several inhomogeneous Poisson models are fitted considering polynomial forms of the intensity function \( P_n(x, y), n = 0, \ldots, 8 \), as well as the non-parametric estimation based on kernel smoothing (denoted NP). We are specially worried about overfitting and interested on capturing only the level of intensity variation that is enough to let the points interact with each other. Fitting Area-interaction models is computationally expensive, and so we first select a tentative set of potentially valid spatial trends using inhomogeneous Poisson models.

![Figure 5.7: Comparison between several inhomogeneous Poisson models (from degree 0 to degree 8), showing the effect on AICc, BIC and CV prediction error of considering proportional intensities across patterns. Less is better.](image)

In either case, we allow intensity function to vary proportionally across patterns; it can be seen in Figure 5.7 that this yields a noticeable improvement in performance
for all scoring criterion (specially in the case of cross-validated residuals). For this reason, we only consider proportional versions of each model from this point. In this context, we obtain formal results based on classical likelihood ratio $\chi^2$ tests, for each of the nested Poisson models (Table 5.7); in summary, improvements on performance are considered to be statistically significant up to 6th degree. More precisely, specially large reductions in deviance are achieved at levels $\{P_2, P_4, P_6\}$, and no clear differences are detected across groups.

<table>
<thead>
<tr>
<th>Model (x, y)</th>
<th>Dev. (\times 10^{-16})</th>
<th>(p)-value</th>
<th>Dev. (\times 10^{-16})</th>
<th>(p)-value</th>
<th>Dev. (\times 10^{-16})</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_1)</td>
<td>111.360</td>
<td>&lt;2.20</td>
<td>155.998</td>
<td>&lt;2.20</td>
<td>261.820</td>
<td>&lt;2.20</td>
</tr>
<tr>
<td>(P_2)</td>
<td>319.750</td>
<td>&lt;2.20</td>
<td>224.482</td>
<td>&lt;2.20</td>
<td>531.540</td>
<td>&lt;2.20</td>
</tr>
<tr>
<td>(P_3)</td>
<td>113.090</td>
<td>2.20</td>
<td>63.226</td>
<td>6.080</td>
<td>163.570</td>
<td>&lt;2.20</td>
</tr>
<tr>
<td>(P_4)</td>
<td>87.010</td>
<td>2.20</td>
<td>69.661</td>
<td>1.210</td>
<td>150.270</td>
<td>&lt;2.20</td>
</tr>
<tr>
<td>(P_5)</td>
<td>24.550</td>
<td>4.14</td>
<td>30.343</td>
<td>3.380</td>
<td>42.950</td>
<td>&lt;2.20</td>
</tr>
<tr>
<td>(P_6)</td>
<td>72.200</td>
<td>5.30</td>
<td>73.708</td>
<td>2.620</td>
<td>130.970</td>
<td>&lt;2.20</td>
</tr>
<tr>
<td>(P_7)</td>
<td>1.190</td>
<td>0.997</td>
<td>15.314</td>
<td>0.0533</td>
<td>5.560</td>
<td>0.6969</td>
</tr>
<tr>
<td>(P_8)</td>
<td>16.460</td>
<td>0.0579</td>
<td>7.528</td>
<td>0.5820</td>
<td>14.640</td>
<td>0.1014</td>
</tr>
</tbody>
</table>

Table 5.7: Results of the analysis of variance for the considered inhomogeneous Poisson models (from degree 0 to degree 8). The reduction in deviance (2 times log-pseudolikelihood ratio) as each nested model is considered, along with resulting \(p\)-values of two-sided $\chi^2$-tests, are shown. Results for different grouping configurations (G40, G85 and all together) are plotted in each column.

When considerings other scoring criterion, some interesting results are found:

- As expected, the pseudolikelihood value (PL) gives preference to higher degree models, even over non-parametric estimations of the intensity function. This has only been considered for comparative purposes.

- Akaike information criterion (AICc) and cross-validation (CV) both match results provided by ANOVA; the former considers 6th degree polynomials to be the best fit for the model, while the latter show preference for the semi-parametric approach but accept 6th degree as suitable (e.g. CV\(^6\) > 0.9). Notice that neither AICc nor BIC have been used to score the semi-parametric fit, as the number of parameters in this model has no real sense (just one parameter corresponding to a proportional covariate).

- BIC gives preference to most parsimonious models. Polynomials of 4th degree model are considered to be or be really close to the best fit (BIC\(^4\)\(_{G85}\) = 0.987 and
Figure 5.8: Score of increasingly complex inhomogeneous Poisson models based on pseudolikelihood (PL), the cross-validated prediction error (CV), Akaike (AICc) and Bayesian (BIC) information criterion. Values are normalized so that they can be properly compared. More precisely, proportional log-polynomial processes are considered (from degree 1 to degree 8), as well as the semi-parametric model based on kernel density estimation (NP). Note that null model (homogeneous Poisson) corresponds to value 0 of performance in all measures and is not shown.

$\text{BIC}_{G40}^1 = 1)$. This is the unique performance scoring measure that contradicts usual techniques, and so we are specially interested on finding if it is suitable.
In general, we may conclude that some parsimonious models – that are rejected by the classical likelihood ratio test – are also considered valid by the introduced scoring criterion (BIC). This might imply these models already generalize the overall behaviour, and capture enough information on the spatial variation to let points interact with each other. Thus, we build a tentative set of possible spatial trends, considering a general selection of both precise and parsimonious models ($P_2$, $P_4$, $P_6$ and NP).

\[
\begin{array}{cccccc}
\text{Model} & \text{KS}_x & \text{KS}_y & \text{KS}_x & \text{KS}_y & \text{KS}_x & \text{KS}_y \\
\hline
P_2(x, y) & 1.7 \cdot 10^{-7} & 1.5 \cdot 10^{-5} & 0.0073 & 0.0019 & 0.0046 & 0.0002 \\
P_4(x, y) & 0.2179 & 0.0170 & 0.5632 & 0.1680 & 0.4803 & 0.1947 \\
P_6(x, y) & 0.7150 & 0.7623 & 0.6562 & 0.9060 & 0.8697 & 0.7278 \\
NP & 0.7250 & 0.7890 & 0.7340 & 0.8620 & 0.7210 & 0.8320 \\
\end{array}
\]

Table 5.8: Resulting \( p \)-values of Komogorov-Smirnov tests for several Poisson models; more precisely, some log-polynomial forms of the spatial trend ($P_2$, $P_4$, $P_6$), as well as the semi-parametric model based on kernel density (NP), are considered.

Validation results for this model based on Kolmogorov-Smirnov tests are shown in Table 5.8. In general, all models are considered suitable except the log-quadratic polynomial form of the intensity. Informal analysis of raw residuals lurking variables confirm this fact (Figure 5.9).

### 5.2.3 Area-interaction Gibbs modelling

In general, several Area-interaction models are fitted (for proportional spatial trends $P_2$, $P_4$, $P_6$, NP and $r \in [0.1, 1.5]$ interaction ranges). As it can be seen in Figure 5.10, resulting interaction parameters highly depend on the selected interaction range. Two important facts may be extracted from this figure:

1. Results are quite consistent with non-parametric analysis in the sense that the interaction behaviour responds to the variation of the range. It can be seen that, when low values are considered ($r \in [0.1, 0.25]$), resulting parameters reflect a slight trend to regularity in the G85 group. For higher values ($r \in [0.3, 1.5]$), estimates indicate that points are a bit over-aggregated. In either case, fitted interaction values are noticeable higher in the case of the G40 group.

2. Estimates of the interaction parameter are almost equal for all tested spatial trends except from that of log-quadratic form; i.e. increasing the adjustment
Figure 5.9: Lurking variables of raw residuals for the $P_2$ Poisson model. Results state that the process is grossly inadequate to model the spatial variation of the intensity.

level in the overall variation of the intensity does not produce changes in inter-point interaction. This may indicate that the parsimonious model ($P_4$), which was initially rejected by classical selection techniques but preferred by one of the considered scoring criterion (BIC), already achieves a good generalization of the spatial trend.

Model profiling outcomes for the different values of the interaction range can be seen in Figure 5.11. If we ignore models corresponding to a log-quadratic spatial trend, some interesting results are shown. Considering pseudolikelihood (PL), no significant differences can be appreciated between age groups: selected interaction ranges indicate a slight clustering behaviour (e.g. $\eta_{G40}^4 = 1.851$ vs. $\eta_{G85}^4 = 1.407$), which is only slightly superior in the case of G40. However, considering our introduced cross-validation measure (CV), fitted parameters are quite consistent with non-parametric results. In these cases, most likely interaction range is considered (a bit) different for each group ($r_{G40}^* = 0.25$ and $r_{G85}^* = 0.2$). Moreover, corresponding fitted parameters reflect a slight trend towards regularity in G85 ($\eta_{G85}^* \sim 0.7$), while a clustering behaviour in G40 ($\eta_{G85}^* \sim 2.1$).

In this context, Figure 5.12 shows a comparison between inhomogeneous Poisson and Area-interaction models, once interaction ranges maximizing pseudolikelihood
Results for both G85 and G40 groups are shown.

(for BIC) or cross-validated prediction error (for CV) have been selected. In either case, results show a noticeable improvement in performance when using the Gibbs model. In general, the performance based on BIC behaves in the same manner as Poisson models: it gives preference to parsimonious models (4th degree). However, CV seems to have an inconsistent behaviour, showing preference for either 2nd or 6th degree forms. Considering this fact in combination with the effect of increasing the degree of the spatial trend, we conclude that an Area-interaction model with $P_4$ intensity function provides a preferable fit for this process.

Validation results for this model are based on residuals. As simulation has not been implemented yet in the context of pattern replication, the overall model is only assessed by means of a informal observation of cumulative sums of residuals (lurking variables). In general, these behave in a similar manner as they do in Poisson models, and so it can be considered that these models provides a reasonable fit for the spatial variation of the intensity.

Inter-point interactions are further checked by using Q-Q plots of residuals; those are based on simulations from the fitted model and so the process is individually tested at each of the considered patterns. Obtained results should be carefully analysed. On one hand, it can be seen that most patterns may be considered a suitable instance from an Area-interaction process, but most of them also fit with pure Poisson behaviour (Figure 5.13). In fact, even some of them seem to yield better fits when inter-point interactions are not considered. This issues will be further analysed in the discussion section of chapter 5.
5.2. Parametric analysis

Figure 5.11: Results of profile pseudolikelihood for the selection of the irregular parameter $r$. Both groups \{G85, G40\} are considered with several inhomogeneous Area-interaction models: \{$P_2$, $P_4$, $P_6$, NP\}. Selection is performed using normalized values of the pseudolikelihood (PL) and the cross-validated prediction error (CV). Best fit in each case is indicated in dashed lines, along with the estimated parameter.
Figure 5.12: Comparison between the performance (on BIC and CV) of Poisson and Area-interaction models. Results for both G85 and G40 are shown; in either case, the improvement is noticeable for all forms of the spatial trend.

5.2.4 Concluding remarks

In what respect to parametric analysis, we extract the following conclusions:

1. Point count is reasonably well represented by a quasi-Poisson log-linear model of the form $n \sim Age + Length$ or $n \sim Age + Area$; in either case, there are signs of overdispersion ($\phi = 10.323$ and $\phi = 10.255$), which indicates a clear departure from Poisson assumption or missing covariates.

2. In both groups, inhomogeneous Poisson model provides a quite reasonable fit, for any log-polynomial spatial trend of degree $n \geq 4$, as well as for the non-parametric estimation based on kernel smoothing. However, it fails to properly model inter-point interaction in some cases, and for this reason also Area-interaction model is considered.

3. Area-interaction model provide better results in both cases (in terms of scoring measures) and these are quite consistent with non-parametric analysis; i.e. the model yields to a regularity behaviour at short ranges in G85 ($r^*_G85 = 0.2$ and $\eta^*_G85 = 0.691$) and to over-aggregation at a bit larger range in G40 ($r^*_G40 = 0.25$ and $\eta^*_G85 = 2.136$).

4. In general, usual measures of performance –likelihood ratio test and AIC– give preference to high order models (e.g. $P_6$). However, employed Bayesian
scoring criterion (BIC) select more parsimonious models that have been shown to be also valid, and so may be interesting to avoid overfitting in this field. Moreover, using the introduced cross-validated prediction error (CV), a kind of group differentiation behaviour arise in parametric analysis.

5. Model validation states that both Area-interaction and inhomogenous Poisson processes may be suitable models for this scenario. Better fits are shown by either one or the other depending on the particular fitted pattern.
Figure 5.13: Q-Q plot of Pearson residuals for Area-interaction and Poisson inhomogenous model with $P_4$ spatial trend, fitted to several patterns. Colors indicate: Poisson superiority (red), Gibbs superiority (green) an equality (blue).
Chapter 6

Conclusions

In this chapter we finally present a brief discussion about results as well as employed methodology; this is based on several comments which were kindly supplied by the supervisors of this Master Thesis, experts on neuroscience from Laboratorio Cajal de Circuitos Corticales (CSIC) and some other authors familiar to spatial statistics. Finally, some additional conclusions on the main contributions of this work and future research lines are provided.

6.1 Main contributions

We consider the main contributions of this work what follows:

1. A handicap of this work lies on the fact that just a few number of publications in the literature have dealt before with the particular scenario of replicated spatial point patterns (about a dozen). In general, those are quite dispersed and do not provide a extensively used and sound framework, but a small catalog of techniques that should be considered independently. In that sense, the first contribution of this work is to provide a comprehensive review on this topic, so it can be used in future works. This survey complements an extensive introduction to spatial point processes theoretical foundations, which has been written with a didactic purpose to allow initiates to learn from the very basics to a level that prepares them to conceptually understand state-of-the-art works. In that sense, it can be also considered a contribution of this work.

2. In practical terms, spatstat package has been used as the unique tool for performing the different methods and algorithms which have been necessary to employ during the analysis. However, as we have introduced before, it lacks of explicit support for replicated spatial point processes; thus, it has been necessary to implement some of those techniques which are described in chapter
3, mainly: (i) aggregation of summary characteristics, (ii) cross-validation of non-parametric intensity functions, (iii) group differentiation based on Besag’s $L$-function and (iv) model fitting for replicated patterns (along with cross-validation of models). In either case, implementations have been done trying to maintain the flexible framework of **spatstat**, so these can be easily extended and reused in future works. In addition, it has been also necessary to provide some additional code to consider the custom edge-correction method, though these are ad-hoc methods and cannot be easily reused.

3. In what concerns to methodological contributions, it has been seen that works on pattern replication are not usually worried about overfitting and so we have focused on this issue by trying to use both BIC and cross-validation prediction errors in model fitting. The former has been shown to be a valid alternative to AIC, while the latter provide interesting results when is used to select irregular parameters.

4. Finally, the case study may be considered the final contribution of this work. An interesting scientific question related with neuroscience has been studied using this novel approach, and some preliminary conclusions are extracted which provide a good starting point for the extension of the work. These are specified in more detail in the next section and have been recently presented at an international conference on spatial statistics and its applications [66]. In summary, the study provides some formal results on the effect of ageing on dendritic spine distribution over dendrite surface (a reduction in clustering among points), as well as two tentative models (inhomogeneous Poisson and Area-interaction) for being used for inference and simulation.

### 6.2 Biological interpretation

Obtained results have been interpreted by experts on the application field, looking for possible biological justifications. In that sense, three main conclusions have been extracted:

1. **Short-range interaction is consistent with spines variable morphology.** Results (mainly non-parametric ones) show a slight trend towards regularity at $r \sim [0.1, 0.3]$ which is common for both G40 and G85 groups. However, we cannot consider the existence of a clear minimum nearest-neighbor distance (hardcore process). The main conjecture is that this smooth inhibition among points is a consequence of spine morphology (highly variable in both shape and volume). There might by other reasons such as a possible competition for “resources” (e.g. synapses with other neurons). However, the range of interaction is quite
consistent with spines’ approximate size and so we consider available space as the most likely reason for this regularity behaviour.

2. **Mid-range interaction is consistent with the aggregation of synapses.** Both parametric and non-parametric results show a slight trend towards clustering at $r \sim [1.0, 1.5]$, specially in the case of G40 group. The main conjecture is that this may arise from a patchy distribution (instead of a homogeneous distribution) of axon terminals forming synapses with spines. That means the clustered behaviour may be a consequence of the existence of axon terminals grouped together (interacting with the dendrite), or caused by spines exposure to common signals coming from a single neuron.

3. **Age-dependent differences may be relate to cognitive deficits.** Both parametric and non-parametric results show the existence of some differences between age-groups, in what respect to inter-point interactions. These are not so significant and may be subject to criticism, but are still consistent with changes associated to cognitive deficits of normal aging. That means we may relate the clustering behaviour to neuronal activity (which is assumed to be higher in a young brain such as G40), and so suppose that clustering is lost as an effect of ageing. As the regularity behaviour has been related to spines size and shape, it is logical that this is still present in both groups.

In general, these interpretations provide a good starting point for extending the work to consider more complex scenarios. However, it should be taken into account that only samples from a few neurons of two different individuals have been considered, and so it is not possible to state definitive conclusions.

### 6.3 Discussion on data limitations

As it would be expected from any case study, dealing with real data sampling and preprocessing implies assuming some risks on the reliability of the obtained results. This is specially usual on biology related applications, in which analysed structures are often complex and data extraction cannot be always automatized. In our particular case, data have imposed some limitations that must be taken into consideration when analysing results.

#### 6.3.1 Data extraction

As we have already explained, data sampling is performed using Confocal Laser Scanning Microscopy. This is an already established technique which has been repeatedly employed in other neuroscience applications. As far as we know, achieved precision
Chapter 6. Conclusions

(which is in the order of a few nm) is enough to ensure the reliability of the results. Furthermore, the same data have been used before in another related work [13], and showed to provide valid results. However, final raw data have been extracted by manually identifying spines’ three-dimensional locations, and so we should assume that some kind of bias has been introduced by the human operator. In fact, variation of the overall intensity along the width of the patterns (around the dendrite) seems to be quite unnatural; dendrites’ absolute orientation is lost during the extraction of sample material, and so it is unlikely that this higher density on the sides of the dendrites is biologically justified. We have conjectured that @Imaris camera point-of-view may have caused these artifacts, and so all results, and specially those related with intensity variation on the Y axis, should be subjected to criticism.

6.3.2 Preprocessing stages

In addition to raw data, preprocessing stages also introduce some bias that should be taken into account. On one side, straightening operation clearly modify inter-point Euclidean distances at locations where polyline components are aligned. However, we are interested on geodesic distances throughout dendrite surface, and those are almost equal in both cases, i.e. only a small bias towards larger distances is present. On the other side, registering and scaling operations affect both intensity variation and inter-point distances; spatial trend could be still considered proportional across patterns, but modifying inter-point distances may yield to invalid results. As a consequence, we should consider non-parametric analysis (which do not require pattern scaling) to be more reliable, and again subject to criticism obtained results.

6.4 Discussion on employed methodology

Due to initial inexperience, several assumptions have been done in what respect to the employed framework and some other decisions taken during the analysis. This mainly affect the use of planar representations and some issues regarding parametric analysis.

6.4.1 Planar representations

Three-dimensional point patterns can be seen as an alternative approach to the study of this particular problem. That scenario was considered at the early stages of the analysis but discarded based on two main reasons: (i) spines are fixed to dendrite and so there is few sense on considering the additional degree of freedom and (ii) we are mainly interested on finding how the available space over dendrites’ surface is allocated among spines. As we have seen, preprocessing stages impose some bias
on the results that must be taken into account, and so there is also a counterpart on using planar representations to the representation of this kind of point processes.

6.4.2 Parametric analysis

In what respect to parametric analysis, it is necessary to consider that Poisson inhomogeneous is already a reasonable model for the distribution of dendritic spines over dendrite surface. Departure from Poisson assumption has been detected in both groups, but this is not as significant as it could be expected for the requirement of inter-point interaction modelling (only a slight difference from local envelopes of significance $\alpha = 0.05$). The use of an Area-interaction model respond to the possibility of improving these obtained results, but also to an academic purpose, i.e. the study of Gibbs point processes; the choice of the particular model is justified by its capability of representing both moderate clustering and regularity behaviours. Thus, further work should be done on determining whether modelling this random phenomenon really requires the consideration of inter-point interactions and also whether Area-interaction is the most adequate model in this scenario (since no other Gibbs models have been tested).

In general, we consider three main possible weaknesses in our parametric analysis using Area-interaction processes. (i) First, results are not so robust, in the sense that there exist quite variability on parameter estimations and most likely configurations, and those not always yields interesting results. This may be due to artifacts arising from quadrature randomisation or to the irregular shape of the window. (ii) Second, estimates of the interaction parameter are, in general, not so far from Poisson theoretical value. Furthermore, values for those interaction ranges close to zero are known to be unstable and so parametric results at these regions should be assumed carefully. (iii) Finally, selection results show some improvement with respect to the reference model, but validation by means of Q-Q plots states once more that Poisson inhomogeneous is also a suitable model in this scenario.

Apart from these issues, it must be noticed that, though introduced scoring criterion (BIC and CV) provide quite interesting selection outcomes, these have not been proven to be valid in any published work; for this reason, current results should be considered preliminary and may be subject to criticism.

6.5 Concluding remarks and future work

Spatial point processes have been proven to be a useful tool for the analysis of the number and distribution of random patterns of points throughout the space. Within this field the replicated scenario has just started to be explored. This particular approach is promising as it makes use of available data to provide more reliable results,
account for generalization and some other unique capabilities. In this context, we have contributed to the literature on the field with a deep introduction to theoretical foundations of spatial point processes, along with a comprehensive review on replicated spatial point patterns.

Moreover, we provide an implementation of some of these novel techniques, which can be easily extended and reused in future works. Then those are applied to the study of a scientific problem related with neuroscience: the spatial distribution of dendritic spines over dendrite surface. During the analysis, a few novel ideas are applied and results provide both a tentative model and some preliminary conclusions, which can be used as a starting point for future works in the field. More precisely, the following extensions are still to be considered:

1. Much work is to be done in what concerns to pattern modelling. In this Master Thesis, we have suggested a possible process and obtained quite good results, but there are many other possibilities. One straightforward improvement is the consideration of variable interaction ranges as it is shown in [37], which may be quite interesting, and potentially consistent with spines size and shape variability.

2. Related to that is the field of marked point patterns [50, 51], which allow us to attach properties to each of the points in the process and consider them during the analysis. In that sense, there are currently available data regarding spines length, surface, volume or even a three-dimensional model of the object, which might be used to extract shape or class markers. The latter is being currently studied, by means of clustering techniques for 3D models.

3. The real expressive power when considering replicated point patterns lies on the use of mixed-effects, allowing model parameters to randomly vary across samples [18, 50, 78]. This is also related with the first point and has not been exploited yet, though some of these approaches have been reviewed in this work.

From a more general point of view, the long-term goal of this work is to find a way of bringing spatial point processes closer to the field of machine learning, in order to adapt existing algorithms to deal with the spatial distribution of points. Up to the moment, the field has only been considered from the approach of statistics, which provide a whole set of tools but lacks of a general framework that could be easily adapted to multiple scenarios. In this context, we are particularly interested on probabilistic graphical models (PGM), because of their known flexibility and generalization capabilities. Related to this, nothing but some preliminary ideas have been published yet and so great efforts need to be put on this research line.
Bibliography


