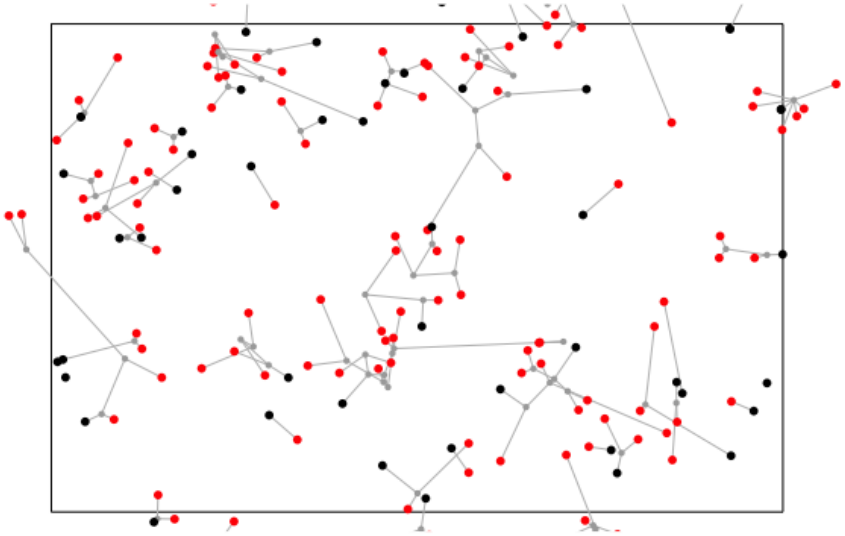

A Spatial Cluster Process Model for the Regeneration of Epidermal Nerve Fiber Patterns

Master's thesis in Engineering Mathematics and Computational Science, MSc

CHUPEI LIN

DEPARTMENT OF MATHEMATICAL SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

A Spatial Cluster Process Model for the Regeneration of Epidermal Nerve Fiber Patterns

CHUPEI LIN

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden 2025

A Spatial Cluster Process Model for the Regeneration of Epidermal Nerve Fiber Patterns
CHUPEI LIN

Supervisor: Aila Särkkä, Department of Mathematical Sciences
Examiner: Serik Sagitov, Department of Mathematical Sciences

Master's Thesis 2025
Department of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A visualization generated in R illustrating the spatial pattern of epidermal nerve fibers (ENFs) within an observation window, simulated using the branching cluster (BC) model developed in this thesis.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

A Spatial Cluster Process Model for the Regeneration of Epidermal Nerve Fiber Patterns

CHUPEI LIN
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Epidermal nerve fibers (ENFs) play a key role in sensory function and are valuable indicators to detect different stages of diabetic neuropathy. Previous studies developed spatial point process models to describe the spatial structure of ENFs and the distribution of their end points. However, modeling how ENFs regrow after damage is still an open problem. In this thesis, we develop a branching cluster (BC) model to describe ENF regeneration by modeling multi-level branching points and end points based on entry points. The model aims to reconstruct nerve patterns in damaged areas, following hypotheses proposed by neurologists. According to these hypotheses, ENF regrowth involves two steps: existing nerve fibers grow into the damaged region, and new ENFs emerge within it. To evaluate the model, we apply a global envelope test using Ripley's K-function to examine the spatial pattern of end points in the damaged region. The results suggest that the BC model captures the observed structure well within small distances. Coverage analysis also shows that the model can regenerate ENF patterns that reflect the healing process. Overall, the BC model offers a statistical framework that incorporates full branching structures to describe and simulate ENF regrowth in damaged skin areas.

Keywords: epidermal nerve fibers (ENFs), point process, cluster process, branching process, damage.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Aila Särkkä, for her continuous support, guidance, and encouragement throughout the course of this thesis. Her insightful feedback and thoughtful advice have been invaluable to my academic and personal growth. I am also thankful to my examiner, Professor Serik Sagitov, for his constructive comments and for taking the time to evaluate my work. Finally, I wish to thank my friends and family for their unwavering support during my studies.

Chupei Lin, Gothenburg, May 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

BC model	Branching Cluster model
CSR	Complete Spatial Randomness
ENFs	Epidermal Nerve Fibers
L-BFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno
MLE	Maximum Likelihood Estimation
NOC model	Non-orphan Cluster model
PDF	Probability Density Function
PMF	Probability Mass Function
QQ plot	Quantile-Quantile plot
UCC model	Uniform Cluster Center model

Contents

List of Acronyms **ix**

List of Figures **xiii**

1 Introduction **1**

2 Methods for Model Construction **5**
2.1 Point processes 5
2.1.1 Spatial point process 5
2.1.2 Cluster process 6
2.2 Goodness-of-fit 7
2.2.1 Ripley's K-function 7
2.2.2 Envelope test 9
2.2.3 Shift plot 11
2.3 Branching process - Galton-Watson process 11

3 Synthetic Data **13**
3.1 Treelike nerve structure 13
3.2 Model for synthetic data 14
3.3 Tools for coverage analysis 16
3.3.1 Number of end points 16
3.3.2 Reactive territory 16

4 The Spatial Cluster Process Model **19**
4.1 Spatial pattern of entry points 20
4.2 Branching process for ENF structure generation 21
4.2.1 Segment length (L): 22
4.2.2 Branching angle (Θ): 22
4.3 Inverse transformation of nerve structures 23

5 Results and Analysis **27**
5.1 Fixed values for model construction 27
5.2 Parameter estimation 27
5.2.1 The length of line segment 27
5.2.2 Branching probability 28
5.3 Goodness-of-Fit Assessment 30
5.4 Coverage analysis 31

5.4.1 Comparison: coverage analysis for a point process incorporating the base model 34

6 Discussion and Conclusion **37**
6.1 Conclusion 37
6.2 Future work 38

Bibliography **39**

List of Figures

3.1 The treelike nerve structure, where the black points are entry points, gray points are branching points, red points are end points, and gray lines are line segments connecting nerve points 13

3.2 An example of synthetic data with the *base model* in an observation window W (10×20 rectangle). Each ENF contains an entry point (black dots), first branching points (gray dots) and end points (red dots). 14

3.3 An illustration of the ENF structure generated by the *base model*: L_1 and L_2 are the length of line segments, and Θ_1 and Θ_2 denote the angles of the line segments relative to the horizontal axis. 15

3.4 Reactive territory of an ENF colored in gray: convex hull reactive territory (left); disc-based reactive territory (right). 17

4.1 A simulation generated from the *BC model*. W is a 10×20 rectangle observation window outlined with black lines. E represents the damaged circular region, indicated by red lines, while C is a smaller circular region within E, marked by pink lines. Outside the red circle, synthetic ENFs are generated using the *base model*. Nerves in the gap G between the red circle zone E and the pink circle zone C extend from the existing nerves, whereas inside the zone C newly regenerated nerves are formed. The black dots denote the entry points, gray dots are the branching points, and red dots represent the end points. . . . 20

4.2 The ENF structure simulated by the *BC model*. Black dots represent entry points, gray dots indicate branching points, and red dots correspond to end points. L denotes the length of the line segments, while Θ_0 and Θ refer to the branching angles. Y represents the number of end points in the tree structure originating from the entry point, and X represents the number of end points in the tree structure originating from a nerve point (either a branching point or an end point). . . 23

4.3 An ENF generated from existing nerves using the *BC model* after applying the inverse transformation. The red circle indicates the damaged area. The nerve structure inside the red circle represents the simulated ENF resulting from the transformation. 25

5.1 Group-wise pooled $L(r) - r$ functions with global envelopes for the end points. The solid curve represents the $L(r)-r$ function calculated from the synthetic data, while the shaded envelope is generated from simulations. The dashed curve indicates the average $L(r)-r$ function across all simulations. In total, 1000 synthetic patterns were analyzed, with 2500 simulations conducted for each pattern. 31
5.2 Shift plot comparing the *BC model* and the synthetic data in terms of the number of end points observed within the damaged area E. . . 32
5.3 The ratio (the relative area covered by regenerated ENFs from the *BC model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the convex hull. 33
5.4 The ratio (the relative area covered by regenerated ENFs from the *BC model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the disc-based approach. 33
5.5 Shift plot comparing the *point process with the base model* and the synthetic data regarding the number end points observed within the damaged area E. 35
5.6 The ratio (the relative area covered by regenerated ENFs from the *point process with the base model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the convex hull. 36
5.7 The ratio (the relative area covered by regenerated ENFs from the *point process with the base model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the disc-based approach. 36

1

Introduction

Epidermal nerve fibers (ENFs) are thin sensory fibers originating from root ganglion cells. These fibers travel through the dermis and extend into the epidermis, the outermost layer of the skin, where they play a crucial role in transmitting sensory signals such as pain, heat, and cold. With the development of staining techniques proposed by Wang et al. (1990) and confocal microscopy studies, Kennedy, Wendelschafer-Crabb and Johnson (1996) confirmed the existence of ENFs. The spatial structure of individual ENFs in the epidermis consists of three types of points: entry points, where the fibers enter the epidermis; branching points, where fibers split to form intricate patterns within the epidermis; and end points, where the fibers terminate at varying levels within the epidermal layer (Kennedy, Wendelschafer-Crabb and Johnson, 1996). Understanding the spatial distribution of these structural points provides valuable insights into potential pathological changes affecting sensory function.

Many studies have highlighted the importance of ENFs in diagnostics, particularly for individuals with diabetic neuropathy (Leong, 2005; Waller et al., 2011; Myllymäki et al., 2012). Investigating the spatial characteristics of ENFs across different stages of diabetic neuropathy is of great significance. Current studies on the spatial patterns of nerve fibers suggest that ENFs exhibit clustering properties, while several researchers (e.g., Kennedy et al., 1996) have noted a significant reduction in ENF density in subjects with diabetic neuropathy. The relevance of analyzing the spatial distribution of nerve points has also been increasingly recognized. Recent investigations have focused more specifically on the spatial analysis of ENF entry and end points. For instance, in skin samples from patients with small fiber sensory neuropathy (SFSN), Leong (2005) quantified changes in ENF spatial distribution, demonstrating increased clustering of ENF entry points. Furthermore, existing studies consistently report that the spatial distribution of ENF entry and end points in individuals with diabetic neuropathy tends to be more clustered than in healthy subjects (e.g., Kennedy et al., 1996; Waller et al., 2011).

Several studies have sought to develop spatial models of ENFs to distinguish different neuropathy statuses, extending previous studies from descriptive analysis to mathematically structural modeling. Existing models primarily focus on the relationship between entry and end points. These models first construct a point process to simulate the spatial patterns of entry points, then create end points based on each entry point. Such models effectively capture the clustered spatial distribution of end points in a 2-d space. Olsbo et al. (2013) introduced a two-stage model, the non-orphan cluster (NOC) model, to describe the spatial distribution of end points, assuming a homogeneous Poisson process for entry points. Andersson, Guttorp and

Särkkä (2016) later proposed the uniform cluster center (UCC) model based on the NOC model, successfully highlighting differences in ENF spatial distributions between healthy and diabetic patterns. Andersson, Rajala and Särkkä (2018) then studied components of the epidermal nerve fibers separately and explored the statistical properties. Additionally, Garcia, Guttorp and Ludwig (2020) proposed a sequential marked point process model that accounted for interactions within entire nerve trees, incorporating both entry and end points as interconnected structures. However, the increased model complexity reduced interpretability.

Andersson, Rajala and Särkkä (2018) emphasized that detailed information on the tree structure could assist neurologists in classifying different stages of diabetic neuropathy. In later studies, researchers argued that incorporating branching points into ENF modeling could provide more detailed insights and improve the accuracy of nerve constructions. Konstantinou and Särkkä (2021) extended the previous work to three-dimensional space and introduced a model that incorporates the first branching point into both the NOC and the UCC model, achieving improved performance in describing the spatial structure of ENFs in both healthy and diabetic subjects.

While previous spatial models primarily focused on describing the structure of ENF, this thesis aims to model the regeneration of ENFs in an damaged area. Based on neurologists' hypotheses, in a region of skin where all nerves have been removed due to damage, intact ENFs surrounding the damaged area would extend toward the center, and new ENFs would regenerate within the affected region. To model this process, the *branching cluster (BC) model* is developed to capture the structure of ENFs. The proposed model aims to provide neurologists with a deeper understanding of healthy nerve regeneration after skin damage by offering a detailed representation of all branching points. Furthermore, it could be applied to explore potential differences in nerve growth between healthy and diabetic conditions.

The study is based on synthetic data generated as follows. Initially, entry point patterns are generated using a Poisson process, followed by the simulation of the first branching points and end points according to the model developed by Konstantinou and Särkkä (2021), which is referred to as the *base model* throughout this thesis. To mimic nerve damage, the nerves and end points within a selected part of the study region are removed. Building on this setup, the BC model that incorporates all branching points is developed to represent the growth patterns of ENFs following damage. For nerves extending from intact ENFs surrounding the damaged area, a set of entry points is selected along the boundary of the damaged region, and nerve growth is modeled using an inverse transformation rule. Newly formed ENFs within the damaged region originate from new entry points generated by a Poisson process and grow according to the proposed model. The BC model is constructed using a branching process combined with probabilistic distributions governing the placement of branching points and end points.

The rest of the thesis is organized as follows. Section 2 reviews methods for constructing the spatial model, including the Poisson process, cluster process and branching process, along with tools for assessing goodness-of-fit. Section 3 introduces the simulation of synthetic data and the tools used to analyze skin coverage. The proposed BC model is described in details in Section 4, and the corresponding

results evaluating the performance of the model are presented in Section 5. Finally, Section 6 provides a summary of this thesis and outlines potential directions for future research.

2

Methods for Model Construction

This section outlines the methodologies used to construct and evaluate the spatial model. Section 2.1 introduces the definition of point processes, which are well-suited for modeling spatial point patterns. The concept of cluster processes is also reviewed in this subsection. Subsequently, methods for assessing the goodness-of-fit of a spatial point process model are discussed in Section 2.2, with particular focus on Ripley's K-function and envelope tests. Section 2.2 also explains the construction and interpretation of shift plots. Finally, Section 2.3 provides a brief overview of the basic concepts of branching processes.

2.1 Point processes

The spatial model for probabilistically describing the structure of ENFs, consisting of entry points, branching points, and end points, is developed using spatial point processes. The definitions and notations in this thesis mainly follow the book by Baddeley, Rubak and Turneret (2016). The 2-d Euclidean space is denoted by $\mathbb{R}^2$.

2.1.1 Spatial point process

A spatial point process is suitable for modeling point patterns in two or higher dimensions. A *Spatial point process* in $\mathbb{R}^2$, denoted by $\mathbf{X}$, is a random mechanism whose outcomes are point patterns, including the location and the number of points. The book (Baddeley, Rubak and Turneret, 2016) gives a more rigorous definition of point processes. A point process is often analyzed within an observation window W, despite the fact that the generated point patterns exist over the entire spatial domain. For any region B in $\mathbb{R}^2$, the random variable describing the number of points in B is denoted as $n(\mathbf{X} \cap B)$.
In this study, the point process used for the spatial model is assumed to be *locally finite*, meaning the number of points in any bounded region $B \subseteq \mathbb{R}^2$ is finite, i.e. $n(\mathbf{X} \cap B) < \infty$. Furthermore, the point process is assumed to be *simple*, meaning no two points coincide at the same location. The point process is also assumed to be *stationary* (translation invariant) and *isotropic* (invariant under rotation). It means with a window W, the statistical properties of a point process within W do not depend on the location or orientation of the observation window. A more formal definition of *stationary and isotropic* is:

- A point process $\mathbf{X}$ is called *stationary* (or *strictly stationary*), if $\mathbf{X}$ and $\mathbf{X} + v$ have the same distribution for any translation vector v in $\mathbb{R}^2$.

- A point process $\mathbf{X}$ is called *isotropic*, if $\mathbf{X}$ and $r\mathbf{X}$ have the same distribution for any rotation r around the origin.

The intensity of a point process, denoted as $\lambda(u)$ at a spatial location $u \in B$, is a location-dependent function that represents one of the crucial spatial characteristics of point processes. It is defined through the relation:
for any region B, the expected number of points of a point process $\mathbf{X}$ falling within B satisfies:

$$\mathbb{E}[n(\mathbf{X} \cap B)] = \int_B \lambda(u)du.$$

In the special case of a stationary point process, the intensity $\lambda(u)$ is constant and thus represents a uniform number of points per unit area.
Poisson process is the fundamental point process model, serving as the foundation for constructing statistical models used to analyze spatial point patterns. The *homogeneous Poisson process*, also called *complete spatial randomness* (CSR), is characterized by the following properties:

- Poisson-distributed counts: the number of points of the point process $\mathbf{X}$ in a given region B follows a Poisson distribution;
- Homogeneous intensity: $\mathbb{E}[n(\mathbf{X}\cap B)] = \lambda \cdot |B|$, where λ is the constant intensity;
- Independence: if $B_1, B_2, ...$ are disjoint regions of space, then the number of points in $B_1, B_2, ...$ are independent random variables;
- Conditional property: given that n points observed in a region B, then those n points are independently and uniformly distributed in B.

In statistical tests, the homogeneous Poisson process often serves as the null hypothesis.

2.1.2 Cluster process

A *cluster process* $\mathbf{X}$ is a two-step stochastic process for generating clustered point patterns. In the first step, the spatial pattern of *parent* points is constructed by a point process $\mathbf{Y}$. In the second step, *offspring* points x_{ij} are generated by a random pattern around each *parent* point y_i. The *parent* points y_i are usually unobserved, while the observed *offspring* points form the cluster process $\mathbf{X}$. In theory, reproduction from the second step could be extended over multiple generations, and each *offspring* may in turn produce further *offspring*. This concept of multi-generational clustering was originally proposed by Neyman and Scott (1958) and later explored by Kingman(1977). For a *Neyman-Scott process* (which is a cluster process), the following assumptions are satisfied:

- Poisson parents: the *parent* points are from a Poisson process;
- Independent cluster: different clusters are independent of each other;
- Identically distributed clusters: different clusters have the same distribution, if based on the same *parent* point;
- Independent *offspring* points within a cluster: given a *parent* point, its *offspring* points are independently and identically distributed.

A typical example is *Thomas cluster process* (Thomas, 1949; Diggle, 1979). Its *parent* points are generated according to a Poisson process. Each *parent* point then produces a Poisson-distributed number of *offspring* points, whose locations are determined by an isotropic Gaussian distribution. Specifically, the displacement of

each *offspring* point from its *parent* point is drawn from a bivariate normal distribution with mean zero and covariance matrix $\sigma^2 I$, meaning the translation vectors follow $N(0, \sigma^2 I)$ independently along both coordinate axes.
In this thesis, the *offspring* points in the new model are generated according to a different set of rules than those in the Thomas cluster process. A detailed construction of the new model is provided in Section 4.

2.2 Goodness-of-fit

This subsection introduces a summary statistic: Ripley's K-function, and envelope tests for assessing the goodness-of-fit of a model using K-function. Thereafter, the shift plot for comparing two distributions that can assist evaluating the performance of a model is recalled.

2.2.1 Ripley's K-function

In spatial statistics, Ripley's K function is commonly used to analyze spatial correlation. The K-function is a cumulative summary of the second-order property for a point process $\mathbf{X}$. Assuming that $\mathbf{X}$ is *stationary* and *isotropic*, the K-function is defined using the expected number of points within the distance r from an arbitrary point of the process namely:

$$K(r) = \frac{1}{\lambda}\mathbb{E}[\text{ number of points within distance r of } x \mid x \in \mathbf{X}],$$

where λ is the intensity of the process.
For the *homogeneous Poisson point process*, the expected number of points within distance r from a arbitrary point is $\lambda\pi r^2$. Divided by λ, $K(r) = \pi r^2$ is often used as a reference value when analyzing the correlation. If a point pattern is generated from a clustered process, $K(r) > \pi r^2$; while $K(r) < \pi r^2$ is consistent with regularity. In practice, an unbiased estimate of the empirical K-function, can be computed as:

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{I}\{||x_i - x_j|| \leq r\} e_{ij}(r)$$

where n is the number of points observed in window W, and $||x_i - x_j||$ denotes the Euclidean distance between the points x_i and x_j. The estimated intensity of the point process is $\hat{\lambda} = (n-1)/|W|$. The $e_{ij}(r)$ is an edge correction weight.
Edge effects occur when unobservable points outside the observation window W influence the spatial distribution of points inside W. Ignoring edge effects leads to a biased empirical K-function. To mitigate this bias, Ripley's isotropic correction (Ripley, 1976 & 1977) is applied by introducing a weight factor for each pair of points (x_i, x_j). Under the assumption of *isotropy*, the weight is computed as the fraction of the perimeter of a circle centered at x_i with a radius equal to the distance to x_j, which lies within the observation window W. This edge correction ensures that the estimated K-function remains a non-decreasing function of r and leads to more accurate statistical inference of spatial patterns.

A commonly used variance-stabilized and centered version of the K-function is the L-function given by:

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

For a *homogeneous Poisson process*, the theoretical L-function simplifies to $L(r) = r$. This makes the interpretation of the transformed function straightforward: if $L(r) - r$ is above the horizontal line at zero, the point process exhibits clustering; if $L(r) - r$ is below the horizontal line at zero, the points are regularly distributed. This interpretation holds for all r, providing an intuitive way to assess spatial patterns. However, since the L-function is cumulative over r, interpreting it at a specific distance r_{i} requires accounting for the accumulated effects. A positive derivative at r_{i} suggests clustering, while a negative derivative indicates regularity in the point pattern at r_{i}.

Additionally, the square root transformation stabilizes the variance of the summary statistic, making $L(r)$ more suitable for statistical analysis. Although the exact expression of the variance for $\hat{K}(r)$ is complex, due to the dependent contributions from spatial point pairs, it is approximately proportional to r^2. In contrast, the variance of $\hat{L}(r)$ remains roughly constant as a function of r, resulting in less deviation and a more stable estimation.

While the K-function and its transformed version L-function effectively describe spatial correlations for the point pattern from one sample, pooled K-function estimates can provide a characterization of the general spatial structure of a point process from several samples (Andersson, Guttorp and Särkkä, 2016). The pooled K-function can represent a group of samples/simulations through introducing a weight factor. Such as in this study, empirical K-functions, i.e. $\hat{K}(r)$, are estimated for multiple samples/simulations and then averaged to enhance stability. This approach aligns with prior empirical studies by Schladitz et al. (2003), Myllymäki et al. (2012), Diggle (2014), and Anderssons et al. (2016). The procedure follows two steps. First, the K-function is computed for each of the m samples. Secondly, the weighted average function across all samples is computed as:

$$\overline{K}(r) = \sum_{k=1}^{m} w_k \hat{K}_k(r).$$

There are various recommendations regarding the choice of weights. Some have advised using point-number-weights unless there is a significant difference in intensity between the samples (Diggle, Mateu and Clough, 2000). Others advocate for squared point-number-weights in all circumstances (Baddeley et al., 1993). The third argument supports using point-number weights universally (Diggle, 2014). In this study, point-number weights are chosen based on the assumptions that samples/simulations are generated with the same intensity for ENF entry points and that the underlying process for creating ENF end points follows the same rules. Consequently, the intensity of end points should be the same across samples.

Let n_k denote the number of points in the observation window $|W|$ for sample k. The corresponding point-number weights are:

$$w_k = \frac{n_k}{\sum_{k=1}^{m} n_k}.$$

The weighted average L-function is then obtained via the square root transformation, with centering applied:

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

This centered and stabilized version with pooled weights is used to evaluate model fit with the methods recalled in the subsequent section.

2.2.2 Envelope test

In the context of spatial statistics, the K-function is commonly used to check the goodness-of-fit of a model. A basic approach is using hypothesis test to assess if the proposed model fits the data, which can be further denoted as two mutually exclusive hypotheses:

H_0: the new model can capture the spatial structure of the points

H_1: the new model cannot capture spatial structure of the points.

In a statistical hypothesis testing, p-value associated is the important metric to determine whether to reject the null hypothesis H_0 in favor of the alternative hypothesis H_1. Given H_0, the p-value represents the probability p of obtaining a value of test statistic T as extreme or more extreme than the observed value (t_{obs}). A test statistic is a numerical value calculated from observed data. For a chosen significance level α (commonly $\alpha = 0.05$), if $p \leq \alpha$, H_0 is rejected. The p-value is formally defined as:

$$p = \mathbf{P}(T > |t_{\text{obs}}| \mid H_0).$$

However, the K-function is a function of r not a single value. In order to reduce the function to a numerical value, one plausible approach is to determine the test statistic as

$$t = \max_{r>0}\{L(r) - r\},$$

which represents the maximum value of the function $L(r) - r$ across all values of r, as been defined in Stoyan and Stoyan's book (1994). Thereafter, a Monte Carlo test using simulations from the null hypothesis can be constructed. To evaluate the goodness-of-fit of a spatial point process, the test statistic is computed for each simulation under H_0, as well as the corresponding t_{obs} from the data. Within a Monte Carlo test, the p-value is computed by dividing the number of test statistics from simulations that are larger than or equal to the observed one by $s + 1$, where s denotes the total number of simulations.

In modern spatial statistics, rather than relying on a single-valued test statistic, it is preferable to use the entire K-function to gain a more comprehensive understanding of point processes. To accomplish this, envelope tests can be performed to assess the goodness-of-fit of the model.

Envelope tests are originally introduced and later substantiated the rationale by Ripley (1977 & 1979). The test envelope, generated from simulated patterns, represents the expected variability of the K-function, denoted as the test statistic $T(r)$ here, under H_0. If the K-function from observed data (i.e. T_{obs}) lies outside the

simulated envelope, the null hypothesis is rejected, indicating the model fails to capture the spatial structure from the observed point patterns. Conversely, if T_{obs} lies within the envelope, the model is considered a plausible fit for the data statistically. In a conventional envelope test, significance is assessed separately at multiple distances r, typically using the minimum and maximum values (or 95% of the width if $\alpha = 0.05$) of the $T(r)$ from simulations. However, this pointwise approach would lead to multiple testing issues, and *data snooping* indicating a specific value of r needs to be chosen before testing when interpreting the results.
The global envelope tests developed and refined by Myllymäki et al. (2017) can be used to address the multiple testing problem raised by point-wise envelope tests. One strategy for generating the global envelope is the *rank envelope test*, which is similar to the methods proposed for Barnard's test with discrete test statistics (e.g. Besag and Clifford, 1989 & 1991).
In the *rank envelope test*, a number s of simulated summary functions $T_i(r)$ i = 1, ..., s (i.e. the K-function), and the observed summary function are ordered according to the so-called *extreme value* R_i. The *extreme value* is a depth measure that represents the *extremeness* of the summary function $T_i(r)$ for i = 1, ..., s+1 including the observed one. R_i can be computed by:

$$R_i := \max\{k : T_{\text{low}}^{(k)}(r) \leq T_i(r) \leq T_{\text{upp}}^{(k)}(r) \text{ for all } r \in I\}.$$

The definition of $T_{\text{low}}^{(k)}(r)$ and $T_{\text{upp}}^{(k)}(r)$ are:

$$T_{\text{low}}^{(k)}(r) = \min_{i=1,\ldots,s+1}^{k} T_i(r)$$

$$T_{\text{upp}}^{(k)}(r) = \max_{i=1,\ldots,s+1}^{k} T_i(r)$$

for each r, The notations $\min^k$ and $\max^k$ mean the k-th smallest or largest values, thus, $T_{\text{low}}^{(k)}(r) \leq T_i(r) \leq T_{\text{upp}}^{(k)}(r)$ means the k-th envelope.

$$k = 1, 2, \ldots, \lfloor (s+1)/2 \rfloor,$$

The expression $\lfloor (s+1)/2 \rfloor$ means the greatest integer less than or equal to $(s+1)/2$
A formal definition of the extreme rank R_i for each $T_i(r)$ is as follows:

1. Denote the ranks of the values $T_i(r)$ from the smallest (as rank 1) to the largest (as rank $s+1$) as $R_i^+(r)$ for all r; and denote the ranks of the values $T_i(r)$ from the largest (as rank 1) to the smallest (as rank $s+1$) as $R_i^-(r)$ for all r;
2. For each r, compute $R_i^*(r) = \min\{R_i^+(r), R_i^-(r)\}$. For $R_i^*(r)$, a smaller value means $T_i(r)$ is more extreme than other summary functions at r.
3. Define the *extreme value* R_i as: $R_i = \min_r R_i^*(r)$, the smallest value across all r.

This definition implies that the *rank envelope test* is a global representation combining all pointwise envelope test for r.
The idea of global comparison makes the envelope tests have the ability to describe the variability of the model globally. For $T_{\text{obs}}(r)$ calculated from the K-function for the observed point patterns, the range of p-values are calculated as follows:

$$p_- = \frac{1}{s+1} \sum_{i=1}^{s+1} \mathbf{1}(R_{\text{obs}} < R_1)$$

$$p_+ = \frac{1}{s+1} \sum_{i=1}^{s+1} \mathbf{1}(R_{\text{obs}} \leq R_1)$$

containing the most liberal and conservative p-values. The results can be interpreted as that if $\alpha \geq p_+$, the null hypothesis is rejected; if $\alpha \leq p_-$, the null hypothesis is not rejected; while, if $p_- < \alpha < p_+$, it is unclear to reject the null hypothesis or not. Myllymäki et al. (2017) recommended at least 2500 simulations for hypothesis testing at the significance level $\alpha = 0.05$ for the *rank envelope test*, based on their investigations.

2.2.3 Shift plot

A useful statistical tool for graphically comparing two distributions, denoted as F and G (Doksum and Sievers, 1976), is the shift function. The main idea is to measure the difference between two distributions regarding the amount that one distribution must be moved to align with the other. If the two random variables $X \sim F$ and $Y \sim G$, then the shift function $\Delta(x)$ is defined as:

$$F(x) = G(x + \Delta(x))$$

which can be rewritten as:

$$\Delta(x) = G^{-1}(F(x)) - x.$$

The shift function satisfies the property that $X + \Delta(X) \sim G$. In particular, if X and Y follow the same distribution, then $\Delta(x) \equiv 0$.
The quantile-quantile (QQ) plot and the shift function are closely associated. In a QQ plot, the quantiles of two samples are plotted against each other. If both samples come from the same distribution, the points should align along the line $y = x$. The shift function $\Delta(x)$ can be measured as the smallest vertical distance between the points in the QQ plot and the diagonal line $y = x$. In order to statistically assess differences between distributions, simultaneous confidence bands for $\Delta(x)$ can be constructed based on the Kolmogorov-Smirnov test (Doksum and Sievers, 1976). The two distributions F and G are considered statistically indistinguishable if the confidence band contains the horizontal line $y = 0$. Compared to directly performing the Kolmogorov–Smirnov test, a significant advantage of using the shift function with a QQ plot is its ability to provide a visual representation of how the distributions differ across different quantiles. This visualization helps interpret whether deviations occur in the center or tails of the distributions, offering deeper insights into their differences.

2.3 Branching process - Galton-Watson process

The notations and definitions for *branching process* are based on the book by Athreya and Ney (2004). The branching process can be viewed as modeling the evolution of a population over generations. Let $\{Z_n; n = 0, 1, 2, \dots\}$ represent the number of individuals in each generation, a Galton–Watson branching process is a discrete-time Markov chain defined on the non-negative integers.

Initially, at time $n = 0$, there are Z_0 individuals. Each individual independently produces a random number of offspring according to the probability function $\{p_k; k = 0, 1, 2, \dots\}$ where $p_k \geq 0, \sum p_k = 1$, forming the next generation. Specifically, the number of individuals in the first generation, denoted as Z_1, is the sum of Z_0 independent identically distributed random variables drawn from the offspring distribution probability law $\{p_k\}$. The process continues recursively, with each individual in nth generation giving rise to offspring in generation $n + 1$. The number of offspring generated by an individual is independent of the other individuals in the same generation and individuals in the previous generations. If the population ever reaches zero (i.e., $Z_n = 0$), it remains at 0 implying extinct in all subsequent generations (i.e., $Z_{n+k} = 0$ for all $k \geq 0$). Therefore, 0 is an absorbing state, and extinction occurs when the process enters this state.
A key analytical tool for studying the branching process is the *probability generating function*. The generating function of a random variable X is defined to be:

$$G(s) = \mathbb{E}(s^X).$$

With its offspring probability mass function (PMF) $\{p_k\}$,

$$G(s) = \sum_k p_k s^k, |s| \leq 1.$$

The function G has the following properties:

- $G(s)$ is strictly increasing and convex over the interval $[0, 1]$;
- $G(0) = p_0$ and $G(1) = 1$.

If the generating function $G(s)$ of X is given, then:

$$\mathbb{E}(X) = G'(1)$$

more generally,

$$\mathbb{E}[X(X-1)...(X-k+1)] = G^{(k)}(1).$$

For two independent random variables X and Y, we have the following equation:

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

3

Synthetic Data

3.1 Treelike nerve structure

Many researchers have suggested that the structure of epidermal nerve fibers can be represented as treelike structure. Andersson, Rajala and Särkkä (2018) described this tree structure as consisting of entry points, branching points, and end points, and the line segments connecting these points. According to findings by neurologists, each nerve fiber has a line segment extending from the entry point. As illustrated in Figure 3.1, each line segment connects a pair of nerve points, referred to as a *parent* point and a *offspring* point. While real line segments tend to be more winding in nature, this thesis simplifies the representation by using straight-line segments. Subsequently, this study assumes that all branching points have exactly two *offspring* points, which reflects the most commonly observed patterns. This assumption is also supported by Andersson, Rajala and Särkkä (2018).

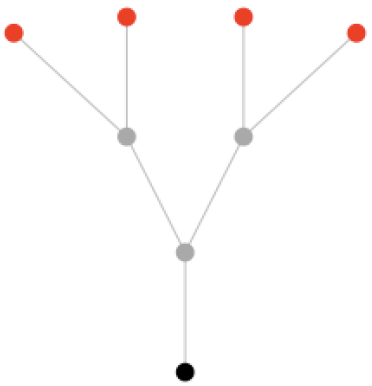

Figure 3.1: The treelike nerve structure, where the black points are entry points, gray points are branching points, red points are end points, and gray lines are line segments connecting nerve points

3.2 Model for synthetic data

In this study, we generate synthetic data by simulating point patterns of end points using the 2-d spatial model developed by Konstantinou and Särkkä (2021), which is also referred to as the *base model* in this thesis. The base model construct a cluster process by modeling the relationships among three key components: the entry point, the first branching point, and the end points. As been demonstrated in Section 2, the end points are the observed points in a cluster process, which are of interest for studying the sensing function of skin. An example of synthetic data is presented in Figure 3.2.

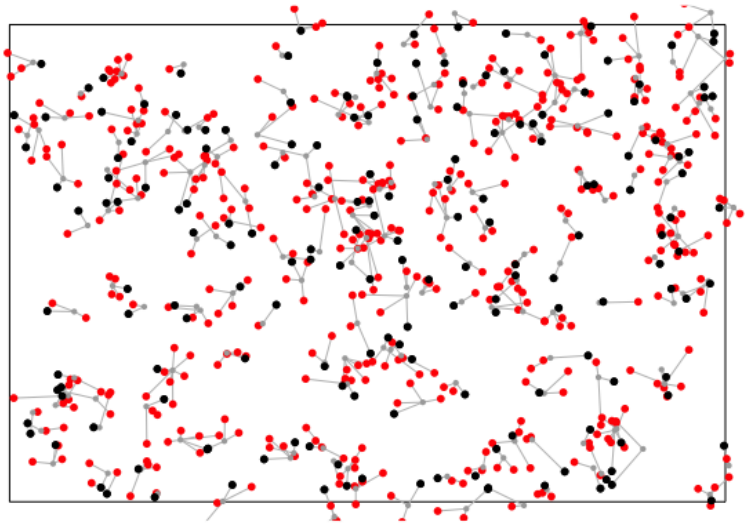

Figure 3.2: An example of synthetic data with the *base model* in an observation window W (10×20 rectangle). Each ENF contains an entry point (black dots), first branching points (gray dots) and end points (red dots).

Olsbo et al. (2013) suggested that a homogeneous Poisson process sufficiently models the distribution of healthy ENF entry points. Accordingly, in this thesis, entry points in the synthetic data are simulated using a homogeneous Poisson process. The base model for describing the spatial structure of epidermal nerve fibers is constructed as follows. First, the first branching point is generated given the location of the entry point. Then, end points are produced as clusters around this branching point. An example of the ENF structure generated by the base model, annotated with the associated random variables for location, is presented in Figure 3.3.

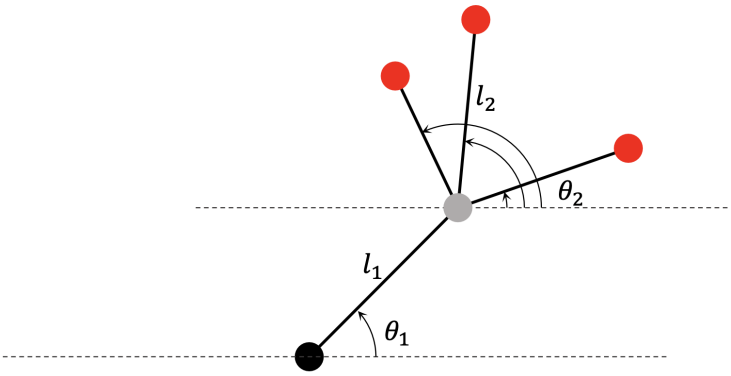

Figure 3.3: An illustration of the ENF structure generated by the *base model*: L_1 and L_2 are the length of line segments, and Θ_1 and Θ_2 denote the angles of the line segments relative to the horizontal axis.

The location of the first branching point is modeled using the UCC model developed by Andersson, Guttorp, and Särkkä (2016), where the distance L_1 to its *parent* point (i.e., the entry point) follows a Gamma distribution. The angle Θ_1, which is formed between the line segment connecting the branching point to the entry point and the horizontal axis, is assumed to follow a continuous uniform distribution. End points are assumed to be uniformly scattered around the first branching point, with their distances L_2 drawn from a Gamma distribution. The number of end points N per ENF is modeled using a negative binomial distribution. The probability mass function of the negative binomial distribution is given by:

$$f(n) = \frac{\Gamma(n+r)}{n!\Gamma(r)} p^r (1-p)^n,\ n = 0, 1, 2, \ldots, p \in [0,1], r \in [0, \infty).$$

To avoid cases where $N = 0$, we define $N \sim \mathrm{NB}(r,p) + 1$, ensuring that each first branching point has at least one end point.

The random variables used in the base model are given by the following distributions:

$$\mathbf{L}_1 \sim \Gamma(\alpha_1, \beta_1),$$

$$\mathbf{L}_2 \sim \Gamma(\alpha_2, \beta_2)$$

$$\Theta_1 \sim U(0, 2\pi)$$

$$\Theta_2 \sim U(0, 2\pi)$$

$$N \sim \mathrm{NB}(r,p) + 1$$

where Θ_2 represents the angle between the line segment connecting the first branching point and its end points and the horizontal axis.

The parameters used for simulating the synthetic data are adapted from the work of Konstantinou and Särkkä (2021), with a modification to match the scale of the study space. Using the base model described above, the synthetic data are generated within an observation window $W \subset \mathbb{R}^2$, defined as a 10×20 rectangle.

While the two-dimensional model effectively captures the spatial patterns of end points and the overall structure of epidermal nerve fibers, including the first branching point, it exhibits notable limitations in the context of this study. Specifically, the base model accounts for only the first branching point per nerve, which does not adequately represent the complex growth processes observed in realistic nerve structures, as described by neurologists. As a result, it lacks the capability to generate growing structures. Nevertheless, since the primary focus of this thesis is on modeling nerve growth in damaged skin areas and spatial patterns of end points, and given that branching points are not considered critical to sensory function, we believe using such synthetic data is sufficient enough in this study.

3.3 Tools for coverage analysis

3.3.1 Number of end points

The area covered by ENFs in the epidermal skin is of interest when neurologists analyze the sensing function of nerves that respond to external stimuli by transmitting nerve signals. Since end points are responsible for sensing external stimuli, the total count of them can be interpreted as a measure of the nerve's sensory capacity.

3.3.2 Reactive territory

Besides the number of end points, a more concrete tool for representing the coverage is the reactive territory of each nerve. Andersson, Guttorp and Särkkä (2016) first proposed and defined the reactive territory as the convex hull formed by the entry point and end points ignoring branching points (see Figure 3.4, left). Under this definition, an epidermal nerve fiber must have at least two end points to have a effective reactive territory.

Alternatively, based on the hypothesis of neurologists, the nerve's reactive territory can also be depicted by placing a disc with fixed radius at each end point, the union of these discs forming the reactive region (see Figure 3.4, right). This method suggests that the end points collectively cover the skin area, reflecting their role in sensing external stimuli. Here, the radius of each disc is calculated as:

$$r = \sqrt{\frac{|W|}{\pi n}},$$

where $|W|$ denotes the total observation sample space area, and n is the number of end points within $|W|$. This definition is based on the idea that nerves approximately cover the entire skin. An alternative method for approximating the radius is to compute the average nearest-neighbor distance, defined as the distance from an arbitrary end point to its closest end point. Besides the two calculation-based methods mentioned above, expert opinion from neurologists could also provide a reasonable estimate.

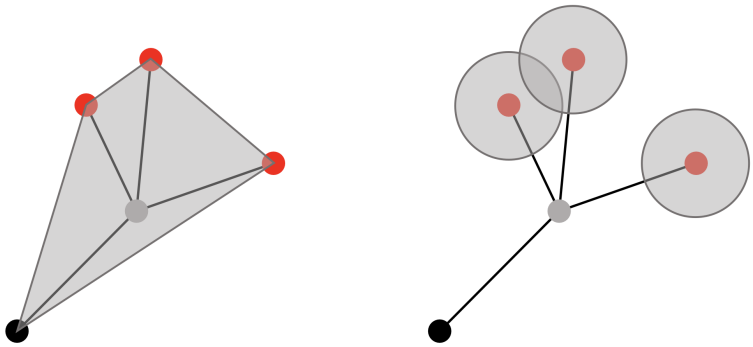

Figure 3.4: Reactive territory of an ENF colored in gray: convex hull reactive territory (left); disc-based reactive territory (right).

4

The Spatial Cluster Process Model

This section introduces a new model referred to as the *branching cluster (BC) model*, a modified Poisson cluster process developed to characterize the spatial structure of healthy epidermal nerve fibers. The BC model is designed to reconstruct the ENFs in a damaged skin area, specifically where all nerve fibers and end points are removed.

According to neurologists' hypotheses, the healing process within a damaged area involves two key stages: first, existing nerves surrounding the damaged area produce branches that extend into it; second, new epidermal nerve fibers are regenerated within a smaller region of the damaged area. Based on this understanding, we adopt two slightly different approaches to simulate the ENF end points. For branches extending from existing nerves, a cluster process with an inverse transformation is employed. For newly generated ENFs, a cluster process is used. These two approaches enable the generation of spatial patterns that illustrate how nerve growth occurs following damage.

Furthermore, the damaged zone is denoted as a subregion $E \subset W$, where W is a 10×20 rectangle observation window. For simplification, E is assumed to be a circular region with radius r. Then, To model the reconstruction of new ENFs after damage, a smaller circular region C, which has the same circle center of E, is distinguished. The gap G between E and C represents a transition zone with width w, where undamaged nerves surrounding the damaged zone can extend and branch into. An example of the study window W and the corresponding zones is illustrated in Figure 4.1.

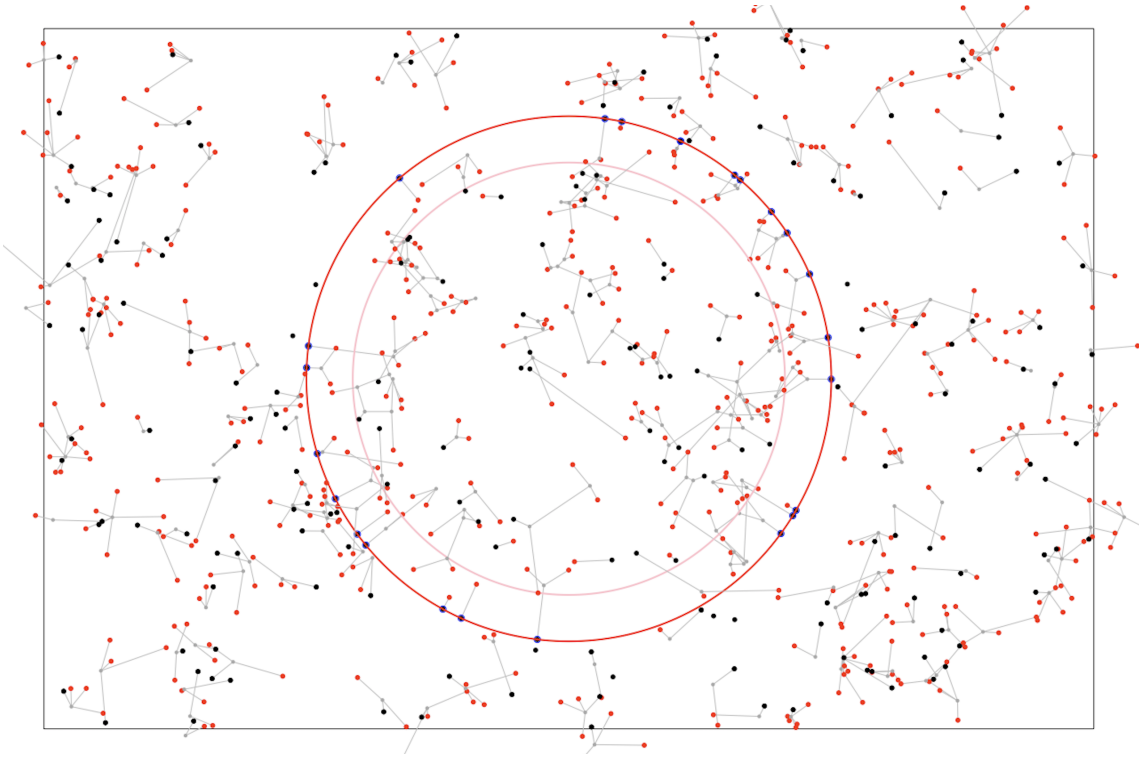

Figure 4.1: A simulation generated from the *BC model*. W is a 10×20 rectangle observation window outlined with black lines. E represents the damaged circular region, indicated by red lines, while C is a smaller circular region within E, marked by pink lines. Outside the red circle, synthetic ENFs are generated using the *base model*. Nerves in the gap G between the red circle zone E and the pink circle zone C extend from the existing nerves, whereas inside the zone C newly regenerated nerves are formed. The black dots denote the entry points, gray dots are the branching points, and red dots represent the end points.

The new spatial model regenerates the nerves in two steps: first, it generates the spatial patterns of entry points within region C using a Poisson process, or selects entry points along the boundary of region E; second, given the entry points, it generates subsequent branching points and end points through a branching process with randomly distributed spatial features for their locations.

To facilitate the initial stages of modeling, the following assumptions and settings are established:

- Each ENF grows independently;
- Structural components of the model are mutually independent.

4.1 Spatial pattern of entry points

Initially, the spatial point pattern of ENF entry points within the observation window W is modeled as a homogeneous Poisson process with constant intensity λ_0. Under the scenario where a part of skin E is damaged and nerves are removed, the simulation of entry points also considers two situations:

1. intact nerves surrounding the damaged zone E;
2. newly generated nerves within the active zone C.

ENFs located near the damaged region have the ability to produce multiple branches that extend into E. Rather than explicitly selecting which surrounding entry points grow into the damaged zone, a set of uniformly located points along the boundary of E is picked to serve as entry points for branching. These entry points are referred to as *pseudo-entry points*. The number of pseudo-entry points, denoted N_e, is modeled as a Poisson distribution:

$$N_e \sim \text{Pois}(\lambda_0|G|)$$

where $|G|$ is the area of the annular transition region $G \subset W$ where nerves extend into. $\lambda_0|G|$ computes the expected mean of the Poisson distribution. Let w denotes the width of ring-shaped gap zone G and r represents the radius of damaged zone E, the expected mean is computed as:

$$\lambda_0|G| = \lambda_0 \times (|E| - |C|) = \lambda_0 \times \pi \times (r^2 - (r-w)^2) = \lambda_0 \times \pi \times (2rw - w^2).$$

In practice, w should be small relative to r. Therefore, the mean number of pseudo-entry points in the gap region G can be approximated as

$$\lambda_0|G| \approx \lambda \times \pi \times 2rw.$$

Furthermore, the first line segment linking pseudo-entry point to its corresponding first branching point is assumed to be perpendicular to the boundary of E and directed inward, toward the center of the damaged area. This reflects a general tendency of existing ENFs to grow centripetally.
Meanwhile, the regeneration of the new ENF entry points within the active zone C is simulated using a homogeneous Poisson process with the same intensity λ_0. The expected number of new entry points emerged in C is calculated as $\lambda_0 \times |C| = \lambda_0 \times \pi \times (r-w)^2$. Unlike the ENFs generated from pseudo-entry points near the boundary, the growing direction of the first line segment for these new ENFs is random.

4.2 Branching process for ENF structure generation

Given the treelike structure of epidermal nerve fibers, the generation of branching points and end points are simulated using a branching process in the BC model. In the branching process, individuals that produce offspring are considered branching points, whereas those that do not are treated as end points. Initially, each entry point generates a first branching point with certainty. Then, for each first branching point, a discrete probabilistic rule is used to govern the generation of offspring:

$$\begin{cases} P(Z=0) = 1 - p_0 \\ P(Z=2) = p_0 \end{cases}$$

where Z denotes the number of offspring. For subsequent branching points, a rule with different probability p is applied:

$$\begin{cases} P(Z=0) = 1 - p \\ P(Z=2) = p \end{cases}$$

The probability distribution implies that each branching point either produces two offspring (resulting in a bifurcation) or none (becoming an end point). To prevent the formation of infinitely growing ENFs, the branching probability must satisfy $p < 0.5$.
The spatial placement of nerve points, including branching points and end points, is determined by a bivariate distribution (L, Θ) relative to their parent points. An example of the ENF structure simulated by the BC model is illustrated in Figure 4.2. Here, L denotes the length of the line segment connecting a nerve point to its parent (which may be either an entry point or a branching point). We refer to *the line segment of a nerve point* as the line segment for which the nerve point serves as the offspring. Then, Θ represents the angular deviation of *the line segment of a nerve point* from the direction of its parent point's line segment, as illustrated in Figure 4.2. It is assumed that offspring points are placed symmetrically around the axis of continuity of the line segment of their parent point. The variables L and Θ are considered independent.

4.2.1 Segment length (L):

The line segment length L follows a log-normal distribution:

$$L \sim \text{Lognormal}(\mu, \sigma^2).$$

The distribution of the random variable is:

$$L = e^{\mu + \sigma V}$$

where V is a standard normal random variable $V \sim \text{N}(0, 1)$. The corresponding probability density function (PDF) is:

$$f_L(l) = \frac{1}{l\sigma\sqrt{2\pi}} \exp(-\frac{(\ln l - \mu)^2}{2\sigma^2})$$

where $\mu \in (-\infty, \infty)$ is expected value (or mean) and $\sigma > 0$ is standard deviation of $\ln(L)$. All segment lengths are created independently from this distribution.

4.2.2 Branching angle (Θ):

The angle Θ is modeled as a uniformly distributed random variable:

$$\Theta_0 \sim U\left(0, \frac{\pi}{2}\right) \quad \text{(offspring from the first branching point)},$$

$$\Theta \sim U\left(0, \pi\right) \quad \text{(offspring from subsequent branching point)}.$$

The choice of ranges reflects the tendency of regenerated ENFs to initially follow a more directional path, while later branches exhibit wider angular spread. The assumption of symmetrical branching about the parent segment is maintained.

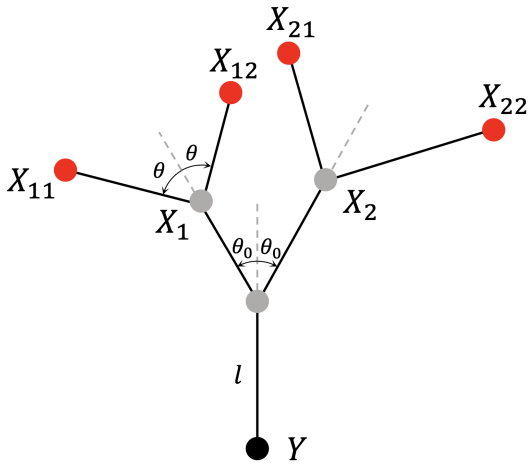

Figure 4.2: The ENF structure simulated by the *BC model*. Black dots represent entry points, gray dots indicate branching points, and red dots correspond to end points. L denotes the length of the line segments, while Θ_0 and Θ refer to the branching angles. Y represents the number of end points in the tree structure originating from the entry point, and X represents the number of end points in the tree structure originating from a nerve point (either a branching point or an end point).

4.3 Inverse transformation of nerve structures

For nerve branches growing from intact ENFs surrounding the damaged zone E, an inverse transformation is applied. Instead of letting the nerves grow towards the center, we first allow them to grow in the opposite direction, taking advantage of the open space. The procedure consists of the following steps:

Step 1 Generate the branching patterns outside of zone E, extending in the direction opposite to the center.

Step 2 Apply a transformation that maps the nerve structures inward toward the center of E.

We define d_{external} as the distance between a simulated point outside zone E (from *Step 1*) and the center of E. The transformed point is located at a distance

$$d_{\text{internal}} = \frac{r^2}{d_{\text{external}}},$$

from the center of E. The transformed point lies along the line passing through the center of E and the corresponding point simulated outside.

As been presented above, in the BC model, the length of the line segment connecting the offspring point and the parent point follows a log-normal distribution. When the inverse transformation is applied, the length of the transformed line segment is altered nonlinearly. Specifically, let L_{external} denote the length of the first line segment outside the damaged zone, and L_{internal} the length of the corresponding line segment after transformation, positioned inside the damaged zone. Then, the

relationship between them is given by:

$$r - L_{\text{internal}} = \frac{r^2}{r + L_{\text{external}}}$$

which leads to:

$$L_{\text{internal}} = \frac{rL_{\text{external}}}{r + L_{\text{external}}}.$$

Using this, the cumulative distribution function (CDF) of L_{internal} is computed as follows:

$$F_{L_{\text{internal}}}(l) = P(L_{\text{internal}} \leq l) = P(\frac{rL_{\text{external}}}{r + L_{\text{external}}} \leq l)$$

$$= P(L_{\text{external}} \leq \frac{rl}{r - l}).$$

With $L_{\text{external}} \sim \text{Lognormal}(\mu, \sigma^2)$, and denoting by Φ the cumulative distribution function of the standard normal distribution, we have:

$$F_{L_{\text{internal}}}(l) = \Phi(\frac{\ln(\frac{rl}{r-l}) - \mu}{\sigma}).$$

The corresponding probability density function is obtained by differentiating:

$$f_{L_{\text{internal}}}(l) = \frac{\mathrm{d}}{\mathrm{d}l}\Phi(\frac{\ln(\frac{rl}{r-l}) - \mu}{\sigma})$$

$$= (\frac{1}{l} + \frac{1}{r - l})\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(\ln(\frac{rl}{r-l}) - \mu)^2}{2\sigma^2}).$$

For subsequent line segments, the distribution becomes more complicated due to the dependence on the geometry and branching structure. Therefore, no analytical form is presented, and simulation methods are preferred. An example of the inverse transformation applied to a nerve, with annotations, is illustrated in Figure 4.3.

A comparative analysis of nerve fiber lengths before and after applying this transformation reveals that regenerated ENFs within the damaged zone C are generally longer than the corresponding nerve segments expanding from existing entry points. This observation aligns with empirical findings, which indicate that regenerating nerves often display increased branching and extended lengths. Such characteristics are believed to compensate for sensory deficits and support the reestablishment of functional connectivity within the affected area.

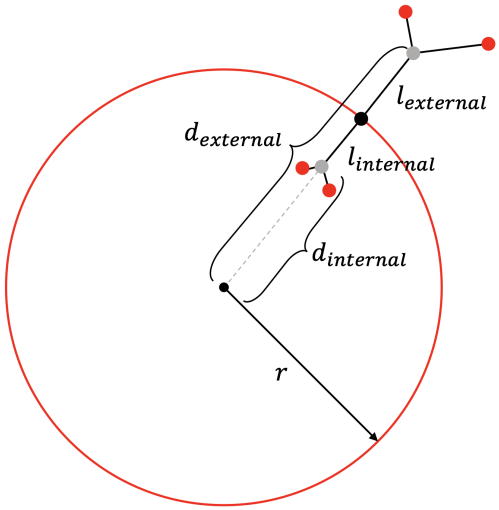

Figure 4.3: An ENF generated from existing nerves using the *BC model* after applying the inverse transformation. The red circle indicates the damaged area. The nerve structure inside the red circle represents the simulated ENF resulting from the transformation.

5

Results and Analysis

5.1 Fixed values for model construction

As demonstrated in Section 4, the BC model for simulating the growing pattern of ENFs within a damaged skin area involves several unknown parameters. Some of them are fixed based on subjective analysis.

For the Poisson process used to model the entry point patterns, the intensity λ_0 is fixed at 2. With this value, the expected number of entry points in W is 400 ($=2 \times 10 \times 20$). The synthetic data is artificially simulated, therefore, the value of λ_0 does not significantly affect the model performance, provided that a sufficient number of entry points are observed within the study window W and damaged zone E. The radius of the circular region E is fixed at 3 in this study. Regarding the gap zone $G \subset E$ allowing existing nerve fibers extend into, the width w is fixed based on the average length of simulated ENFs, which yields a value of 0.53. This choice is consistent with neurologists' description about the reinnervation of intact nerves.

5.2 Parameter estimation

This subsection presents a brief overview of the parameter estimation procedure using the Maximum Likelihood Estimation (MLE) approach. The optimization algorithm called the Limited-memory BFGS (L-BFGS) method from the family of quasi-Newton methods is employed to obtain the optimal parameter values. As been assumed in Section 4, all components in the BC model are independent, which allows each set of parameters to be estimated separately.

For the branching angle Θ, the uniform distributions given in Section 4 are used. No parameter estimation is required.

5.2.1 The length of line segment

For the line segment length L, two likelihood functions are adopted for parameter estimation, depending on the type of ENF regeneration in the damaged area. Different from the base model developed by Konstantinou and Särkkä (2021), a log-normal distribution with an inverse transformation is used to model the length of the line segments in the generated branches from existing ENFs (or pseudo-entry points). Given in Section 4, the transformed length of the first line segment follows

a distribution with the PDF:

$$f_{L_{\text{internal}}}(l) = (\frac{1}{l} + \frac{1}{r-l})\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(\ln(\frac{rl}{r-l}) - \mu)^2}{2\sigma^2}).$$

Using the length of the first line segments observed in the synthetic data, the estimated parameters are $\mu = -1.309$ and $\sigma = 0.741$. On the other hand, for the regenerated ENFs within the active zone C after damage, the length of line segments is modeled directly using a log-normal distribution: $L \sim \text{lognormal}(\mu, \sigma^2)$. Due to limited availability of detailed information regarding subsequent branching points and segment lengths in the synthetic data, parameter estimation is again performed using the first line segments only. The estimated parameters are $\mu = -1.435$ and $\sigma = 0.682$.

5.2.2 Branching probability

In the branching process, two key parameters need to be estimated: the branching probability p_0 for the first branching point and p for subsequent branching points. Based on prior studies (Andersson, Rajala and Särkkä, 2018), p_0 is approximately 0.65. In this study, we rely on simulated synthetic data that includes only the first branching points. Consequently, detailed hierarchical structural information on the treelike nerve fibers, such as secondary or higher-order branching, is not available. Given this limitation, we focus on analyzing the distribution of the number of end points per ENF, denoted as Y, as a function of the branching probabilities p_0 and p. This approach allows us to estimate the branching probabilities, even in the absence of complete tree structures.
The distribution of Y is defined in terms of the branching probabilities p and p_0 as follows:

$$Y = \begin{cases} 1 & \text{with probability } 1 - p_0 \\ X_1 + X_2 & \text{with probability } p_0 \end{cases}$$

where X_i for i=1,2 represents the number of end points generated from the offspring of the first branching point, if it branches. The variable X_i is further defined by a similar branching structure:

$$X_i = \begin{cases} 1 & \text{with probability } 1 - p \\ X_{i1} + X_{i2} & \text{with probability } p \end{cases}$$

where X_{ij} for j=1,2 represents the number of end points generated from the offspring in the subsequent generation.
This branching process continues for all subsequent offspring points, with each new point in the tree either terminating (with probability $1 - p$) or producing two new points (with probability p). The distribution of the number of end points for an offspring point in the subsequent generation is defined as:

$$X_{ij} = \begin{cases} 1 & \text{with probability } 1 - p \\ X_{ij1} + X_{ij2} & \text{with probability } p \end{cases}$$

All the $X's$, i.e. $X_1, X_2, X_{11}, X_{12}, \ldots$, follow the same offspring distribution rule. A detailed annotation of Y and X representing the number of end points originating from a point, is provided in Figure 4.2.
Given the branching structure descibed above, the probability mass function of Y can be expressed as:

$$p_Y(k) = P(Y = k) = \begin{cases} 1 - p_0 & \text{for } k = 1 \\ P(X_1 + X_2 = k) \cdot p_0 & \text{for } k > 1 \end{cases}$$

where

$$P(X_1 + X_2 = k) = \sum_{i=1}^{k-1} p_X(i) p_X(k - i)$$

and $p_X(i)$ is the probability mass function of the random variable $X's$. Similarly, the probability mass function of $X's$ is given by:

$$p_X(k) = P(X = k) = \begin{cases} 1 - p & \text{for } k = 1 \\ P(X_1 + X_2 = k) \cdot p & \text{for } k > 1 \end{cases}$$

The probability generating function of X, denoted $G_X(z)$, satisfies the equation:

$$G_X(z) = \sum_{j=0}^{\infty} z^j p_X(j) = (1 - p)z + p(G_X(z))^2.$$

Solving this equation yields:

$$G_X(z) = \frac{1 - \sqrt{1 - 4p(1 - p)z}}{2p}.$$

With the binomial expansion:

$$\sqrt{1 - 4p(1 - p)z} = \sum_{k=0}^{\infty} \frac{(-1)^k (2k)!}{2^k (k!)^2} (4p(1 - p)z)^k,$$

we can derive the probability mass function of X as follows:

$$p_X(k) = P(X = k) = \begin{cases} 1 - p & \text{if } k = 1 \\ -\frac{1}{2p} \cdot \binom{n}{k} \cdot (-1)^k \cdot (4p(1 - p))^k & \text{if } k \geq 2 \end{cases}$$

Based on above formulations, the branching probabilities p and p_0 can be estimated using the MLE method from the empirical distribution of the number of end points per ENF, as observed in the synthetic data. The estimated values are $p = 0.283$ and $p_0 = 0.693$.
With these estimated parameters, we proceed to investigate whether the new branching model is capable of generating end point patterns that are consistent with the observed patterns in the synthetic data. This validation is carried out in the following subsections using spatial statistical analysis tools presented in Section 2.

5.3 Goodness-of-Fit Assessment

The goodness-of-fit of the BC model is assessed using a global envelope test based on the K-function. As discussed in Section 2, the stabilized and centered version of the K-function, namely $L(r)-r$, is particularly effective in revealing spatial correlations among points and visualization. The observation window for calculating $L(r)-r$ is the damaged area E, where the extension of surrounding nerves into the damaged region and the regrowth of new epidermal nerve fibers (ENFs) are simulated.

To effectively evaluate the model, we generate 1000 synthetic point patterns with the base model, which are grouped to compute the $L(r)-r$ function. As introduced in Section 2, the group-pooled K-function is calculated using weights based on the number of points in each pattern. This results in a narrow envelope, as shown in Figure 5.1. For each synthetic point pattern, 2500 regeneration of nerves in the damaged area are simulated using the BC model. Based on these simulations, the constructed global envelopes enable the computation of a valid p-interval, in line with the recommendation of Myllymäki et al. (2017).

As shown in Figure 5.1, the BC model successfully captures the spatial dependencies of end points at distances up to approximately 0.5. The $L(r)-r$ curve consistently lies above zero, indicating clustering among end points. Notably, the curve peaks around $r = 0.5$, suggesting that clustering is particularly pronounced at this spatial scale. This pattern implies that end points tend to cluster within a distance of approximately $r \approx 0.5$. Combined with the visualizations in Figure 4.1 and the fact that the average length of an entire simulated ENF is around 0.53, this supports the interpretation that end points are clustered within individual ENFs. However, the model fails to accurately capture the spatial structure at larger distances. Compared to the synthetic data, it tends to generate less regular point patterns at broader spatial scales. This difference may be attributed to the method of selecting pseudo-entry points along the boundary of the damaged area and the construction of nerves directed toward the center of the damaged region.

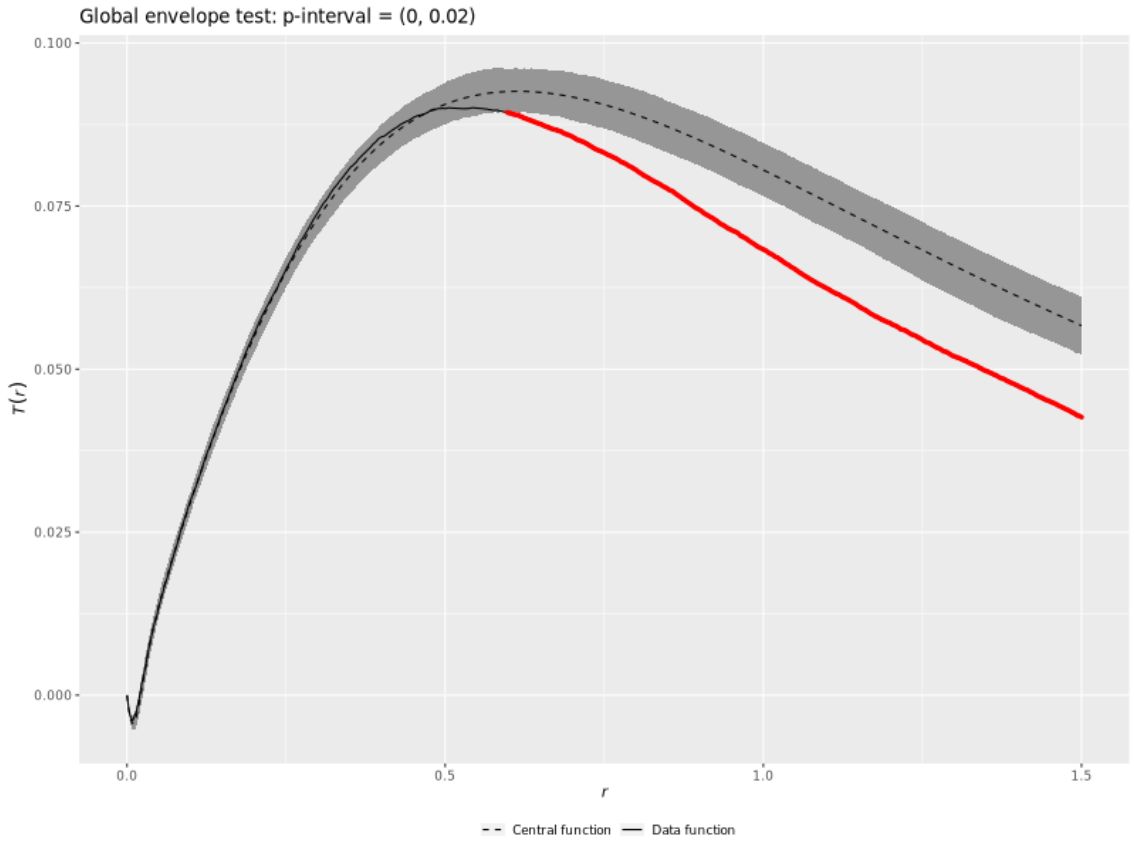

Figure 5.1: Group-wise pooled $L(r) - r$ functions with global envelopes for the end points. The solid curve represents the $L(r) - r$ function calculated from the synthetic data, while the shaded envelope is generated from simulations. The dashed curve indicates the average $L(r) - r$ function across all simulations. In total, 1000 synthetic patterns were analyzed, with 2500 simulations conducted for each pattern.

5.4 Coverage analysis

In addition to the global envelope test, the coverage of regenerated nerve fibers following damage serves as another empirical criterion for evaluating model performance in this study. One method involves comparing the number of end points located within the damaged zone E, as discussed in Section 3. In the shift plot shown in Figure 5.2, the number of end points within E is compared between point patterns simulated from the new branching model and those from the synthetic data. The results indicate that the underlying implicit distributions are statistically similar, as the horizontal zero line lies within the envelope.

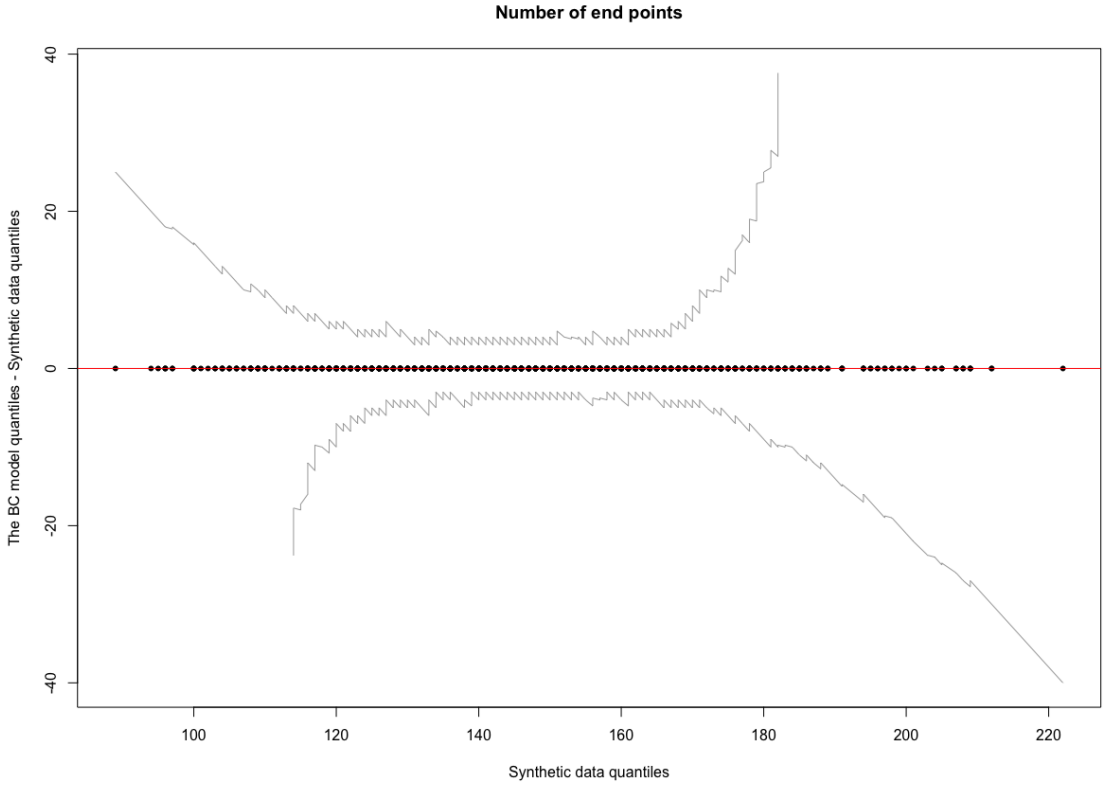

Figure 5.2: Shift plot comparing the *BC model* and the synthetic data in terms of the number of end points observed within the damaged area E.

Furthermore, the area covered by regenerated ENFs within E is analyzed. As outlined in Section 3, the area is estimated by the reactive territory, which is constructed either by the convex hull formed by nerve points or by the union of discs centered at each end point. For the disc-based method, the radius is determined to be approximately 0.253 from calculation. To assess whether the BC model can regenerate ENFs with comparable coverage, the *relative area* is computed as the ratio of reactive territory of ENFs to the total area of the damaged zone E.

Figures 5.3 and 5.4 present comparisons between the regenerated ENFs and the synthetic data. The x-axis shows the relative area covered by synthetic data within E, while the y-axis displays the ratio of the relative area covered by regenerated ENFs to that of the synthetic data. A ratio close to one indicates that the model effectively recovers the sensing function in the damaged area.

The mean coverage ratio is 1.357 using the convex hull method and 1.028 with the disc-based method. While the disc-based coverage ratios are consistently within the range $(0.5,\ 1.5)$, the convex hull method exhibits greater variability, with some ratios exceeding 3. These results suggest that the BC model successfully describe the ENF regrowth process based on neurologists' hypothesis, producing point patterns that closely match those from synthetic data in terms of coverage. Considering both the shift plot and the disc-based coverage analysis, we conclude that the BC model effectively compensates for the loss of ENFs in the damaged area.

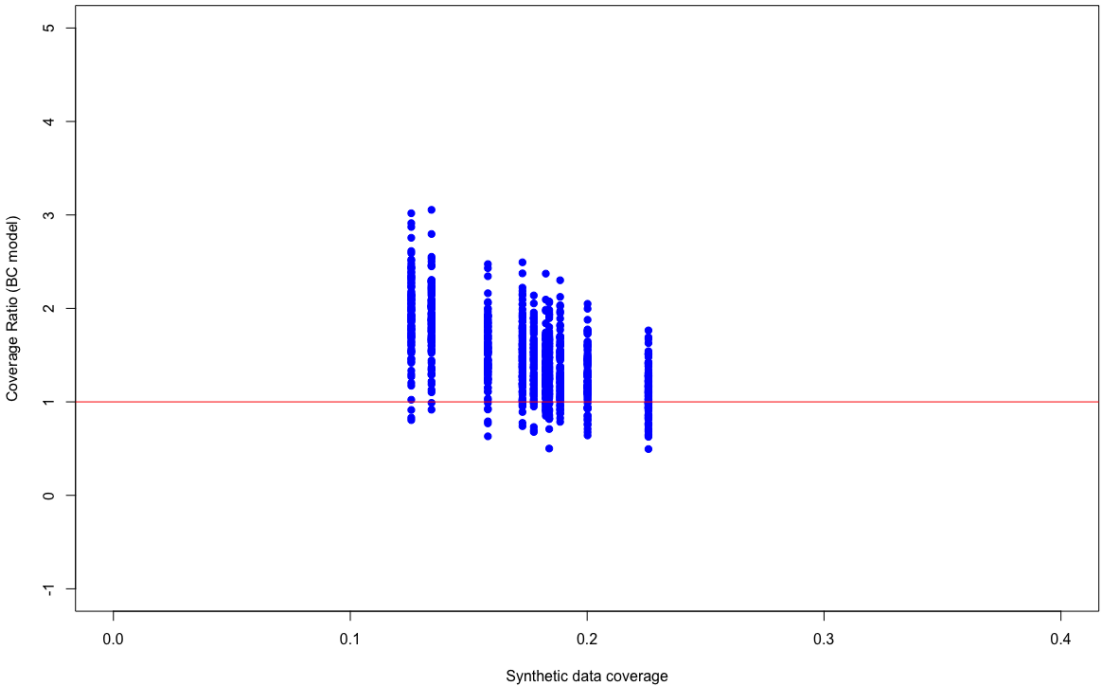

Figure 5.3: The ratio (the relative area covered by regenerated ENFs from the *BC model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the convex hull.

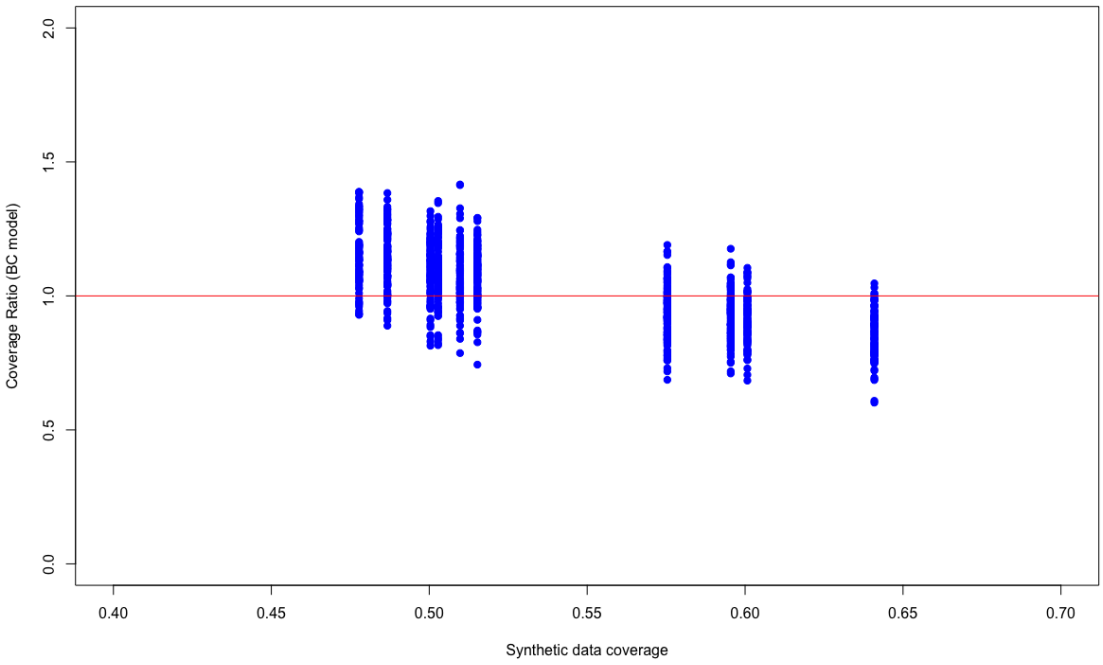

Figure 5.4: The ratio (the relative area covered by regenerated ENFs from the *BC model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the disc-based approach.

5.4.1 Comparison: coverage analysis for a point process incorporating the base model

Additionally, we evaluated the performance of a point process model that modifies the BC model by replacing its branching mechanism with the base model. In this *point process with the base model*, the generation of new entry points within the active zone C and the selection of pseudo-entry points along the edge of the damaged zone E follow the same procedures as in the BC model. However for the regeneration of nerve fibers from these entry points:

- the inverse transformation rule described in Section 4 is still applied for the nerves extending from existing entry points around E;
- the first branching point grows in a random direction from each newly generated entry point in the active zone C, i.e. the UCC model;
- the end points are uniformly clustered around the first branching point of each nerve.

With above implementations, coverage analyses are conducted on both the number of end points and the relative area. In the shift plot of the number of end points generated by the point process with the base model (see Figure 5.5), the horizontal line at zero falls within the confidence envelope, indicating no significant difference between the empirical distribution and the simulated patterns. Regarding skin coverage, both the convex hull and disc-based approaches (see Figures 5.6 and 5.7) suggest that the point process with the base model is capable of generating ENFs patterns that effectively restore coverage in the damaged region. This performance is likely due to the fact that the synthetic data itself is generated using the base model, making the simulated point patterns inherently similar. The main distinction lies in the additional reinnervation of nerves extending from existing entry points around E.

However, the average coverage ratio based on the disc-based method is slightly lower for the point process with the base model compared to the BC model. Additionally, Figure 5.6 shows that the point process with the base model yields more consistent coverage estimates under the convex hull method than the BC model. In conclusion, both models (*the point process with the base model* and the *BC model*) demonstrate the ability to compensate for nerve loss in terms of spatial coverage. Nevertheless, the BC model offers a more mechanistically informative framework by modeling not only the spatial distribution of end points but also the detailed branching structure of ENFs, making it a more comprehensive representation of the nerve regeneration following damage.

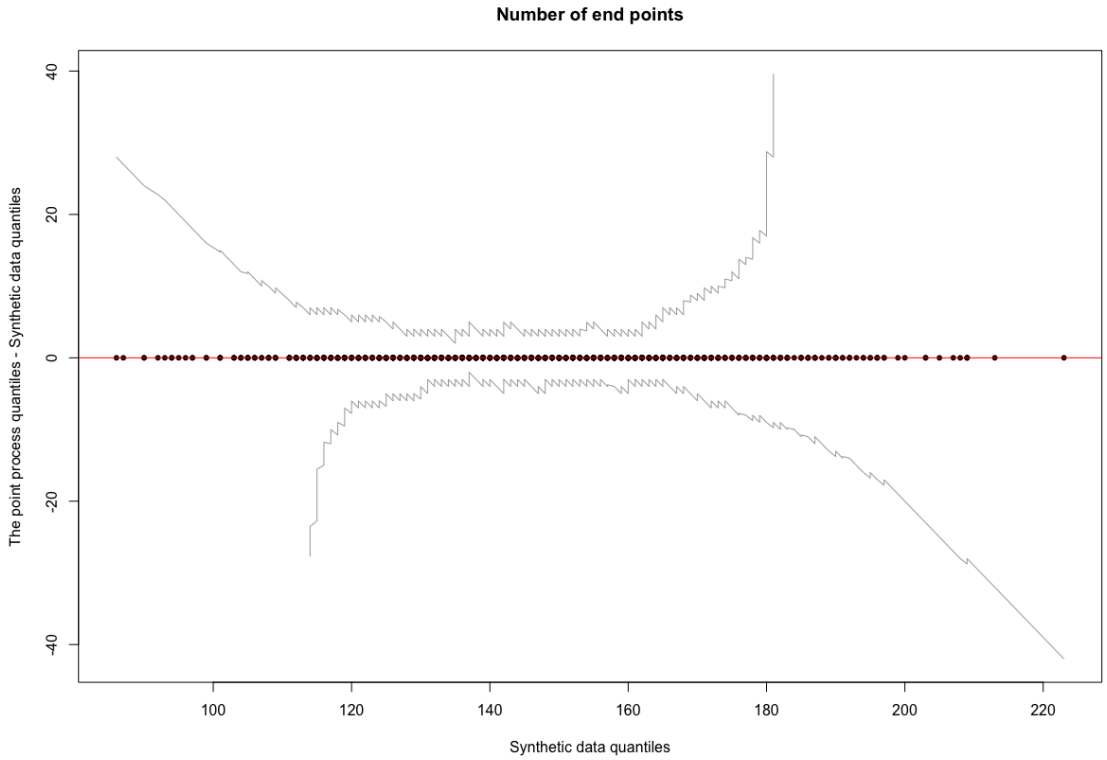

Figure 5.5: Shift plot comparing the *point process with the base model* and the synthetic data regarding the number end points observed within the damaged area E.

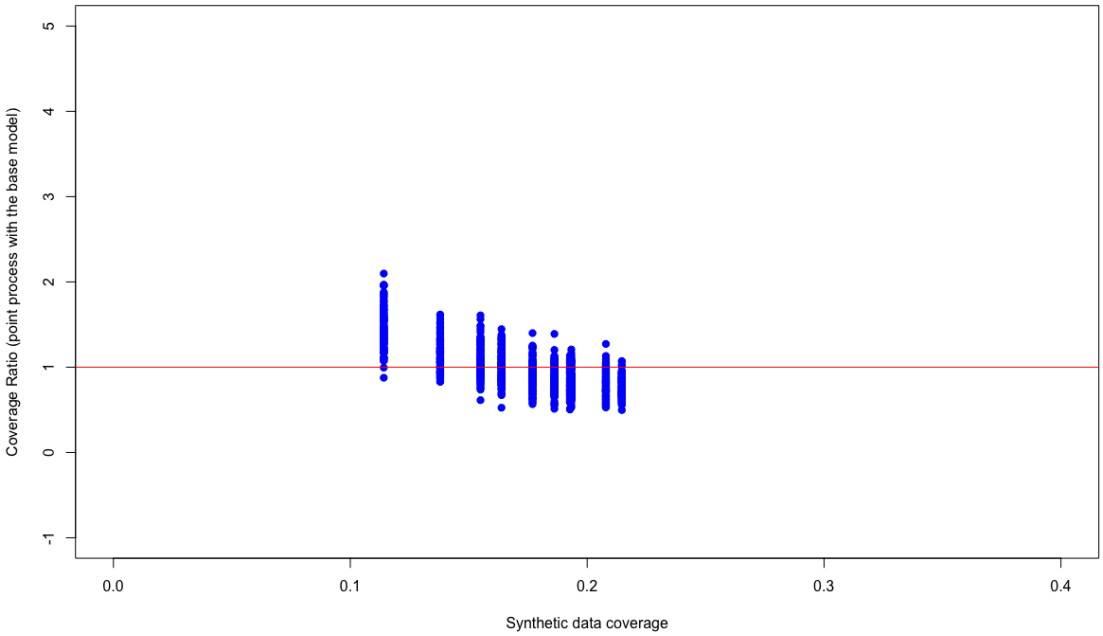

Figure 5.6: The ratio (the relative area covered by regenerated ENFs from the *point process with the base model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the convex hull.

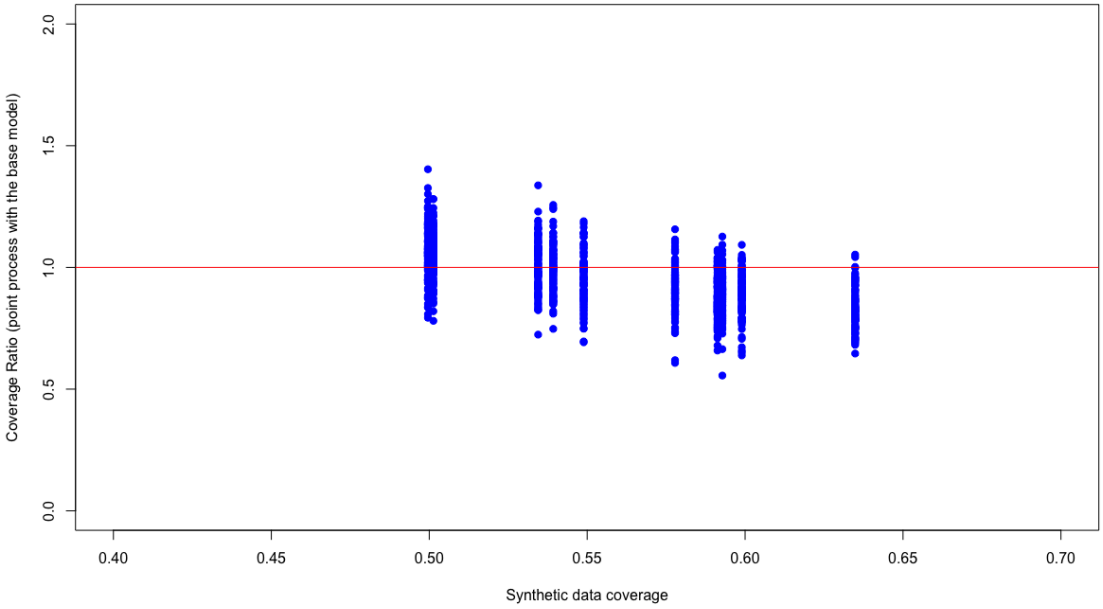

Figure 5.7: The ratio (the relative area covered by regenerated ENFs from the *point process with the base model* divided by that of the synthetic data) versus the relative area of the synthetic data. The reactive territory is defined using the disc-based approach.

6

Discussion and Conclusion

6.1 Conclusion

This thesis has introduced and developed a new spatial model, referred to as the *branching cluster (BC) model*, for simulating the spatial structure of epidermal nerve fibers (ENFs). Previous research on modeling ENFs has primarily focused on modeling nerve structures in terms of entry points and end points, which are of particular relevance in neuropathy studies. For instance, Olsbo et al. (2013) introduced the non-orphan cluster (NOC) model and building on this, Andersson, Guttorp and Särkkä (2016) proposed a modified version, the uniform cluster center (UCC) model. In both cases, end points are modeled given the entry points. Subsequent work by Konstantinou and Särkkä (2021) revealed that end points tend to cluster more systematically around the first branching points than around entry points. They extended the UCC and NOC models by incorporating the first branching points, leading to a model that better captured spatial interactions among end points. In this thesis, a two-dimensional version of the their model (which is called the *base model* here) is used to generate synthetic data.

The primary aim of this thesis is to study how the regrowth of epidermal nerve fibers following damage would look like, which are created by using the neurologists' hypothesis. To this end, we develop the BC model that incorporates the complete branching structure of ENFs, extending upon previously developed spatial models. The methodology begins by generating synthetic data using the base model, from which a part is emptied to simulate a damaged region. The BC model is employed to simulate the expansion of surrounding nerves into the damaged region and the regeneration of new nerves within the damaged region. While the model does not explicitly include time component, the line segments connecting nerve points offers an implicit representation of the growth process during nerves regeneration.

As demonstrated in Section 5 (Results and Analysis), the envelope test demonstrates that the BC model is capable of generating end points with spatial patterns similar to the original synthetic data at small spatial scales, while also preserving the hierarchical structure of multi-level branching points. Moreover, the coverage analysis of the simulated data shows that the model successfully produces end point patterns that cover a comparable area of skin, thereby compensating for sensory loss in the damaged region. The proposed model effectively simulates nerve regeneration patterns in accordance with the neurologists' hypothesis, which is:

- existing nerves surrounding the damaged area extend into it, start to branch and terminate;

- new entry points emerge within the damaged area, and begin to grow.

Overall, the model shows good performance in producing ENFs with spatial characteristics observed in the synthetic dataset.
Another important modeling decision is the application of an inverse transformation to simulate the growth of nerve fibers originating from existing entry points. This transformation ensures that most of the regenerated nerve fibers are directed toward the damaged zone and its center, in line with hypotheses from neurologists.

6.2 Future work

To the best of our knowledge, this study is the first to introduce a cluster process model for ENFs that explicitly incorporates all branching points. However, the current formulation of the model has been designed specifically for simulating nerve regrowth within a damaged area. Its capacity to reconstruct complete nerve patterns across an entire skin region remains untested. Furthermore, the BC model has so far been validated only on synthetic data. To assess the model's ability to generate realistic spatial patterns, evaluation using real clinical data is essential. Using such real data could also help determine whether interactive effects should be included in the model. Previous work by Garcia, Guttorp, and Ludwig (2020) employed a birth-and-death process to account for interactions among entry points, suggesting the importance of such mechanisms. In contrast, the base model developed by Konstantinou and Särkkä (2021) was able to replicate realistic nerve patterns without modeling interaction. Additionally, the current branching process in the BC model assumes a simplified structure in which each branching point produces either zero or two offspring points. Incorporating more flexible and biologically accurate branching mechanisms may improve the model's applicability to real-world nerve regeneration scenarios.

Bibliography

[1] Andersson, C., Guttorp, P., and Särkkä, A. "Discovering early diabetic neuropathy from epidermal nerve fiber patterns". In: *Stat Med* 35.24 (2016), pp. 4427–4442.

[2] Andersson, C., Rajala, T., and Särkkä, A. "Hierarchical models for epidermal nerve fiber data". In: *Statistics in medicine* 37.3 (2018), pp. 357–374.

[3] Athreta, K. B. and Ney, P. E. *Branching processes*. Courier Corporation, 2004.

[4] Baddeley, A., Rubak, E., and Turner, R. *Spatial Point Patterns: Methodology and Applications with R*. Vol. 1. Boca Raton: CRC Press, 2016.

[5] Baddeley, A. J., Moyeed, R. A., Howard, C. V., and Boyde, A. "Analysis of a three-dimensional point pattern with replication". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 42.4 (1993), pp. 641–668.

[6] Besag, J. "Comment on "Modeling Spatial Patterns" by B.D. Ripley". In: *Journal of the Royal Statistical Society: Series B* 39.3 (1977), pp. 193–195.

[7] Besag, J. and Clifford, P. "Generalized monte carlo significance tests". In: *Biometrika* 76.4 (1989), pp. 633–642.

[8] Besag, J. and Clifford, P. "Sequential monte carlo p-values". In: *Biometrika* 78.2 (1991), pp. 301–304.

[9] Diggle, P. J. "On parameter estimation and goodness-of-fit testing for spatial point patterns". In: *Biometrics* (1979), pp. 87–101.

[10] Diggle, P. J. *Statistical Analysis of Spatial and Spatio-temporal Point Patterns*. CRC Press: Boca Raton: Chapman & Hall, 2014.

[11] Diggle, P. J., Mateu, J., and Clough, H. E. "A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns". In: *Advances in Applied Probability* 32.2 (2000), pp. 331–343.

[12] Doksum, K. A. and Sievers, G. L. "Plotting with confidence: Graphical comparisons of two populations". In: *Biometrika* 63.3 (1976), pp. 421–434.

[13] Garcia, N. L., Guttorp, P., and Ludwig, G. "Interacting cluster point process model for epidermal nerve fibers". In: *Spatial Statistics* 35 (2020), p. 100414.

[14] Kennedy, W. R. and Wendelschafer-Crabb, G. "The innervation of human epidermis". In: *Journal of the neurological sciences* 115.2 (1993), pp. 184–190.

[15] Kennedy, W. R., Wendelschafer-Crabb, G., and Johnson, T. "Quantitation of epidermal nerves in diabetic neuropathy". In: *Neurology* 47.4 (1996), pp. 1042–1048.

[16] Kingman, J. "Remarks on the spatial distribution of a reproducing population". In: *Journal of applied probability* 14.3 (1977), pp. 577–583.

[17] Konstantinou, K. and Särkkä, A. "Spatial modeling of epidermal nerve fiber patterns". In: *Statistics in Medicine* 40.29 (2021), pp. 6479–6500.

[18] Leong, T. “First- and second-order properties of spatial point processes in biostatistics”. In: *Ph.D. Dissertation, Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA* (2005).

[19] Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. “Global envelope tests for spatial processes”. In: *Journal of the Royal Statistical Society, Statistical Methodology, Series B* 79.2 (2017), pp. 381–404.

[20] Myllymäki, M., Panoutsopoulou, I. G., and Särkkä, A. “Analysis of spatial structure of epidermal nerve entry point patterns based on replicated data”. In: *Journal of Microscopy* 247.3 (2012), pp. 228–239.

[21] Myllymäki, M., Särkkä, A., and Vehtari, A. “Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates”. In: *Spatial Statistics* 8 (2014), pp. 104–121.

[22] Neyman, J. and Scott, E. L. “A statistical approach to problems of cosmology”. In: *Journal of the Royal Statistical Society: Series B* 20.1 (1958), pp. 1–43.

[23] Olsbo, V., Myllymäki, M., Waller, L. A., and Särkkä, A. “Development and evaluation of spatial point process models for epidermal nerve fibers”. In: *Mathematical biosciences* 243.2 (2013), pp. 178–189.

[24] Ripley, B. D. “The foundations of stochastic geometry”. In: *The Annals of Probability* 4.6 (1976), pp. 995–998.

[25] Ripley, B. D. “Modelling spatial patterns (with discussion)”. In: *Journal of the Royal Statistical Society* Series B.39 (1977), pp. 172–212.

[26] Ripley, B. D. “Tests of ’randomness’ for spatial point patterns”. In: *Journal of the Royal Statistical Society* Series B.41 (1979), pp. 368–374.

[27] Schladitz, K., Särkkä, A., Pavenstädt, I., Haferkamp, O., and Mattfeldt, T. “Statistical analysis of intramembranous particles using freeze fracture specimens”. In: *Journal of Microscopy* 211.2 (2003), pp. 137–153.

[28] Stoyan, D. and Stoyan, H. *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. John Wiley & Sons Ltd, 1994.

[29] Thomas, M. “A generalization of Poisson’s binomial limit for use in ecology”. In: *Biometrika* 36.1/2 (1949), pp. 18–25.

[30] Waller, L. A., Särkkä, A., Olsbo, V., Myllymäki, M., Panoutsopoulou, I. G., Kennedy, W. R., and Wendelschafer-Crabb, G. “Second-order spatial analysis of epidermal nerve fibers”. In: *Statistics in Medicine* 30.23 (2011), pp. 2827–2841.

[31] Wang, L., Hilliges, M., Jernberg, T., Wiegleb-Edström, D., and Johansson, O. “Protein gene product 9.5-immunoreactive nerve fibres and cells in human skin”. In: *Cell and tissue research* 261 (1990), pp. 25–33.

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY