

RESEARCH ARTICLE

Statistical modeling of diabetic neuropathy: Exploring the dynamics of nerve mortality

Konstantinos Konstantinou^{1,2}  | Farnaz Ghorbanpour³ | Umberto Picchini^{1,2} | Adam Loavenbruck⁴ | Aila Särkkä^{1,2}

¹Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden

²Department of Mathematical Sciences, University of Gothenburg, Gothenburg, Sweden

³Department of Mathematical Sciences, Allameh Tabataba'i University, Tehran, Iran

⁴Department of Neurology, Kennedy Laboratory, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence

Konstantinos Konstantinou, Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden.

Email: konkons@chalmers.se

Funding information

Chalmers AI Research Centre; Vetenskapsrådet, Grant/Award Number: 2019-03924

Diabetic neuropathy is a disorder characterized by impaired nerve function and reduction of the number of epidermal nerve fibers per epidermal surface. Additionally, as neuropathy related nerve fiber loss and regrowth progresses over time, the two-dimensional spatial arrangement of the nerves becomes more clustered. These observations suggest that with development of neuropathy, the spatial pattern of diminished skin innervation is defined by a thinning process which remains incompletely characterized. We regard samples obtained from healthy controls and subjects suffering from diabetic neuropathy as realisations of planar point processes consisting of nerve entry points and nerve endings, and propose point process models based on spatial thinning to describe the change as neuropathy advances. Initially, the hypothesis that the nerve removal occurs completely at random is tested using independent random thinning of healthy patterns. Then, a dependent parametric thinning model that favors the removal of isolated nerve trees is proposed. Approximate Bayesian computation is used to infer the distribution of the model parameters, and the goodness-of-fit of the models is evaluated using both non-spatial and spatial summary statistics. Our findings suggest that the nerve mortality process changes as neuropathy advances.

KEYWORDS

approximate Bayesian computation, dependent thinning, epidermal nerve fibers, reactive territory, spatial point process

1 | INTRODUCTION

Epidermal nerve fibers (ENFs) are dendroidal thin sensory nerve fibers in the outermost layer of the human skin, called the epidermis. They enter, grow and branch in the epidermis until they terminate. Throughout the paper, the entry points will be referred to as base points and the termination points as end points. The nerve fibers transfer signals such as heat and pain recorded by the end points to the central nervous system. Diabetic neuropathy is a disorder in which elevated blood sugar and related processes in the body damage ENFs all along their course, from the dorsal root ganglion near the spinal cord to the skin, negatively affects their functionality, and over time causes attrition of ENFs. Progression of ENF dysfunction and loss is characterized, respectively, by neuropathic pain and loss of sensation.¹ While damaged nerves

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

may heal and regrow with sustained improvement of blood sugar, this regrowth is very slow and often incomplete. It is therefore important to detect the neuropathy at the earliest stage possible, to prevent ENF damage before it occurs.

The diagnostic capabilities of ENFs have been established in several studies. More specifically, neuropathy progression decreases the ENFs spatial intensity and total coverage of the ENFs in the epidermis.¹⁻³ In addition, the two-dimensional spatial structure of the base and end points of subjects suffering from diabetic neuropathy tend to be more clustered than the structure of healthy controls.^{2,4-6} Furthermore, the nerve fibers in subjects with diabetic neuropathy tend to branch fewer times before terminating than the nerve fibers in healthy subjects.⁷

A considerable amount of earlier research on the spatial structure of ENFs concentrated on modelling the spatial structure. The planar locations of the base and end points are treated as realisations of two-dimensional spatial point processes, and point process models have been developed for the end points conditioned on the empirical base point locations. For instance, the non-orphan cluster (NOC) model⁶ and the uniform cluster centre (UCC) model² are point process models of this nature. In the UCC model, the direction of the end point clusters with respect to their corresponding base points is uniformly distributed, while in the NOC model the clusters are constructed towards open space. A sequential marked point process model³ and a continuous time birth-and-death process model⁸ were proposed to capture possible interactions between the nerve trees. In addition, recently, some models for the three-dimensional spatial structure have been suggested.^{9,10}

In this paper, we will focus on the underlying process that guides the morphological changes in the spatial structure of the nerve trees as diabetic neuropathy advances. We have skin samples from healthy subjects and subjects suffering from either mild or moderate diabetic neuropathy. The mild point patterns consisting of the base and end points of ENFs are treated as spatial thinnings of the healthy point patterns, and the moderate patterns as thinnings of the mild patterns. For this purpose, different spatial thinning schemes are proposed. Since we do not have a closed-form expression for the likelihood function for such thinning models, we suggest an approximate Bayesian computation (ABC) approach to estimate the parameters of the model. Finally, the models are evaluated using Ripley's K function, mark correlation function, and some non-spatial summary statistics. Our findings indicate that nerve mortality does not occur completely at random. To the best of our knowledge, this is the first study investigating nerve loss due to diabetic neuropathy using spatial thinning models.

The paper is organised as follows. In Section 2, the ENF data set is described and a brief introduction to point processes and spatial thinning operations is given in Section 3. In Section 4, the proposed thinning schemes are described. Our findings are presented in Section 5 and further discussed in Section 6. The code used to perform the statistical analyses is available at <https://github.com/Konstkon/ENFthin>.

2 | DATA

The epidermal nerve fiber data we have available are obtained using suction induced skin biopsies, a medical procedure where a skin sample is taken, mounted on a slide and stained for imaging.^{11,12} Then, confocal microscopy is used to manually trace the base points, which are the entry locations of the ENFs in the epidermis, the branching points, which are the locations where the nerve branches within epidermis, and the end points, which are the locations where the nerve fibers terminate. Two skin blister specimens were taken from six different body parts (foot, calf, thigh, hand, forearm and upper back) from each subject in the study resulting in three to six images (usually four) per subject and body part. The degree of diabetic neuropathy was determined based on the nerve counts of samples from the six body locations and the results of a nerve conductivity test. Based on those factors the subjects were categorized into suffering from mild, moderate or severe diabetic neuropathy. Therefore, the degree of diabetic neuropathy, that is, healthy, mild, moderate, or severe, is known for each subject. The original spatial point patterns are three dimensional, with the scale in the z -direction being significantly smaller than the scale in the x and y directions. However, since we are interested in the coverage of the ENFs on the skin, we concentrate on the two dimensional projections of the patterns.

Here, we limit our analysis to the data collected from the feet of 32 healthy subjects, eight subjects with mild diabetic neuropathy, and five subjects with moderate diabetic neuropathy. The choice of the body part is motivated by the observation that changes in the ENFs morphology occur at the earliest stage in distant body parts such as feet.¹³ We have left out the group with severe diabetic neuropathy as those samples contain very few nerves. The data consisting of the base and end point locations are treated as realisations of spatial point processes in \mathbb{R}^2 observed in the window W with area of approximately 330×432 microns. In total, 112 healthy, 28 mild diabetic and 13 moderate diabetic skin samples are included in the analysis. From now on, we refer to the three groups as healthy, mild, and moderate, respectively. An

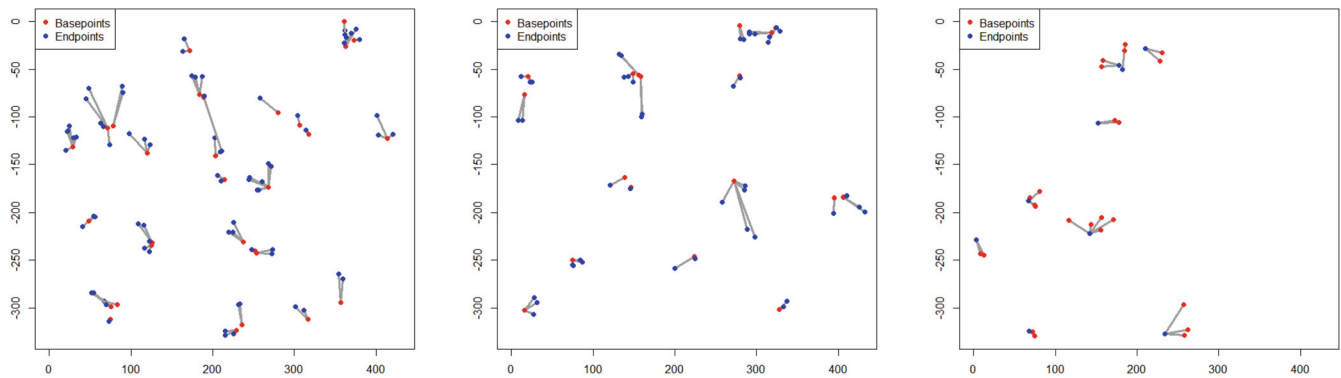


FIGURE 1 Examples of ENF patterns for healthy (left), mild (middle) and moderate (right) diabetic samples. Red colour denotes the base points and the blue colour denotes the end points of the nerve fibres.

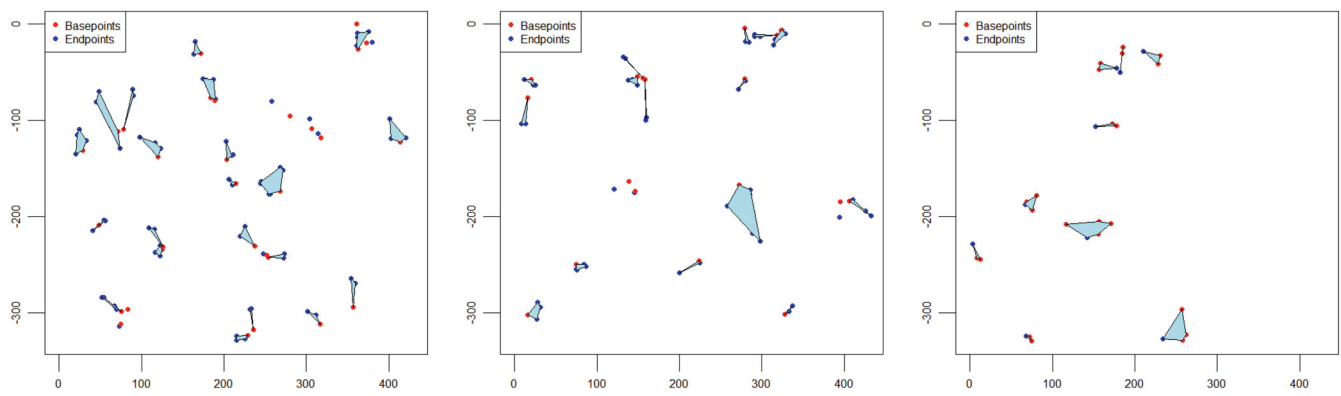


FIGURE 2 Reactive territories of the skin samples obtained from the same samples as in Figure 1. Red points represent the base and blue points the end points of the nerve fibres.

example of a randomly chosen ENF sample from each group is displayed in Figure 1, where the different types of points are represented with different colours.

The area of the skin that the ENFs cover can be described by reactive territories.² The term “reactive territory” of a nerve tree refers to the area of the skin that falls within the sensory field of the nerve tree. Two definitions are given:² A reactive territory of a nerve is (1) the smallest disc covering the projected base, all branching, and end points belonging to the nerve tree or (2) the convex hull determined by the locations of the projected end points and base points belonging to the nerve tree. Here, we use the convex hull definition of the reactive territory. The reactive territories of the ENF samples in Figure 1 are presented in Figure 2. Note that a nerve tree has to have at least two end points to have a positive reactive territory area.

It is well established that the degree of neuropathy and the area of (positive) reactive territory are negatively correlated, i.e. as the degree of neuropathy increases the area of the skin covered by the ENFs decreases.^{1,2} This is illustrated for our data in the left panel of Figure 3. The decrease in ENF coverage is translated into neuropathic pain and loss of sensation, the main symptoms of the neuropathy. The corresponding result for the reactive territories of the individual trees is illustrated in the right panel of Figure 3. As expected, the reactive territories for the healthy nerve trees are on average larger than the reactive territories for the diabetic nerve trees. On the other hand, there are no significant differences between the reactive territories between the mild and moderate diabetic nerve trees.

3 | METHODS FOR SPATIAL POINT PROCESSES

Point patterns consisting of ENF base and end points projected into the plane are regarded as realisation of two dimensional spatial (marked) point processes. In this section, we give some definitions and notations (mainly from Iljan et al¹⁴)

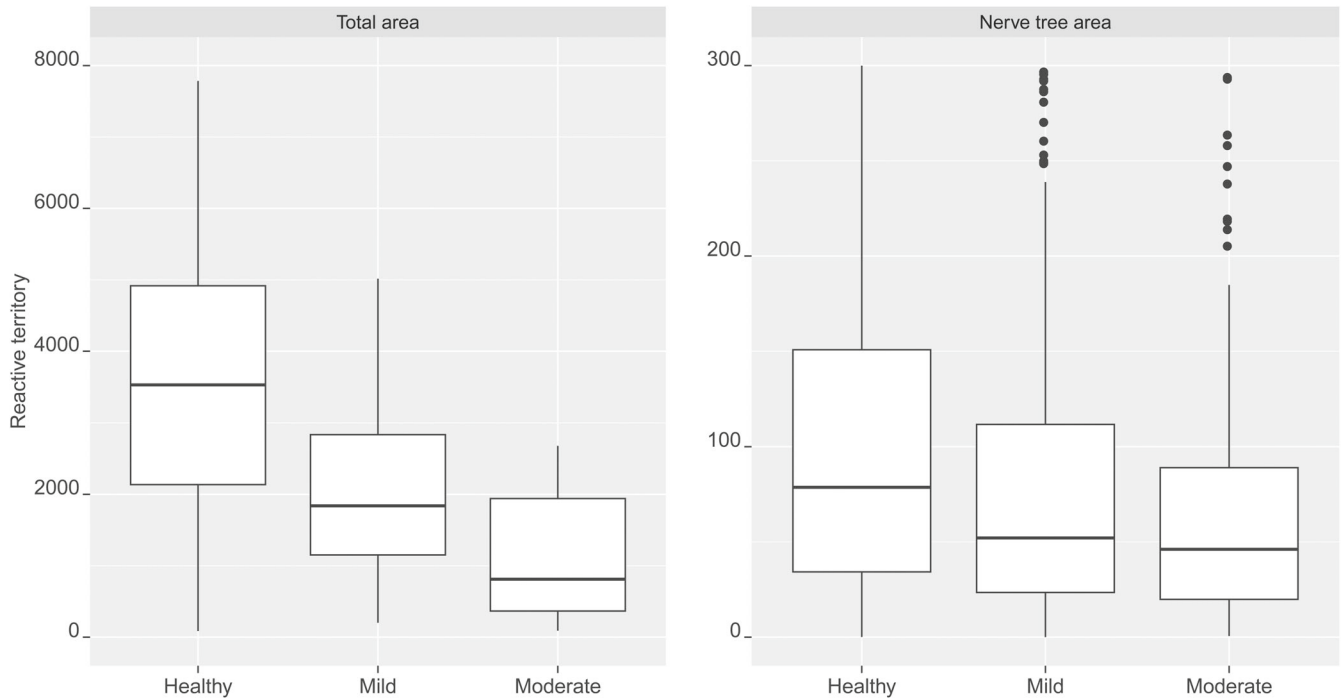


FIGURE 3 The total area of the skin covered by the reactive territories (left) and reactive territory area per nerve tree (right), in samples from the healthy subjects and the subjects with mild or moderate diabetic neuropathy.

for spatial unmarked and marked point processes and introduce some summary functions. Furthermore, we recall some thinning operations for point processes. For more rigorous treatment of the topic, the reader is referred to the literature.¹⁴⁻¹⁷

3.1 | Spatial point processes

Spatial point processes are mathematical models for point patterns. A spatial point process X is defined as the random set of locations in a spatial domain D where the events of interest occur. Usually, the point process is observed in an observation window $W \subset D$. We refer to this as a point pattern in the observation window W . In this work, the locations, where the nerves enter the epidermis and where they terminate, are treated as realisations of spatial point processes in a rectangular observation window $W \subset \mathbb{R}^2$. The point processes are assumed to be locally finite, that is for every bounded subset B in the Borel set $\mathcal{B}(\mathbb{R}^2)$, the number of points of the process that lie in B is finite. Further, the point processes are assumed to be simple, that is at any location there is at most one point of the process. Lastly, the point processes are assumed to be stationary (translation invariant) and isotropic (rotation invariant).

Sometimes additional characteristics, marks, are attached to each point in the point pattern. Such marked point patterns often provide a deeper insight into the underlying physiological processes.¹⁴ Let $M \subset \mathbb{R}$ be a mark space with the mark $m_i \in M$ attached to the point $x_i \in X$. A realisation of the corresponding marked point process X_M is then

$$\{(x_i, m_i), i = 1, \dots, n\} \subset W \times M,$$

where n is the observed number of points. We assume that X_M is stationary and isotropic, that is, invariant under translations and rotations of the point locations, respectively.

3.2 | Ripley's K function

Here, we review Ripley's K function that is used to describe the second-order properties of a point process.¹⁸ For stationary and isotropic point processes in \mathbb{R}^2 , Ripley's K function has a straightforward interpretation. In particular, $\lambda K(r)$, λ being

the intensity (the mean number of points per unit area) of the process, gives the expected number of further points of the process in the disc with radius r centred at an arbitrary point of the process. For the Poisson point process,

$$K(r) = \pi r^2, \quad r \geq 0.$$

The homogeneous Poisson point process corresponds to complete spatial randomness, and therefore it is often treated as a reference model.^{14,19} Observed values of the K function that are smaller or larger than their theoretical values under complete spatial randomness indicate regularity or clustering, respectively. Since there are points of the process outside the observation window that may interact with the points inside the window, an edge correction term is needed when estimating the K function. An approximately unbiased estimator of the K function is given by

$$\hat{K}(r) = \frac{|W|}{n^2} \sum_{i=1}^n \sum_{j \neq i} w(x_i, x_j) \mathbf{1}\{\|x_i - x_j\| \leq r\}, \quad r \geq 0,$$

where n is the total number of the observed points in W , $\|x_i - x_j\|$ denotes the Euclidean distance between the points x_i and x_j , $\mathbf{1}\{A\}$ is the indicator function equal to 1 when event A is true and zero otherwise, and $w(x_i, x_j)$ is an edge correction term. We used the translation correction $w(x_i, x_j) = \frac{1}{|W_{x_i} \cap W_{x_j}|}$, where W_{x_i} is the translated window $W_{x_i} = \{z + x_i : z \in W\}$ and $|\cdot|$ denotes the two dimensional Lebesgue measure. In this work, we use the variance stabilized and centred variant of the K function,¹⁴ defined by

$$L(r) - r = \sqrt{\frac{K(r)}{\pi}} - r, \quad r \geq 0, \quad (1)$$

which for the Poisson process equals zero. Therefore, positive values of this centred function indicate clustering and negative values regularity.

Our data are hierarchically structured into groups (healthy, mild, moderate), subjects within the groups, and samples from the subjects. Since we are interested in the average spatial structure of the ENFs in each group, the overall group-wise K and L functions need to be estimated. This can be achieved as follows. For group g , we initially estimate the sample-wise summary functions K_{ij} for sample $j \in \{1, \dots, m_i\}$ of subject $i, i \in \{1, \dots, N_g\}$. Let \hat{K}_{ij} denote the corresponding estimator.

Then, the subject specific mean function \bar{K}_i can be obtained as a weighted mean of the functions K_{ij} . An unbiased estimator for the subject-wise \bar{K}_i function for each subject i is given by

$$\hat{\bar{K}}_i(r) = \sum_{j=1}^{m_i} w_{ij} \hat{K}_{ij}(r).$$

Similarly, the subject-wise functions are weighted to obtain the group-wise function \bar{K}_g for the group g . A group-wise estimator for \bar{K}_g is given by

$$\hat{\bar{K}}_g(r) = \sum_{i=1}^{N_g} w_i \hat{\bar{K}}_i(r). \quad (2)$$

Square point number weights are used to compute the subject-wise and group-wise estimates, since the point patterns from different samples and subjects cannot be assumed to have the same intensity.^{5,9,20} Let n_{ij} denote the number of points in sample j of subject i , and let $n_i = \sum_{j=1}^{m_i} n_{ij}$ be the total number of points in the samples from subject i . Then, the square point number weights for the group-wise and subject-wise K functions are given by

$$w_i = \frac{n_i^2}{\sum_{k=1}^N n_k^2}, \quad w_{ij} = \frac{n_{ij}^2}{\sum_{k=1}^{m_i} n_{ik}^2}.$$

This procedure is illustrated in a sample dataset in Appendix B. Second-order properties of point processes with marks can also be investigated. The mark correlation function for marked point processes is discussed in the following section.

3.3 | Mark correlation function

The mark correlation function describes the second-order characteristics for point processes with quantitative marks. It can be defined as the (conditional) expectation

$$k(r) = \frac{\mathbb{E}[f(m_i, m_j) | \|x_i - x_j\| = r]}{\mu^2},$$

where μ is the mean of the considered mark distribution and $f(m_i, m_j)$ is a so-called test function.²¹ Here, we use $f(m_i, m_j) = m_i m_j$. Therefore, if the marks are uncorrelated, k equals 1. Values less than 1 indicate negative correlation and values greater than 1 positive correlation. The mark correlation function can be estimated by kernel estimation, namely

$$\hat{k}(r) = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n m_i m_j \cdot w_{ij}}{\bar{m}^2}, \quad (3)$$

where \bar{m} is the mean of the observed marks,

$$w_{ij} = \frac{e_b(r - |x_i - x_j|)}{|W_i \cap W_j|},$$

and e_b is a kernel function with bandwidth b .²¹ Generally, the choice of the kernel function is less critical than choosing the bandwidth in an optimal way.²² Therefore, we used the Epanechnikov kernel function as it is often used for estimating the mark correlation function. The bandwidth is chosen by using Stoyan's rule-of-thumb,²¹ which is the default in the R package spatstat.

Since our data are replicated, we need to estimate the mark correlation function for each sample and then pool all the estimates to obtain the subject-wise and group-wise estimates in a similar fashion as we estimate the K function above. The mark correlation (Equation 3) and the L function (Equation 1) will be used throughout the paper to assess the goodness of fit of the proposed model. The inference method we chose, on the other hand, requires an informative summary function of the data. To avoid using the same summary functions for inference and model evaluation, we used a different summary function for inference, which is described in the following section.

3.4 | Empty space function

The *empty space distribution function* $F(r) : [0, \infty) \rightarrow [0, 1]$ is related to the probability that an arbitrary point $x \in \mathbb{R}^2$ has an empty disc of radius r around it. It is defined as

$$F(r) = 1 - P(X(b(x, r)) = 0),$$

where $X(b(x, r))$ is the random number of points of the process in the disc centered at x with radius r , denoted by $b(x, r)$. Let $\{y_i\}_{i=1}^t$ be t points randomly sampled within $W_{\ominus r} := \{x \in W : \min(\|x - x_b\|) \geq r, x_b \in \partial W\}$, where ∂W is the boundary of W . Then, an unbiased estimator for the empty space function F ¹⁴ is given by

$$\hat{F}(r) = \frac{1}{t} \sum_{i=1}^t \mathbf{1}\{\min(\|y_i - x\|) \leq r, x \in X \cap W\}. \quad (4)$$

4 | MODELING THE ENF THINNING PROCESS

Thinned point processes provide a class of models for point patterns that are caused by random removal of events. A thinning operation defines a rule which determines which points in a point process X should be deleted, to obtain a thinned point process X_{thin} . Thinning operations can be divided into the following three different types.¹⁴

- **Independent p -thinning:** In p -thinning each point in the point process X is deleted with constant probability $1 - p$, $p \in [0, 1]$, independently of its location and on the other points in X . The parameter p is called the “retention probability”.
- **Independent $\pi(x)$ -thinning:** $\pi(x)$ -thinning is a generalization of p -thinning where the retention probability depends on the location x of the point, that is for all $x \in X$, the deterministic function $\pi(x)$ gives the probability that $x \in X_{thin}$. As p -thinning, $\pi(x)$ thinning is statistically independent, that is the deletion or non-deletion of any particular point does not depend on the operation on the other points.
- **Dependent thinning:** More general thinning strategies can be constructed if we let the retention probability to depend on the other points in the point process X , that is for every point $x \in X$ the retention probability is given by a function $\pi(x|X)$.

As neuropathy advances, the number of nerve trees (base points) and end points decreases. Here, we suggest a thinning scheme to describe the biological process behind these changes. It is believed that whole ENF trees die and, in addition, some individual nerve endings may disappear or appear, the latter being caused by the existing nerve fibers branching and creating new end points in order to compensate for the loss of nerves. In addition, the spatial pattern of base and end points becomes more clustered as the neuropathy advances as illustrated in Figure 5. Both the base and end point patterns are clustered as their corresponding centered L functions (Equation 1) are positive, except at very small distances. Note that the clustering of end points increases from healthy to mild and from mild to moderate patterns but the clustering of base points increases only from healthy to mild.

Below, we first illustrate how independent random thinning schemes, namely an independent p -thinning, applied to healthy samples do not result in patterns similar to the observed mild patterns. Then, we propose a dependent thinning strategy, where the probability for a point to be retained depends on the other points in X , and suggest an approximate Bayesian computation approach for the inference under this model.

4.1 | Independent random p -thinning

A natural starting point is to investigate whether mild patterns can be constructed by randomly removing either complete nerve trees or individual nerve end points from healthy patterns, that is, that there is no underlying location-based mechanism guiding nerve removal. However, if the process X is stationary, then Ripley's K and L functions are invariant under independent random thinning operation, and therefore the summary functions of X and of the thinned process X_{thin} are identical. As mild patterns of base and end points are more clustered than the healthy patterns, we expect the independent random thinning to be unsuitable for capturing the spatial structure of the mild diabetic patterns.

To confirm this, we chose as our null models two different independent random p -thinning models and applied them to the healthy base and end point patterns. For each model this was performed by (i) estimating the probability p as the ratio between the corresponding mean mild group and mean healthy group intensities $\bar{\lambda}_M$ and $\bar{\lambda}_H$, and (ii) by randomly and independently removing either end points or base points together with all the connected end points with probability $1 - p$. Then, we constructed 95% global envelopes (see Appendix A) based on 2500 independent summary curves obtained via simulations from an independent thinning model with the estimated retention probability \hat{p} (see Figure 4). Since the data curves (in red) fall completely outside the envelopes, we can conclude that, as expected, the thinned normal patterns fail to capture the structure present in the mild patterns.

4.2 | Dependent thinning

It can be seen in Figure 5 that the mild base and end point patterns are more clustered than the corresponding healthy patterns. Therefore, base/end points should be removed from healthy patterns such that the resulting patterns are more clustered. Below, we suggest a parametric thinning strategy, where isolated nerve trees are more likely to be removed than non-isolated ones.

Let B^M and E^M denote the base and end point patterns for the targeted mild diabetic sample observed in W and n_B and n_E be the numbers of points in B^M and E^M , respectively. Now, let $B^H = \{y_j, m(y_j)\}$ and $E^H = \{x_i, m(x_i)\}$, with $j = 1, \dots, n'_B$ and $i = 1, \dots, n'_E$, denote the marked base and end point patterns for a healthy sample, with mark $m(y_j)$

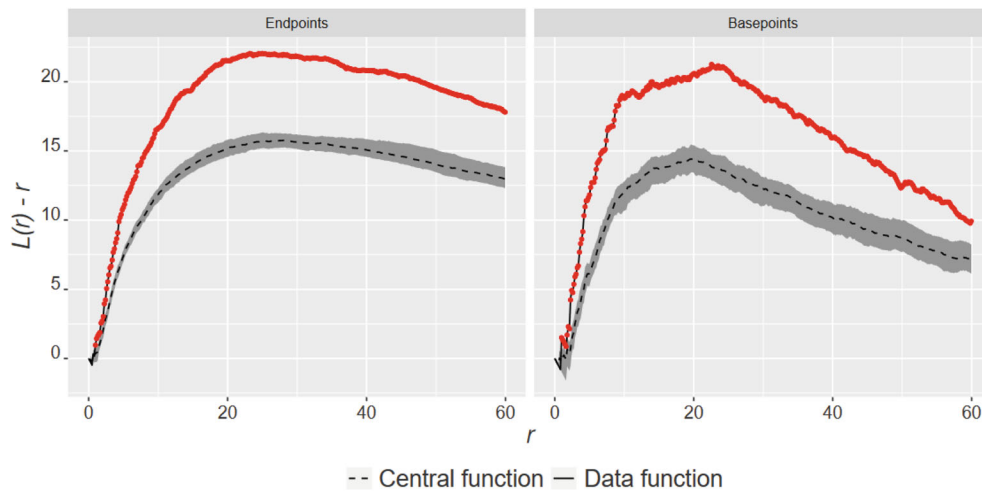


FIGURE 4 Group-wise centred L functions based on the data (red, solid) for the mild patterns with 95% global envelopes (grey) based on independent random p -thinning of end points (left) or base points (right) of the healthy patterns.

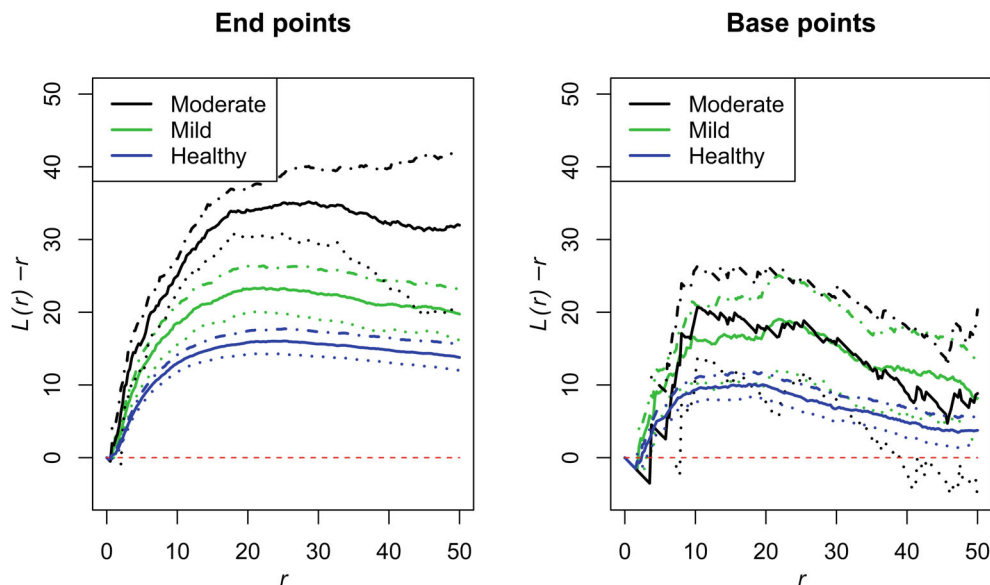


FIGURE 5 Group-wise pooled $L(r) - r$ functions, for the end points (left) and base points (right) of the healthy (blue), mild (green) and moderate (black) groups with 95% pointwise bootstrap envelopes (dashed lines). The red dashed line corresponds to the complete spatial randomness.

giving the Euclidean distance to the closest other base point y_k , $k \neq j$, with n'_B and n'_E being the numbers of points in B^H and E^H , respectively.

We thin the pattern B^H to exactly n_B base points according to an iterative thinning scheme (see Algorithm 1), with $\pi(y_j) = f(m(y_j); \theta)$, where θ is a scale parameter and $f(\cdot; \theta)$ is given by

$$f(m; \theta) \propto e^{-\theta^2 m^2}. \quad (5)$$

By definition, $f(\cdot; \theta)$ favors the removal of isolated points, that is, points with large marks m , since the retention probabilities $\pi(y_j)$ decrease with increasing distance m (see Figure 6). Moreover, the removal probabilities are proportional to $1 - e^{-\theta^2 m^2}$, and therefore, the larger the value of θ , the closer this thinning strategy is to independent thinning. For each y_j that is removed, we remove all the end points connected to it. The resulting base and end point patterns are denoted by y^* and x^* , respectively.

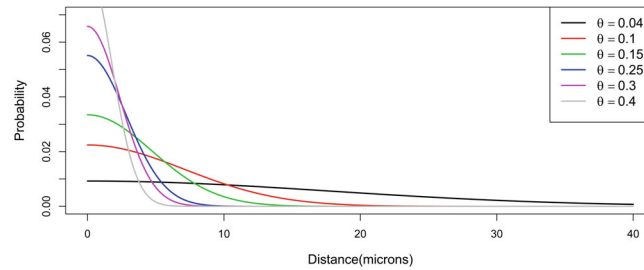


FIGURE 6 The density function defined in Equation (5) for different values of the parameter θ .

Algorithm 1. Dependent thinning model $\mathcal{M}(\theta|B^H, n_B)$

Input: Healthy basepoint pattern $B^H = \{y, m(y)\}$ and corresponding endpoint pattern $E^H = \{x, m(x)\}$, parameter θ , desired number of basepoints n_B .

Output: Simulated mild diabetic basepoint pattern y^* and corresponding endpoint pattern x^* .

repeat

Let $I = \{1, \dots, n'_B\}$ be an index set.

For $y_j \in y, j \in I$, calculate $m_j = m(y_j) > 0$

For $y_j \in y, j \in I$, calculate $f_j = \frac{1 - e^{-\theta^2 m_j^2}}{\sum_k 1 - e^{-\theta^2 m_k^2}}$

Sample with replacement an index $l \in I$ using the weights $f_j, j \in I$.

Remove y_l from y . Call y^* the resulting basepoint pattern.

From the endpoint pattern x , remove all points that are connected to y_l . Call x^* the resulting endpoint pattern.

Update the basepoint pattern $y = y^*$ and the endpoint pattern $x = x^*$

Update n'_B , the number of points in y .

until $|B^H| = n_B$

For a mild diabetic sample with n_B base points, the aforementioned thinning model is applied to all healthy patterns that have at least $n_B + 5$ base points. The number of such healthy patterns for a mild pattern $a \in \{1, \dots, 28\}$ is denoted by N_a . We regard the healthy patterns as independent observations from the healthy population even though we have several (independent) samples from each subject. This seems reasonable since there is more variation between the samples within a subject than between subjects.²³ Hence for each mild diabetic sample a , a healthy pattern is sampled with replacement from the N_a healthy samples, and thinned to exactly the same number of base points as in pattern a . This is repeated several times. Then, the simulated mild patterns are used to construct the group-wise L function estimates. Throughout the remainder of this paper, this model will be denoted as $\mathcal{M}(\theta) = \mathcal{M}(\theta|B^H, n_B)$, with parameter θ , and is detailed in Algorithm 1.

4.3 | Inference using approximate Bayesian computation

To infer plausible values of the scale parameter θ controlling the retention probabilities, we used approximate Bayesian computation (ABC).^{24,25} This is a family of algorithms suitable for Bayesian inference when the likelihood function associated to a statistical model $\mathcal{M}(\theta)$ is unavailable in closed form, or is computationally too expensive to approximate, but given a parameter vector θ , it is possible to simulate artificial data from $\mathcal{M}(\theta)$. The simplest ABC method is the acceptance-rejection sampling,²⁶ that, for given data y and (vector of) summary statistics thereof $S(y)$, samples from an approximation $P_\epsilon(\theta|S(y))$ of the posterior distribution $P(\theta|S(y))$. This is performed by (i) proposing a parameter $\theta^* \sim P(\theta)$ sampled from its prior distribution $P(\theta)$; (ii) conditionally on θ^* , simulate an artificial dataset y^* as $\mathcal{M}(\theta^*) \rightarrow y^*$, to be read as the output of a “run” of model $\mathcal{M}(\theta^*)$; (iii) reduce both y^* and y to a low-dimensional set of summary statistics $S(y^*)$ and $S(y)$, respectively, and evaluate their proximity using some distance (eg, Euclidean) $\|S(y^*) - S(y)\|$; and finally, (iv)

retained θ^* if $\|S(y^*) - S(y)\| < \epsilon$, for some small $\epsilon > 0$, and reject it otherwise. The procedure (i)–(iv) is iterated until N parameter values have been accepted. A pseudocode for the ABC rejection sampler is given in Algorithm 2 below. Each accepted parameter is a draw from the approximate posterior

$$P_\epsilon(\theta|S(y)) \propto P(\theta) \int \mathbf{1}_{\|s^* - s\| < \epsilon} P(s^*|\theta) ds^*,$$

where in the integrand we have used the shorthand notations $s^* = S(y^*)$ and $s = S(y)$. This algorithm is computationally inefficient when the posterior is very dissimilar to the prior, resulting in many rejections. Instead of fixing ϵ , we can simulate a large number of θ values and choose the most appropriate of these as described below.

Algorithm 2. ABC rejection sampler

Input: prior $P(\theta)$, model $\mathcal{M}(\theta)$, summaries $S(\cdot)$, threshold $\epsilon > 0$, positive integer N .

Output: posterior draws $(\theta_1, \dots, \theta_N)$.

for $i \leftarrow 1, \dots, N$ **do**

repeat

 Draw from prior $\theta^* \sim P(\theta)$

 Simulate $\mathcal{M}(\theta^*) \rightarrow y^*$

 Compute $S(y^*)$

until $\|S(y^*) - S(y)\| < \epsilon$

$\theta_i \leftarrow \theta^*$

end for

More sophisticated ABC methods taking into account information about the previously accepted draws for θ have been suggested, both in an MCMC framework^{27,28} and as sequential Monte Carlo algorithms.^{29–31} However, as our model includes only the scaling parameter θ , the simple ABC rejection based method described above was found to be suitable enough.

Notice that $P_\epsilon(\theta|S(y))$ coincides with $P_\epsilon(\theta|y)$, when $S(\cdot)$ is a sufficient statistic for θ . On the other hand, sufficient statistics are generally unavailable and therefore, in practice, ABC always returns approximate inference even in the limit when $\epsilon = 0$. It is therefore crucial to construct appropriate (“informative”) summary statistics that are able to retain information about θ . As a rule of thumb, it is suggested that the length of the vector $S(y)$ (which is the same as the length of $S(y^*)$) should be the same as the length of θ .³²

In our case, as we only have the scaling parameter θ in Equation (5), we construct a single summary statistic. The K or L function could be chosen as the summary statistic in the ABC algorithm but since the centred L function will be used to evaluate the goodness-of-fit of the thinning model, we chose to use the empty space function F_y instead. However, we did not use the entire function but we considered the summary function previously used for point process models for the locations of sweat glands.³³ In particular, we used $s = S(y) = \min(\{r : \hat{F}_y(r) = 0.3\})$, where $\hat{F}_y(r)$ is the estimator given in Equation (4). The summary statistic for the observed data is $s = S(y) = \min(\{r : \hat{F}_y(r) = 0.3\})$, where $\hat{F}_y(r)$ is the estimator given in Equation (4), and y are the observed mild diabetic basepoint patterns. Similarly, for generic simulated mild base point patterns y^* obtained using $\mathcal{M}(\theta^*|B^H, n_B)$, as defined in Section 4.2, we computed the summary statistics $s^* = S(y^*) = \min(\{r : \hat{F}_{y^*}(r) = 0.3\})$. Notice that, even though $\mathcal{M}(\theta^*|B^H, n_B)$ generates a simulated endpoint pattern x^* , the inference is solely based on the simulated and empirical mild basepoint patterns y^* and y .

4.4 | ABC inference using simulated data

A simulation study was conducted to assess the performance of the inference method. For this purpose, healthy data were simulated from a Matérn cluster process using parameters estimated from the data using the minimum contrast method.¹⁴ The simulated parent pattern, was represented by the base points, and the simulated daughter pattern, was represented by the end points. Then, the proposed dependent thinning was applied to the simulated pattern for different chosen values for θ , and the corresponding realisations were used to obtain the empirical summary statistics used in the ABC method.

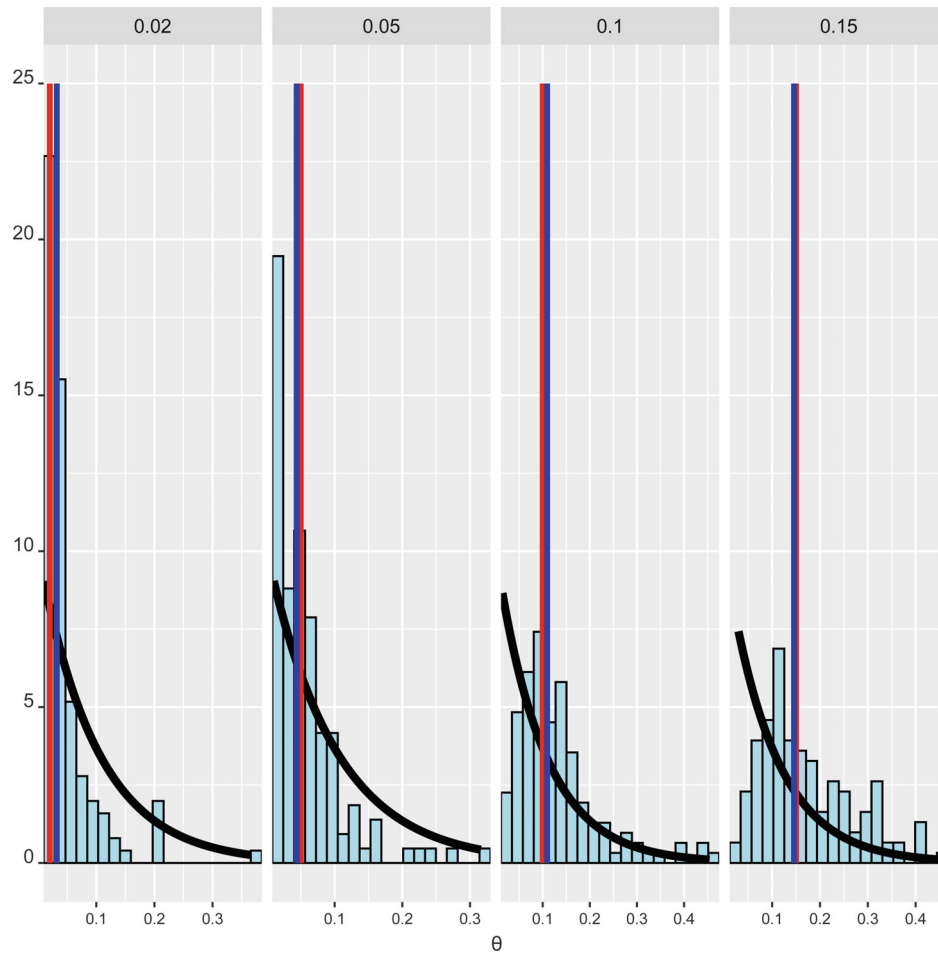


FIGURE 7 Histograms of ABC posterior draws for θ together with the posterior median (solid blue line). The prior density (solid black line) is for an exponential (10) distribution truncated to values larger than 0.01. The true values of θ (solid red lines) are also reported on the panel headings.

Each pattern was thinned such that $n_B = 14$ parent points (and their daughter points) remained. The posterior distributions for θ are displayed in Figure 7 together with an exponential prior $P(\theta) = \text{Exponential}(10)$ for θ . The true values of the data-generating θ are included for comparison. We observe that the true parameter value θ^0 is well identified when $\theta^0 < 0.1$, and while for larger values the posterior mode is close to the true value, the posterior uncertainty increases, that is as the value of θ^0 increases it becomes progressively more challenging to identify it (see Table 1). This is due to the fact that as θ^0 increases the dependent thinning scheme approaches independent random thinning, as shown in Figure 6, and as a result, the influence of the parameter on the realisations diminishes. Inference results have been obtained using the “reference table” version of the ABC rejection algorithm, which does not require the threshold ε to be prefixed in advance, and it is illustrated as follows: (i) we simulated 1,330,000 parameters independently from the prior $\text{Exponential}(10)$, and conditionally on these draws we simulated correspondingly 1,330,000 data sets $y^* = \mathcal{M}(\theta^*)$; (ii) we calculated $S(y^*)$ on each of these data sets and $S(y)$ for the observed data, and (iii) accepted those θ^* s for which the corresponding distances $\|S(y) - S(y^*)\|$ were smaller than the 0.1-th percentile of all 1,330,000 ABC distances. We used the `abc` function in the R `abc` package³⁴ to carry out the computations. In next section we consider inference on real ENF data.

5 | RESULTS

In this section, the dependent thinning model is fitted to the ENF data and the goodness-of-fit of the model is investigated by comparing the thinned and target patterns with respect to some spatial and non-spatial summary statistics. Initially, we compare the structure in the thinned healthy and mild patterns, and then, in the thinned mild and moderate patterns.

TABLE 1 Posterior median and 95% credibility intervals CI_θ for the parameter θ in the simulation study, where the target patterns have 14 base points and the true value of θ varies between 0.02 and 0.15.

True value of θ	Median	CI_θ
0.02	0.028	[0.011, 0.195]
0.05	0.038	[0.011, 0.211]
0.10	0.119	[0.037, 0.384]
0.15	0.150	[0.046, 0.404]

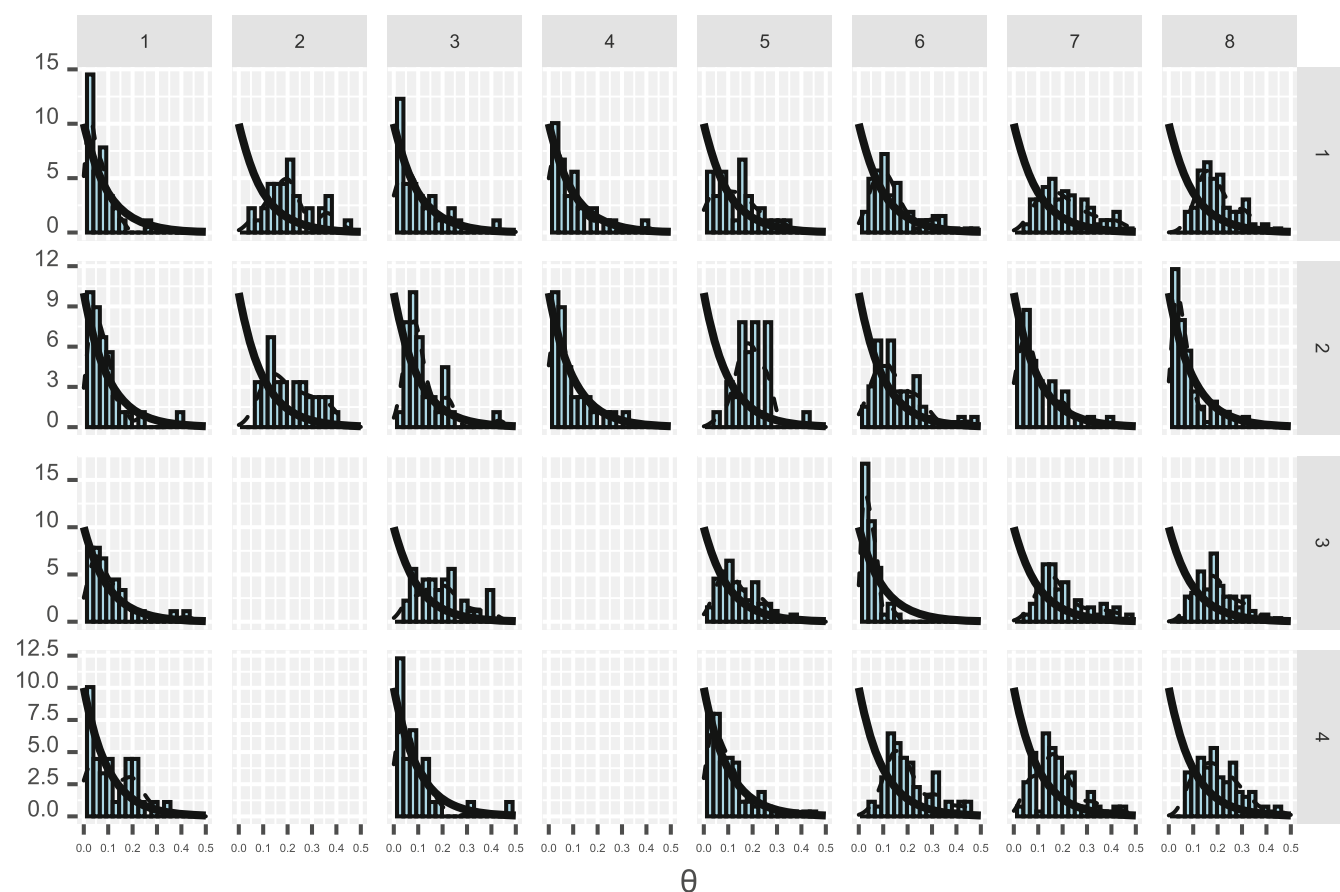


FIGURE 8 ENF data: Histograms of ABC posterior draws for θ for each mild diabetic sample obtained by thinning healthy patterns. The prior densities are plotted with solid black lines. Each column corresponds to one subject (total eight subjects), and each row corresponds to different samples from one subject. The number of samples for each subject are varying from 2 to 4.

5.1 | Healthy versus mild

We applied the thinning strategy introduced in Section 4.2 and removed whole nerve trees (base points and end points connected to it) from the healthy patterns. The posterior distributions of θ parameter based on the thinned healthy patterns are displayed for each mild diabetic pattern in Figure 8. As the model favors the removal of isolated trees if the θ is small, and corresponds to independent thinning when θ is large, an exponential (10) prior was chosen for θ . Initially, we used a uniform prior which gave posterior distributions with large variance. We believe that this choice made the ABC method more efficient, which resulted in better inference. For several mild diabetic samples, the bulk of the posteriors $P_\epsilon(\theta|S(y))$ is located around very small θ values indicating that, on a typical healthy pattern, isolated nerve trees are favored to be removed from the healthy patterns in order to obtain patterns similar to these mild patterns. On the other hand, for some mild diabetic samples, the posterior $P_\epsilon(\theta|S(y))$ is centred at "large" values, indicating that randomly

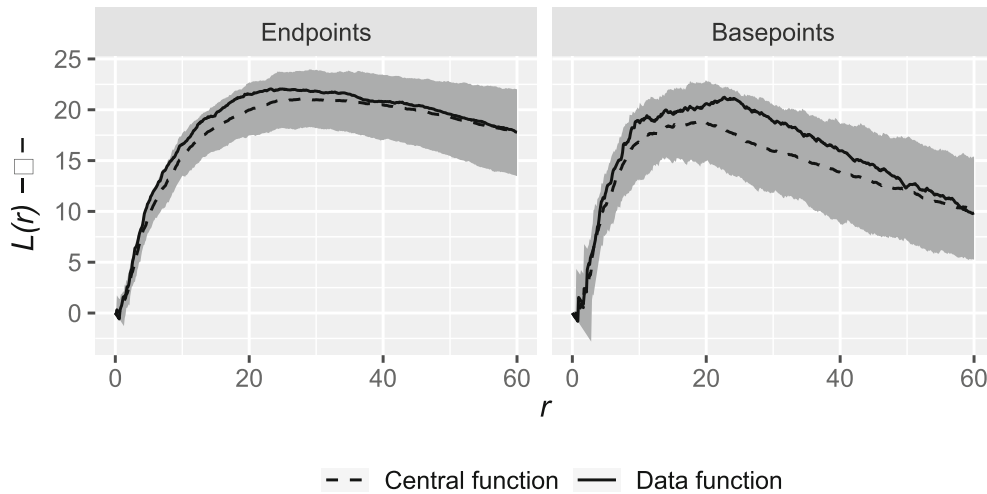


FIGURE 9 Group-wise centered L functions with 95% global envelopes for the end points (left) and base points (right) constructed from 2500 simulations from the posterior predictive distribution of the thinning model. The solid curves are the centered L functions estimated from the mild data.

Algorithm 3. Posterior predictive bands

Input: Basepoint and endpoint healthy patterns $\{B^H, E^H\}_j$ with $j = 1, \dots, 112$,

Number of basepoints n_B^i in the mild diabetic samples $i = 1, \dots, 28$,

Posterior sample $\theta^i = (\theta_1^i, \dots, \theta_N^i)$ for the mild diabetic samples $i = 1, \dots, 28$,

A user-specified summary function $T(\cdot)$

Output: 95% global envelopes constructed from 2500 simulations from the posterior predictive distribution.

for $s \leftarrow 1, \dots, 2500$ **do**

for $i \leftarrow 1, \dots, 28$ **do**

 Sample with replacement a value θ from θ^i

 Sample with replacement an index $j \in \{1, \dots, 112\}$

 Simulate $\mathcal{M}(\theta \mid B_j^H, n_B^i) \rightarrow (y^*, x^*)$

 Compute $T_i = T(y^*, x^*)$

end for

Calculate \bar{T}_g^s using (T_1, \dots, T_{28}) as described in Equation (2)

end for

Construct 95% global envelopes (Appendix A) using $(\bar{T}_g^1, \dots, \bar{T}_g^{2500})$

thinning the healthy patterns is sufficient to capture the structure in the targeted mild diabetic sample. The latter patterns contained rather large number of nerve trees indicating that the neuropathy is in an early stage, and hence cannot be detected from the nerve patterns yet. Also for a few mild diabetic samples, the posterior $P_e(\theta|S(y))$ coincides with the prior. In particular, the inference quality is low for the patterns with small number of nerve trees or for the patterns where most of the nerve trees are located close to the edge of the observation window.

The results regarding the spatial structure of the end and base points of the thinned healthy patterns are presented in Figure 9. The proposed thinning scheme creates patterns that capture both the end point and base point spatial structure very well as the 95% global envelopes completely cover the empirical centered L curves. The envelopes are based on 2500 simulations from the posterior predictive distribution of the thinning model. In other words, for each mild diabetic neuropathy sample, we simulate data y^* using θ^* , with θ^* sampled from $P_e(\theta|S(y))$, which are then used to calculate the mean $L(r) - r$ for the simulated mild diabetic neuropathy group. Notice that each of the 2500 simulations is generated by selecting one of the approximately 1300 posterior draws available for each mild diabetic sample and one of the 112 healthy diabetic samples at random. Pseudocode for this procedure is given in Algorithm 3, where $T(\cdot)$ is the summary statistic, for example, $T(\cdot) = L(\cdot)$ or $T(\cdot) = K_{mm}(\cdot)$, specified by the user.

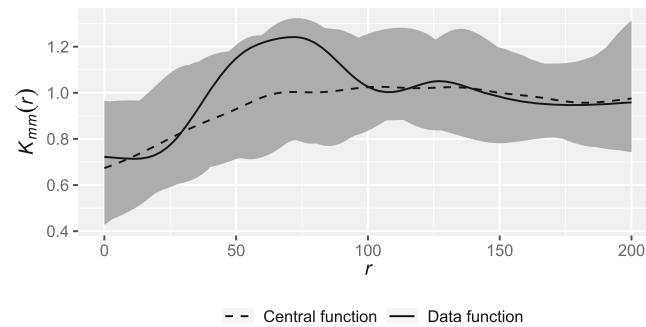


FIGURE 10 Group-wise mark correlation function for the base points marked by the areas of their reactive territories with 95% global envelopes constructed from 2500 simulations from the posterior predictive distribution of the thinning model. The solid curve is the mark correlation function estimated from the mild data.

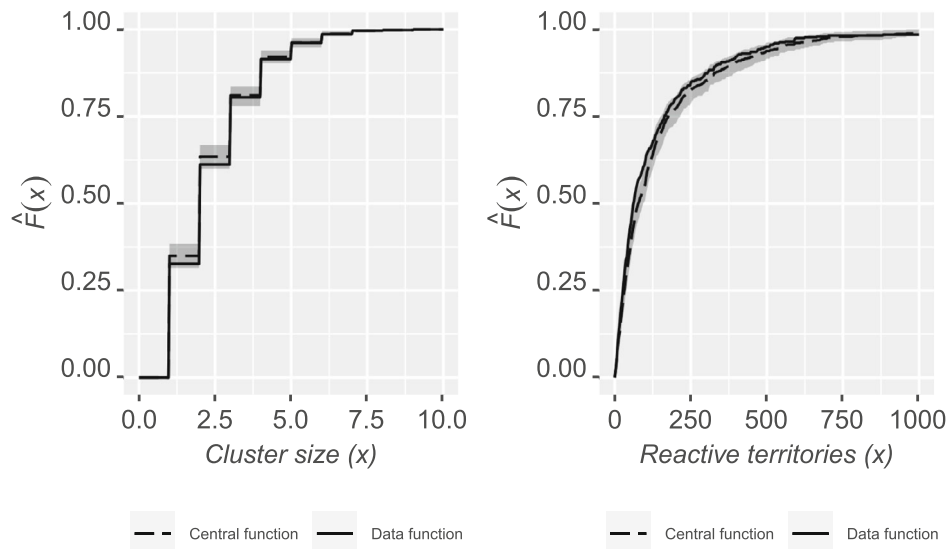


FIGURE 11 Cumulative distribution functions of the cluster size (left) and the total area of reactive territories (right) with 95% global envelopes constructed from 2500 simulations from the posterior predictive distribution of the model. The solid curves are the corresponding cumulative distribution functions estimated from the mild data.

Moreover, for each nerve tree we calculated the area of its reactive territory—convex hull determined by the base and its end points—which was then attached to each base point as a mark. To deal with zero areas, that is, nerve trees with only one end point, the length between the end and the base point was used instead. This length is much smaller than a typical area of a reactive territory. Figure 10 illustrates the mark correlation function with 95% global envelopes constructed using simulations from the posterior predictive distribution as explained above. The thinning model captures even the mark correlation structure between the sizes of the reactive territories well.

We also computed some non-spatial summary statistics, namely the cluster size distribution, that is, the cumulative distribution of the number of end points per nerve tree, and of the total area of the reactive territories to evaluate the dependent thinning model, illustrated in Figure 11. The model seems to capture even these characteristics well.

5.2 | Mild versus moderate

When applied to the healthy patterns, the suggested dependent thinning approach seems to be able to produce patterns similar to the empirical mild patterns. A natural question is whether we can obtain patterns similar to the observed moderate patterns by thinning mild patterns in a similar manner. It can be seen in Figure 5 that the end points in the moderate patterns are more clustered than in the mild patterns but since there is not such a big difference in clustering

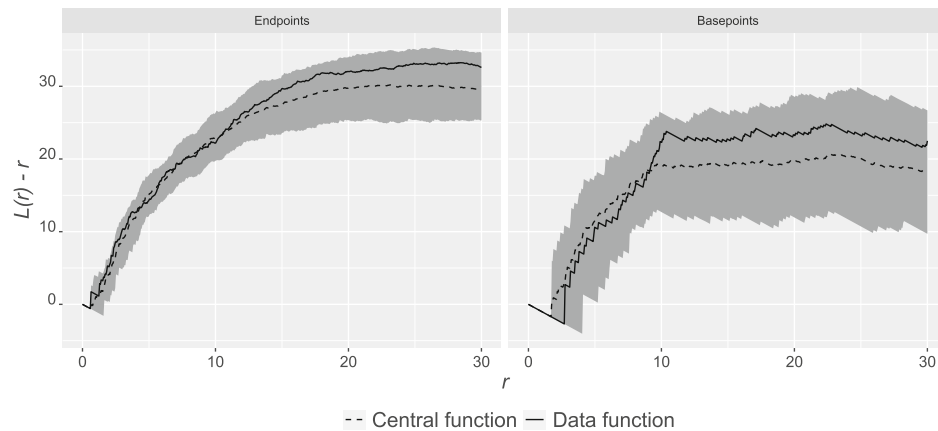


FIGURE 12 Group-wise centered L functions with 95% global envelopes for the end points (left) and base points (right) constructed from 2500 simulations from the independent thinning model. The solid curves are the centered L functions estimated from the moderate data.

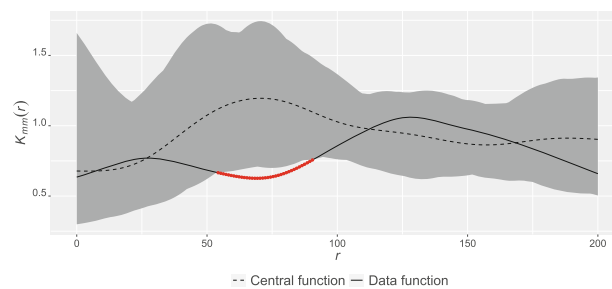


FIGURE 13 Group-wise mark correlation function for the base points marked by the areas of their reactive territories with 95% global envelopes constructed from 2500 simulations from the independent random thinning model. The solid curve is the mark correlation function estimated from the moderate data.

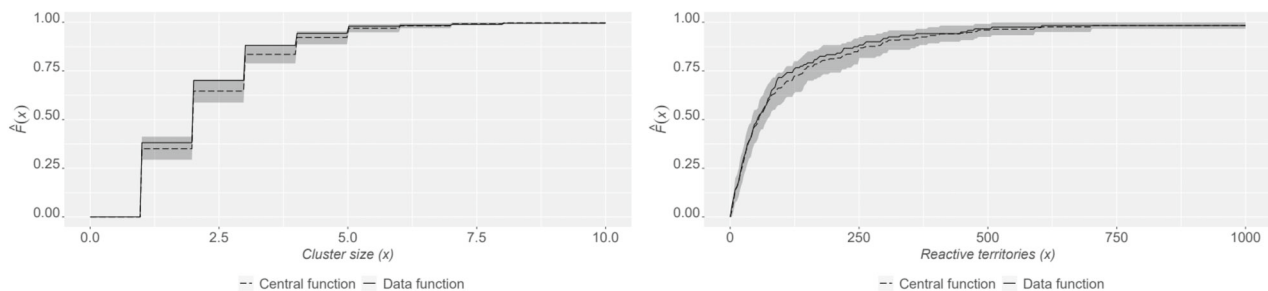


FIGURE 14 Cumulative distribution functions of the cluster size (left) and the total area of reactive territories (right) with 95% global envelopes constructed from 2500 simulations from the independent random thinning model. The solid curves are the corresponding cumulative distribution functions estimated from the moderate data.

of the base points in the two groups, independent random thinning could be appropriate for the base points. Therefore, we randomly thinned the mild patterns by removing nerve trees to the number of nerve trees, that is, base points, in the observed moderate patterns. Note that only the base points, not the end points, are randomly thinned. No parameters need to be estimated in this case as nerve trees are randomly thinned with equal probabilities.

The group-wise centered L functions with 95% global envelopes for the base and end point patterns after independent thinning of nerve trees are given in Figure 12. We observe that the empirical centered L functions lie within the envelopes for the base point patterns indicating a good fit of the model. Even the overall structure of the end point patterns is captured quite well by the model. However, the mark correlation of the sizes of the reactive territories is not completely caught by the independent thinning model, see Figure 13. On the other hand, as seen in Figure 14, the end point cluster size

distribution and the distribution of the area of the reactive territories are quite well described by the model even though in the former case the moderate data based distribution is very close to the upper boundary of the interval.

6 | DISCUSSION

The biological process that guides the physiological changes in the epidermal nerve fiber structure of the neuropathy was investigated. The effects of varying severity of the underlying neuropathic condition, diabetes, were also analyzed. For this purpose, we treated the ENF samples from the feet of healthy patients, and from patients with mild and moderate diabetic neuropathy, as realizations of spatial point processes occurring in response to progressive pathologic severity. As the spatial intensity of the ENFs decreases with progression of neuropathy, the mild diabetic neuropathy patterns can be considered as spatial thinnings of the healthy patterns, and the moderate diabetic neuropathy patterns as spatial thinnings of the mild patterns. Therefore, we proposed spatial thinning models for the changes that occur in the ENF structure in neuropathy as the as diabetes becomes more severe. To the best of our knowledge, this is the first study that considers such spatial thinning models to investigate the nerve removal as a result of varying severity of diabetes, and corresponding severity of neuropathy.

Two spatial thinning models were investigated, an independent random p -thinning and a dependent thinning scheme. The scale parameter controlling the retention probability in the latter model was estimated by using approximate Bayesian computation (ABC), which is a very flexible methodology that is applicable whenever the likelihood function is unavailable or is computationally expensive to evaluate but it is feasible to simulate from a computer model.

We focused first on the nerve removal from the healthy patterns to obtain patterns similar to the observed mild patterns. An independent random p -thinning seems insufficient in modeling the nerve removal process at this earliest stage of diabetic neuropathy. Therefore, a more complex thinning model that favored the removal of isolated nerves in order to increase the overall clustering in the patterns, was proposed. Measured by the L function, mark correlation function, and some non-spatial summary statistics, the model was able to describe the change from healthy to mild diabetic neuropathy very well. On the other hand, the independent removal of nerve trees was enough to model the nerve mortality in mild diabetic neuropathy patients to obtain patterns similar to the observed moderate patterns.

Our original hypothesis was that first, whole ENF trees die due to neuropathy and then, some individual nerve endings may disappear or appear. In our study, it was enough to remove entire nerve trees and no additional removal or addition of individual nerve endings was needed. However, we can see in Figure 14 that the independent thinning of mild patterns only barely covered the corresponding curve estimated from the moderate patterns and in Figure 13 that the fit of the mark correlation function was not perfect in this case. Therefore, even though the suggested models describe the data sufficiently well, they could still be improved.

ACKNOWLEDGEMENTS

The authors thank William R. Kennedy's group (University of Minnesota) for blister immunostaining, quantification and morphometry of the ENF data. The authors also thank the Swedish Research Council for financially supporting the project. Umberto Picchini acknowledges funding from the Swedish National Research Council (Vetenskapsrådet 2019-03924) and the Chalmers AI Research Centre. We are very grateful for the reviewers for their careful reading of the manuscript and thoughtful comments that led to an improved version of the paper.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

Unfortunately, we are not able to share the data publicly.

ORCID

Konstantinos Konstantinou  <https://orcid.org/0000-0003-3549-5851>

REFERENCES

1. Kennedy WR, Wendelschafer-Crabb G, Johnson T. Quantitation of epidermal nerves in diabetic neuropathy. *Neurology*. 1996;47(4):1042-1048.

2. Andersson C, Guttorp P, Särkkä A. Discovering early diabetic neuropathy from epidermal nerve fiber patterns. *Stat Med*. 2016;35(24):4427-4442.
3. Ghorbanpour F, Särkkä A, Pourtaheri R. Marked point process analysis of epidermal nerve fibres. *J Microsc*. 2021;283(1):41-50.
4. Waller LA, Särkkä A, Olsbo V, et al. Second-order spatial analysis of epidermal nerve fibers. *Stat Med*. 2011;30(23):2827-2841.
5. Myllymäki M, Panoutsopoulou IG, Särkkä A. Analysis of spatial structure of epidermal nerve entry point patterns based on replicated data. *J Microsc*. 2012;247(3):228-239.
6. Olsbo V, Myllymäki M, Waller LA, Särkkä A. Development and evaluation of spatial point process models for epidermal nerve fibers. *Math Biosci*. 2013;243(2):178-189.
7. Andersson C, Rajala T, Särkkä A. Hierarchical models for epidermal nerve fiber data. *Stat Med*. 2018;37(13):141-168.
8. Garcia NL, Guttorp P, Ludwig G. Interacting cluster point process model for epidermal nerve fibers. *Spat Stat*. 2020;35:100414.
9. Konstantinou K, Särkkä A. Spatial modeling of epidermal nerve fiber patterns. *Stat Med*. 2021;40(29):6479-6500.
10. Konstantinou K, Särkkä A. Pairwise interaction Markov model for 3D epidermal nerve fibre endings. *J Microsc*. 2022;288(1):54-67.
11. Wendelschafer-Crabb G. Epidermal nerve fiber densities in six body locations of normal and diabetic subjects. *J Peripher Nerv Syst*. 2005;10:104.
12. Panoutsopoulou IG, Wendelschafer-Crabb G, Hodges JS, Kennedy WR. Skin blister and skin biopsy to quantify epidermal nerves: a comparative study. *Neurology*. 2009;72(14):1205-1210.
13. Kennedy WR, Nolano M, Wendelschafer-Crabb G, Johnson TL, Tamura E. A skin blister method to study epidermal nerves in peripheral nerve disease. *Muscle Nerve: Official J Am Assoc Electrodiagnost Med*. 1999;22(3):360-371.
14. Illian J, Penttinen A, Stoyan H, Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns*. Hoboken: John Wiley & Sons; 2008.
15. Diggle P. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Boca Raton: CRC Press, Taylor Francis Group; 2014.
16. Møller J, Waagepetersen R. *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: CRC Press; 2004.
17. Chiu SN, Stoyan D, Kendall WS, Mecke J. *Stochastic Geometry and its Applications*. Hoboken: John Wiley & Sons; 2013.
18. Ripley BD. The second-order analysis of stationary point processes. *J Appl Probab*. 1976;13:255-266.
19. Miles RE. On the homogeneous planar Poisson point process. *Math Biosci*. 1970;6:85-127.
20. Diggle PJ, Mateu J, Clough HE. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Adv Appl Probab*. 2000;32(2):331-343.
21. Stoyan D, Stoyan H. *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics*. Vol 302. Chichester: John Wiley & Sons Incorporated; 1994.
22. Silverman BW. *Density Estimation*. London: Chapman and Hall; 1986.
23. Myllymäki M, Särkkä A, Vehtari A. Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates. *Spat Stat*. 2014;8:104-121.
24. Sisson SA, Fan Y, Beaumont M. *Handbook of Approximate Bayesian Computation*. Boca Raton: CRC Press; 2018.
25. Marin JM, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. *Stat Comput*. 2012;22(6):1167-1180.
26. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol*. 1999;16(12):1791-1798.
27. Marjoram P, Molitor J, Plagnol V, Tavaré S. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci*. 2003;100(26):15324-15328.
28. Picchini U. Inference for SDE models via approximate Bayesian computation. *J Comput Graph Stat*. 2014;23(4):1080-1100.
29. Sisson SA, Fan Y, Tanaka MM. Sequential Monte Carlo without likelihoods. *Proc Natl Acad Sci*. 2007;104(6):1760-1765.
30. Beaumont MA, Cornuet JM, Marin JM, Robert CP. Adaptive approximate Bayesian computation. *Biometrika*. 2009;96(4):983-990.
31. Del Moral P, Doucet A, Jasra A. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat Comput*. 2012;22(5):1009-1020.
32. Fearnhead P, Prangle D. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J R Stat Soc Series B Stat Methodology*. 2012;74(3):419-474.
33. Kuronen M, Myllymäki M, Loavenbruck A, Särkkä A. Point process models for sweat gland activation observed with noise. *Stat Med*. 2021;40:2055-2072.
34. Csillery K, Francois O, Blum MGB. Abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol*. 2012;3(3):475-479.
35. Myllymäki M, Mrkvička T, Grabarnik P, Seijo H, Hahn U. Global envelope tests for spatial processes. *J R Stat Soc Series B Stat Methodology*. 2017;79(2):381-404.
36. Narisetty NN, Nair VN. Extremal depth for functional data and applications. *J Am Stat Assoc*. 2016;111(516):1705-1714.

How to cite this article: Konstantinou K, Ghorbanpour F, Picchini U, Loavenbruck A, Särkkä A. Statistical modeling of diabetic neuropathy: Exploring the dynamics of nerve mortality. *Statistics in Medicine*. 2023;42(23):4128-4146. doi: 10.1002/sim.9851

APPENDIX A. GLOBAL ENVELOPE TESTS

Global envelope tests are graphical Monte-Carlo tests based on functional or multivariate summary statistics.³⁵ Let T_1, T_2, \dots, T_s be d -dimensional vectors containing the discretizations of a functional statistic, such as the K or L function, at d points, with T_1 denoting the data vector and T_2, \dots, T_s the vectors obtained under the model corresponding to a null hypothesis (“null model”). An envelope is defined as the band bounded from below by the d -dimensional vector T_{low} and from above by the d -dimensional vector T_{upp} . Then, a $100(1 - \alpha)\%$ global envelope is a set $(T_{low}^\alpha, T_{upp}^\alpha)$ such that the probability that T_i falls outside the envelope in any of the d points is α . In the recent point process literature, such tests are widely used to assess the goodness-of-fit of spatial point process models, as well as constructing global confidence bands from a set of functions obtained from the posterior predictive distribution. Construction of a global envelope $(T_{low}^\alpha, T_{upp}^\alpha)$ depends on the measure used to rank the extremeness of the vectors T_1, T_2, \dots, T_s as well as on the significance level α . Here, we used the extreme rank length measure (ERL) to construct 95% global envelopes.^{35,36} Using the ERL measure the different statistics are initially ranked for each value r in the discretization grid. Then, the number of r values where the statistic is extreme is taken into account to construct the ERL measure E_i for each statistic T_1, T_2, \dots, T_s , with $E_i < E_j$ interpreted as T_i is more extreme than T_j . Now let $E_{(\alpha)} \in \mathbb{R}$ be the largest E_i such that

$$\sum_{i=1}^s \mathbf{1}(E_i < E_{(\alpha)}) \leq \alpha s$$

and let $I_{(\alpha)}$ be the set of vectors less than or as extreme as $E_{(\alpha)}$. Then, the $100(1 - \alpha)\%$ global envelope based on the ERL measure is given by

$$\left(T_{low\ k}^{(\alpha)}, T_{upp\ k}^{(\alpha)} \right) = \left(\min_{i \in I_{(\alpha)}} T_{ik}, \max_{i \in I_{(\alpha)}} T_{ik} \right) \quad \text{for } k = 1, \dots, d.$$

APPENDIX B. SAMPLE DATASET

Here, we provide a sample dataset (see Figure B1) consisting of four ENF samples obtained from two subjects. The aim is to clarify the notation used in the paper through an illustrative example. The weights to calculate the subjectwise $\hat{K}_i(r)$ summary functions for the endpoint patterns (blue points) are computed as follows:

$$n = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} \# \text{ blue points in sample 1 of subject 1} & \# \text{ blue points in sample 2 of subject 1} \\ \# \text{ blue points in sample 1 of subject 2} & \# \text{ blue points in sample 2 of subject 2} \end{pmatrix} = \begin{pmatrix} 40 & 25 \\ 19 & 6 \end{pmatrix}$$

$$w = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} = \begin{pmatrix} \frac{40^2}{40^2 + 25^2} & \frac{25^2}{40^2 + 25^2} \\ \frac{19^2}{19^2 + 6^2} & \frac{6^2}{19^2 + 6^2} \end{pmatrix} = \begin{pmatrix} 0.719 & 0.281 \\ 0.909 & 0.91 \end{pmatrix}.$$

The weights to calculate the groupwise $\hat{K}_g(r)$ summary function for the endpoint patterns are computed as follows:

$$n = \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} = \begin{pmatrix} \# \text{ blue points in samples of subject 1} \\ \# \text{ blue points in samples of subject 2} \end{pmatrix} = \begin{pmatrix} 40 + 25 \\ 19 + 6 \end{pmatrix} = \begin{pmatrix} 65 \\ 25 \end{pmatrix}$$

$$w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \frac{65^2}{65^2 + 25^2} \\ \frac{25^2}{65^2 + 25^2} \end{pmatrix} = \begin{pmatrix} 0.871 \\ 0.129 \end{pmatrix}.$$

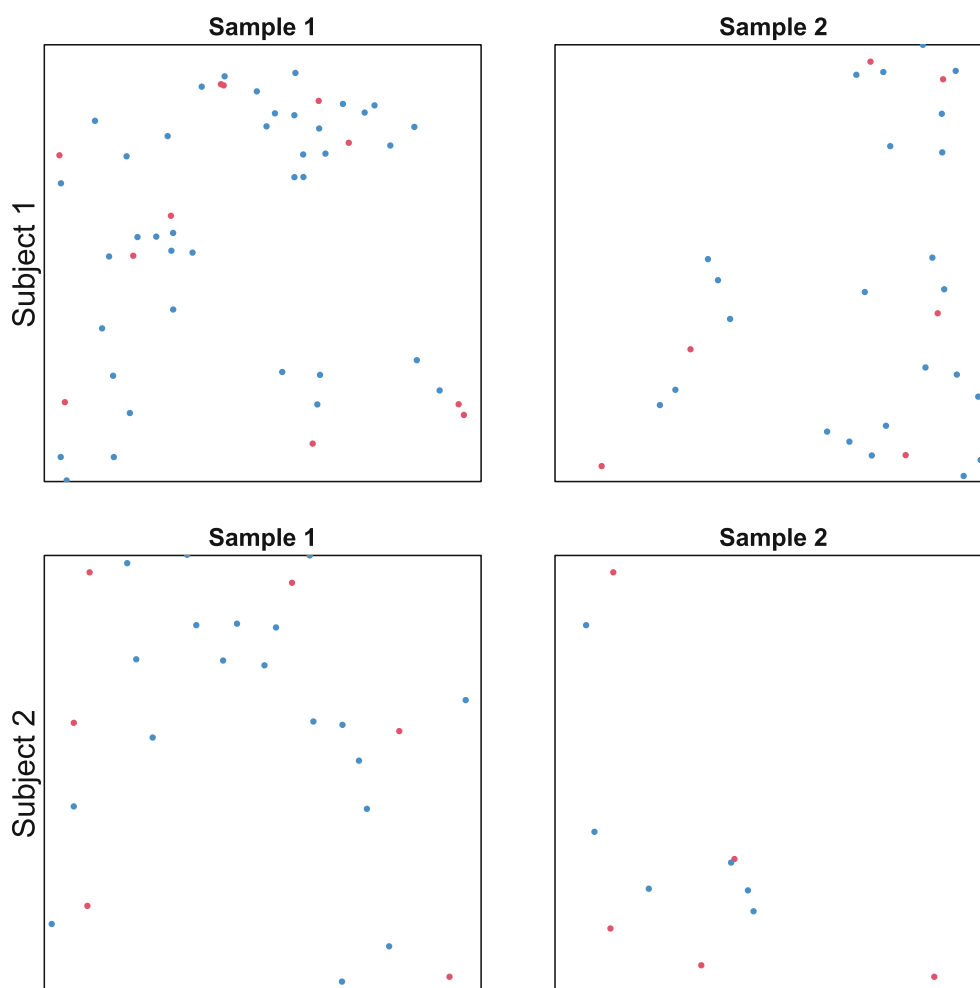


FIGURE B1 Sample dataset of four ENF samples from two subjects. Red points represent the base points, and blue points represent the end points of the nerve fibres.