Author(s): Myllymäki, Mari & Särkkä, Aila & Vehtari, Aki

Title: Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates

Year: 2014

Version: Post print

Please cite the original version:

Myllymäki, Mari & Särkkä, Aila & Vehtari, Aki. 2014. Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates. Spatial Statistics. Volume 8. 104-121. ISSN 2211-6753 (printed). DOI: 10.1016/j.spasta.2013.07.006.

Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates

Mari Myllymäki[a,1,*], Aila Särkkä[b], Aki Vehtari[a]

[a] *Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P. O. Box 12200, FI-00076 AALTO, Finland*
[b] *Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, S-41296 Gothenburg, Sweden*

Abstract

In this paper we propose a method on how to incorporate the effect of non-spatial covariates into the spatial second-order analysis of replicated point patterns. The variance stabilizing transformation of Ripley's K function is used to summarize the spatial arrangement of points, and the relationship between this summary function and covariates is modelled by hierarchical Gaussian process regression. In particular, we investigate how disease status and some other covariates affect the level and scale of clustering of epidermal nerve fibers. The data are point patterns with replicates extracted from skin blister samples taken from 47 subjects.

Keywords: Epidermal nerve fiber, functional data analysis, Gaussian process, K function, replicated point pattern, spatial point process

1. Introduction

There is a wide range of methods for analysing and modelling spatial point patterns (see e.g. Cressie [1, Chapter 8], Diggle [2] and Illian et al. [3]). However, the majority of the methods are designed to analyze a single point pattern, while more and more replicated point pattern data are collected and need to be analyzed. In this paper, we suggest a flexible non-parametric approach to study the relationship between non-spatial covariates and spatial

*Corresponding author. Tel: +358 50 512 4490. Fax: +358 9 470 23182
Email addresses: mari.myllymaki@aalto.fi (Mari Myllymäki), aila@chalmers.se (Aila Särkkä), aki.vehtari@aalto.fi (Aki Vehtari)

Accepted to Spatial Statistics(2013), http://dx.doi.org/10.1016/j.spasta.2013.07.006

second-order structure based on replicated data. The motivation comes from neurological studies, where samples are taken from subjects for whom some covariate information is available.

Some tools for analyzing replicated point patterns are already available. Diggle et al. [4] and Baddeley et al. [5] introduced how some summary statistics can be estimated from replicated point patterns, and further applied these pooled summary statistics to test whether spatial patterns are different in different groups. Such methods were further used e.g. in Schladitz et al. [6] and Redenbach et al. [7]. Hahn [8], moreover, proposed a non-parametric test for direct comparison of two or more point patterns. Some model-based inference for replicated point patterns can be found in Diggle et al. [9], Eckel et al. [10], Bell and Grunwald [11], Illian and Hendrichsen [12], and Illian et al. [13]. In the first four papers Gibbs point process models were fitted to replicated data using the pseudo-likelihood method, whereas Illian et al. [13] used log Gaussian Cox process models in a Bayesian setting. In the three latter papers some spatial covariates were added in the model for the first-order intensity.

In this paper, we suggest an approach to model the dependence of the second-order structure of a point pattern on some non-spatial covariates, when the data are replicated point patterns. The spatial arrangement of points is summarized using the variance stabilizing transformation of Ripley's K function [14]. Then, hierarchical Gaussian process regression is used to study the relationship between the covariates and the summary function. This model is a flexible non-parametric model, where we do not need to assume a priori linear or any other particular form of dependence between the summary function and the covariates. Myllymäki et al. [15] studied the effect of non-spatial covariates on the same summary function. However, the mixed model approach they proposed allows only limited flexibility.

Gaussian processes have been widely used in spatial statistics, especially in geostatistics, where they are often called Gaussian random fields or Gaussian functions, see e.g. Matheron [16], Cressie [1], Chilés and Delfiner [17], Diggle and Ribeiro [18] and Gelfand et al. [19], and in spatial epidemiology, see e.g. Banerjee et al. [20]. For point processes, log Gaussian Cox processes have been considered, see e.g. Møller et al. [21] and Rue et al. [22]. The term "Gaussian process model" is used in the machine learning literature, see Rasmussen and Williams [23] where the machine learning perspective is comprehensively summarized. O'Hagan [24] was one of the first ones to consider Gaussian processes in a general probabilistic modelling context similar

to the one we use in the current paper. Following the guidelines in Rasmussen and Williams [23] and Neal [25], we employ hierarchical Gaussian models in a Bayesian framework, where a Gaussian process serves as a prior for a latent function.

Our particular aim is to study the spatial structure of epidermal nerve fibers (ENFs), which are thin nerve fibers in the epidermis, the outmost living layer of the skin. We will use the hierarchical Gaussian regression model to study covariate effects on the spatial pattern of locations of ENF *entry* and *end* points. The entry points are the locations where the trunks of the nerves enter the epidermis, and the end points are the terminal nodes of the nerve fibers in the epidermis. The data are samples from both healthy and diseased subjects, where the latter suffer from mild or moderate diabetic neuropathy. The interest in the second-order property of the patterns stems from neurological studies [26, 27]. While Kennedy et al. [26] reported diminished numbers of ENFs per surface area as well as reduced summed length of all ENFs per volume (or area) in subjects suffering from diabetic neuropathy, Kennedy et al. [27] observed that nerve fiber loss due to diabetic neuropathy seems to result in more clustered pattern of ENFs. This observation of clustering was quantified by Waller et al. [28] based on second-order analysis of ENF entry points extracted from suction-induced skin blister images [27] from the thighs of four subjects. They analyzed one sample from a healthy subject, and two samples from each of the three diseased subjects, one of which suffered from mild, one from moderate and one from severe diabetic neuropathy. Our aim is to quantify the observation of increased clustering due to diabetic neuropathy based on much larger data, accounting for effects of other covariates and intra-subject and inter-subject variation.

The rest of the paper is organized as follows. The data are described in Section 2. Section 3 introduces the statistical methods, which are specified for the ENF data in Section 4. The results for the ENF data are presented in 5 and Section 6 is for further discussion.

2. Data

Two skin blister specimens were obtained from the right calf of 47 subjects using the suction skin blister method, see Wendelschafer-Crabb et al. [29] and Panoutsopoulou et al. [30]. Among these subjects 32 were healthy and had no symptoms or history of peripheral neuropathy, and 15 suffered from mild or moderate diabethic neuropathy. Age, gender and body mass index (BMI)

of each subject were also recorded. (Body mass index is defined as the weight (kg) divided by the squared height (m).)

From three to six images (usually four) with a surface area of approximately 432×330 microns were taken from the two blisters, typically two images from each blister. For each image, ENFs were immunostained, imaged confocally, and traced to determine ENFs in the epidermis, see more details of the technique in Panoutsopoulou et al. [30] and also in Waller et al. [28]. In our analysis, we combine all the images from one subject ignoring the fact that they originate from two blisters since there should be no blister effect present (personal communication with William R. Kennedy; Panoutsopoulou et al. [30]). Moreover, our own inspection of the data confirmed that samples taken from the same blister did not appear to be more similar to each other than samples taken from two different blisters from the same subject.

We study the locations of ENF entry and end points traced from the images. The original data are three dimensional with small variation in the z-direction, but we focus on the spatial pattern of the ENF coverage across the skin and analyze the two dimensional projection of the patterns as Waller et al. [28], Myllymäki et al. [15] and Olsbo et al. [31]. Figure 1 shows examples (two healthy and two diseased) of the ENF entry and end point patterns. Samples having less than ten entry points were excluded from the spatial analysis of entry points. In the remaining patterns the number of entry points varied between 10 and 53 (mean 22.6), and the number of end points between 16 and 160 (mean 63.3).

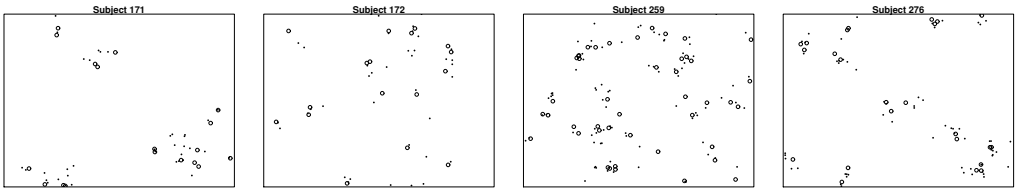

Figure 1: Four point patterns of ENF entry (circles) and end (small black dots) points taken from the right calf of Subjects 171, 172 (healthy), 259 and 276 (diseased). The window size is 432×330 microns.

3. Methods

This section introduces the methods for the second-order analysis of replicated point patterns, first the summary function and then the hierarchical Gaussian process model. We further present the covariance functions employed in the ENF application in Section 4. We assume throughout the paper that we have N subjects ($s = 1, \ldots, N$) for which J_s point pattern samples ($j = 1, \ldots, J_s$) and covariate information are available.

3.1. Summary statistic

We regard ENF entry and end point patterns (see Figure 1) as realizations of stationary and isotropic spatial point processes $\Phi = \{z_i\}$, where $z_i \in \mathbb{R}^2$, and use Ripley's K function [32] to summarize the spatial second-order structure. Let λ denote the intensity (mean number of points per unit area) of the stationary point process. Then $\lambda K(r)$ gives the mean number of additional points of the process within distance $r > 0$ from a typical process point. As an estimator for the K function, we use

$$\hat{K}(r) = \frac{|W|^2}{N_\Phi^2} \sum_{k=1}^{N_\Phi} \sum_{l=1, l\neq k}^{N_\Phi} \frac{\mathbf{1}(\|z_k - z_l\| \leq r)}{|W_{z_k} \cap W_{z_l}|}, \tag{1}$$

where N_Φ is the number of observed points of the process Φ in the window $W \subseteq \mathbb{R}^2$, and $\|z_k - z_l\|$ denotes the distance between z_k and z_l. Furthermore, $|W_{z_k} \cap W_{z_l}|$ is the translational edge correction term, the area of the intersection of W_{z_k} and W_{z_l}, where W_z is the translated window $W_z = \{s+z : s \in W\}$ (see e.g. Illian et al. [3]).

When we have replicates, we may want to estimate the overall average of the K functions of all samples of all subjects. The estimator (1) can be used to obtain the K function for each sample resulting in $\hat{K}_{sj}(r)$. Then, the sample specific functions $\hat{K}_{sj}(r)$ can be pooled together to construct subject specific K functions, and these can further be combined to the overall pooled K function for the data, namely

$$\hat{K}_{\text{pooled}}(r) = \frac{1}{\sum_{s=1}^{N} n_s^2} \sum_{s=1}^{N} n_s^2 \left(\sum_{j=1}^{J_s} \frac{n_{sj}^2}{\sum_{l=1}^{J_s} n_{sl}^2} \hat{K}_{sj}(r) \right), \tag{2}$$

where the number of points in the sample j of subject s, n_{sj}, and $n_s = \sum_{j=1}^{J_s} n_{sj}$ define weights for $\hat{K}_{sj}(r)$ [9, 6].

Besag [33] suggests the following variance stabilizing transformation of the K function,

$$L(r) = \sqrt{K(r)/\pi}. \tag{3}$$

The L function allows a more readily interpretable diagnostic tool than the K function, since the centred L function, $L(r) - r$, can be plotted against r, and the resulting curve can be compared to zero (expected value under complete spatial randomness). That is, under clustering, $L(r) - r$ tends to be larger than zero, whereas under regularity, typically $L(r) - r$ is less than zero. An estimator for the L function is obtained by applying the square root transformation (3) to the empirical K function (1). Similarly, the estimated pooled centred L function is obtained as

$$\hat{L}_{\text{pooled}}(r) - r = \sqrt{\hat{K}_{\text{pooled}}(r)/\pi} - r, \tag{4}$$

where $\hat{K}_{\text{pooled}}(r)$ is given in (2).

3.2. Hierarchical Gaussian process regression

In the following, using the terminology in Rasmussen and Williams [23], we propose a hierarchical Gaussian process model for the L functions and describe the inference in the general case, where we have data from several subjects, replicates from each subject and covariate information. We also introduce methods for investigation of the covariate effects: the posterior predictive L functions indicate how the L function is affected by the covariates, and the average predictive comparison, which we specify for the L functions, serves as a formal investigation of the significance of the covariate effects.

3.2.1. Model

Assume that the centred L function has been estimated through (1) for each sample. Let $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ be the vector of observations (target values) containing the estimated centred L function values for the chosen distances $r_1, \ldots, r_R$ for all samples of all subjects. Denote the corresponding inputs by $\boldsymbol{X} = \{\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,D})^T\}_{i=1}^n$, where $\boldsymbol{x}_i \in \mathbb{R}^D$. The input $\boldsymbol{x}_i$ contains the information on the covariates and specifies also the distance r,

the subject and the sample for the observation y_i. We use the shorthand notation

$$\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_D)^T = (r, \text{covariate}_1, \ldots, \text{covariate}_C, s, j)^T$$

(similarly for $\boldsymbol{x}'$), where C is the number of covariates, $D = C+3$, and s and j are the identity codes for the subject and the sample, respectively. The covariate values and s are of course the same for all y_i from the same subject.

Dependence between the observations and inputs is modeled by a (non-parametric) latent function $f(\boldsymbol{x})$ which has the value $f_i = f(\boldsymbol{x}_i)$ at $\boldsymbol{x}_i$. A set of latent variables at $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is denoted by $\boldsymbol{f} = (f_1, \ldots, f_n)^T$. The latent function $f(\boldsymbol{x})$ represents an unknown functional description of the dependence between the observations and inputs, where the observations are noisy realizations of this underlying function. Due to the hierarchical structure in the data (several subjects and replicates from each subject), it is natural to assume that the centred L function consists of four components: a general level $f^{(1)}$ for a subject with specific covariates, a subject specific effect $f^{(2)}$ modelling the subject's individual contribution (due to other effects than the observed covariates), a correlated noise component $f^{(3)}$ modelling the variation due to the sample within the subject, and an uncorrelated noise component. The first three components form a hierarchical latent Gaussian process

$$\begin{aligned} f(\boldsymbol{x}) &= f^{(1)}(\boldsymbol{x}) + f^{(2)}(\boldsymbol{x}) + f^{(3)}(\boldsymbol{x}) \\ &= f^{(1)}(r, \text{covariates}) + f^{(2)}(r, s) + f^{(3)}(r, s, j), \end{aligned} \tag{5}$$

where the second row describes inputs on which the components depend. Each function $f^{(t)}$, $t = 1, 2, 3$, is a priori assumed to be a real-valued Gaussian process (GP), which is characterized by its mean $m_t(\boldsymbol{x})$ and covariance function $k_t(\boldsymbol{x}, \boldsymbol{x}'|\theta_t)$ with parameter vector θ_t. Consequently, the sum (5) is also a Gaussian process, namely

$$\begin{aligned} f(\boldsymbol{x})|\theta \;\sim\; & GP(m_1(\boldsymbol{x}) + m_2(\boldsymbol{x}) + m_3(\boldsymbol{x}), \\ & k_1(\boldsymbol{x}, \boldsymbol{x}'|\theta_1) + k_2(\boldsymbol{x}, \boldsymbol{x}'|\theta_2) + k_3(\boldsymbol{x}, \boldsymbol{x}'|\theta_3)), \end{aligned} \tag{6}$$

where $\theta = (\theta_1, \theta_2, \theta_3)$.

The finite-dimensional distributions of a Gaussian process are all Gaussian and, thus,

$$\boldsymbol{f}|\theta \sim N(m(\boldsymbol{X}), \Sigma_\theta), \tag{7}$$

where $m(\boldsymbol{X}) = (m(\boldsymbol{x}_1), \ldots, m(\boldsymbol{x}_n))^T$ is the known mean vector (obtained as the sum of the three mean vectors) and Σ_θ is the covariance matrix whose entries are determined by the corresponding covariance functions.

The fourth model component, the uncorrelated noise, corresponds to the observation model

$$p(\boldsymbol{y}|f, \sigma^2) = \prod_{i=1}^{n} N(y_i|f_i, \sigma^2). \tag{8}$$

That is, given the latent function f in (6), the y_i's are assumed to be conditionally independent and normally distributed with mean f_i and variance σ^2.

Gaussian process regression aims at reconstructing the latent function f. That is, the interest is in the posterior predictive distribution of f, after observing the data. For Bayesian inference, we need to specify priors for the means and the covariance functions of $f^{(t)}$, $t = 1, 2, 3$, and for the error variance σ^2 in (8).

We assume known prior means. For the components $f^{(2)}$ and $f^{(3)}$ zero-mean is a reasonable assumption, whereas the prior mean of $f^{(1)}$ should be chosen depending on the application: the covariate effects are inspected as deviations from this mean level. Note that having a known prior mean corresponds to modelling normalized data (data minus the mean) by a zero-mean Gaussian process.

For the covariance functions, we use parametrized models $k_t(\boldsymbol{x}, \boldsymbol{x}'|\theta_t)$, since they are parsimonious and our data are not likely to provide enough information to estimate covariance functions non-parametrically from the data. The choice of the parametric forms for the covariance functions k_t, $t = 1, 2, 3$, in (6) depends on the application, but some general rules can be given for them: Since the first component $f^{(1)}$ represents the highest level of hierarchy giving the mean L function for subjects with specific covariates, it can be assumed to be quite smooth. The Gaussian (or squared exponential) covariance function, which belongs to the Matérn family and is infinitely many times differentiable [20, 1, 23], is often a reasonable choice. The second model component $f^{(2)}$ may also be assumed to be quite smooth, since it represents the mean L function of a subject. However, since empirical L functions are never really smooth, it is reasonable to assume that the third component $f^{(3)}$, being at the lowest level of the hierarchy, is less smooth. Each of these components are considered below in detail for the ENF application.

Note that in our application, we allow dependencies between all covariates

and r. Thus, all covariates appear in the same model component $f^{(1)}$ and their effects are incorporated into the covariance function $k_1(\boldsymbol{x}, \boldsymbol{x}'|\theta_1)$. The correlation is then allowed to decrease in a different way for distance r and for each of the covariates by letting each of them have its own (length-scale) parameter controlling the range of correlation. If joint effects of covariates were excluded from the model, the first component would reduce to the sum

$$f^{(1)}(r, \text{covariates}) = f^{(1,1)}(r, \text{covariate}_1) + \cdots + f^{(1,C)}(r, \text{covariate}_C),$$

where $f^{(i,c)}$, $c = 1, \ldots, C$, are Gaussian processes.

Once the covariance functions have been chosen, hyperpriors are to be specified for their parameters. The parametric model (typically) determines the smoothness of the Gaussian process, whereas its parameters control the range of correlation and variance.

3.2.2. Inference and posterior predictive L functions

Because both the prior (7) and the likelihood (8) are Gaussian, we can integrate out the latent function analytically and perform numerical integration for the hyperparameters only. The log marginal likelihood given the hyperparameters equals

$$\begin{aligned} \log p(\boldsymbol{y}|\boldsymbol{X}, \theta, \sigma^2) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_\theta + \sigma^2 I| - \\ &\quad \frac{1}{2}(\boldsymbol{y}^T - m(\boldsymbol{X}))(\Sigma_\theta + \sigma^2 I)^{-1}(\boldsymbol{y} - m(\boldsymbol{X})), \end{aligned} \tag{9}$$

where θ collects all the parameters of f, and Σ_θ is the covariance matrix in (7). Combining the prior distributions with (9) we can then obtain the marginal posterior distribution $p(\theta, \sigma^2|\boldsymbol{y}, \boldsymbol{X})$ for the hyperparameters.

The main interest is in the posterior distribution of the latent function f (and of $f^{(t)}$, $t = 1, 2, 3$), namely

$$p(f|\boldsymbol{y}, \boldsymbol{X}) = \int p(f|\boldsymbol{y}, \boldsymbol{X}, \theta, \sigma^2) p(\theta, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) \, \mathrm{d}\theta \, \mathrm{d}\sigma^2, \tag{10}$$

with any input values (not necessarily the observed ones) $\boldsymbol{X}$. The mean of the posterior distribution of $f^{(1)}$ for a virtual subject with specific covariate values gives the mean behavior of the centred L function and, for example, the 5% and 95% r-wise quantiles of the posterior distribution can be used to characterize the uncertainty. Effects of the covariates on the L function

can then be inspected by varying the values of the covariates. By calculating (10) for $f^{(1)}$, $f^{(1)} + f^{(2)}$ and $f^{(1)} + f^{(2)} + f^{(3)}$ for the subjects in our data, we can inspect different components of the model and compare them to the curves estimated from the data.

The posterior distribution (10) can be calculated by Monte Carlo integration over the hyperparameters. One possibility would be to find a maximum a posteriori estimate of the hyperparameters [23], but this optimization does not account for the uncertainty in the hyperparameters. Rue et al. [22] used the grid and central composite design sampling methods, but these work well only for a low number of hyperparameters. Since in our application we have many hyperparameters and also subject-specific parameters, we use a Markov Chain Monte Carlo (MCMC) method to sample from $p(\theta, \sigma^2|\boldsymbol{y}, \boldsymbol{X})$ and to obtain the posterior distribution (10).

3.2.3. Average predictive comparisons

In order to investigate the significance of the covariate effects on the L function, we propose to use average predictive comparisons introduced by Gelman and Pardoe [34]. This approach works only for a single outcome, and we need to choose a summary statistic that reduces the information contained in the L function (on a given interval) to a single number. Examples of such summary statistics are the maximum summary statistic

$$d_{\max} = \max_{r \in [r_{\min}, r_{\max}]} (L(r) - r) \tag{11}$$

and the integral summary statistic

$$d_{\text{int}} = \int_{r_{\min}}^{r_{\max}} (L(r) - r)\, \mathrm{d}r, \tag{12}$$

where $r_{\min}$ and $r_{\max}$ define the range of distances under consideration.

Let d_{stat} be the chosen statistic, e.g. $d_{\max}$ or d_{int}. Letting u be the value of one of the covariates and v represent all the other covariates, the average predicted comparison gives the expected difference in the outcome associated with a unit difference in u. Note that the covariate effects are incorporated into the model component $f^{(1)}$ and, therefore, we are interested in the outcome d_{stat} calculated from the L function given by this component. For an increasing transition from $u^{(1)}$ to $u^{(2)}$, we focus on the expected change in d_{stat} and our average predictive comparison looks at the mean value (and

distribution) of

$$\delta_u(u^{(1)} \to u^{(2)}, v, \theta_M) = \frac{g(d_{\text{stat}}|u^{(2)}, v, \theta_M) - g(d_{\text{stat}}|u^{(1)}, v, \theta_M)}{u^{(2)} - u^{(1)}}$$

over the distribution of the covariates and the estimated model parameters θ_M. Here, $g(d_{\text{stat}}|u, v, \theta_M)$ is a value of d_{stat} for a centred L function given u, v and θ_M.

Let $\theta_M^1, \ldots, \theta_M^B$ be a sample from the posterior distribution of the model parameters θ_M, and let $(u_s, v_s), s = 1, \ldots, N$, denote the values of all covariates for all subjects s. As an estimator for the average expected change in our outcome, following Gelman and Pardoe [34], we use

$$\hat{\Delta}_u = \frac{1}{B} \sum_{b=1}^{B} \hat{\Delta}_u(\theta_M^b) \tag{13}$$

with

$$\hat{\Delta}_u(\theta_M^b) = \frac{\sum_{s_1=1}^{N} \sum_{s_2=1}^{N} w_{s_1,s_2} \left(g(d_{\text{stat}}|u_{s_2}, v_{s_1}, \theta_M^b) - g(d_{\text{stat}}|u_{s_1}, v_{s_1}, \theta_M^b) \right) \text{sign}(u_{s_2} - u_{s_1})}{\sum_{s_1=1}^{N} \sum_{s_2=1}^{N} w_{s_1,s_2} (u_{s_2} - u_{s_1}) \text{sign}(u_{s_2} - u_{s_1})}, \tag{14}$$

where w_{s_1,s_2}'s are weights, multiplying by $\text{sign}(u_{s_2} - u_{s_1})$ guarantees that only increasing transitions of u are considered ($\text{sign}(u_{s_2} - u_{s_1})$=1 if $u_{s_2} - u_{s_1} > 0$ and -1 if $u_{s_2} - u_{s_1} < 0$) and $g(d_{\text{stat}}|u_{s_1}, v_{s_1}, \theta_M^b)$ is the statistic d_{stat} calculated from a sample drawn from the distribution of $f^{(1)}$ given the covariate values and the parameter vector θ_M^b. We use the weights $w_{s_1,s_2} = 1/(1 + (v_{s_1} - v_{s_2})^T \Sigma_v^{-1} (v_{s_1} - v_{s_2}))$, where Σ_v is the covariance matrix of v, suggested by Gelman and Pardoe [34] to give more weight to those changes in the covariates that are in the support of the joint distribution of the data. Using the Mahalanobis weights corresponds to the use of a simplified Gaussian approximation to the joint distribution of the covariates v.

For each covariate u, we have chosen to present the distribution of (14), i.e. the distribution of the difference in d_{stat} with a positive unit difference in the covariate u. Alternatively, means and standard errors could be presented [34].

3.2.4. Subject specific predictive comparison

In addition, we define subject specific predictive comparisons as those that focus on the change in the outcome d_{stat} with v being fixed to the observed

covariate values of the subject. The distribution of the expected change in d_{stat} in subject s is estimated by the distribution of

$$\hat{\Delta}_{u,s}(\theta_M^b) = \frac{\sum_{s_2=1}^{N} w_{s,s_2} \left(g(d_{\text{stat}}|u_{s_2}, v_s, \theta^b) - g(d_{\text{stat}}|u_s, v_s, \theta^b) \right) \text{sign}(u_{s_2} - u_s)}{\sum_{s_2=1}^{N} w_{s,s_2} (u_{s_2} - u_s) \text{sign}(u_{s_2} - u_s)}. \tag{15}$$

The distribution of (15) for each subject can be illustrated e.g. by plotting its mean value and error bars showing some quantiles of the distribution.

3.3. Piecewise polynomial covariance functions with compact support

For the ENF application, we employ compactly supported covariance functions (see e.g. Rasmussen and Williams [23]), which have the advantage that the covariance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ becomes zero when the distance between them exceeds a certain threshold. As a consequence, the covariance matrix will become sparse by construction leading to computational advantages.

We will employ covariance functions that are in the class of piecewise polynomial covariance functions $k_{ppD,q}(h|\theta)$ of order q, which are positive definite in $\mathbb{R}^D$. These covariance functions are $2q$-times continuously differentiable and the corresponding processes are q-times continuously mean-square differentiable [23]. The order q plays a similar role as the smoothness parameter of the well-known Matérn covariance functions; the larger the order q is, the smoother are the realisations of the process.

We have chosen to use the first and third order piecewise polynomial covariance functions which are defined as

$$k_{ppD,1}(h;\theta) = \sigma_{pp}^2 \max(0, 1-h)^{j+1}((j+1)h+1) \tag{16}$$

and

$$\begin{aligned} k_{ppD,3}(h;\theta) &= \sigma_{pp}^2 \max(0, 1-h)^{j+3} \Big((j^3 + 9j^2 + 23j + 15)h^3 \\ &\quad + (6j^2 + 36j + 45)h^2 + (15j + 45)h + 15 \Big)/15, \end{aligned} \tag{17}$$

respectively, for $j = \lfloor \frac{D}{2} \rfloor + q + 1$. The parameter vector θ consists of the variance parameter σ_{pp}^2 and the length-scale (or range) parameters ϕ_d, $d = 1, \ldots, D$, which enter the covariance function through h, namely

$$h = h(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{\sum_{d=1}^{D} \frac{(x_d - x_d')^2}{\phi_d^2}}. \tag{18}$$

The length-scale parameter ϕ_d controls how fast correlation decreases in the input direction d.

4. Analysis of the ENF data

We investigate which of the covariates, age, gender, BMI and disease status, if any, affect the second-order structure of the ENF entry and end point patterns in samples taken from right calves of the subjects. The solid lines in Figures 2 and 3 show the estimated centred L functions for the ENF entry and end point patterns, respectively, for four subjects. The functions indicate that both entry and end point patterns are clustered. The entry points typically also have some small-scale inhibition (dip at short distances in the centred L functions). The pooled $L(r) - r$ functions (dashed lines) estimated from the entry and end point patterns using (4) are also shown.

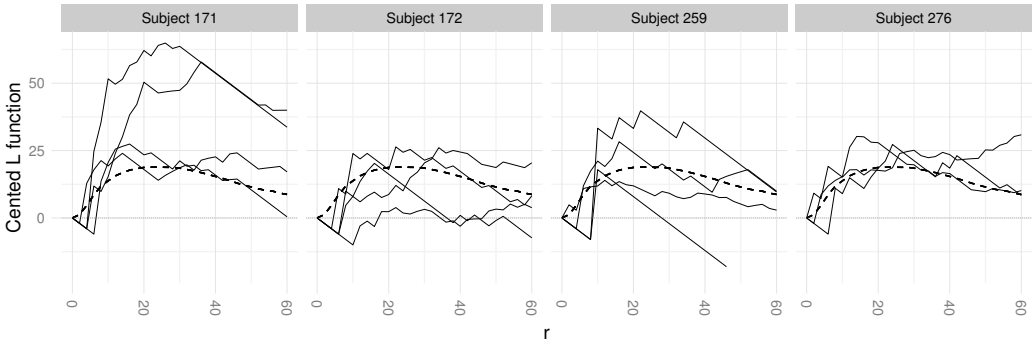

Figure 2: The estimated centred L functions (solid lines) of each ENF entry point pattern of Subjects 171, 172 (healthy), 259 and 276 (diseased), and the overall pooled function (4) estimated from all the entry point patterns of all subjects (thick dashed line). The horizontal zero line represents complete spatial randomness.

4.1. Specification of the model and priors

We employ the Gaussian process model introduced in Section 3.2. The model consists of four components: covariate effect, subject specific effect, correlated noise (sample specific effect) and uncorrelated noise. Below we specify the three first components of the model. The uncorrelated noise term was discussed in Section 3.2.

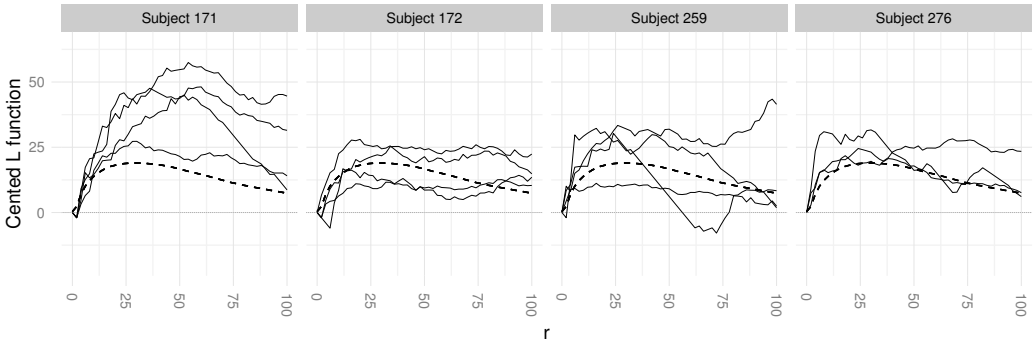

Figure 3: The estimated centred L functions (solid lines) of each ENF end point pattern of Subjects 171, 172 (healthy), 259 and 276 (diseased), and the overall pooled function (4) estimated from all the end point patterns of all subjects (thick dashed line). The horizontal zero line represents complete spatial randomness.

Disease status and gender have been coded as 0 (normal, male) and 1 (diseased, female), and all inputs are normalized to have zero mean and unit variance.

4.1.1. The first component

The first Gaussian process $f^{(1)}$ models the effect of age, gender, BMI and disease status together with the distance r. This component has the pooled $L(r) - r$ function estimated from all the samples, see (4), as its prior mean and we are interested in deviances from this average level. Note that the posterior distribution of $f^{(1)}$ is not restricted to the average level and will depend on the values of the covariates.

The effects of the covariates are incorporated in the covariance function $k_1(\boldsymbol{x}, \boldsymbol{x}'|\theta_1)$, where $\boldsymbol{x} = (r, \text{age}, \text{gender}, \text{BMI}, \text{status})^T$. We assume that $f^{(1)}$ is quite smooth and use the compactly supported piecewise polynomial function (17) ($q = 3$) with h defined in (18). Note that since $D = 5$ and $q = 3$, we have $j = 6$. We denote the variance parameter of (17) by σ_1^2 and the length-scale parameters of r and the four covariates by ϕ_{1d}, $d = 1, \ldots, 5$, respectively. Thus,

$$\theta_1 = (\phi_1, \sigma_1^2) = (\phi_{11}, \phi_{12}, \phi_{13}, \phi_{14}, \phi_{15}, \sigma_1^2).$$

This prior covariance structure is illustrated in Figure 4 (left) showing that the values of $f^{(1)}$ within a subject are correlated. There is also correlation between subjects due to similar covariate values. Within a sample the

values are most correlated for nearby r values.

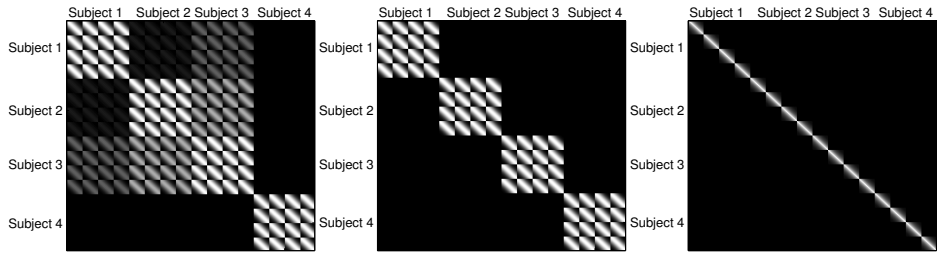

Figure 4: Illustration of the prior covariance structures of (from left) $f^{(1)}$, $f^{(2)}$, and $f^{(3)}$. Each figure shows a part of the prior covariance matrix for four subjects where the data are organized subject wise, within subject sample wise and within sample from smallest to largest value of r. Lighter color means higher correlation, and black means no correlation.

4.1.2. The second component

The Gaussian process $f^{(2)}$ models the subject specific effect and is also assumed to be quite smooth. Its prior mean is zero and its values are correlated between different values of r within a subject. Formally, the correlation structure is specified as a product of the chosen covariance function and the indicator function that equals 1 if the two observations are from the same subject and 0 otherwise. This guarantees that there is no correlation between subjects. As the covariance function we take the function (17) with the variance parameter σ_2^2 and

$$h = h(r, r') = |r - r'|/\phi_2,$$

where ϕ_2 is the length-scale parameter. Having $D = 1$ and $q = 3$ leads to $j = 4$. We denote $\theta_2 = (\phi_2, \sigma_2^2)$.

Note that the prior distribution for this component is the same for each subject, but after observing the data, a posteriori, the latent function is subject specific. Figure 4 (middle) illustrates the prior covariance structure of this component.

4.1.3. The third component

The process $f^{(3)}$ is the correlated noise component which is assumed to be less smooth than the first two components. Its prior mean is zero and its values are correlated between different values of r within a sample. Since

the variability of the L functions taken from the samples of a subject seems to differ from subject to subject, see Figures 2 and 3, we allow a different variance parameter of $f^{(3)}$ for each subject $s = 1, \ldots, N$.

The covariance between different values of r within sample j of subject s is determined by the covariance function (16) with $q = 1$, the variance parameter σ^2_{3s} and

$$h = h(r, r') = |r - r'|/\phi_3,$$

where ϕ_3 is the length-scale parameter. Since $D = 1$ and $q = 1$, j becomes 2. Formally, the correlation between different samples is set to zero by multiplying the covariance function (16) by the indicator function that equals 1 if $(s, j) = (s', j')$ and 0 otherwise. Figure 4 (right) illustrates this correlation structure.

The parameter vector for $f^{(3)}$ is $\theta_3 = (\phi_3, \sigma^2_{3s}, s = 1, \ldots, N)$. The variance parameters are controlled by assuming them to origin from a joint prior distribution.

4.1.4. Hyperpriors

Our model consists of the Gaussian observation model (8), the latent Gaussian process (6) discussed in detail above, and specification of the priors for the parameters of the observation and latent models. The hyperprior structure of the model can be summarized as

$$\begin{aligned}
\text{hyperpriors:} \quad & \sigma^2 \sim p(\sigma^2), \\
& \theta_1 = (\phi_1, \sigma_1^2) \sim p(\phi_1)p(\sigma_1^2), \\
& \theta_2 = (\phi_2, \sigma_2^2) \sim p(\phi_2)p(\sigma_2^2), \\
& \theta_3 = \{\phi_3, \sigma^2_{3s}, s = 1, \ldots, N\} \sim p(\phi_3) \prod_{s=1}^{N} p(\sigma^2_{3s}|s^2_\sigma) \\
\text{hyper-hyperprior:} \quad & s^2_\sigma \sim p(s^2_\sigma)
\end{aligned}$$

where $p(\cdot)$ denotes a distribution.

Following the recommendation by Gelman et al. [35], we use the half-Student-t distribution with four degrees of freedom for all the length-scale parameters and for the square root of the variance parameters σ_1^2 and σ_2^2. The scale parameter of the half-Student-t distribution is set to one for σ_1 and σ_2 to have close to uniform mass on the interval from 0 to 1. (Since the data are normalized for the analysis, the sum of the variances of different

components is approximately 1.) For the length-scale parameters, the scale is set to hundred to allow both small and large length scales.

For the subject-specific hyperparameters σ^2_{3s}, $s = 1, \ldots, N$, of $f^{(3)}$, we chose the scaled-inverse-χ^2 distribution with scale parameter s^2_σ and four degrees of freedom, and to s^2_σ we assign the scaled-inverse-χ^2 prior distribution with scale parameter 1 and one degree of freedom. Finally, for the error variance σ^2, we use the scaled-inverse-χ^2 prior distribution with scale 0.01 and one degree of freedom.

4.2. Inference

The marginal posterior distribution is

$$p(\theta, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) \propto p(\sigma^2)p(\phi_1)p(\sigma^2_1)p(\phi_2)p(\sigma^2_2)p(\phi_3)p(s^2_\sigma) \prod_{s=1}^{N} \left[p(\sigma^2_{3s}|s^2_\sigma)\right] p(\boldsymbol{y}|\boldsymbol{X}, \theta, \sigma^2),$$

where the marginal likelihood $p(\boldsymbol{y}|\boldsymbol{X}, \theta, \sigma^2)$ given in (9) may be written as

$$p(\boldsymbol{y}|\boldsymbol{X}, \phi_1, \sigma^2_1, \phi_2, \sigma^2_2, \phi_3, \{\sigma^2_{3s}, s = 1, \ldots, N\}, \sigma^2).$$

That is, the marginal likelihood does not depend on s^2_σ given $\{\sigma^2_{3s}, s = 1, \ldots, N\}$.

To obtain a sample from the posterior distribution of the hyperparameters, we run an MCMC simulation updating in turns the hyper-hyperparameter s^2_σ and the rest of the parameters. To update s^2_σ, we use slice sampling [36] with ten iterations at each step (saving only the last one of the ten), since this step is very fast and using many iterations leads to better mixing. The conditional posterior distribution of s^2_σ, from which we sample, is

$$p(s^2_\sigma|\theta_{\text{hyper}}, \boldsymbol{y}, \boldsymbol{X}) \quad \propto \quad p(\theta, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) \propto p(s^2_\sigma) \prod_{s=1}^{N} p(\sigma^2_{3s}|s^2_\sigma),$$

where all the distributions are scaled inverse-χ^2 distributions and θ_{hyper} denotes the hyperparameters of the model (including σ^2). Therefore,

$$\log(p(s^2_\sigma|\theta_{\text{hyper}}, \boldsymbol{y}, \boldsymbol{X})) \propto \log p_{\chi^2}(s^2_\sigma; 0.1, 4) + \sum_{s=1}^{N} \log p_{\chi^2}(\sigma^2_{3s}; s^2_\sigma, 4),$$

where $p_{\chi^2}(\cdot\,; s^2, \nu)$ is the density of the scaled inverse-χ^2 distribution with scale parameter s^2 and ν degrees of freedom. We actually sample from the distribution of $\log(s_\sigma^2)$, using the univariate slice sampling method.

The conditional distribution for the remaining parameters, θ_{hyper}, given s_σ^2 is

$$p(\theta_{\text{hyper}}|s_\sigma^2, \boldsymbol{y}, \boldsymbol{X}) \propto p(\sigma^2)p(\phi_1)p(\sigma_1^2)p(\phi_2)p(\sigma_2^2)p(\phi_3) \prod_{s=1}^{N} \left[p(\sigma_{3s}^2|s_\sigma^2)\right] p(\boldsymbol{y}|\boldsymbol{X}, \theta, \sigma^2), \tag{19}$$

which can be sampled using standard MCMC methods for the Gaussian process models [25]. In updating the hyperparameters, we use No-U-Turn Hamiltonian Monte Carlo sampling (HMC-NUTS) [37]. Our implementation utilizes sparse matrix computations and methods implemented earlier in the GPstuff toolbox [38].

4.3. Computational details and sensitivity analysis

For the analysis of the L function data, one needs to choose the values of the distance r where the summary function is estimated. The denser the spacing between the values is, the more accurately the summary function is presented. A too dense spacing, on the other hand, leads to computational burden. Therefore, we plotted the estimated summary characteristic for different choices of r-values to evaluate for which sparseness the essential features of the characteristic were still present. This led us to r-values $0, 8, 16, \ldots, 56$ with step size $d_r = 8$ between consecutive values of r for the ENF entry point data, and r-values $0, 12, 24, \ldots, 96$ with $d_r = 12$ for the end point data. The maximum r-values were chosen as the distances, where the L functions level down.

Using the end point data and $d_r = 12$, we studied the effect of our prior choices. Varying the hyperpriors had only small influence on the size of the covariate effects: If a smaller length-scale parameter was used in the Student-t prior distribution of the length-scale parameters of the model, i.e. the hyperprior was less vague, the covariate effects of age, gender and BMI became slightly larger (length-scale parameters slightly smaller). The effect of disease status stayed approximately the same.

In addition to the choices $d_r = 8$ (entry) and $d_r = 12$ (end), we performed the same analysis using the values $d_r = 4$ (entry) and $d_r = 8$ (end). The conclusions made based on the two different choises of d_r were very similar.

A disadvantage of the rather complicated model adopted to the structure of the data is that the simulation from the posterior distribution is computationally heavy, taking days to run. We run 3000 updates in the MCMC procedure for $d_r = 12$ (end), and for $d_r = 8$ (end), $d_r = 8$ (entry) and $d_r = 4$ (entry), the number of updates was 1500. We investigated the convergence of the chains visually and made convergence diagnostics by calculating potential scale reduction factors (PSRF) for all parameters, see Brooks and Gelman [39]. By throwing away the burn-in, where the step-size parameter ϵ of the HMC-NUTS algorithm was adapted [37], and thinning the chain, we finally used 500 updates in calculating the posterior distribution (10) and investigations of covariate effects.

To see if the computational load could be reduced, we fit also the simpler model where the variability of the L functions was assumed to be the same within each subject, i.e. $\sigma_{3s}^2 = \sigma_3^2$. The square root of this variance was assigned the same Student-t prior distribution as σ_1 and σ_2, 600 updates were run from the posterior distribution using HMC-NUTS and 100 updates thrown away as the burn-in. The results were obtained faster. The conclusions on the covariate effects were similar in the two models. The difference of the two models was how they assigned the variability due to individual samples to the second and third component of the model. The simpler model resulted in larger subject specific effects than the more complex model, which assigned almost all variation to the third component.

We compared the two models for the end point data by estimating the expected predictive performance for a hypothetical subject [40]. Since the full leave-one-subject-out cross-validation, which may be considered as the standard approach for assessing the predictive performance, would be computationally costly, we approximated it by the method called ghosting [41] or mixed predictive checking [42]: we used the full data posterior of the hyperparameters to calculate the (log) predictive distribution (10) of f for each of the 47 hypothetical subjects which shared the covariate values of subjects in our data. Since, the original model had higher predictive performance than the simpler model, the results in the following section are based on the original model.

5. Results for the ENF data

We constructed posterior mean predicted centred L functions for virtual subjects with different covariate values and made average predictive compar-

isons (see Section 3.2.3) to study how the covariates affect the second-order structure of ENF entry and end point patterns. Since we did not observe any effect of the covariates (not even disease status) in the entry point analysis, we only show the results for the end point data below.

5.1. Predictions for L functions for end points

Figures 5 and 6 show the mean L functions for the end point patterns together with 5% and 95% r-wise posterior quantiles for male and female, and healthy and diseased subjects of different age and BMI. All end point patterns are clustered. The cluster radius (r value corresponding to the maximum value of the centred L function) in diseased patterns is about 35-40 microns, which is slightly larger than the cluster radius 25-30 microns in healthy patterns. Furthermore, the clusters in the diseased patterns have more end points relative to the total number of points than the clusters in healthy patterns (diseased curve is above the healthy curve).

The centred L functions in Figures 5 and 6 reveal that the difference between healthy and diseased patterns is clearer for women than for men. Furthermore, the difference is more easily seen for younger subjects and subjects with high BMI than for older subjects and subjects with low BMI.

The posterior mean predicted curves $f^{(1)}$ (general level), $f^{(1)} + f^{(2)}$ (subject specific level) and $f^{(1)} + f^{(2)} + f^{(3)}$ (sample specific level) are shown in Figure 7 for four subjects. We can see that all the components together (thin black solid lines) capture well the form of the data curves (thick grey solid lines). The correlated noise-component $f^{(3)}$ seems to take most variation in itself, which is apparently associated with the large variation between the curves within subjects. After covariate effects have been taken into account, the subject specific effect is estimated to be quite small, which is in accordance with our previous studies [15].

5.2. Average predictive comparisons for end points

We next performed the average predictive comparisons in order to investigate the significance of the differences observed in the predicted L functions. Initially, we experimented with the two deviation statistics (11) and (12), but since the results for our data were very similar independently of which one was used, we decided to use the simpler measure (11).

Violinplots in Figure 8 show the results from the average predictive comparisons for each of the covariates for the ENF end point data. We observe that the diseased pattern is clearly more clustered than the healthy pattern

(the mass of the distribution of (14) lies clearly above zero). Note, however, that since the plots are based on $d_{\max}$, not on the entire L function, we can only say that the relative number of points per cluster has changed.

The effects of age, gender and BMI are not evident. To see whether these covariates have significant effects in the diseased case, we also performed average predictive comparisons only for the diseased subjects. As Figure 9 shows, the effect of BMI is slightly higher among the diseased subjects than among all the subjects (compare to Figure 8).

We further investigated how the probability of increased clustering (measured by $d_{\max}$) when changing from a healthy subject to diseased subject depends on the age, gender and BMI of the subject. We made subject specific predictive comparisons for the disease status and summarized the distribution of $\Delta_{u,s}(\theta_M^b)$ in (15) by its probability mass lying above zero, namely

$$\hat{p}_{u,s} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(\Delta_{u,s}(\theta_M^b) > 0). \tag{20}$$

These probabilities are plotted in Figure 10. There is a tendency that the probability is slightly smaller for subjects with high age and for subjects with low BMI. This confirms the observation made based on the predicted L functions that the difference between healthy and diseased patterns seems to be harder to detect in old subjects and in subjects with low BMI than in subjects with other covariate values.

6. Discussion

We suggest hierarchical Gaussian process regression as a flexible non-parametric method for studying the effect of non-spatial covariates on the second-order structure of spatial point patterns. The spatial structure is summarized by Ripley's K function which is modelled by Gaussian process regression based on replicated data. The work has been motivated by the desire to find out whether epidermal nerve fiber patterns, more precisely entry and end point patterns, are more clustered in subjects suffering from diabetic neuropathy than in healthy subjects. In addition to disease status (healthy, diabetic neuropathy), age, gender, and BMI are included in the analysis as covariates.

The neurologists believe that the nerve fiber loss due to diabetic neuropathy does not result in random removal of nerves, rather the remaining nerves

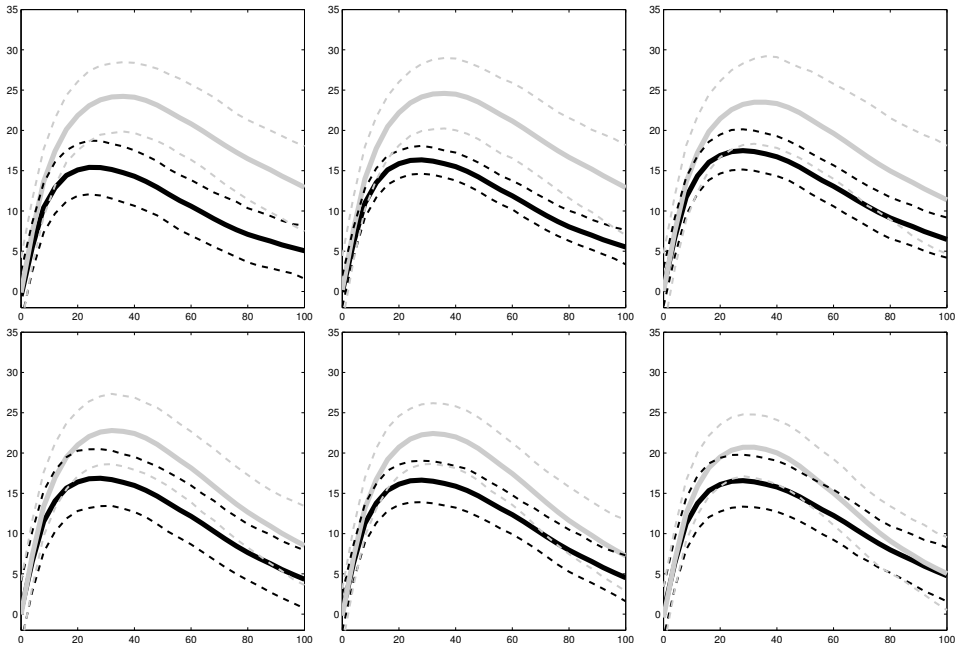

Figure 5: Mean prediction curves (solid lines) and 5% and 95% r-wise posterior quantiles (dashed lines) for the ENF end point data for comparison of disease status (black = healthy; grey = diseased) for female (first row) and male (second row) subjects. Columns from left to right: Age 30, 45, 60. BMI is fixed to 25.

seem to be arranged in more clustered patterns than the patterns before the nerve loss. This would suggest that both the entry and the end point patterns would be more clustered in diseased subjects than in healthy subjects. However, we did not see that the disease status would affect clustering of the entry points in samples taken from calves but the diseased end point patterns were clearly more clustered. More precisely, the diseased end point patterns had more points in clusters relative to the total number of points than the healthy patterns and, moreover, there is slight indication that their cluster radius would be somewhat larger than that of the healthy patterns. The latter observation may be explained by the remaining ENFs tending to grow longer in order to compensate for the loss of nerves.

Thus, according to our study, the ENF end points carry more information about clustering of ENFs, and have therefore better diagnostic capabilities than the entry points. This may be due to the number of end points being

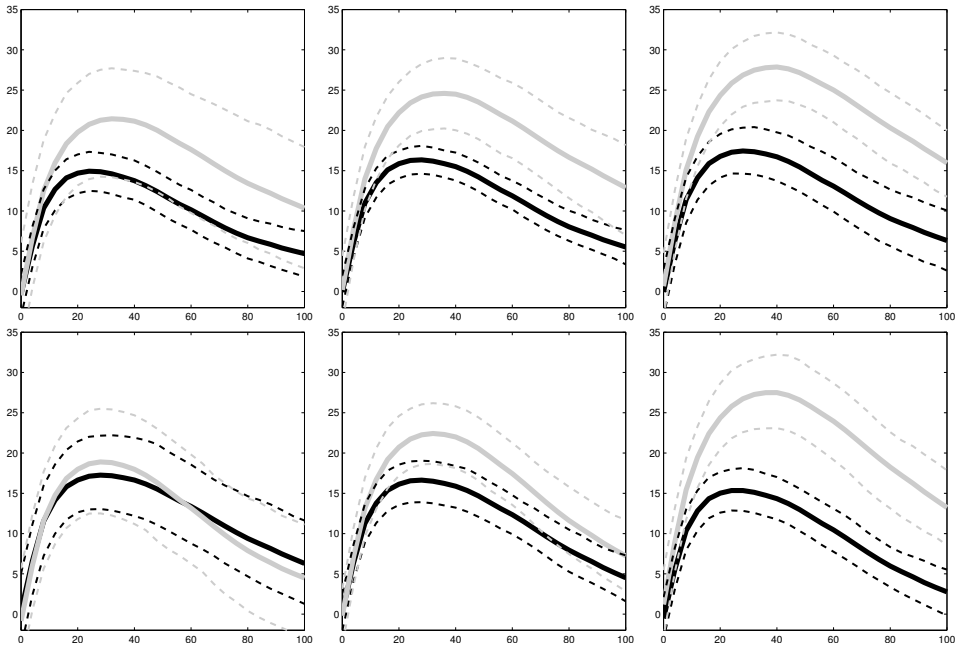

Figure 6: Mean prediction curves (solid lines) and 5% and 95% r-wise posterior quantiles (dashed lines) for the ENF end point data for comparison of disease status (black = healthy; grey = diseased) for female (first row) and male (second row) subjects. Columns from left to right: BMI 20, 25, 30. Age is fixed to 45.

larger than the number of entry points. In the thigh data studied by Waller et al. [28], where some evidence for the diseased entry point patterns being more clustered than the healthy patterns was found, the number of entry points was higher than in the calf data analysed here.

We have assumed that the entry and end point patterns are realizations of stationary point processes, and used the homogeneous K function to describe the second-order structure of the patterns. As pointed out by Waller et al. [28], the underlying physiology may affect the locations of the ENF entry points causing heterogeneities due to the so-called dermal papillae which makes the bottom of the epidermis uneven. Therefore, the inhomogeneous K function [43] could be used instead. However, as in Waller et al. [28], the data sets in this study are relatively small and do not reveal any obvious heterogeneities.

In addition to the K functions for entry points and end points, the bi-

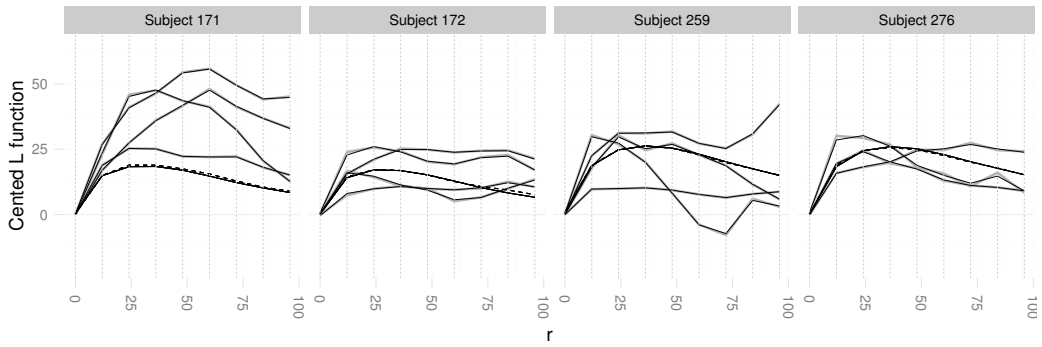

Figure 7: The estimated sample specific centred L functions (thick grey solid lines) of the ENF end points of Subjects 171, 172 (healthy), 259 and 276 (diseased). The posterior mean of the model components: general level f_1 (thick solid line), subject specific level $f_1 + f_2$ (thick dashed line) and sample specific level $f_1 + f_2 + f_3$ (thin black solid lines). Vertical dashed lines show the distances used for inference.

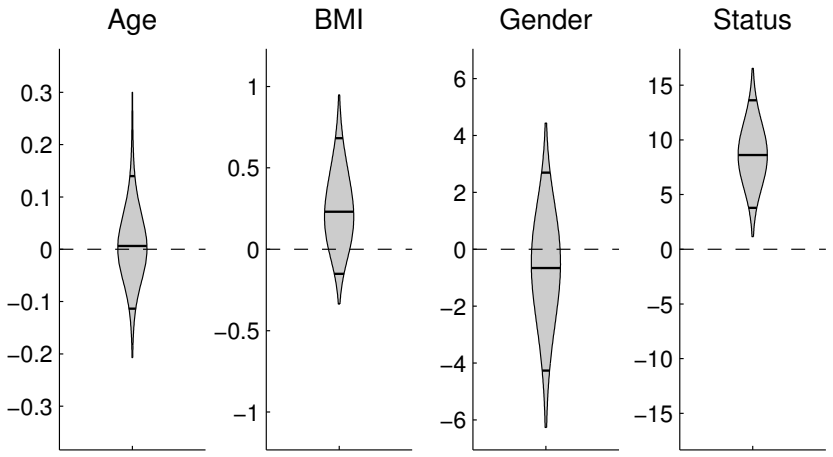

Figure 8: Average predictive comparisons. Violinplots characterize the distributions of (14) for the covariates for ENF end point data. The distributions for gender and disease status are for the changes from male to female and from healthy to diseased. The horizontal lines in the "violins" give the estimated 5%, 50% and 95% quantiles of the distributions.

variate K function (entry points-end points) could be estimated and modeled by the same hierarchical model (a priori) as the univariate functions. If the complete fiber pattern is of interest, a similar hierarchical model could also

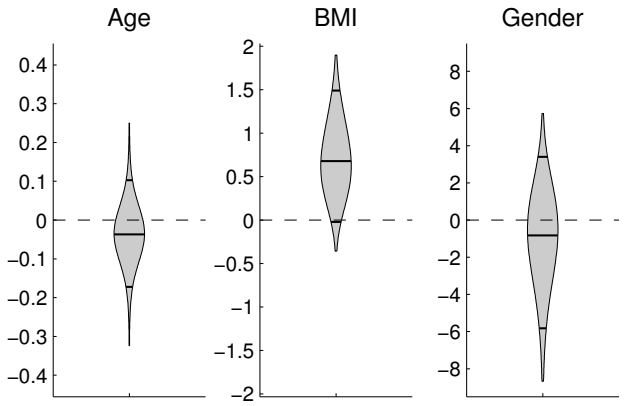

Figure 9: Average predictive comparisons. Violinplots characterize the distributions of (14) for the covariates for diseased subjects for the ENF end point data. The distribution for gender is for the change from male to female. The horizontal lines in the "violins" give the estimated 5%, 50% and 95% quantiles of the distributions.

be applied to the fiber K function.

We modelled the L functions using the Gaussian process regression. Gaussian Markov random fields (GMRFs), which have computational advantages over the Gaussian processes if the dimensions of inputs are low and there are no high level interactions, have been used in similar regression studies [22]. However, our model included interactions between all the covariates, which was accomplished easily in the Gaussian process approach. If the joint effects of the covariates were excluded from the model a priori, then a GMRF approach would be a competitive alternative for fast inference. Yue and Loh [44] used such a latent non-parametric GMRF model in estimation of the pair correlation function (derivative of the K function). Note that, since our observation model is Gaussian, we can calculate the marginal likelihood of the hyperparameters exactly and, thus, there is no need for analytical approximations, neither in the Gaussian process nor in the GMRF approach.

The hierarchical modelling approach we propose was motivated by the ENF data, but because of the flexibility and non-parametric nature of the Gaussian process model, basically the same model (a priori) can be used in other applications. For example in social networks, the question of interest may be whether closeness and relationship (relative or friend) have effect on clustering (in time) of communication between people.

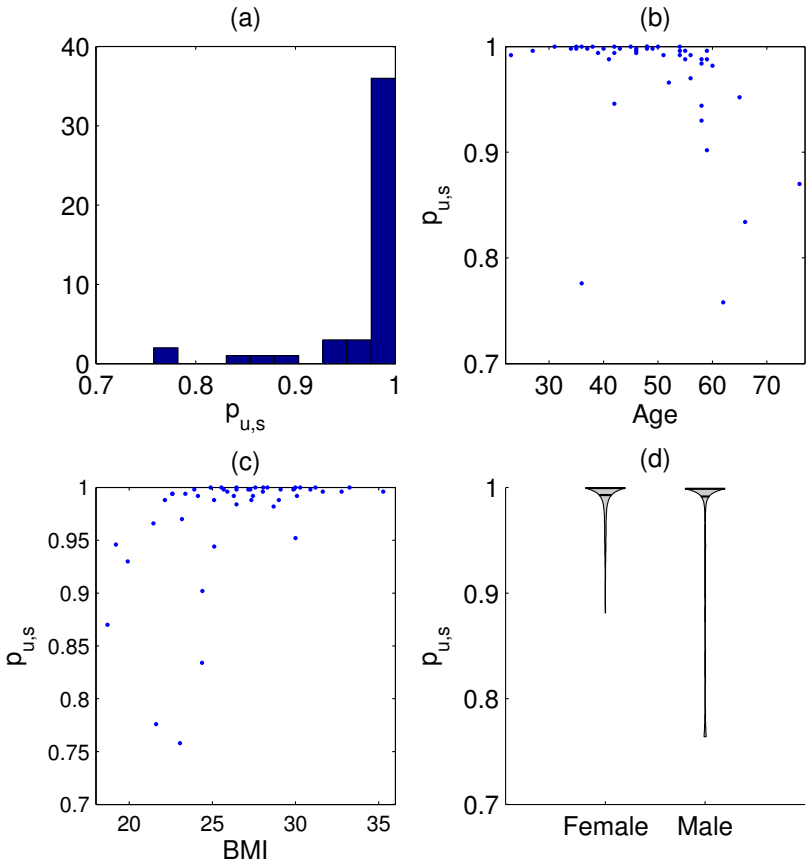

Figure 10: Subject specific predictive comparisons for the ENF end point data for disease status. (a) Histrogram of subject specific probabilities in (20) for positive change in the outcome $d_{\max}$ with a change from healthy to diseased. (b) The probabilities plotted against age. (c) The probabilities plotted against BMI. (d) Violinplots of the probabilities for female and male subjects.

In this paper, the second-order structure of spatial point patterns based on replicated data with non-spatial covariate information was modeled. The second-order analysis only looks at a specific feature of the data, which was in our case motivated by the neurologists' hypothesis. The point patterns may, however, have other interesting features and therefore, next steps will include developing suitable point process models based on replicated data where also non-spatial covariates are available. Point process models presented in Olsbo

et al. [31] provide a basis for developing such models for ENF patterns.

Acknowledgements

The authors thank William R. Kennedy, Gwen Wendelschafer-Crabb and Ioanna G. Panoutsopoulou (University of Minnesota) for blister immunostaining, quantification and morphometry of the ENF data, Sakari Cajanus (Aalto University) for data processing, Ville Tolvanen (Aalto University) for discussions on implementation of the HMC-NUTS algorithm to our data and two anonymous reviewers for their comments which led to improvement of the paper. The research has been funded by the Academy of Finland (project numbers 250860 and 218248), and by the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, the Royal Swedish Academy of Sciences, the Royal Society of Arts and Sciences in Gothenburg, and the Wilhelm and Martina Lundgren's Foundation. Furthermore, Aila Särkkä would like to thank the Department of Statistics at the University of Washington for its hospitality during the time this paper was completed.

References

[1] N. A. C. Cressie, Statistics for Spatial Data, revised ed., Wiley, New York, 1993.

[2] P. J. Diggle, Statistical Analysis of Spatial Point Patterns, 2 ed., Arnold, London, 2003.

[3] J. Illian, A. Penttinen, H. Stoyan, D. Stoyan, Statistical Analysis and Modelling of Spatial Point Patterns, Wiley, Chichester, 2008.

[4] P. J. Diggle, N. Lange, F. M. Beneš, Analysis of variance for replicated spatial point patterns in clinical neuroanatomy, Journal of the American Statistical Association 86 (1991) 618–625.

[5] A. J. Baddeley, R. A. Moyeed, A. Boyde, Analysis of a three-dimensional point pattern with replication, Applied Statistics 42 (1993) 641–66.

[6] K. Schladitz, A. Särkkä, I. Pavenstädt, O. Haferkamp, T. Mattfeldt, Statistical analysis of intramembranous particles using freeze fracture specimens, Journal of Microscopy 211 (2003) 137–153.

[7] C. Redenbach, A. Särkkä, J. Freitag, K. Schladitz, Anisotropy analysis of pressed point processes, Advances in Statistical Analysis 93 (2009) 237–261.

[8] U. Hahn, A studentized permutation test for the comparison of spatial point patterns, Journal of the American Statistical Association 107 (2012) 754–764.

[9] P. J. Diggle, J. Mateu, H. E. Clough, A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns, Advances in Applied Probability 32 (2000) 331–343.

[10] S. Eckel, F. Fleischer, P. Grabarnik, M. Kazda, A. Särkkä, V. Schmidt, Modelling tree roots in mixed forest stands by inhomogeneous marked Gibbs point processes, Biometrical Journal 51 (2009) 522–539.

[11] M. L. Bell, G. K. Grunwald, Mixed models for the analysis of replicated spatial point patterns, Biostatistics 5 (2004) 633–648.

[12] J. Illian, D. K. Hendrichsen, Gibbs point process models with mixed effects, Environmetrics 21 (2010) 341–353.

[13] J. B. Illian, S. H. Sørbye, H. Rue, D. K. Hendrichsen, Using INLA to fit a complex point process model with temporally varying effects – a case study, Journal of Environmental Statistics 3 (2012) 1–29.

[14] B. D. Ripley, The second-order analysis of stationary point processes, Journal of Applied Probability 13 (1976) 255–266.

[15] M. Myllymäki, A. Särkkä, I. G. Panoutsopoulou, Analysis of spatial structure of epidermal nerve entry point patterns based on replicated data, Journal of Microscopy 247 (2012) 228–239.

[16] G. Matheron, The intrinsic random functions and their applications, Advances in Applied Probability 5 (1973) 439–468.

[17] J.-P. Chilés, P. Delfiner, Geostatistics: Modeling Spatial Uncertainty, Wiley, New York, 1999.

[18] P. J. Diggle, P. J. Ribeiro, Model-based Geostatistics, Springer, New York, 2007.

[19] A. E. Gelfand, P. J. Diggle, M. Fuentes, P. Guttorp (Eds.), Handbook of Spatial Statistics, 1 ed., CRC Press, Boca Raton, 2010.

[20] S. Banerjee, B. P. Carlin, A. E. Gelfand, Hierarchical Modeling and Analysis for Spatial Data, 1 ed., Chapman & Hall/CRC, Boca Raton, 2004.

[21] J. Møller, A. R. Syversveen, R. P. Waagepetersen, Log Gaussian Cox processes, Scandinavian Journal of Statistics 25 (1998) 451–482.

[22] H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion), Journal of the Royal Statistical Society, Series B 71 (2009) 319–392.

[23] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, 2 ed., MIT Press, Cambridge, 2006.

[24] A. O'Hagan, Curve fitting and optimal design for prediction, Journal of the Royal Statistical Society, Series B 40 (1978) 1–42.

[25] R. M. Neal, Regression and classification using Gaussian process priors (with discussion), in: J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), Bayesian Statistics, 1 ed., Oxford University Press, 1999, pp. 475–501.

[26] W. R. Kennedy, G. Wendelschafer-Crabb, T. Johnson, Quantitation of epidermal nerves in diabetic neuropathy, Neurology 47 (1996) 1042–1448.

[27] W. R. Kennedy, M. Nolano, G. Wendelschafer-Crabb, T. L. Johnson, E. Tamura, A skin blister method to study epidermal nerves in peripheral nerve disease, Muscle and Nerve 22 (1999) 360–371.

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

[29] G. Wendelschafer-Crabb, W. R. Kennedy, D. Walk, Morphological features of nerves in skin biopsies, Journal of the Neurological Sciences 242 (2005) 15–21.

[30] I. Panoutsopoulou, G. Wendelschafer-Crabb, J. S. Hodges, W. R. Kennedy, Skin blister and skin biopsy for quantifying epidermal nerve fibers: a comparative study, Neurology 72 (2009) 1205–1210.

[31] V. Olsbo, M. Myllymäki, L. A. Waller, A. Särkkä, Development and evaluation of spatial point process models for epidermal nerve fibers, Mathematical Biosciences 243 (2013) 178–189.

[32] B. D. Ripley, Modelling spatial patterns, Journal of the Royal Statistical Society, Series B 39 (1977) 172–212.

[33] J. E. Besag, Comment on 'Modelling spatial patterns' by B. D. Ripley, Journal of the Royal Statistical Society. Series B (Methodological) 39 (1977) 193–195.

[34] A. Gelman, I. Pardoe, Average predictive comparisons for models with nonlinearity, interactions, and variance components, Sociological Methodology 37 (2007) 23–51.

[35] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian Data Analysis, 2 ed., Chapman & Hall/CRC, Boca Raton, 2004.

[36] R. M. Neal, Slice sampling, The Annals of Statistics 31 (2003) 705–767.

[37] M. D. Hoffman, A. Gelman, The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, arXiv:1111.4246 [stat.CO] (2011).

[38] J. Vanhatalo, J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, A. Vehtari, Gpstuff: Bayesian modeling with Gaussian processes, Journal of Machine Learning Research 14 (2013) 1175–1179.

[39] S. P. Brooks, A. Gelman, General methods for monitoring convergence of iterative simulation, Journal of Computational and Graphical Statistics 7 (1998) 434–455.

[40] A. Vehtari, J. Ojanen, A survey of Bayesian predictive methods for model assessment, selection and comparison, Statistics Surveys 6 (2012) 142–228.

[41] E. C. Marshall, D. J. Spiegelhalter, Approximate cross-validatory predictive checks in disease mapping models, Statistics in Medicine 22 (2003) 1649–1660.

[42] E. C. Marshall, D. J. Spiegelhalter, Identifying outliers in Bayesian hierarchical models: a simulation-based approach, Bayesian Analysis 2 (2007) 409–444.

[43] A. J. Baddeley, J. Møller, R. Waagepetersen, Non- and semi-parametric estimation of interaction in inhomogeneous point patterns, Statistica Neerlandica 54 (2000) 329–350.

[44] Y. R. Yue, J. M. Loh, Bayesian nonparametric estimation of pair correlation function for inhomogeneous spatial point processes, Journal of Nonparametric Statistics 25 (2013) 463–474.