Diagnostic value of epidermal nerve fibre patterns

Aila Särkkä

Mathematical sciences
Chalmers University of Technology and the University of Gothenburg
Gothenburg, Sweden

March 17, 2015

Collaborators

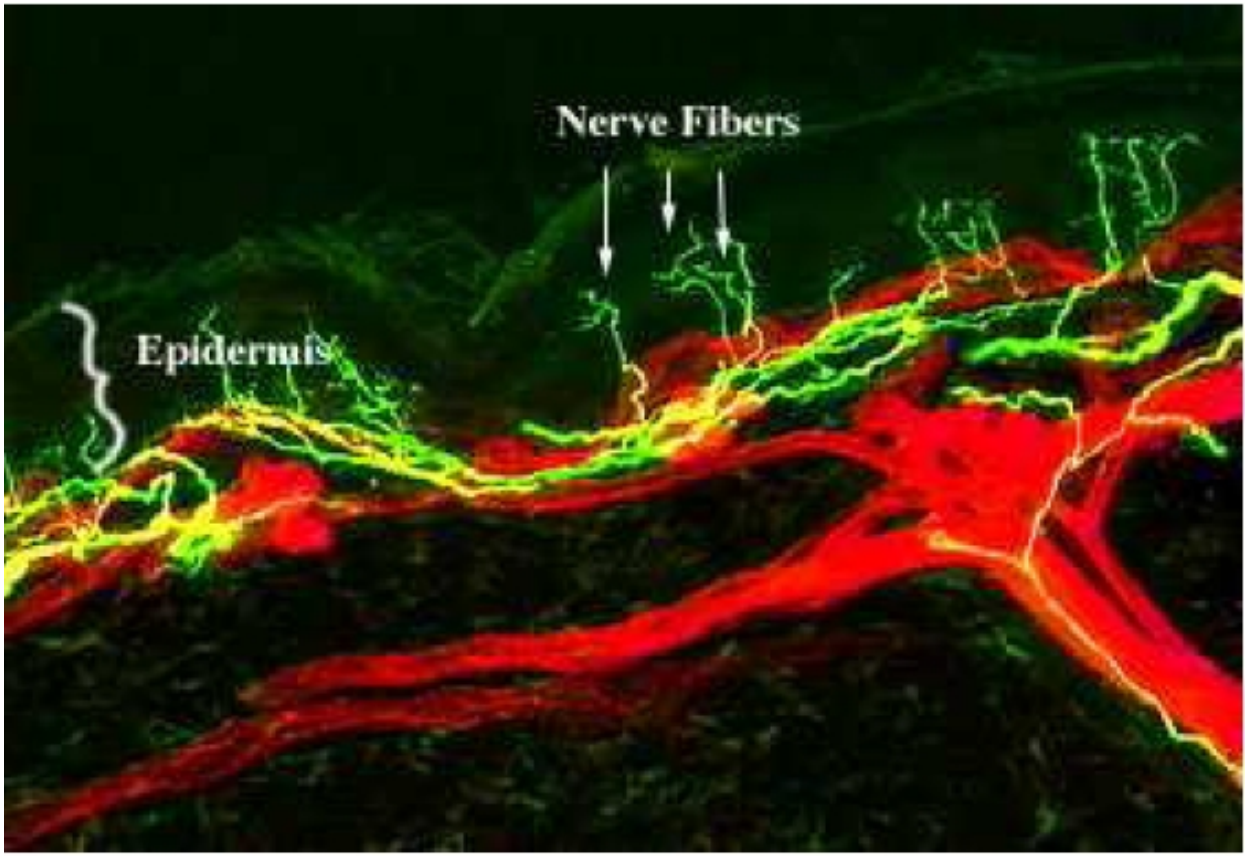
- Mari Myllymäki and Aki Vehtari, Aalto University, Finland
- William R. Kennedy, Ioanna G. Panoutsopoulou, and Gwen Wendelschafer-Crabb, University of Minnesota

Outline

- Motivation: what are epidermal nerve fibers (ENFs)?
- Question: how is the ENF pattern affected by some covariates?
- Spatial analysis by using the second-order summary statistic Ripley's K function
  $\rightarrow$ How to include covariates?
  $\rightarrow$ Hierarchical Gaussian process regression
- Results for ENF data
- Current/future work

Epidermal nerve fibers

ENFs are thin nerve fibers in the epidermis (the outmost living layer of the skin)



Diagnostic value of ENFs

Kennedy *et al.* (1996, 1999): Subjects with diabetic neuropathy have

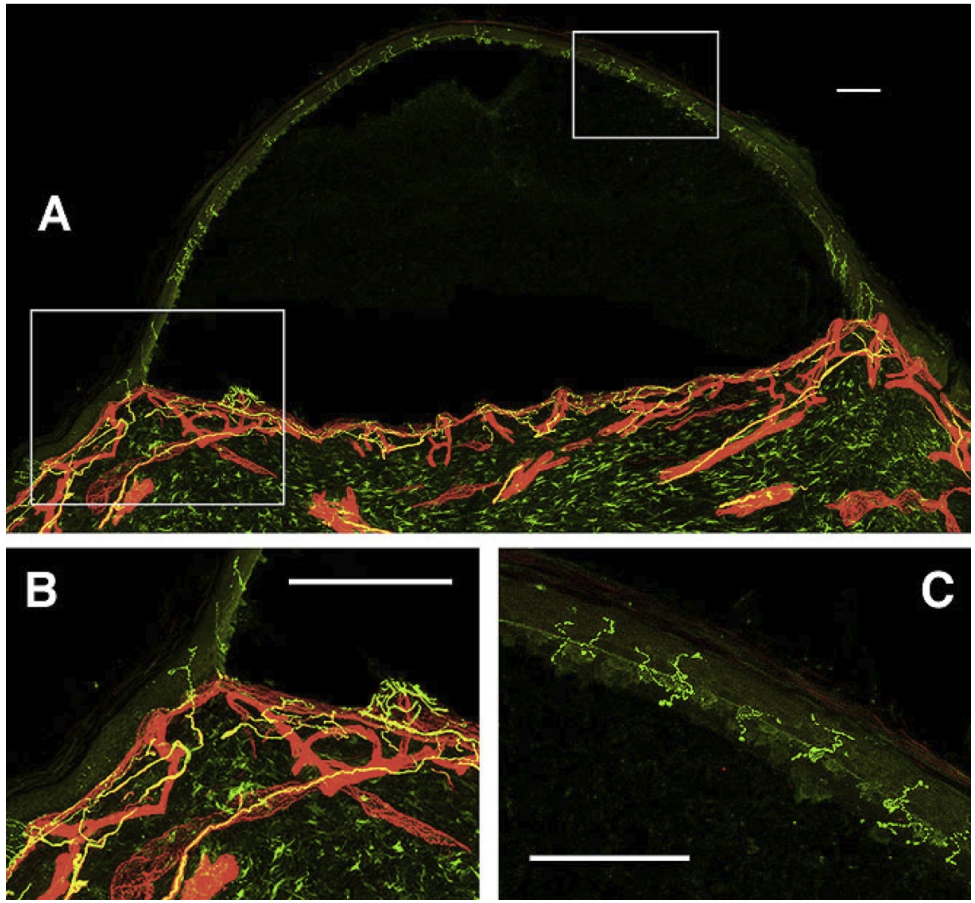
- less ENFs per surface area
- shorter summed length of ENFs per volume
- more clustered ENF pattern

than healthy subjects.

Data

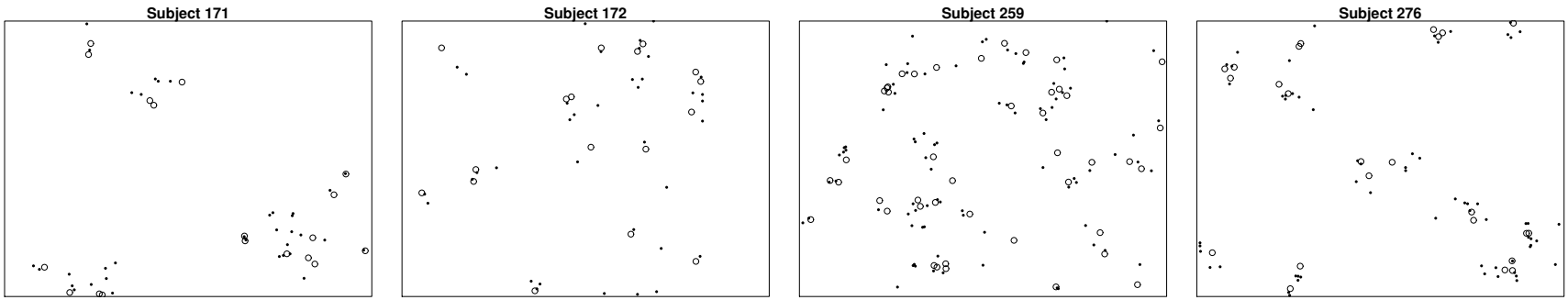
- 32 healthy subjects and 15 subjects with (mild or moderate) diabetic neuropathy
- Two skin blisters (3-6 samples) from calf of each subject (replicates)
- Age, gender and body mass index (BMI) of each subject available

Skin blister



Spatial pattern of ENF entry (base) and end points

We concentrate on base points, which are the locations, where the nerves enter the epidermis, and end points, which are the locations of the termination points of ENFs.



Locations of base points (open circles) and end points (small black dots) for two healthy and two diseased subjects

Remark

The data are fibre patterns in 3D (with z direction much smaller than x and y directions) but we have looked only at the spatial pattern of base points and end points in 2D

- end points of ENFs sense heat and pain, and play, therefore, a more important role than the ENFs themselves (fibers can be omitted)
- our focus is on the spatial pattern of ENF coverage across the skin (2D projection appropriate)
- point patterns of ENF base and end points regarded as realizations of stationary spatial point processes

Our data are point patterns with replicates and non-spatial covariates

Question

Observation: Subjects with diabetic neuropathy tend to have more clustered ENF patterns than healthy subjects.

Question: Is the (second-order) spatial structure of ENF base and end points affected by the disease status and other covariates (age, gender and BMI)?

Second-order summary statistic

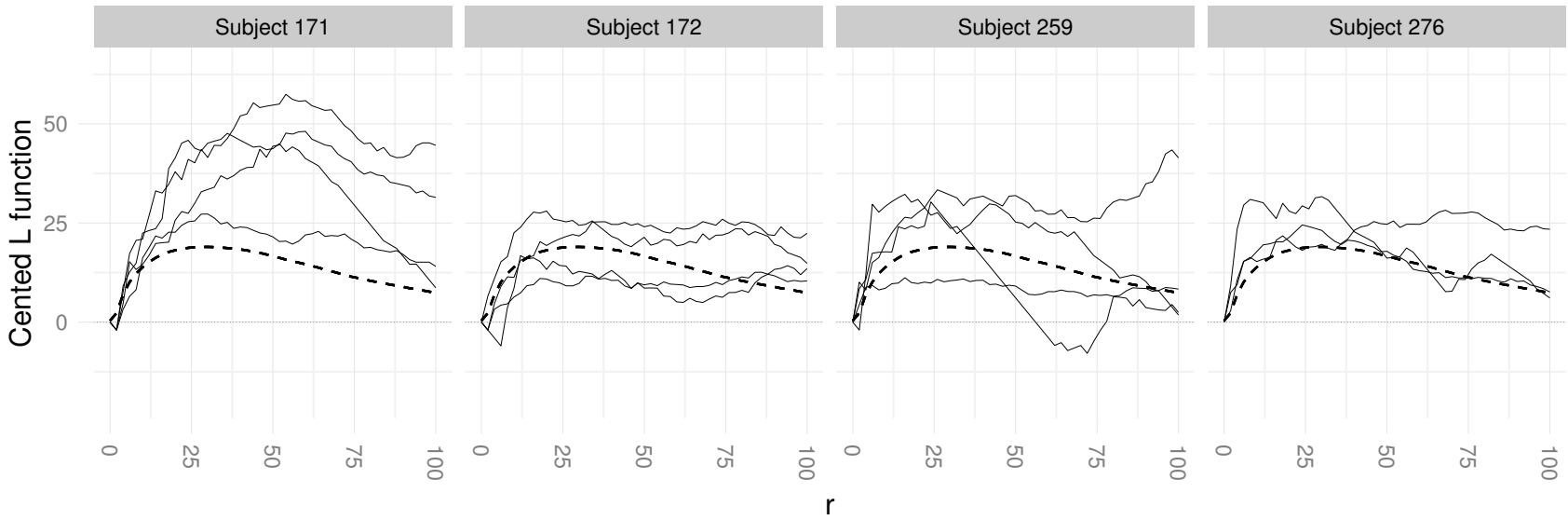
Ripley's K function: $\lambda K(r)$ is the expected number of further points within distance r from an arbitrary point of the process, where λ is the intensity (mean number of points per unit area) of the process (Ripley, 1977).

We use a variance stabilizing and centered version of the K function (Besag, 1977), namely

$$L(r) - r = \sqrt{K(r)/\pi} - r,$$

which equals 0 under complete spatial randomness. Values less than zero indicate regularity and values larger than zero clustering.

Individual $L(r) - r$ functions for end points



Subject 171 and Subject 172 are healthy, the other two diseased

How to include non-spatial covariates?

- Pooled summary statistics for groups (see e.g. Diggle *et al.* 1991; Baddeley *et al.* 1993; Hahn, 2012)
- Summary statistic modeled by using linear mixed models (Myllymäki *et al.* 2012)
- Summary statistic modeled by using hierarchical Gaussian process regression

Hierarchical Gaussian process regression model for centered L function

- Flexible non-parametric models for making inference about the relationship between some characteristics (centered L function) and covariates
- We do not need to assume linear or any other particular form of dependence between the characteristics and covariates, a priori
- Bayesian approach

Hierarchical model

Our model is

$$y_{sjk} = L_{sj}(r_k) - r_k = f_1(\mathbf{x}_s, r_k) + f_2(s, r_k) + f_3(s, j, r_k) + \epsilon_{sjk},$$

where

- f_1 models the effect of age, gender, BMI and disease status (collected in $\mathbf{x}_s$) together with distance r
- f_2 models the subject-specific effect
- f_3 models the sample-specific effect
- latent function $f = f_1 + f_2 + f_3$
- ϵ_{sjk}'s are independent and $\sim N(0, \sigma^2)$

Hierarchical model: f_1 (covariates) and f_2 (subject-specific effect)

f_1 is a Gaussian process (GP) with

- mean $\overline{L(r_k) - r_k}$
- covariance function having an own length scale parameter for each covariate (age, gender, BMI, disease status) and for r
- values of f_1 are correlated within a subject and also between subjects due to similar covariate values.

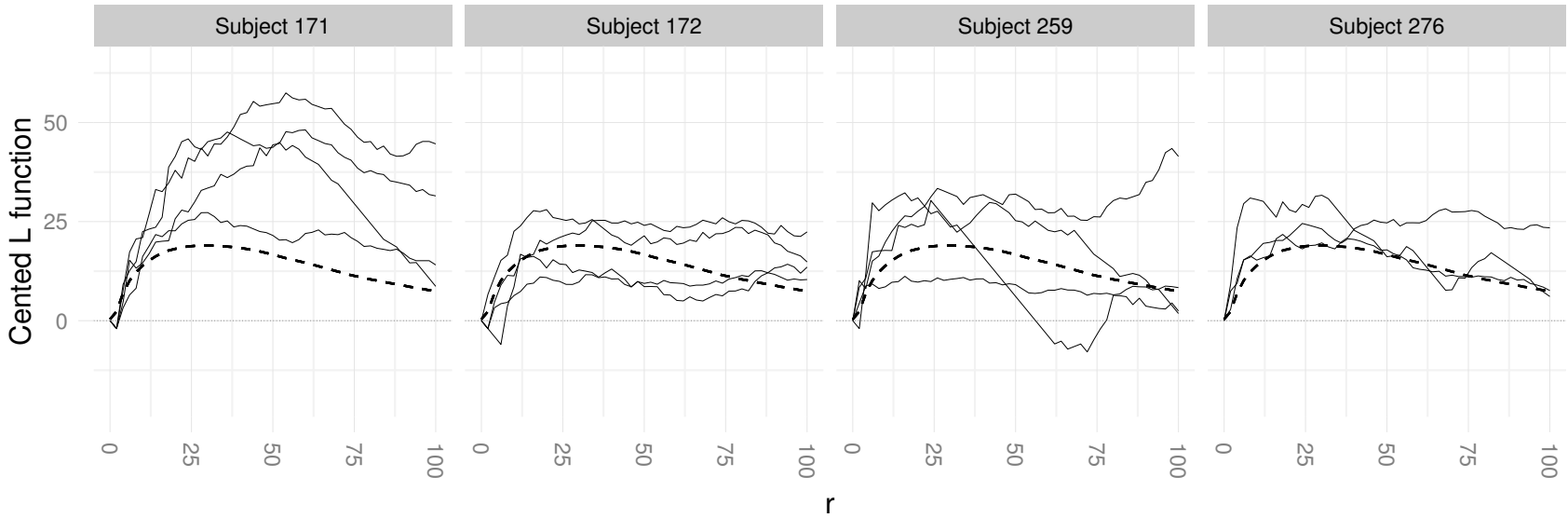
f_2 is a GP with

- mean zero
- covariance function, which is a priori the same for each subject

Hierarchical model: f_3 (sample-specific effect)

f_3 is a GP with

- mean zero
- covariance function, where the variance parameter σ^2_{3s} is allowed to vary from subject to subject
- values of f_3 are correlated only within a sample



Hierarchical model

- Observation model $\mathbf{y}|f, \sigma^2 \sim \prod_i N(y_i|f, \sigma^2)$
- GP prior

$$f(\mathbf{x})|\theta \sim GP(m(\mathbf{x}), k_1(\mathbf{x}, \mathbf{x}'|\theta_1) + k_2(\mathbf{x}, \mathbf{x}'|\theta_2) + k_3(\mathbf{x}, \mathbf{x}'|\theta_3))$$

- Hyperpriors

$$\begin{aligned}
\sigma^2 &\sim p(\sigma^2) \\
\theta_1 = (\phi_1, \sigma_1^2) &\sim p(\phi_1)p(\sigma_1^2) \\
\theta_2 = (\phi_2, \sigma_2^2) &\sim p(\phi_2)p(\sigma_2^2) \\
\theta_3 = \{\phi_3, \sigma_{3s}^2, s = 1, \ldots, N\} &\sim p(\phi_3) \prod_{s=1}^{N} p(\sigma_{3s}^2|s_\sigma^2)
\end{aligned}$$

- Hyper-hyperprior $s_\sigma^2 \sim p(s_\sigma^2)$

Choices

- Spacing of r values ($r = 0, 12, 24, ..., 96$ for end points)
- Piecewice polynomial compactly supported covariance functions (less smooth for f_3 than for the first two components)
- Half-Student t and scaled inverse χ^2 priors for hyperparameters and for s_σ^2

Bayesian inference and posterior predictive L functions

Since f and the likelihood are Gaussian, we can integrate out the latent function and obtain the *log marginal likelihood*

$$\log p(\mathbf{y}|\mathbf{X}, \theta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|K + \sigma^2 I| - \frac{1}{2}\mathbf{y}^T(K + \sigma^2 I)^{-1}\mathbf{y},$$

where $\theta = (\theta_1, \theta_2, \theta_3)$ collects all the parameters of f and K is the covariance matrix.

The posterior distribution of the latent function f_1

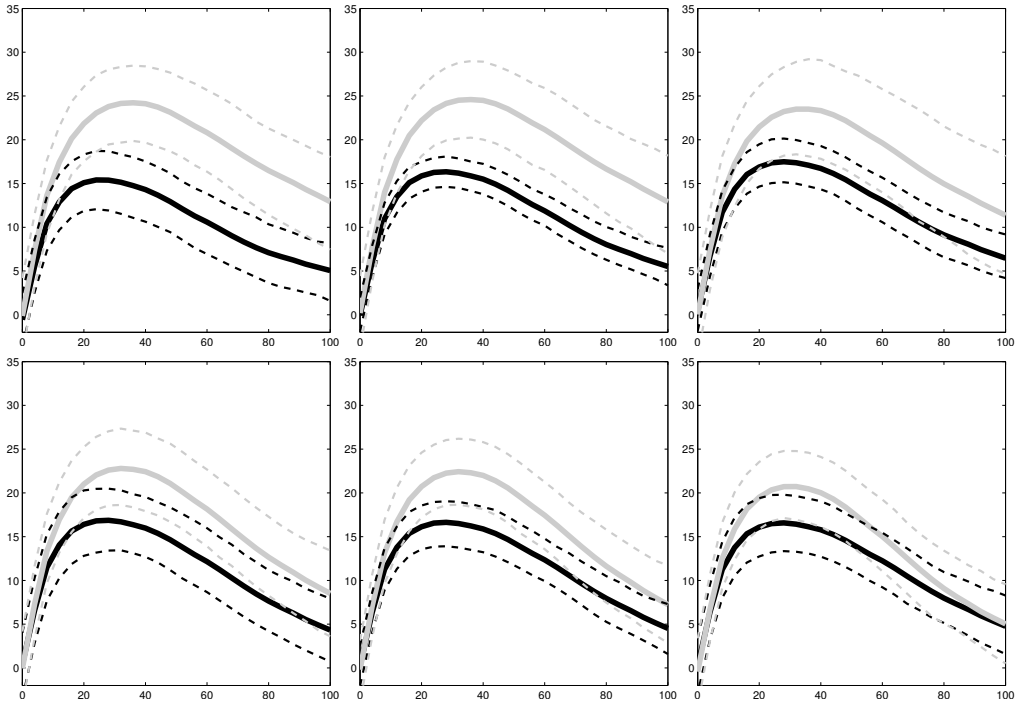
$$p(f_1|\mathbf{y}, \mathbf{X}) = \int p(f_1|\mathbf{y}, \mathbf{X}, \theta, \sigma^2)p(\theta, \sigma^2|\mathbf{y}, \mathbf{X})\mathrm{d}\theta\mathrm{d}\sigma^2$$

can be obtained by Monte Carlo integration over the hyperparameters

Computations

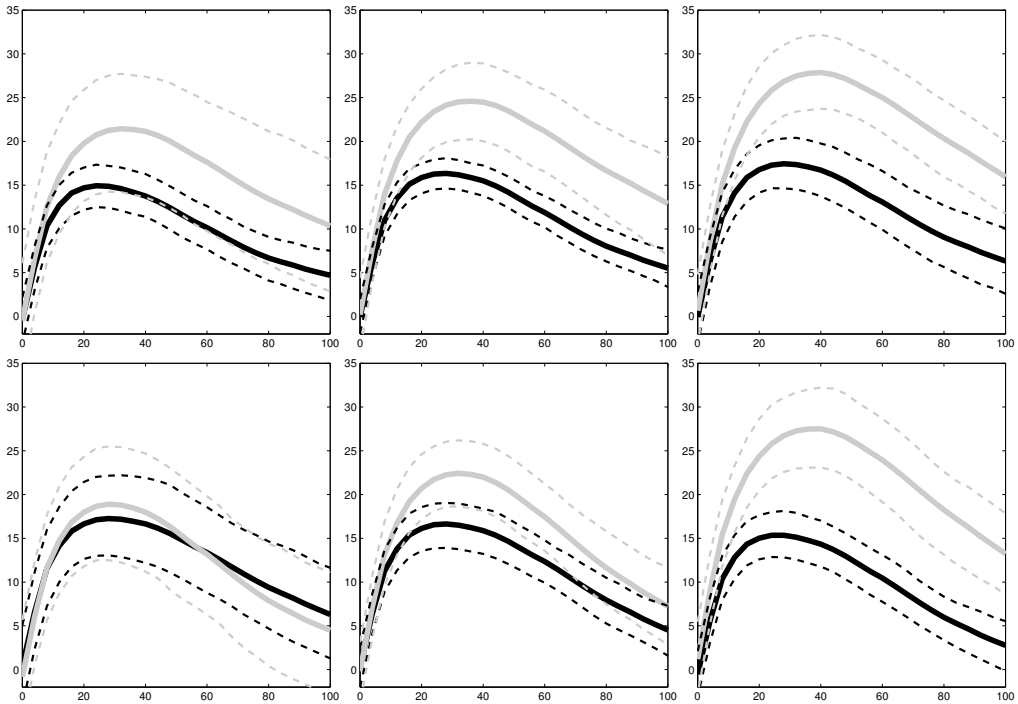
- To obtain the posterior distribution of the parameters, we run an MCMC simulation updating in turns the hyper-hyperparameter and the hyperparameters
- For sampling the Matlab toolbox GPstuff (Vanhatalo et al., 2013) is used

Mean prediction centred L curves (mean of the posterior predictive distribution of f_1) for end points



Female (first row), male (second row)
From left to right: Age 30, 45, 60; BMI is fixed to 25
Healthy (black), diseased (grey)

Mean prediction centred L curves for end points



Female (first row), male (second row)
From left to right: BMI 20, 25, 30; Age is fixed to 45
Healthy (black), diseased (grey)

Results

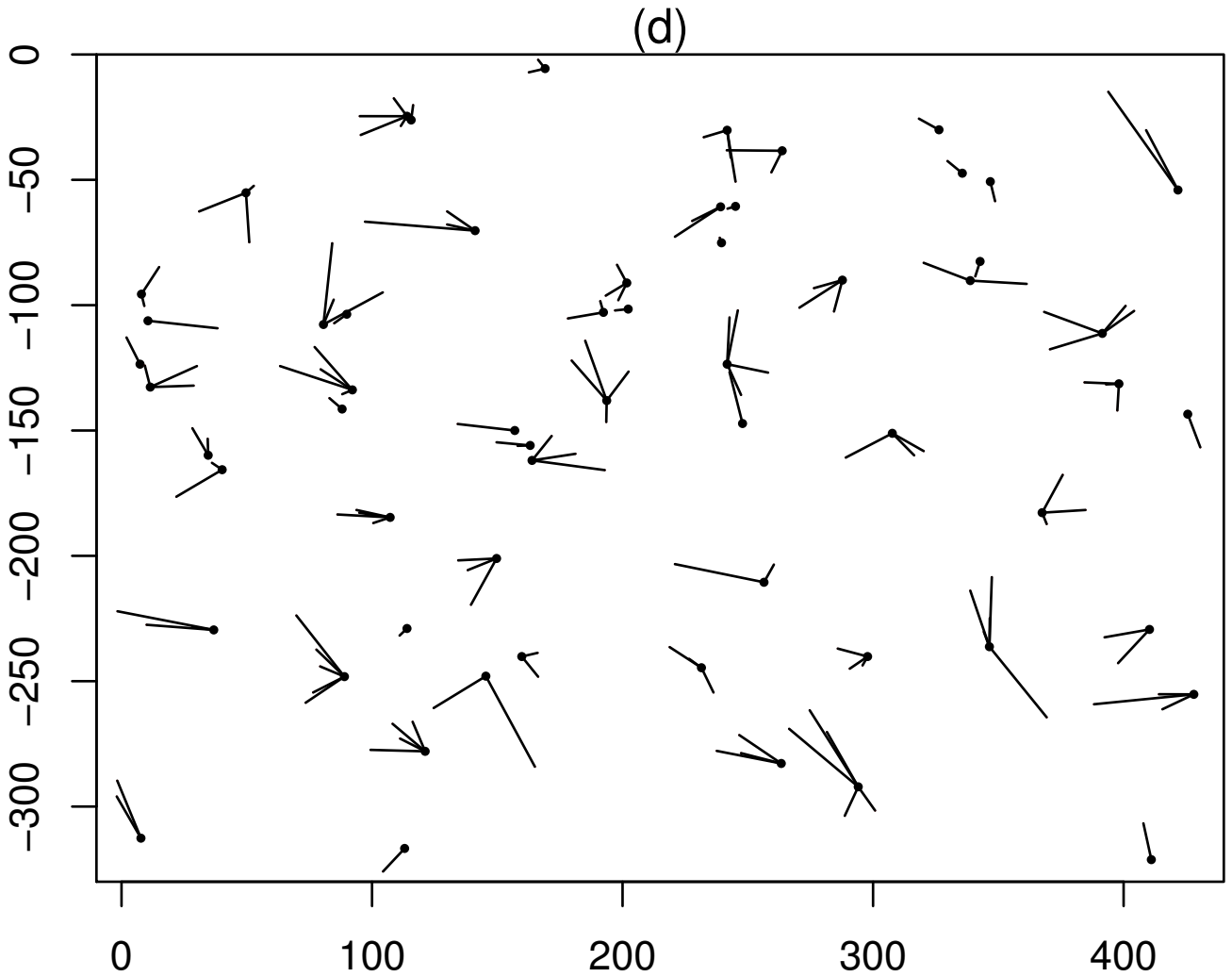
Base points: covariates (including disease status) do not seem to have any effect on the ENF pattern

End points

- diseased patterns clearly more clustered than healthy ones
- difference between healthy and diseased patterns is clearer for women than for men
- difference between healthy and diseased patterns is more easily seen for younger subjects and subjects with high BMI than for older subjects and subjects with low BMI
- effects of age, gender and BMI not evident

Current/future work

- 2D and 3D spatial point (and fibre) process models for ENFs
- Spatio-temporal models for ENF growth
- How to use replicates and include non-spatial covariates in the models?


(d)
0
−50
−100
−150
−200
−250
−300
0
100
200
300
400

Thank you!

References

- A. J. Baddeley, R. A. Moyeed, A. Boyde, Analysis of a three-dimensional point pattern with replication, Applied Statistics 42 (1993) 641-666.
- 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.
- 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.
- A. Gelman, I. Pardoe, Average predictive comparisons for models with nonlinearity, interactions, and variance components, Sociological Methodology 37 (2007) 23-51.
- U. Hahn, A studentized permutation test for the comparison of spatial point patterns, Journal of the American Statistical Association 107 (2012) 754-764.

References (continues)

- W. R. Kennedy, G. Wendelschafer-Crabb, T. Johnson, Quantitation of epidermal nerves in diabetic neuropathy, Neurology 47 (1996) 1042-1448.
- 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.
- 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.
- B. D. Ripley, Modelling spatial patterns, Journal of the Royal Statistical Society, Series B 39 (1977) 172-212.
- 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.