Spatial point process models and empirical analysis of epidermal nerve fibers

Lance A. Waller
Department of Biostatistics and Bioinformatics
Rollins School of Public Health, Emory University

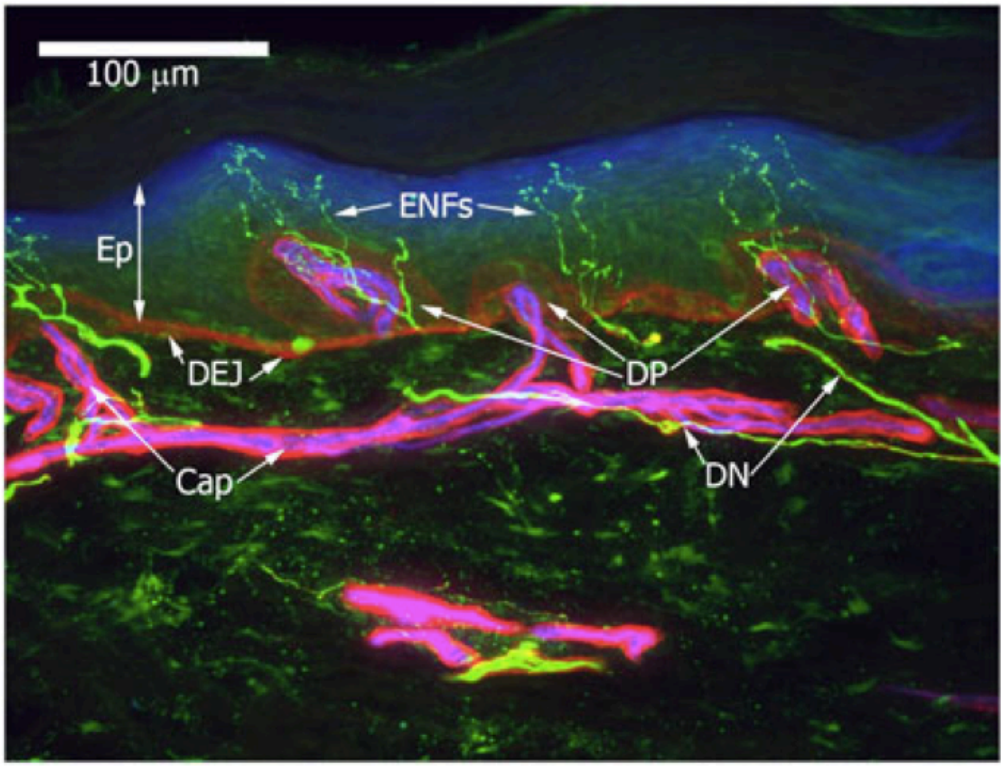
Viktor Olsbo, Aila Särkkä (Chalmers), Mari Myllymäki (Aalto), Ioanna Panoutsopoulou, William R. Kennedy, and Gwen Wendelschafer-Crabb (Minnesota)

A collection of collaborative work supported in part by NINDS R21 NS046258, Swedish Research Council, and the Swedish Foundation for Strategic Research.

What is an ENF?

- Thin, unmyelinated sensory nerve fibers based in dorsal root ganglia cells.
- Branching and ending in the epidermis (think of trees).
- Imaged via confocal microscopy since the 1990s (Kennedy and Wendelschaffer-Crabb 1993)
- Sensor network for heat, cold, pain.
- Goal: full coverage, compensate for loss.
- Innervation and denervation are dynamic processes.
- Pathology: Neuropathy

ENFs (side view)



What can we measure?

- Number of base points per unity area.
- Number of branches per base point.
- Length of branches (individual and summed).
- Direction of branches.
- All of these matter in understanding the nervous system.

What do we want to do?

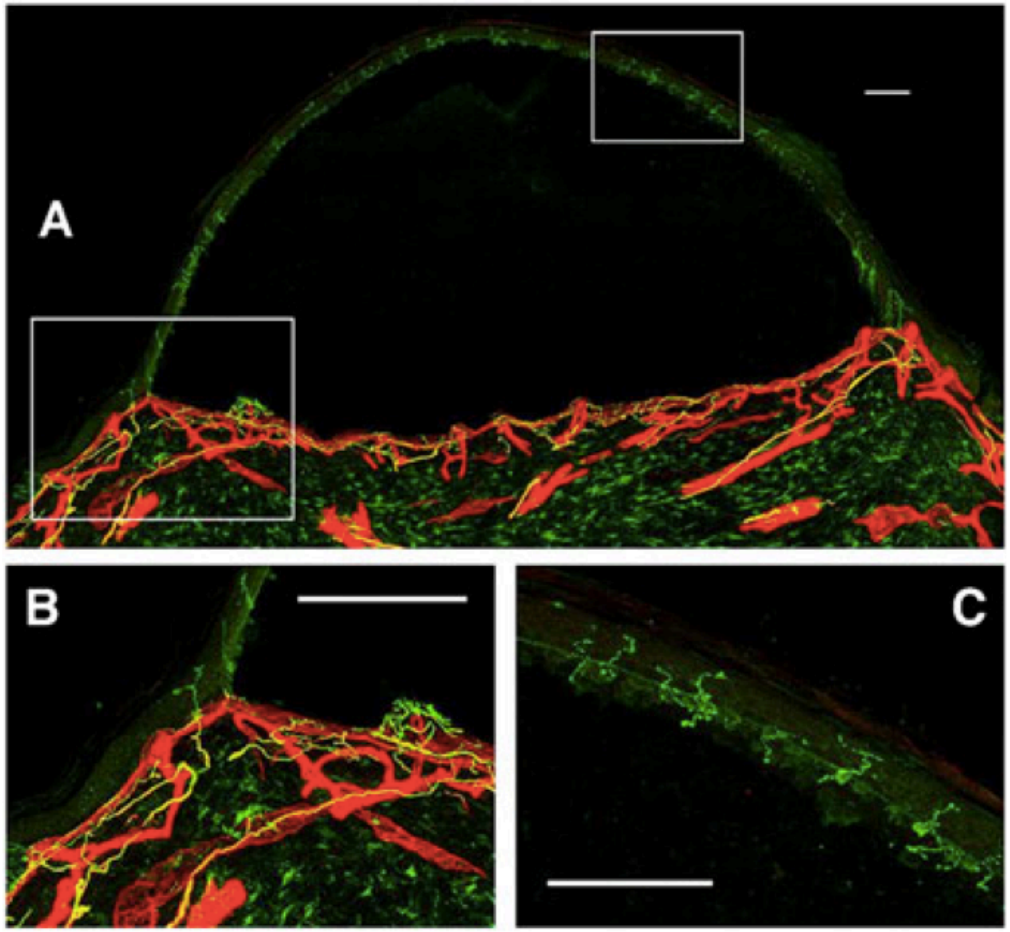
Two goals today:

- Can we detect a diagnostic change in spatial pattern?
  - Number of ENFs per unit area goes down with neuropathy.
  - Qualitative description of spatial pattern becoming "more clustered" as neuropathy progresses.
- Can we construct a stochastic model of ENF structure?
  - Can we generate a model with realistic structure of ENFs?
  - Can we generate realistic distributions for the number/length/direction of branches?
  - Can we estimate these features from data?

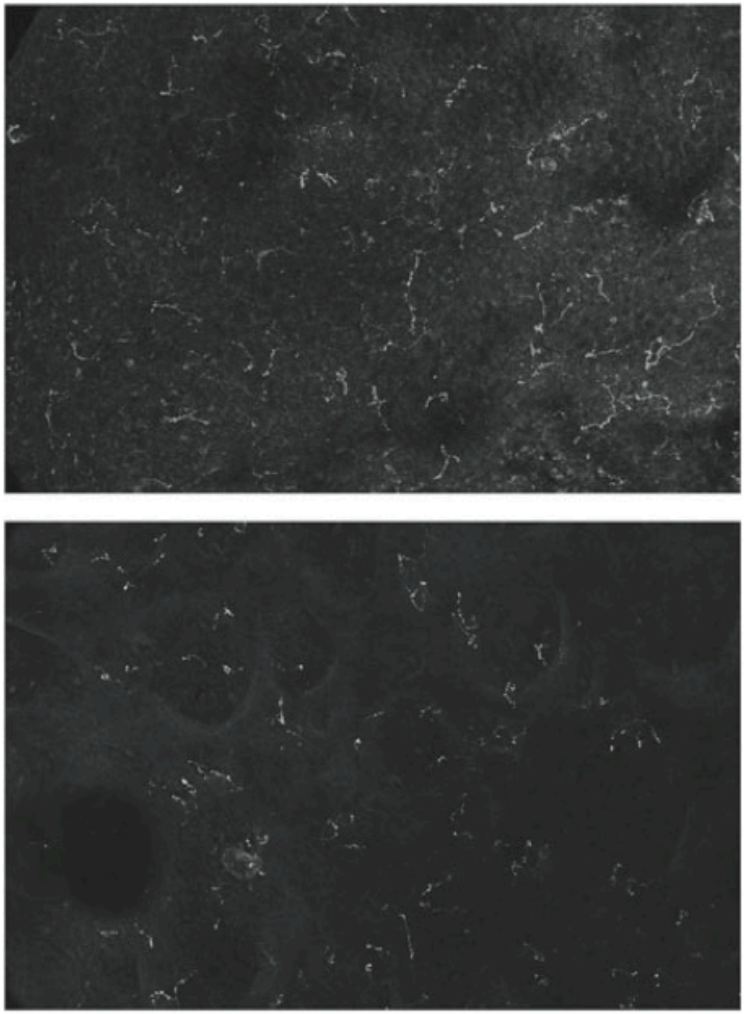
What data do we have?

- Skin punch biopsy gives side view and number of ENFs per unit area.
  - Clear association with progression of diabetic neuropathy.
- Blister biopsy separates epidermis from dermis with ENFs intact.
  - Allows bird's-eye view of spatial arrangement of ENFs.
- Three dimensional tracing of ENFs
  - Adds (small) third dimension, but not used in our work to date.

Blister biopsy, side view



Top view (non-diabetic (top), and severe diabetic (bottom))



What analytic tools do we have?

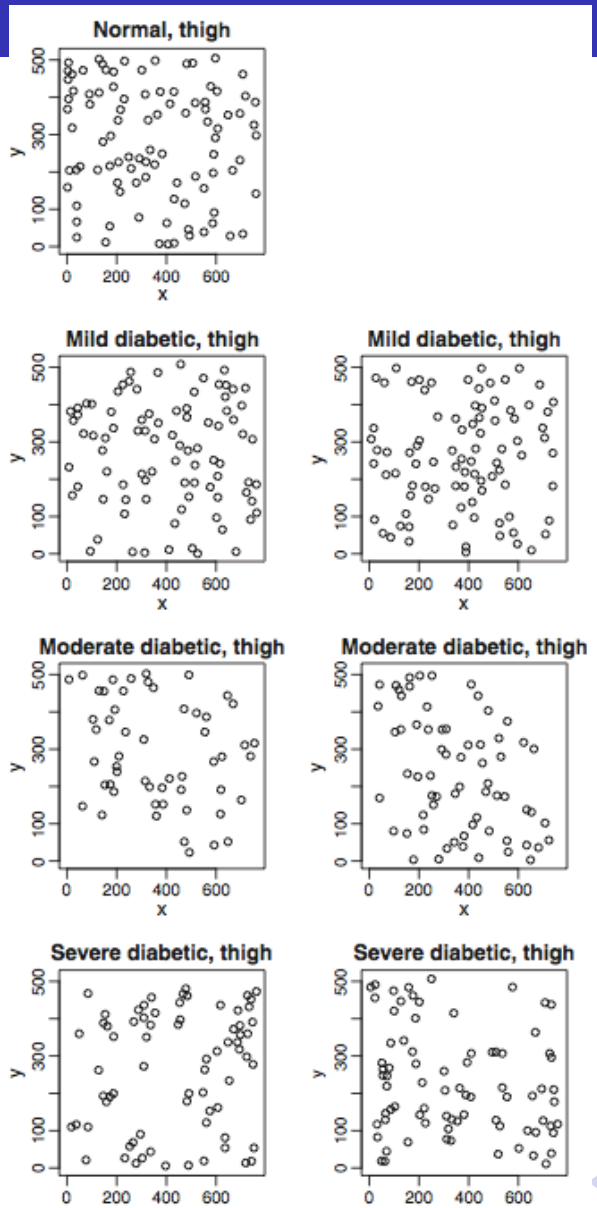
Spatial point and fiber processes:

- Points: How to measure "more clustered" (and at what distance)?
  - Second order properties: K function and pair correlation function.
  - Note: Not interested in *where* clusters happens, want summary of how much clustering overall.
  - We will summarize parters of ENF "base" points (the point where the ENF enters the epidermis, before branching).
- Fibers: How to describe number/length/direction of ENF branches.
  - Allows bird's-eye view of spatial arrangement of ENFs.
  - We will model base points and end points (two linked point processes).
  - Two hierarchical (multi-stage) models of ENF structure.

Study 1: Second order properties

- Waller et al. (2011, *Stat in Med*).
- Seven patients (one non-diabetic, two mild diabetics, two moderate diabetics, two severe diabetics).
- Blister biopsy from thigh.
- Base locations extracted from image.
- We want to quantify the amount of clustering.
- We want to compare patients with different disease severity.

Study 1 data


Normal, thigh
Mild diabetic, thigh
Mild diabetic, thigh
Moderate diabetic, thigh
Moderate diabetic, thigh
Severe diabetic, thigh
Severe diabetic, thigh
x
y
0
200
400
600
100
300
500

Basic elements of spatial point processes

- Poisson process as point of reference (complete spatial randomness, CSR). Two magical properties:
  - Number of events follows a Poisson distribution.
  - Given number of events, locations of events follows a uniform distribution.
- First order property: the *intensity* represents the expected number of events per unit area.
- Second order property: Summarizes inter-relationships between events.
- Bartlett: Without additional information, cannot identify whether first or second order properties drive observed pattern.

Statistical tools

- Estimate homogeneous intensity, λ, via $\frac{\text{number of events}}{\text{area}}$.
- Estimate heterogeneous intensity via non-parametric density estimators (e.g., kernels). But not today.
- Second order: Ripley's K function
- $K(h) = \frac{E[\#\text{ of events within } h \text{ of a } \textit{randomly} \text{ chosen event}]}{\lambda}$
  - Under CSR, $K_{CSR}(h) = \pi h^2 =$ area of circle with radius h.
  - Under regularity, $K(h)$ tends to be $< \pi h^2$.
  - Under clustering, $K(h)$ tends to be $> \pi h^2$.
- Besag transformation: $L(h) = \{K(h)/\pi\}^{1/2}$
- Under CSR, $L(h) - h = 0$.

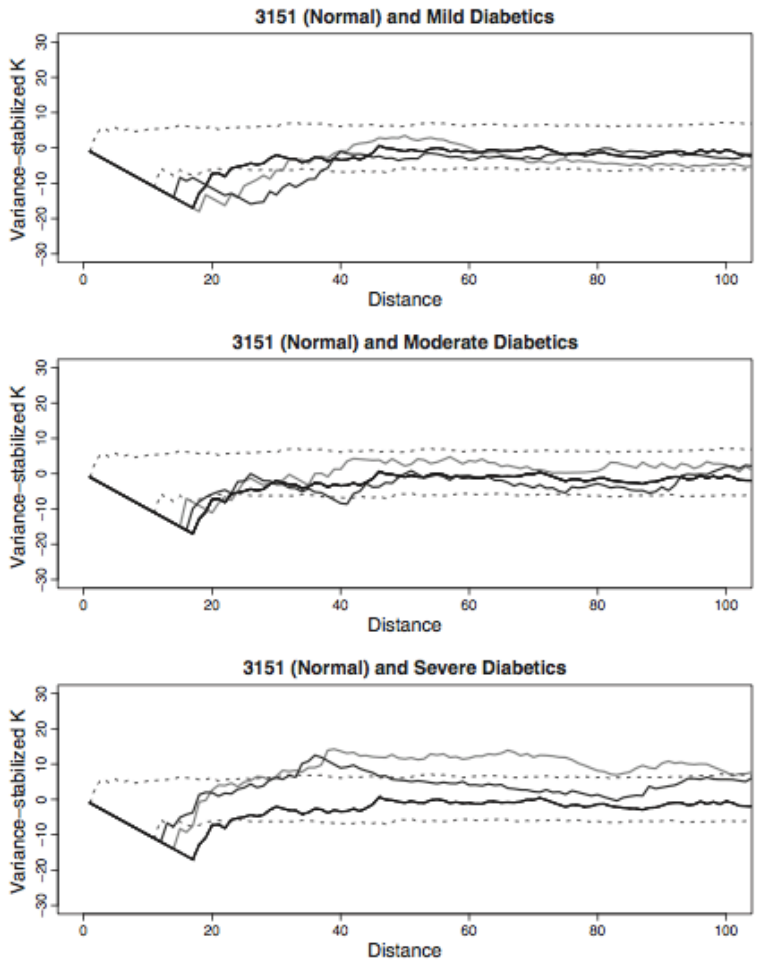
Estimation of $K(h)$

- $K(h) = \dfrac{E[\#\text{ of events within } h \text{ of a } \textit{randomly} \text{ chosen event}]}{\lambda}$
- Ripley's edge-corrected estimate:

$$\widehat{K}(h) = \frac{1}{N\widehat{\lambda}} \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} w(\mathbf{x}_i, \mathbf{x}_j)^{-1} I(\|\mathbf{x}_i - \mathbf{x}_j\| \leq h), \ h > 0,$$

  - $\widehat{\lambda}$ = the estimated intensity
  - $\mathbf{x}_i$ = the location of the ith event $(i = 1, \ldots, N)$
  - $I(\cdot)$ = the indicator function.
  - $w(\mathbf{x}_i, \mathbf{x}_j)$ = the proportion of the circumference of a circle centered at $\mathbf{x}_i$, passing through $\mathbf{x}_j$ that is within the study area.
- Simulate 500 realizations of CSR with the same sample size, calculate $\widehat{K}(h)$ and plot envelopes based on 2.5th and 97.5th percentiles at each h.

Non-diabetic (black) and diabetics (grey)



Features

- Some inter-individual variability, but consistency between repeat samples from the same individual.
  - Combine estimates from same individual:

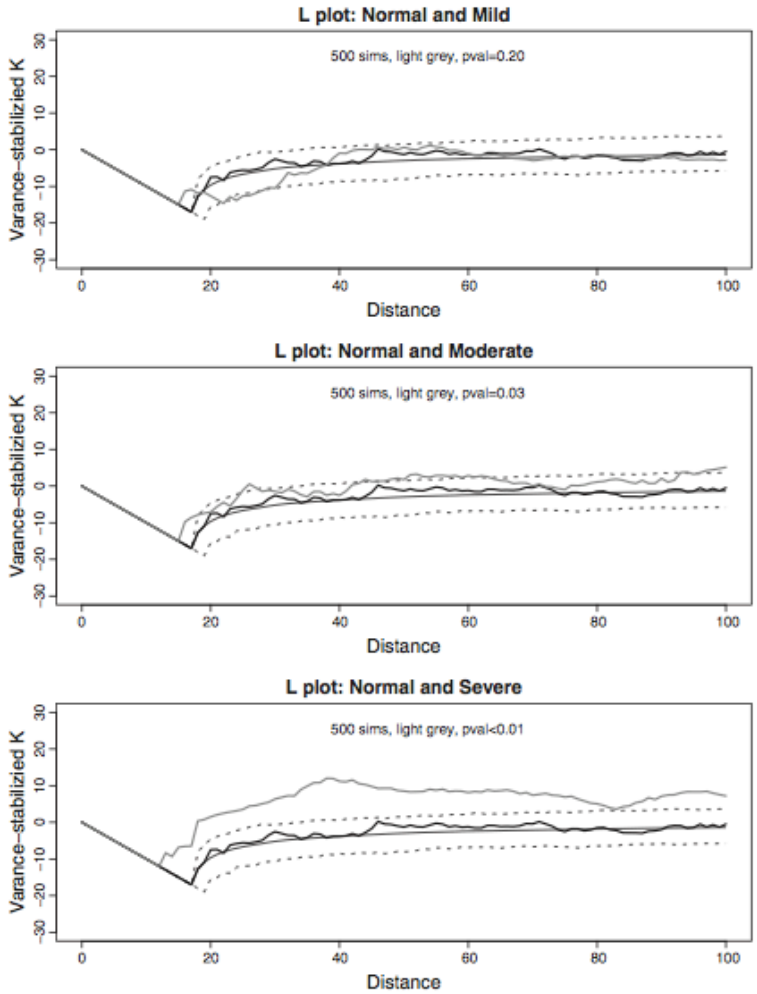
$$\hat{K}(h) = \frac{n_1 \hat{K}_1(h) + n_2 \hat{K}_2(h)}{n_1 + n_2}$$

- Clear increased clustering for severe diabetic.
- Initial linear decline indicates an absence of observed events within very short distances of one another.
- Is CSR the appropriate null hypothesis?
- Primary question is whether patterns from diabetics are different from non-diabetics, not different from CSR.

Hard-core process

- Process containing zone around each event wherein no other events occur.
- Matérn type II hard-core process with radius R. No events allowed within R of any other event.
- Straightforward to simulate (reject any proposed events within R of any other events).
- Simulate 500 realizations, calculate $\widehat{K}(h)$ and plot 2.5th and 97.5th percentiles at each h.

Non-diabetic (black) and diabetics (grey)



Features

- Non-diabetic K functions fall within the hard-core envelopes.
- Identifies spatial scale (17.28 units) of the "territory" around each ENF trunk. (Interesting neurologic finding).
- Monte Carlo test based on test statistic:

$$\max_{30 \leq h \leq 60} [\hat{L}_{diab}(h) - \hat{L}_{non}(h)]$$

- Significant for moderate and severe diabetics.

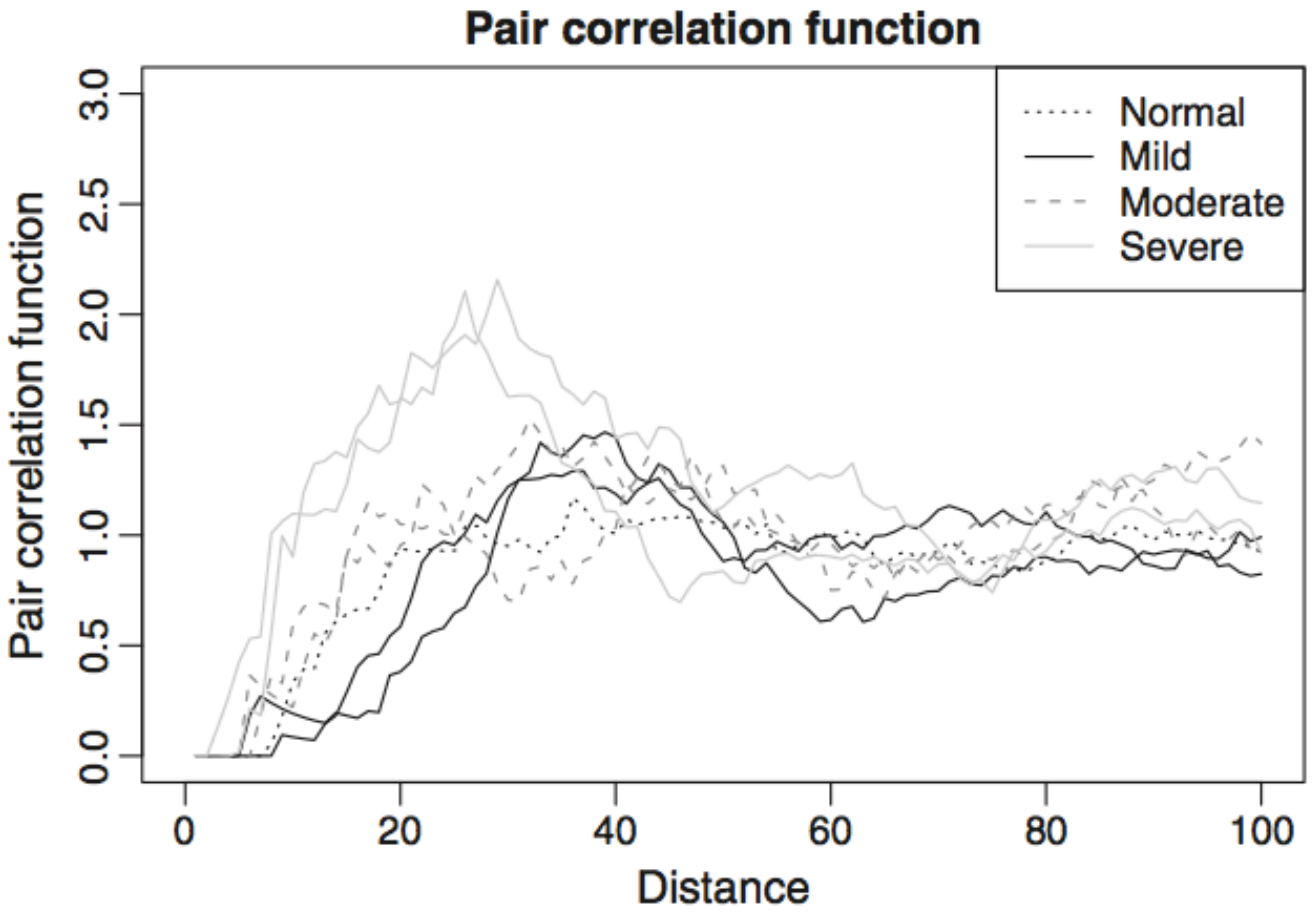
Pair correlation function

- K function is a cumulative function ("clustering up to distance h").
- What about clustering *at* distance h?
- The pair correlation function defined as $g(h) = \frac{1}{2\pi h}\frac{dK(h)}{dh}$.
- $g(h) = \lambda^2\rho(h)$ where $\rho(h)$ is the second-order product density.
- Estimated by

$$\hat{\rho}(h) = \frac{1}{2\pi h}\sum_{i=1}^{N}\sum_{j=1,i\neq j}^{N} w(\mathbf{x}_i,\mathbf{x}_j)^{-1}k(h-||\mathbf{x}_i-\mathbf{x}_j||)$$

for kernel function $k(\cdot)$.

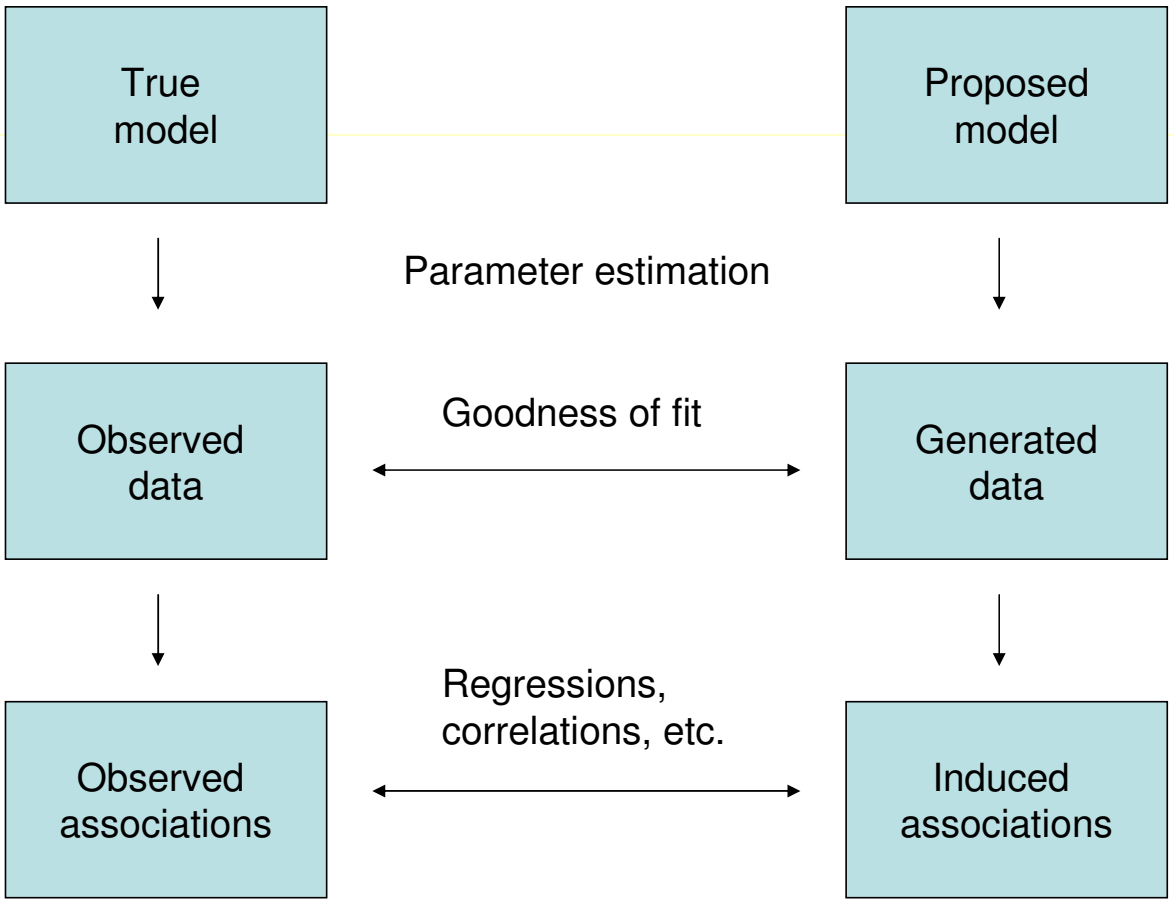
Estimated pair correlation functions



Conclusions for Study 1

- Quantifies qualitative suggestion of clustering increasing with disease severity.
- Clear impact of severe diabetes, more subtle for moderate, difficult to detect for mild.
- Better sensitivity and specificity as we refine inference to focus on questions of interest.
  - CSR a place to start.
  - Hard-core process a little clearer.
  - Monte Carlo tests (via simulation of null model).
  - Pair correlation suggests scale of clustering decreasing as well.
- Moving from basic statistics toward modeling of underlying process.

Mathematical modeling and statistics



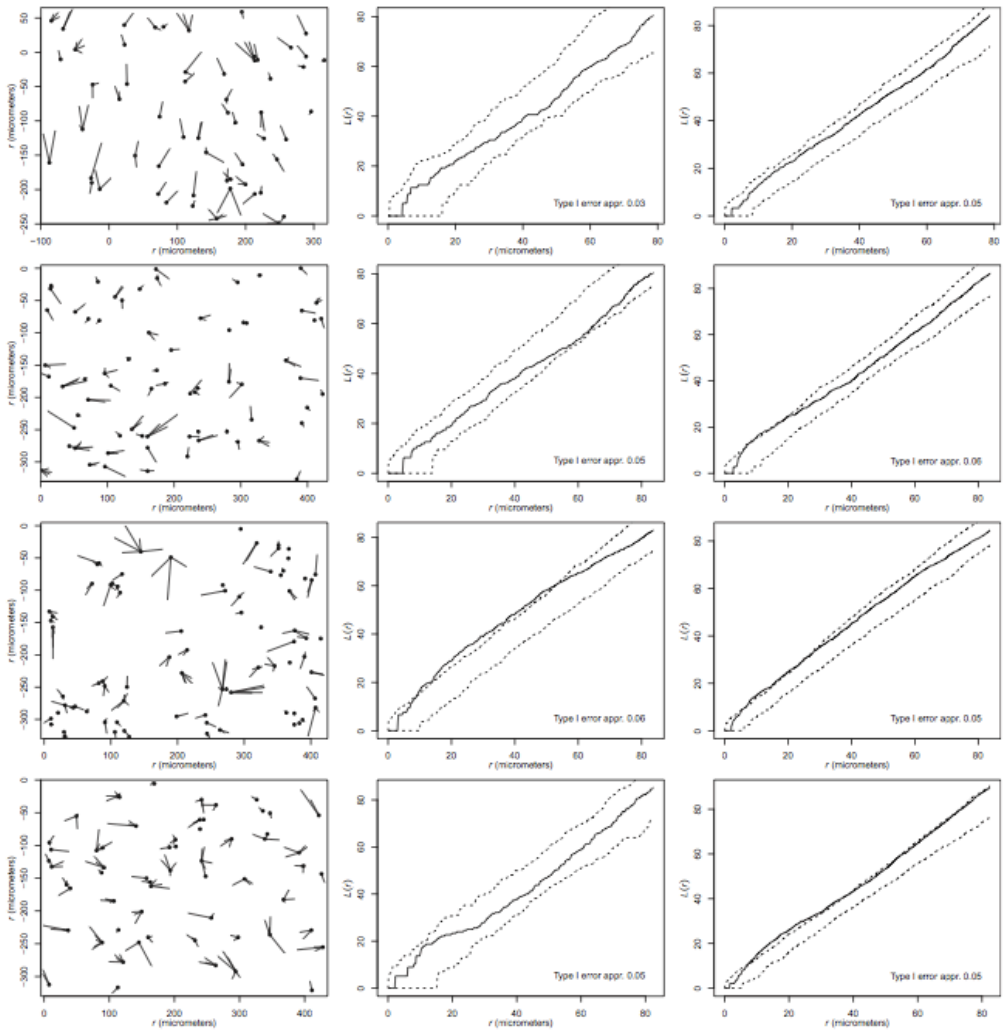
Study 2: Modeling the structure of ENFs

- Olsbo et al. (2013, *Mathematical Biosciences*).
- Goal: Build a stochastic model that generates ENF patterns to describe:
  - Number of branches
  - Direction of branches
  - Length of branches
- Allow inter-subject variability in parameters.
- Data: Base and end points of ENFs
  - base = tree trunk
  - end = tip of a branch

Data description

- Blister biopsies from thigh for four (non-diabetic) subjects.
- Observe base points, end points, and connect base to end points.
- First descriptive statistics:
  - Compare base points to CSR via the K function.
  - Compare end points to CSR via the K function.

CSR comparison for base and end points



Features

- Directionality for Subject 171 (top line)?
- Inter-subject variation in direction, length.
- Base points largely follow CSR.
- End points potentially clustered?

Two models

- Two stage Poisson (TSP) model
  - Descriptive model
  - Motivation: Uniform coverage of skin by end points
- Non-orphan cluster (NOC) model
  - Mechanistic model
  - Motivation: Hierarchical description of ENF structure.

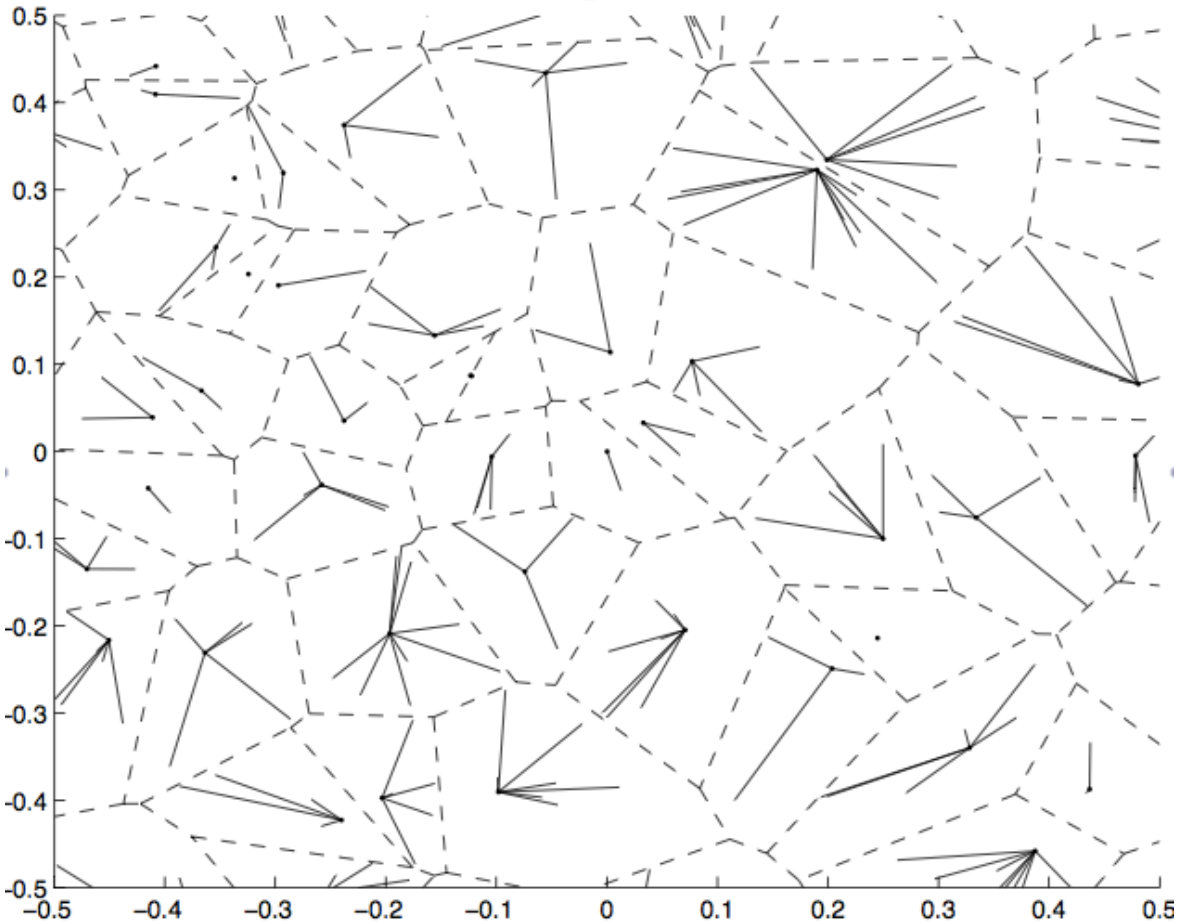
Two stage Poisson model

- Motivation: If goal is to provide uniform coverage across the skin, should end points follow uniform distribution?
- Three components:
  - Base points follow spatial point process Φ_b.
  - End points follow spatial point process Φ_e.
  - Every event in Φ_e is linked to one event in Φ_b.
- To begin: let Φ_b, Φ_e be homogeneous Poisson processes (CSR).

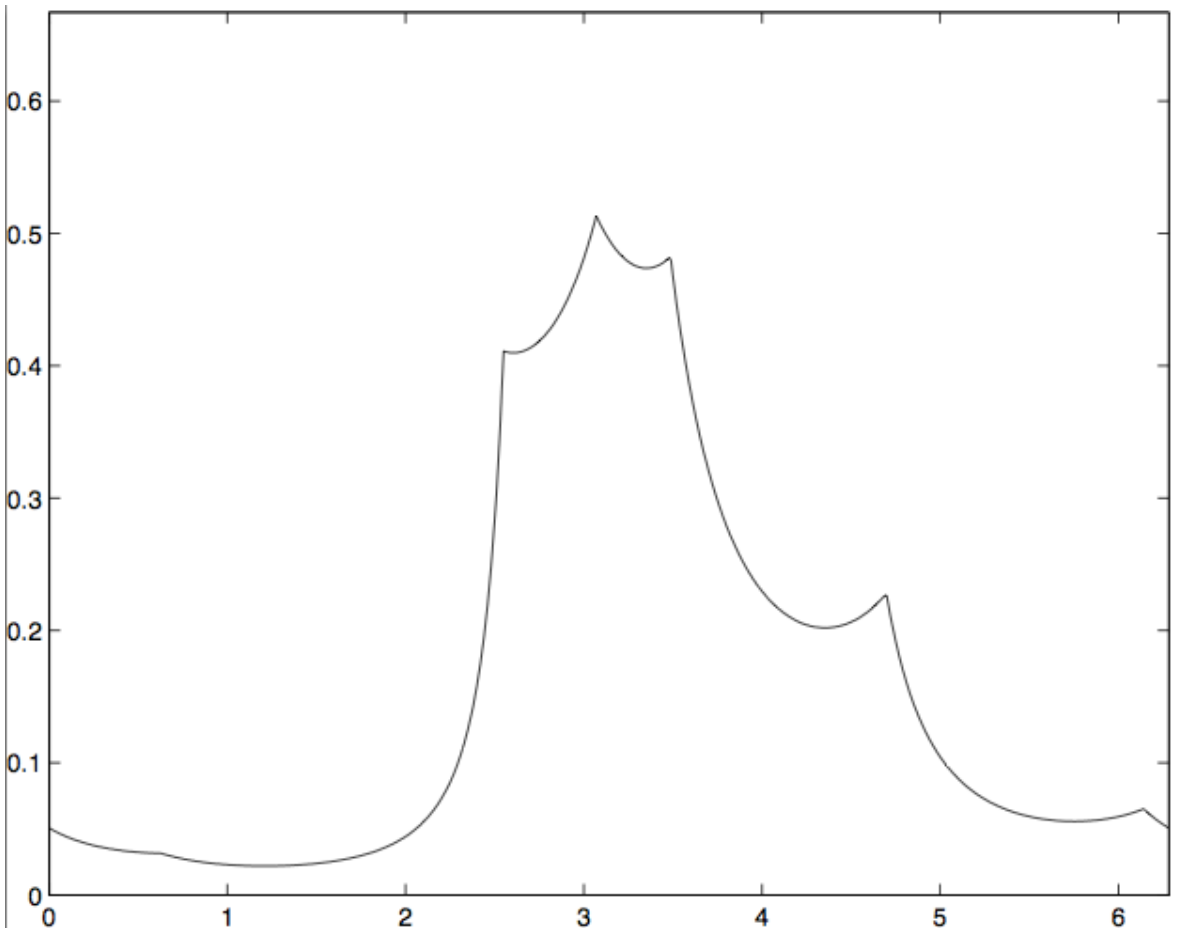
Linking ends to bases

- First thought: "Grumpy neighbor process": Link each event in Φ_e to nearest event in Φ_b.
- "Keep your kids out of my yard!"
- No overlapping branches (recall, none observed).
- All end points for a particular base point fall within the Voronoi cell for that base point.
- Marginal distribution of branch direction is uniform.
- Conditional distribution of branch direction (given observed base events) had attractive feature: branches tend to grow toward empty space.
- Similar model in cell phone utilization (Foss and Zuyev, 1996, *Adv. Appl. Prob.*).

Grumpy Neighbor Process



Conditional angle distribution



Generalization yields TSP

- Allow end points to link to nearest base with high probability, but also allow link to other base points (with probability decreasing with distance).
- Connection probability: $f(m, \theta)$ where $m =$ mark denoting the rank distances to base points ($m = 0$ for nearest).
- Note: TSP model not a direct model of generating end points for each base point.
- Developed in reverse from goal of uniform coverage of the area by end points and observation that base points are also uniform.
- Assess fit and adequacy of model through simulation.
- Summarize observed features of physiologic interest: spatial distributions, numbers, length, direction of branches.

Some theoretical results

- Foss and Zuyev (1996) provide framework for deriving moments of functions of end points connected to a typical base point for GNP.
- Olsbo et al. (2013) extend to include the connection probability (mark distribution).

$$E(\text{number of branches per base}) = \frac{\lambda_e}{\lambda_b}$$

  - Note: does not depend on m (which base).

$$E(\text{branch length}) = \frac{\lambda_e}{\lambda_b^{3/2}\pi^{1/2}} \sum_{m=0}^{\infty} f(m;\theta)\frac{\Gamma(m+3/2)}{\Gamma(m+1)}$$

- Second order moments (variances and covariances) derivable but more involved.

Completing the TSP model

- Φ_b, Φ_e are homogeneous Poisson processes with intensities λ_b, λ_e.
- Stoyan and Stoyan (1994) provide density of nearest neighbor distances given m.
- We propose a generalized Poisson distribution (additional shape parameter) for m.
- Taken together provides the unconditional branch length distribution.

Next steps

- TSP has appealing theory, but doesn't model structure directly.
- We can measure ENF features, but they are not inherent in the model structure.
- TSP also generates parents with no children.
- Can we build a more mechanistic model?

Non-orphan cluster (NOC) model

- Model base point process Φ_b as homogenous Poisson process.
- Define a number of end points for each base point.
- Distribute children around the parent.
- Similar to Poisson cluster process (or, more generally, a shot-noise process), but our realization includes both parents (bases) and their children (ends) *and the links between them*.
- To complete the model we need a distribution of number, direction, and distance of ends for each base.

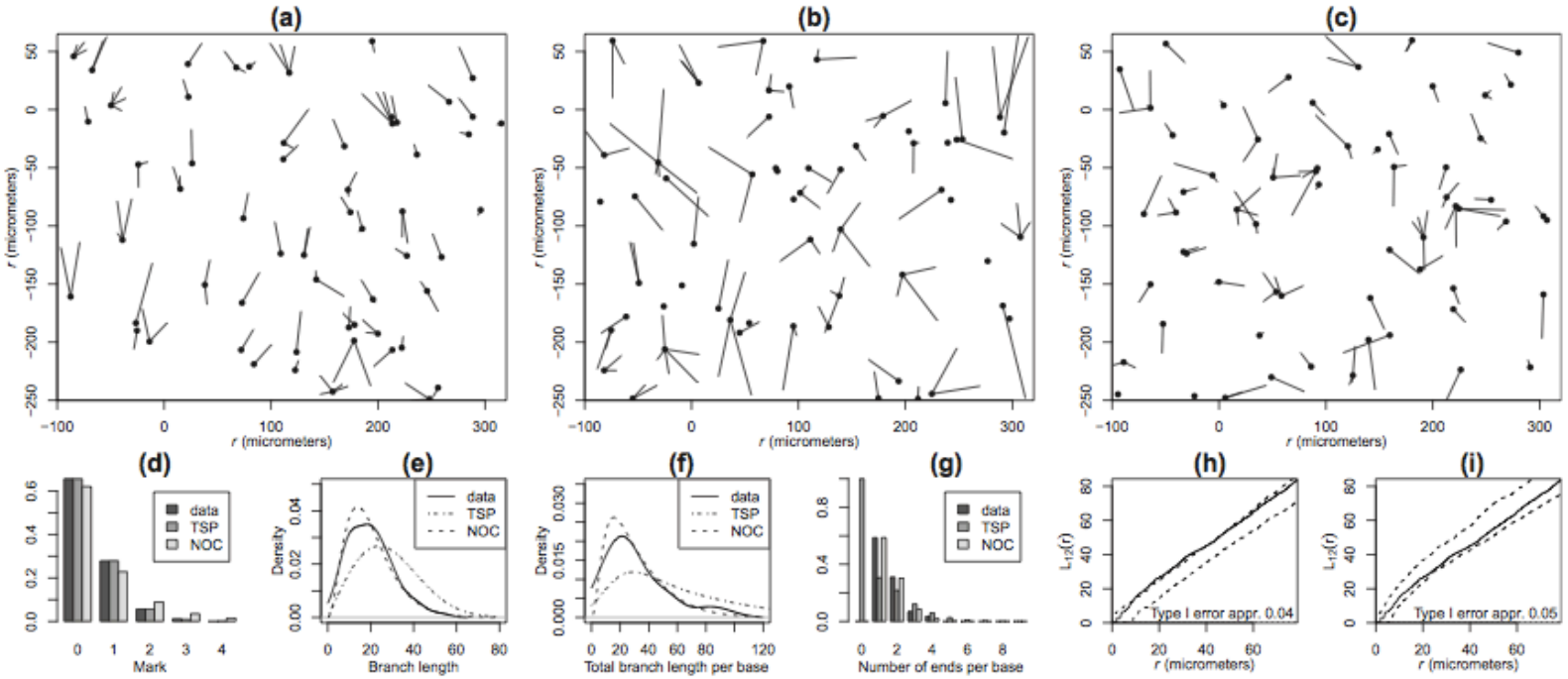
NOC specification

- Φ_b a homogeneous Poisson process with intensity λ_b.
- Number of branches per base follows a generalized Poisson distribution with shape parameter and $\Pr(m = 0) = 0$.
- Direction (for each base point) follows a von Mises distribution with mean direction *opposite* the direction to the nearest neighboring base to the parent (aim for open space).
- Branch lengths follow a gamma distribution.

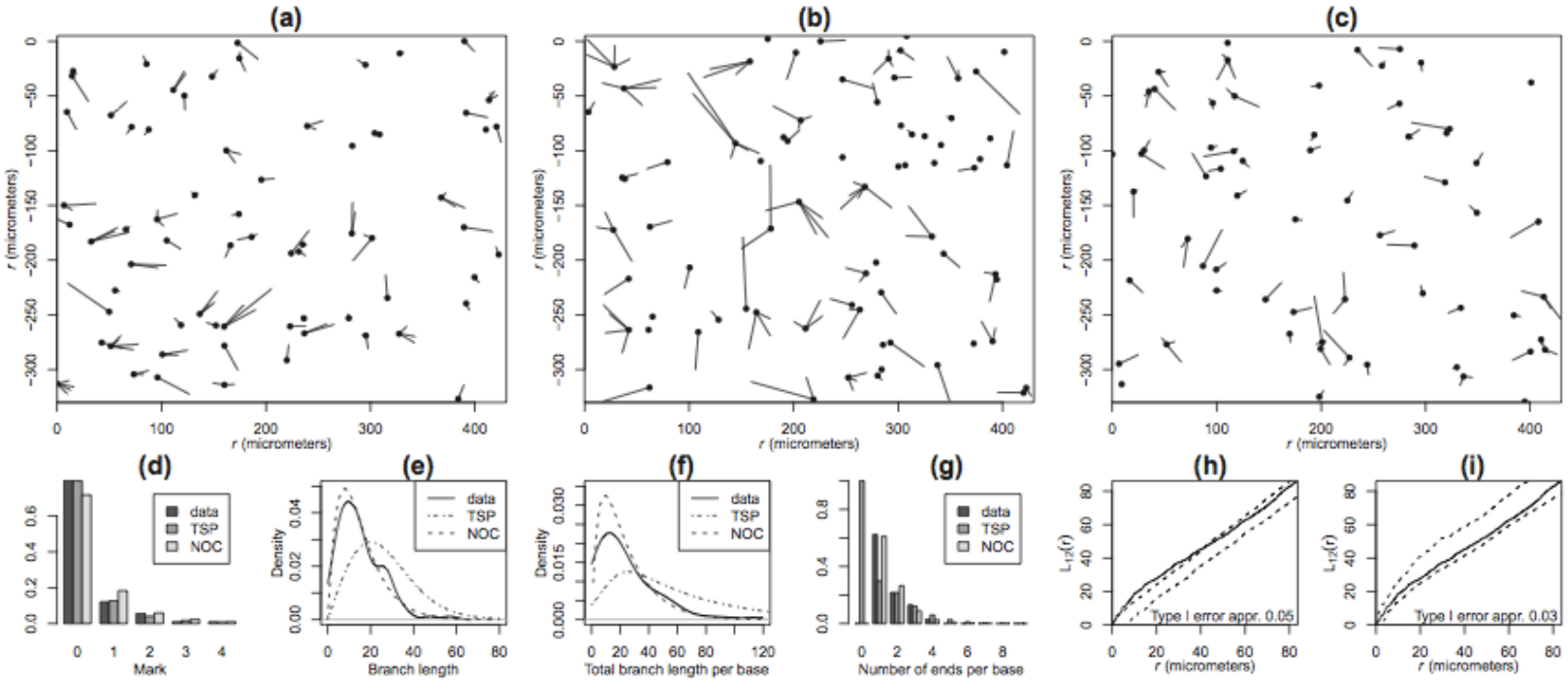
Parameter estimation and simulation

- Poisson process assumption: MLE for λ_b and λ_e are number of events divided by total area.
- TSP requires estimates of two parameters in the generalized Poisson distribution (MLEs).
- NOC requires estimates for number of branches, direction, and lengths.
  - Maximize log "likelihood" functions for conditional distributions.
- Simulate realizations from fitted processes and compare simulated feature distributions to empirical distributions in the data.

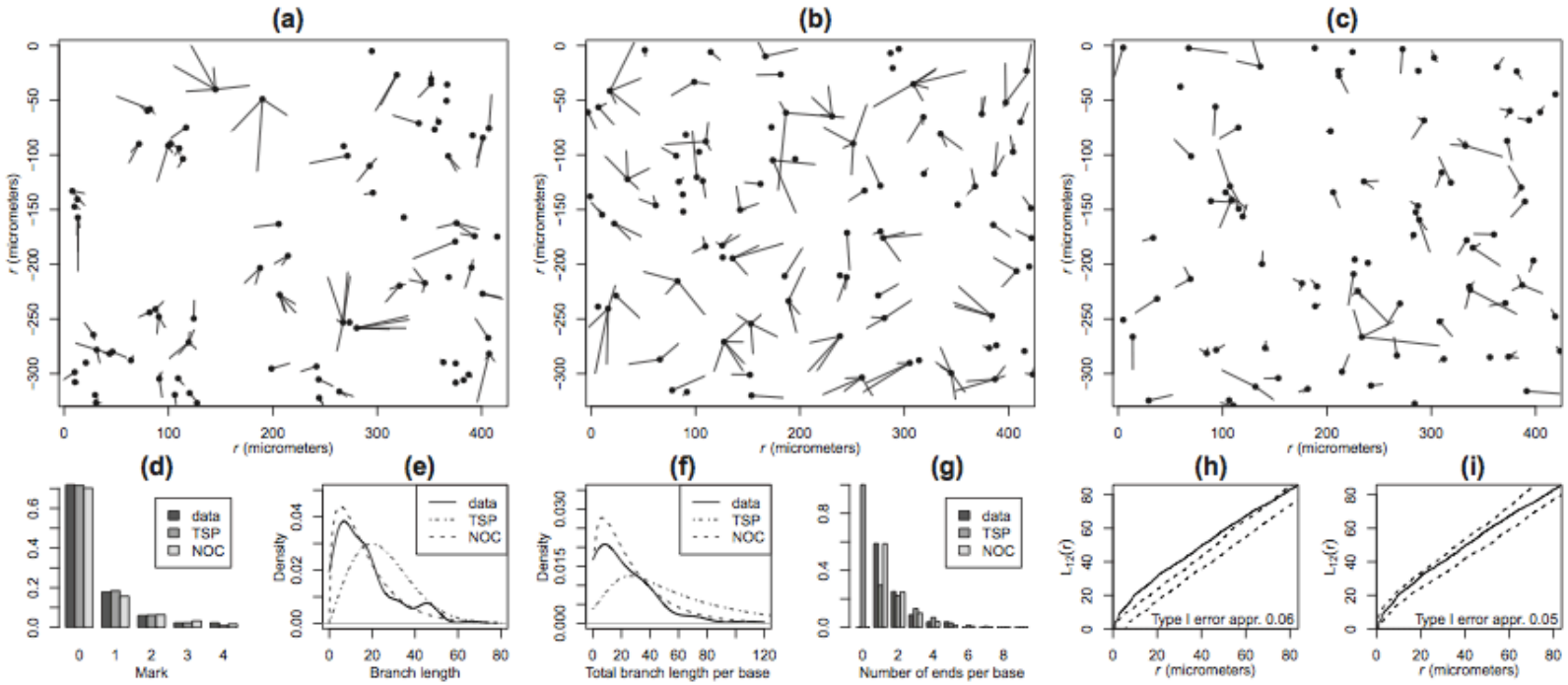
Subject 171


(a)
(b)
(c)
r (micrometers)
(d)
(e)
(f)
(g)
(h)
(i)
data
TSP
NOC
Mark
Density
Branch length
Total branch length per base
Number of ends per base
L12(r)
Type I error appr. 0.04
Type I error appr. 0.05

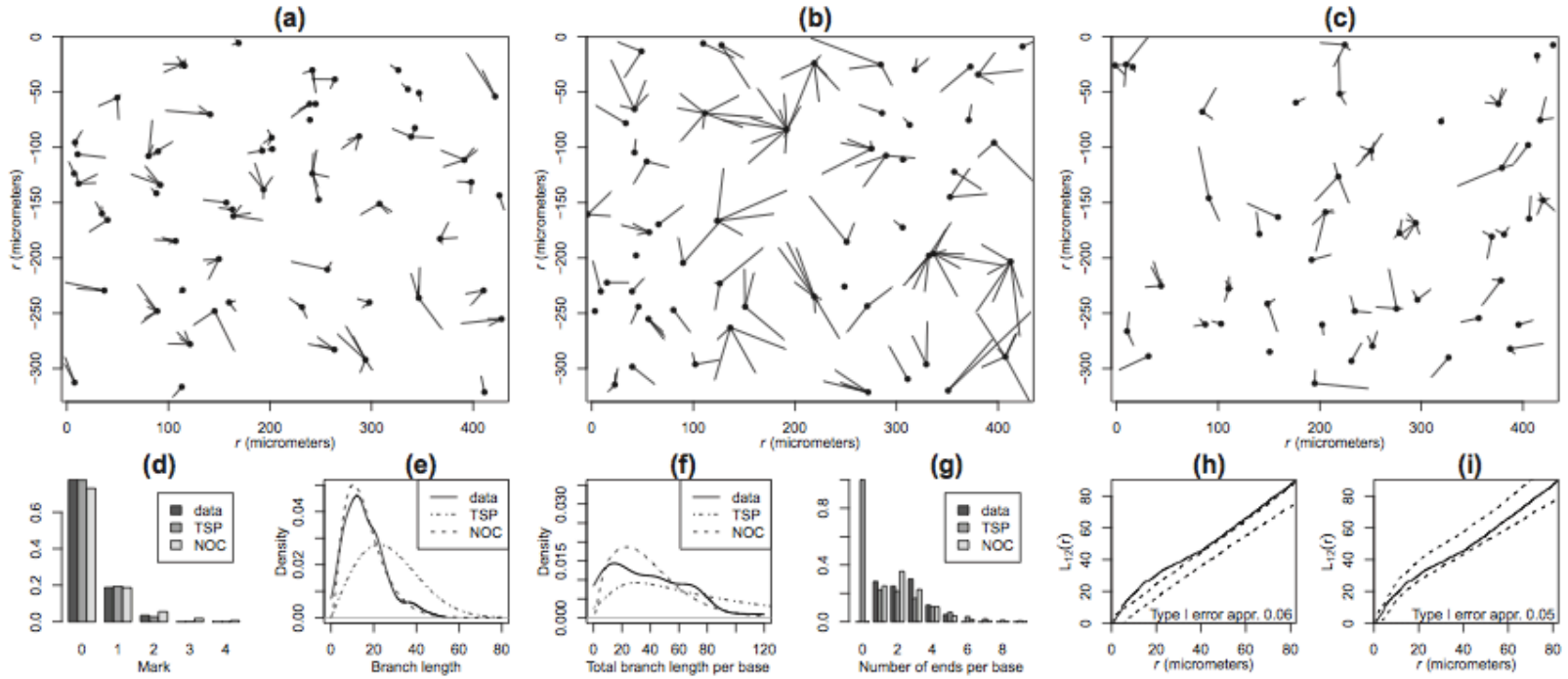
Subject 224


(a)
(b)
(c)
r (micrometers)
(d)
(e)
(f)
(g)
(h)
(i)
data
TSP
NOC
Mark
Branch length
Density
Total branch length per base
Number of ends per base
Type I error appr. 0.05
Type I error appr. 0.03

Subject 230


(a)
(b)
(c)
r (micrometers)
r (micrometers)
(d)
(e)
(f)
(g)
(h)
(i)
data
TSP
NOC
Mark
Density
Branch length
Total branch length per base
Number of ends per base
Type I error appr. 0.06
Type I error appr. 0.05

Subject 256


(a)
(b)
(c)
r (micrometers)
(d)
(e)
(f)
(g)
(h)
(i)
data
TSP
NOC
Mark
Density
Branch length
Total branch length per base
Number of ends per base
Type I error appr. 0.06

Findings

- TSP captures mark distribution (connection probabilities) well but tends to overestimate branch length.
- NOC provides better match for number of branches, individual branch length, total branch length per base.
- NOC a better match for second order properties than TSP.
- No strong evidence for preferred direction away from neighboring base points.

Conclusions

- Spatial point processes provide valuable analytic tools for description and inference for ENFs.
- However, requires link between descriptive summaries and models of ENF structure to answer more specific questions.
- Close collaboration with neurologists required to make sure the questions we answer are of interest.
- Simulation provides a powerful tool, but also some creative insight for application.
- Statistical analysis provides insight into structure and pattern in ENFs within disease classes.
- Future work: Adding time to model innervation/denervation process.

Collaborators

- Emory University: Traci Leong
- Chalmers Institute of Technology: Viktor Olsbo, Aila Särkkä
- Aalto University: Mari Myllymäki
- University of Minnestoa: Ioanna Panoutsopoulou, William R. Kennedy, and Gwen Wendelschafer-Crabb

References

- Waller et al. (2011) Second-order spatial analysis of epidermal nerve fibers. *Stat in Med*.
- Olsbo et al. (2013) Development and evaluation of spatial point process models for epidermal nerve fibers. *Mathematical Biosciences*.