Challenges in spatial point pattern analysis

Aila Särkkä

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

Background

In early spatial point process literature point patterns were

- small
- observed in 2D
- with simple interaction structures
- with no repetitions available
- realizations of stationary and isotropic point processes.

Nowadays, it is more and more common that point patters are

- large
- observed in 3D or in space and time
- with complicated interaction structures
- with repetitions
- realizations of point processes that are not stationary/isotropic.

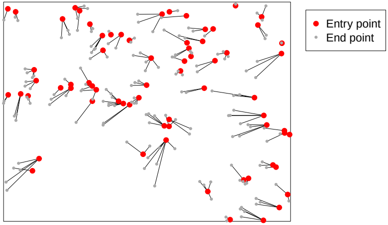
Two examples of complicated point pattern data

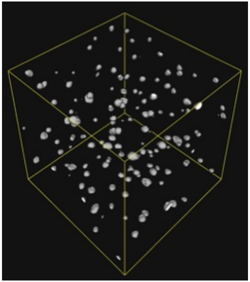
Locations of entry and end points of epidermal nerve fibers

- 3D, repetitions, unusual cluster process, local anisotropy

Locations of air bubbles in polar ice

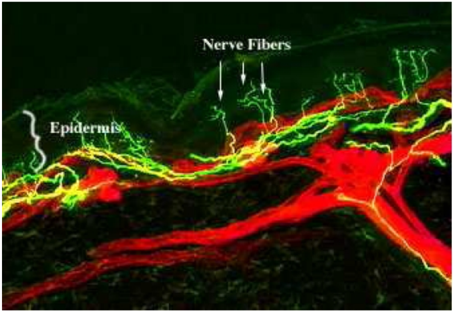
- 3D, repetitions, anisotropy, non-stationarity, noise present





Epidermal nerve fibers

Epidermal nerve fibers (ENFs) are thin sensory nerve fibers in the epidermis, which is the outermost living layer of the skin.



Entry (base) points: locations where the ENFs penetrate the epidermis

End points: locations of the terminations of ENFs

Diagnostic value of ENFs

ENF structure changes due to so-called small fiber neuropathies, such as diabetic neuropathy.

Dr. Kennedy's group at the University of Minnesota has noticed that subjects with diabetic neuropathy have

- less ENFs per surface area
- smaller summed length of ENFs per volume
- more clustered nerve patterns

than healthy subjects.

Data

- 32 healthy volunteers and of 20 subjects with diabetes.
- Six body parts.
- Two blisters, approximately of the size 330 microns by 432 microns (by 20-50 microns), taken from each body part. (Here, only 2D data.)
- From three to six images were taken from the two blisters, typically two images from each blister. (Here, blister effect ignored.)
- Non-spatial covariates age, gender, and BMI are known for each subject.

Point process summary statistics

- Both entry point patterns and end point patterns from subjects with diabetic neuropathy tend to be more clustered than healthy patterns
  - Waller et al. (2011): only entry points and data from thigh
  - Myllymäki et al. (2013): data from calf and foot, covariates included
  - Andersson et al. (2016): data from foot, only diabetic subjects with "mild" neuropathy (ENF density within the normal range) included, reactive territory introduced
- Point process models could reveal more detailed differences between healthy and neuropathic patterns
  $\rightarrow$ construct models for entry and end points

Models for end points: cluster process

- We assume stationarity.
- Locations of entry points given.
- End point locations are modeled as clusters "around" the entry points. For a given entry point the model for an end point cluster contains
  - the number of points in the cluster (a discrete valued random variable)
  - the distance between an end point and the entry point (a positive random variable)
  - the angle of an end point (a random variable taking values in $[0, 2\pi)$).
- We have suggested two models, non-orphan cluster (NOC) model (Olsbo et al., 2013) and uniform cluster center (UCC) model (Andersson et al., 2016)

Models for end points

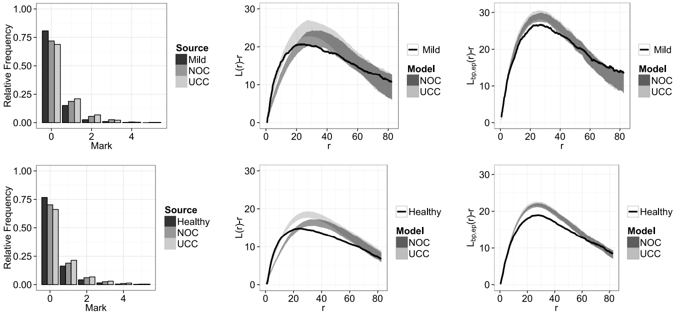
NOC

- Number of points per cluster: Jonquiére distribution or negative binomial distribution (for number - 1).
- Distance between an end point and its entry point: Gamma distribution.
- Angle of an end point: von Mises distribution, with mean angle θ_0 and concentration parameter κ. The mean angle is chosen so that end points favour directions towards open space, i.e. avoid the direction of the nearest other entry.

UCC

- Number of end points and distance between end point and its entry as in NOC
- von Mises for the angles: each end point cluster has a mean angle that is uniformly distributed on $[0, 2\pi)$. Points in end point clusters have still clustered directions but no preferred direction.

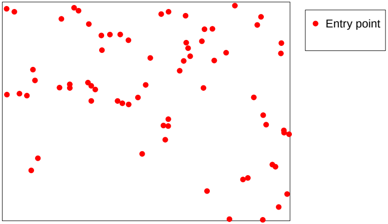
Are NOC and UCC good models?

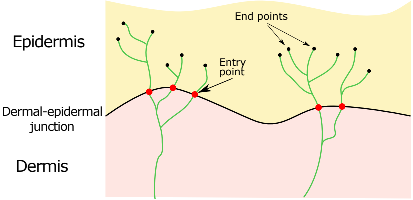


From left to right: "mark" distribution, where the mark is the number of entry points that are closer to an end point than the end point's own entry point is, Ripley's L function for the end points, and the cross L function for the entry and end points. Top row: Mild diabetic neuropathy; Bottom row: Healthy.

Model for entry points

Neyman-Scott cluster process for the entry points works quite well according to our preliminary results.

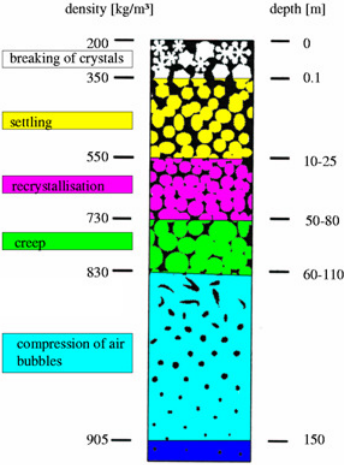




Final model: (3D) cluster process for end points where the parent (entry) points are clustered. Model for the end points should be improved by taking dependence on the other end and entry points into account.

Air bubbles in polar ice

- Polar ice has information on the climate of the past.
- To be able to interprete ice core records, one has to know how old the ice is.
- There are some theories connecting the dynamics of glaciers to the age of ice.
- Question: How can we estimate the deformation in polar ice?
- Method: Polar ice is compacted snow. If we go deep enough, the air pores are isolated in the ice.
  → Study the anisotropy (deformation) of these air inclusions in the ice samples.



Two ice data sets

Talos Dome, Antarctica (Talos)

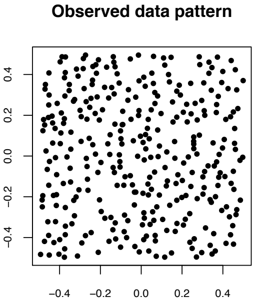
- ice from depths 153m, 353m, and 505m
- 14 samples ($1.0 \times 1.0 \times 1.0\text{cm}^3$, 350-750 bubbles) per depth
- volume preserving spherical compression, where the ice is compressed in z direction and stretched isotropically in the xy plane

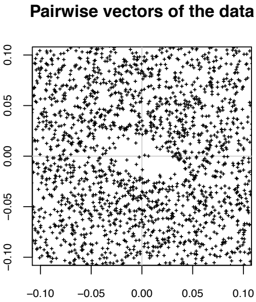
Dronning Maud Land, East Antarctica (EDML)

- ice from depths 151m, 749m, and 915m
- 30 samples ($3 \times 3 \times 1.7\text{cm}^3$, 2000-3000 bubbles) per depth
- volume preserving transformation
- compression accompanied by a lateral flow

How to visualize anisotropy?

A point pattern (left) and its pairwise difference vectors for each point pair, a Fry plot (right).





The pairwise difference vectors not rotationally invariant
→ pattern not isotropic

Set-up

- The underlying point process X_0 is regular, stationary and isotropic
- We assume that the observed point pattern is a realization of a point process X, which we obtain by

$$X := TX_0 = \{Tx : x \in X\},$$

where $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is an invertible linear mapping, and since X_0 is isotropic, can be decomposed as

$$T = RC$$

where R is a rotation matrix and C is a diagonal scaling matrix that compresses and stretches the dimensions.

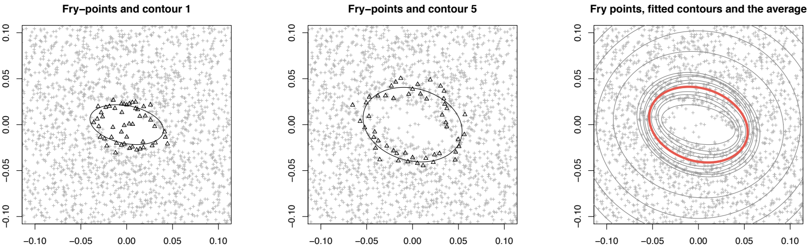
- Note that X is stationary

Estimating the transformation in two steps

- First, we estimate the rotation R by fitting ellipsoids to the contours of directed cumulants of the difference vectors, i.e. fitting ellipsoids to the Fry plot.
- Second, we estimate the scaling C by transforming the back-rotated data as in Redenbach et al. (2009).

Illustration of the ellipsoid fitting in 2D

54 directions



Left: Ellipsoid fitted to the nearest Fry points
Middle: Ellipsoid fitted to the 5th nearest Fry points
Right: Mean ellipsoid

Estimating scaling

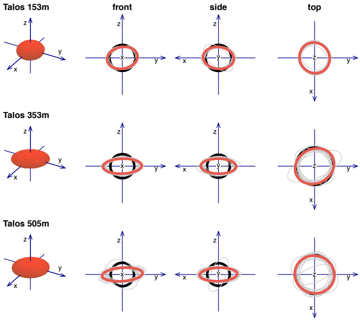
- Back-rotate $\tilde{\mathbf{x}} = \hat{R}^{-1}\mathbf{x}$ so that we get approximately $C\mathbf{x}_0$, where $\mathbf{x}_0$ is the original isotropic pattern.
- We assume volume preservation, i.e. $|T| = 1$.
- We follow Redenbach et al. (2009): A grid of scaling parameters $\{\mathbf{d}_1, ..., \mathbf{d}_\mathbf{m}\}$ is chosen. Then, the data are back-transformed by $C(\mathbf{d}_\mathbf{i})^{-1}$ (for each i) and the final estimate for the scaling is given by the $\mathbf{d}_\mathbf{i}$ that minimizes

$$T_\mathbf{d} = \int_{r_1}^{r_2} (|\hat{K}_{x,\mathbf{d}}(r) - \hat{K}_{z,\mathbf{d}}(r)| + |\hat{K}_{y,\mathbf{d}}(r) - \hat{K}_{z,\mathbf{d}}(r)| + |\hat{K}_{x,\mathbf{d}}(r) - \hat{K}_{y,\mathbf{d}}(r)|)\, dr,$$

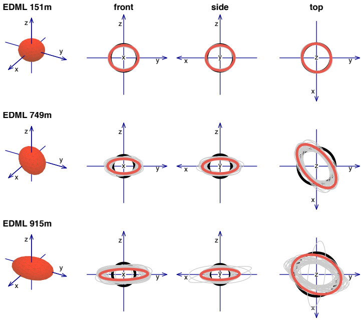
where $\hat{K}_{x,\mathbf{d}}(r)$ is an estimator for the conical K function (a directional version of Ripley's K function) in direction of the x axis etc.

Results for the ice data

Talos



EDML



What next?

- The new samples (EDML) contain extra relaxation (noise) bubbles that do not give any information on the motion of the ice but disturb the anisotropy analysis
  $\rightarrow$ We have developed an MCMC approach and a variational Bayes approach to classify the bubbles into two classes, real and noise bubbles
  $\rightarrow$ To do: Perform the anisotropy analysis and classification simultaneously
- Data show some layering of the ice (intensity of the air pores varies from layer to layer) due to seasonal changes
  $\rightarrow$ To do: Include inhomogeneity

References

- Andersson C, Guttorp P, and Särkkä A. Discovering early diabetic neuropathy from epidermal nerve fiber patterns. *Statistics in Medicine* **35**(24) (2016), 4427-4442.
- Myllymäki M, Särkkä A, and Vehtari A. Hierarchical second-order analysis of replicated spatial point patterns with non-spatial covariates. *Spatial Statistics* **8** (2014), 104-121.
- Olsbo V, Myllymäki M, Waller LA, and Särkkä A. Development and evaluation of spatial point process models for epidermal nerve fibers. *Mathematical Biosciences* **243** (2013), 178-189.
- Rajala T, Särkkä A, Redenbach C, and Sormani M. Estimating geometric anisotropy in spatial point patterns. *Spatial Statistics* 15 (2016) 100-114.
- Redenbach C, Särkkä A, Freitag J, and Schladitz K. Anisotropy analysis of pressed point processes. *Advances in Statistical Analysis* 93(3) (2009) 237-261.
- Waller LA, Särkkä A, Olsbo V, Myllymäki M, Panoutsopoulou IG, Kennedy WR, and Wendelschafer-Crabb G. Second-order spatial analysis of epidermal nerve fibres. *Statistics in Medicine* **30**(23) (2011) 2827-2841.