

# Lecture 12: Point processes

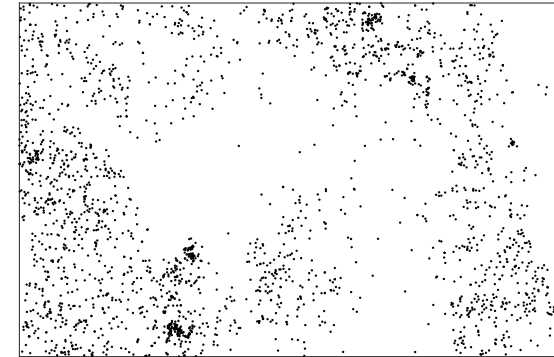
## Spatial Statistics and Image Analysis

David Bolin  
University of Gothenburg

Gothenburg  
May 15, 2019



### Example: Data



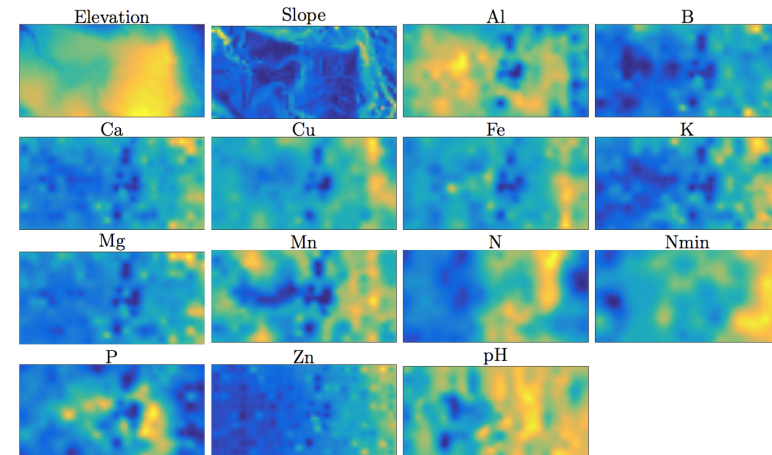
The locations of the tree species *Beilschmiedia Pendula* in the tropical rainforest plot on Barro Colorado Island.

### Types of spatial data

Three main types of data in spatial statistics

- Continuously indexed data.
  - We have observations at some fixed locations  $s_1, \dots, s_N$  of a random field  $X(\mathbf{s})$  where  $\mathbf{s}$  is in some region  $D \subseteq \mathbb{R}^2$ .  $X$  is a random field defined on  $D$ .
- Discretely indexed data.
  - We have observations at some fixed locations  $s_1, \dots, s_N$  of a random field  $X(\mathbf{s})$  where  $\mathbf{s}$  is in some discrete set of locations  $\tilde{D}$ , such as a regular lattice.  $X$  is a random field defined on  $\tilde{D}$ .
- Point process data.
  - We have observations  $s_1, \dots, s_N$  indicating where something occurred. We want to draw conclusions about the process based on these locations, which are now considered to be random.

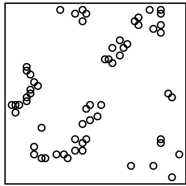
### Example: Covariates



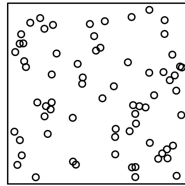
Possible covariates that can be used for drawing conclusions on the association of habitat preferences.

## Different types of point patterns

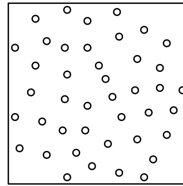
clustered



completely random

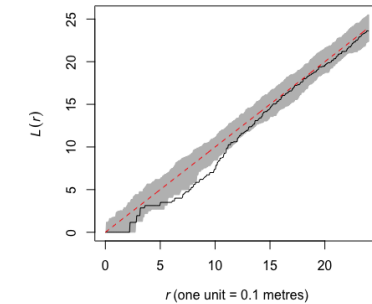
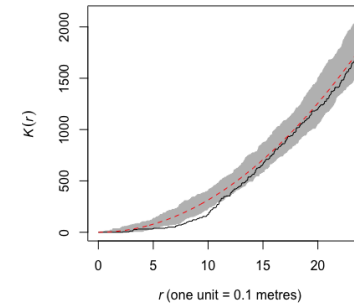
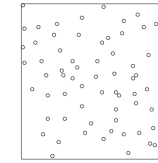


regular

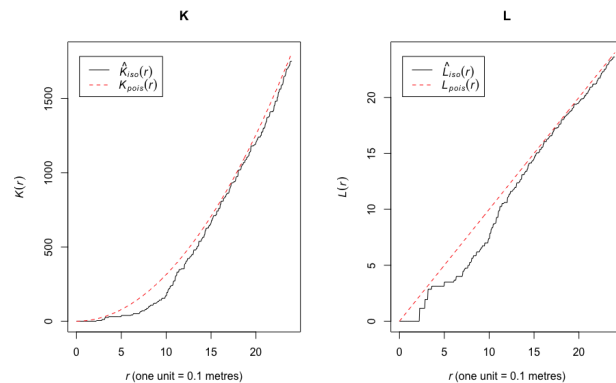
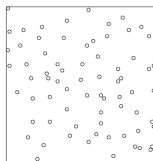


- One of the main questions for point pattern data is usually to determine if we have clustering or repulsion.
- The completely random case corresponds to the Poisson process.
- One typically use the K function to decide if we have clustering or repulsion.

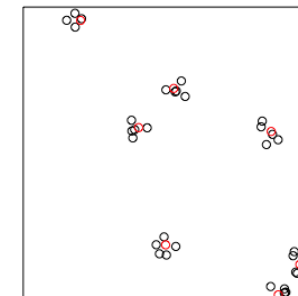
## Example: Swedish pine saplings



## Example: Swedish pine saplings



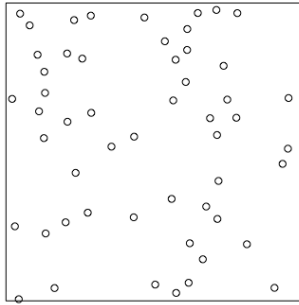
## Neyman-Scott process



## Hierarchical construction

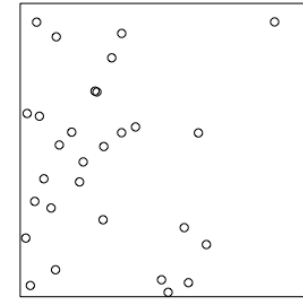
- Simulate a “mother process”  $Z$  as a Poisson process.
- For each point  $z_i \in Z$ , simulate  $n_i$  independent “daughter points”  $x_j$  from a distribution  $\pi(x|z_i)$  and remove  $Z$ .
- For example: Let  $n_i = 5$  and take  $x|z_i$  be uniform in a disc of radius 0.05 centered in  $z_i$ .
- Extension: take  $n_i$  from some distribution  $\pi(n)$ .

## Matérn Type 1 inhibition process



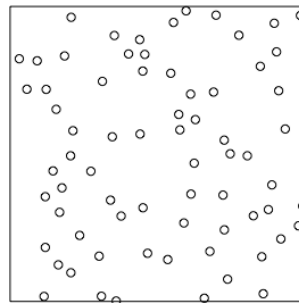
- Simulate a homogeneous Poisson process  $Z$ .
- Delete any point in  $Z$  that lies closer than a distance  $r$  from the nearest other point.
- Thus, pairs of close neighbours annihilate each other.
- $r$  is called the hard core distance.

## Inhomogeneous Poisson process



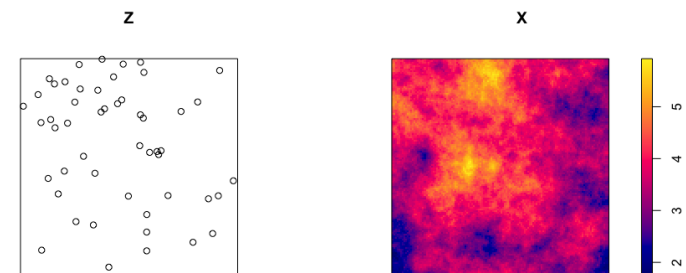
- A poisson process with a spatially varying intensity function.
- Example  $\lambda(x, y) = 100 \exp(-3x)$ .

## Matérn Type 2 inhibition process



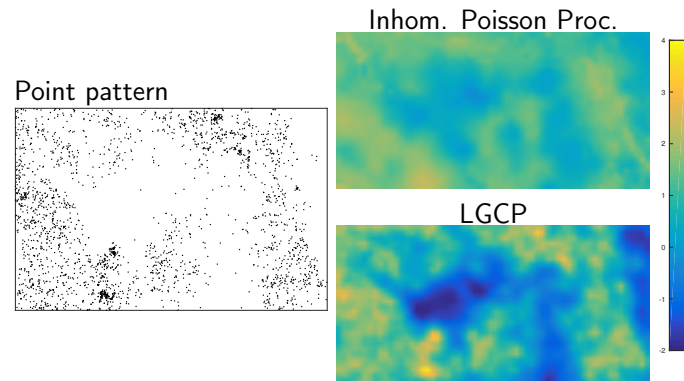
- Simulate a homogeneous Poisson process  $Z$ .
- Mark each point in  $Z$  by "ages", which are independent and uniformly distributed numbers in  $[0, 1]$ .
- Delete any point in  $Z$  that lies closer than a distance  $r$  from another point that has a higher age.
- This process was originally used to model forests.

## Log-Gaussian Cox process



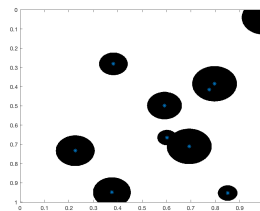
- Hierarchical model, where  $X$  is a Gaussian random field and  $Z|X$  is an inhomogeneous Poisson process where  $\lambda(x, y) = \exp(X(x, y))$ .
- Example:  $X$  is a Gaussian random field with mean 3 and an exponential covariance function.

## Example: Barro Colorado Island



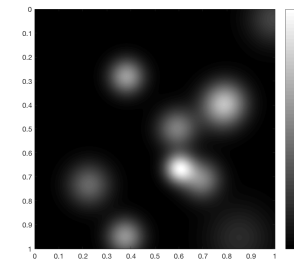
- For the inhomogeneous Poisson process, the intensity function is a regression on the soil covariates.
- For the LGCP the intensity function is the regression plus a mean-zero Gaussian random field.

## Marked point processes and boolean models



- For each point  $z_i$  in the point pattern, assign a random “mark”  $M_i$ .
- Can be used to include more information in spatial point pattern data (such as tree size or age).
- Models like this can also be used to build more complicated spatial models, such as Boolean models:
  - For each  $z_i$ , define a disc  $D_i$  centered at the point, with a random radius  $r_i$ .
  - Define a binary image by letting the pixels have value one in the region  $\cup_i D_i$ , and zero elsewhere.

## Models for grayscale images based on point processes



- We can extend the boolean model to a model for grayscale images by replacing the discs with kernel functions.
- Example: Let the value at location  $\mathbf{s}$  in the image be

$$I(\mathbf{s}) = \sum_i \pi_G(\mathbf{s}; z_i, r_i \mathbf{I})$$

where  $r_i$  are some positive random variables.

## Project report

- Parts 1 and 2 should be documented as lab reports, describing what you did and the results you obtained but it does not need a detailed introduction and discussion.
- The ideal form of the report for Part 3 is in principle a journal paper, containing:
  - 1 Project title, author names, course name, date of report.
  - 2 Abstract/Summary: about 10-15 lines
  - 3 Introduction: Statement of problem, earlier work with references.
  - 4 Data description and source.
  - 5 Methods: Mathematical, statistical, computational, image analysis.
  - 6 Results with tables and figures.
  - 7 Discussion: include if possible here also comparison with results from literature.
  - 8 Conclusions, suggested continued studies.
  - 9 References.

## Deadlines

For Parts 1 and 2:

- A PDF containing these two parts should be submitted via PINGPONG at the latest May 17, 23:55. Include the code as a zip-file.

For Part 3:

- A PDF containing a preliminary version of the report should be submitted together with Parts 1 and 2 at the latest May 17, 23:55. It does not need to be complete.
- The final version of the report should be submitted at the latest May 31, 23:55, together with the code as a zip-file, as a revision of the submitted project in PINGPONG.

## Project seminars

- The lectures and computer exercises next week will be devoted to project seminars.
- In the seminar there will be 7 minutes allotted to each group participant, where you should describe what you have done so far and what you plan to do.
- After this there will be 2 minutes left for discussion for each group.
- The object of the seminar is that you should get feedback from the audience, both to point out parts that are less clear and suggestions on what to do, the goal being to help you to write as good a report as possible.

## Preliminary schedule

### Monday 10:00-12:00

- G14 1 persons, 9 minutes
- G5 2 persons, 16 minutes
- G12 2 persons, 16 minutes
- G20 3 persons, 23 minutes
- G2 1 person, 9 minutes
- G25 1 person, 9 minutes
- G11 1 persons, 9 minutes

### Wednesday 10:00-12:00

- G24 3 persons, 23 minutes
- G9 1 persons, 9 minutes
- G10 3 persons, 23 minutes
- G16 1 persons, 9 minutes
- G3 3 persons, 23 minutes

### Monday 13:15-15:00

- G15 1 person, 9 minutes
- G23 1 person, 9 minutes
- G13 3 persons, 23 minutes
- G6 2 person, 16 minutes
- G4 2 persons, 16 minutes
- G7 1 persons, 9 minutes
- G17 1 persons, 9 minutes

### Wednesday 13:15-15:00

- G8 1 persons, 9 minutes
- G22 2 persons, 16 minutes
- G18 1 persons, 9 minutes
- G1 3 persons, 23 minutes
- G21 2 persons, 16 minutes
- G19 2 persons, 16 minutes

## Exam

- Takes place on June 5: 8:30 - 12:30.
- The exam can give 20 points, and the project reports another 20 points.
- The final grade is based on the sum of the points from the exam and the project.
- You need to have at least 8 points on the exam, as well as 8 points on the project to pass the course.
- The only aid you may bring to the exam is a "Chalmers allowed" calculator (will likely not be needed).