Random patterns and structures in spatial data

Radu Stoica
Université de Lorraine - IUT Charlemagne,
Institut Elie Cartan de Lorraine, Inria - Pasta, RING

radu-stefan.stoica@univ-lorraine.fr

Haapsalu, 2025

Table of contents

Course 1. Introduction: course organisation, data sets and examples

About me

- professor at Université de Lorraine
- mail : radu-stefan.stoica@univ-lorraine.fr
- web page : https://sites.google.com/site/radustefanstoica/
- IECL office : 120
- phone : + 33 6 20 06 29 30

Bibliographical references

• Baddeley, A. J., Rubak, E., Turner, R. (2016). Spatial point patterns. Methodology and applications with R. Chapman and Hall.

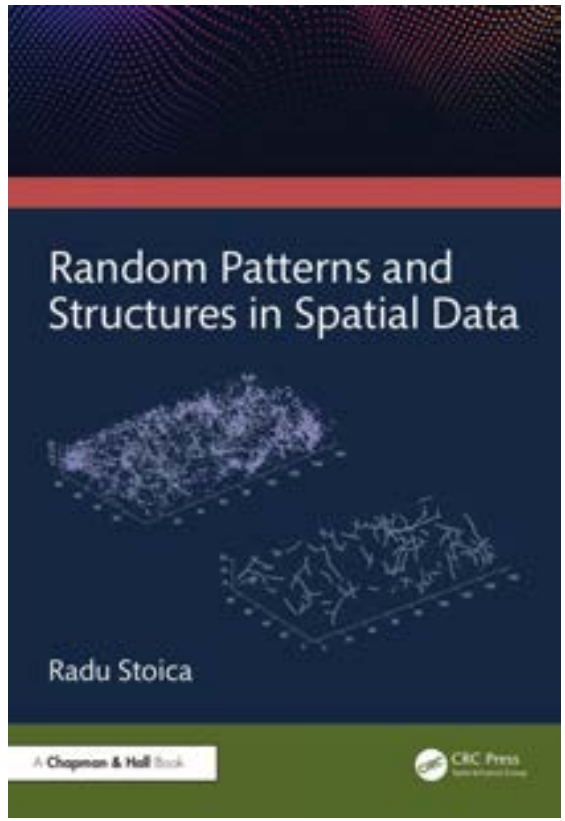
• Chiu, S. N., Stoyan, D., Kendall, W. S., Mecke, J. (2013). Stochastic geometry and its applications. John Wiley and Sons.

• Daley, D. J. and Vere Jones, D. (2003, 2008). An Introduction to the Theory of Point Processes, Volume 1: Elementary Theory and Methods, Volume 2: General Theory and Structure. Springer (2nd edition).

• Diggle, P. J. (2014). Statistical Analysis of Spatial and Spatio-Temporal Point Patterns (3rd edition). Chapman and Hall.

• van Lieshout, M. N. M. (2000). Markov point processes and their applications. Imperial College Press.
• van Lieshout, M. N. M. (2019). Theory of spatial statistics. A concise introduction. Chapman and Hall, CRC Press.
• Martinez, V. J., Saar, E. (2002). Statistics of the galaxy distribution. Chapman and Hall.
• Møller, J., Waagepetersen, R. P. (2004). Statistical inference and simulation for spatial point processes. Chapman and Hall.
• Winkler, G. (2003). Image Analysis, Random Fields and Markov Chain Monte Carlo Methods A Mathematical Introduction. Springer Verlag (2nd edition).

• Stoica, R. S. (2025). Random patterns and structures in spatial data. Chapman and Hall.



available online at `http://www.routledge.com/`

Acknowledgements

K. Adamczyk, D. Astaburuaga, L. Ben Allal, A. J. Baddeley, P. Bermolen, L. Beznea, M. Bodea, F. Bonneau, G. Castellan, G. Caumon, M. Clausel, J.-F. Coeurjolly, F. Cuevas-Pacheco, T. Danelian, N. Dante, Yu. Davydov, M. Deaconu, X. Descombes, N. Emelyanov, A. Fratani, M. Fouchard, E. Gay, D. Gemmerle, N. Gillot, J. Giraud, P. Gregori, C. Gouache, P. Heinrich, D. Hestroffer, L. Hurtado, K. Kieu, R. Kipper, F. Kleinschroth, I. Kovalenko, A. Kretzschmar, M. Kruuse, C. Lantuéjoul, Q. Laporte-Chabasse, A. Lejay, N. Libeskind, M. N. M. van Liehsout, S. Liu, L. J. Liivamägi, O. Lupascu, R. Marchand, V. Martinez, J. Mateu, J. Møller, H. Monod, E. Mordecki, F. Mortier, Z. Pawlas, A. Philippe, C. Reype, A. Richard, F. Rihmi, E. Saar, A. Särkkä, J. Sorce, D. Stoyan, F. Taty, E. Tempel, C. V. Tran, G. B. Valsecchi, A. Vienne, R. van de Weijgaert, R. Waagepetersen, J. Zerubia

and the list is far from being complete ...

Examples: data sets, application domains and related questions

Forestry data (1): the points positions exhibit attraction $\rightarrow$ clustered distribution

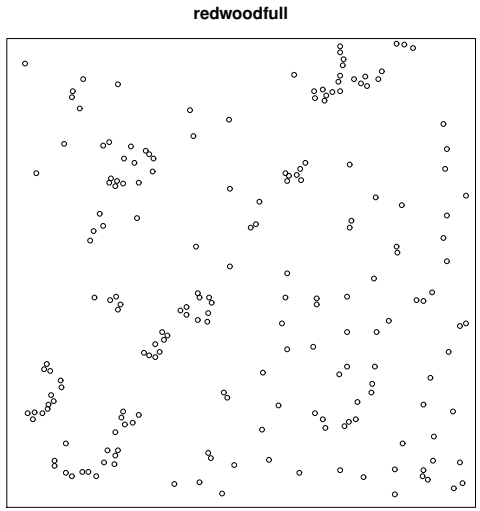


Figure: Redwoodfull data from the **spatstat** package

```
> library(spatstat)
> data(redwoodfull) ; plot(redwoodfull)
```

Forestry data (2): the points positions exhibit neither attraction nor repulsion $\rightarrow$ completely random distribution

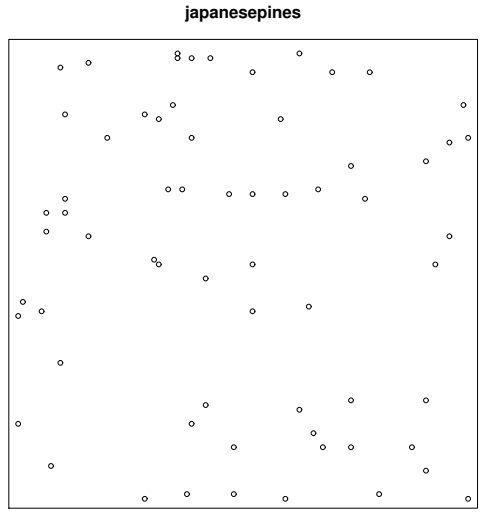


Figure: Japanese data from the **spatstat** package

In order to see all the available data sets

```
> data(package="spatstat")
```

Biological data (1): the points positions exhibit repulsion $\rightarrow$ regular distribution

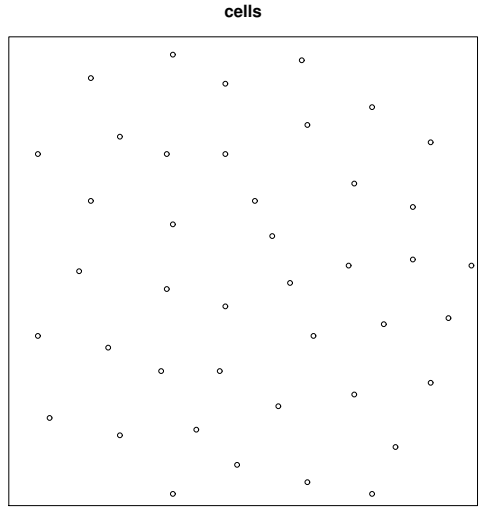


Figure: Cell data from the **spatstat** package

```
> data(cells)
> cells
planar point pattern:  42 points
window:  rectangle = [0, 1] x [0, 1] units
```

Biological data (2): two types of cells exhibiting attraction and repulsion depending on their relative positions and types

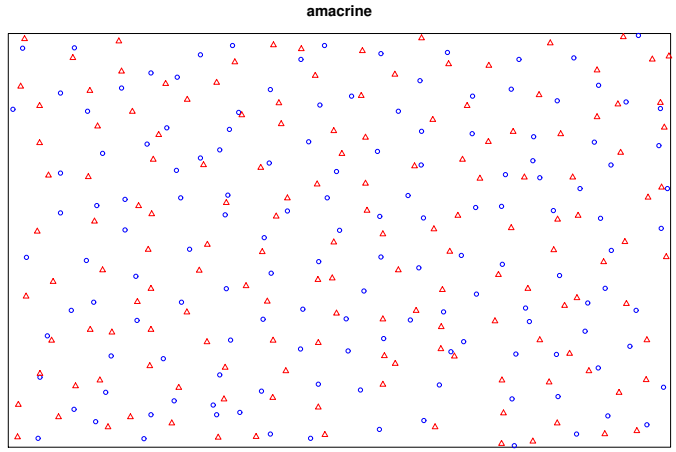


Figure: Amacrine data from the **spatstat** package

```
> data(amacrine) ;
plot(amacrine,cols=c("blue","red"))
```

Geological data: two types of patterns, line segments and points
$\rightarrow$ are these patterns independent ?

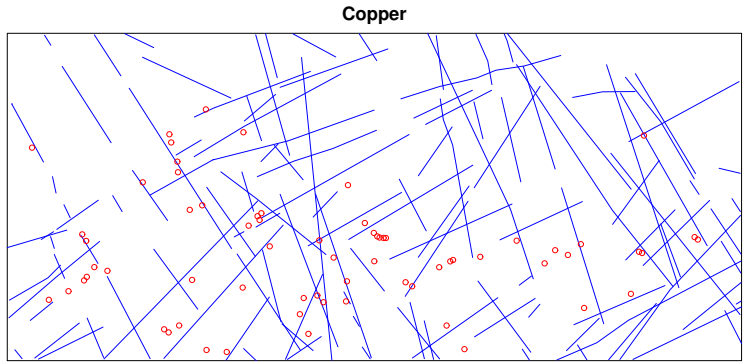


Figure: Copper data from the **spatstat** package

```
> attach(copper) ; L=rotate(Lines,pi/2) ;
P=rotate(Points,pi/2)
> plot(L,main="Copper",col="blue") ;
points(P$x,P$y,col="red")
```

Animal epidemiology: sub-clinical mastitis for diary herds

- points $\rightarrow$ farms location
- to each farm $\rightarrow$ disease score (continuous variable)
- clusters pattern detection : regions where there is a lack of hygiene or rigour in farm management

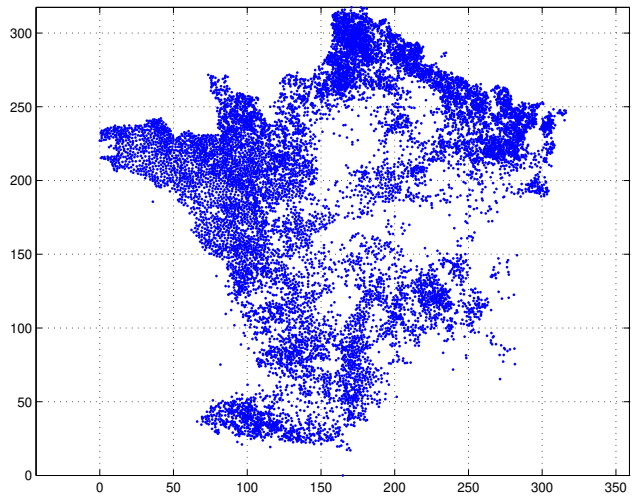


Figure: The spatial distribution of the farms outlines almost the entire French territory (INRA Avignon).

Cluster pattern: some comments

- particularity of the disease: can spread from animal to animal but not from farm to farm
- cluster pattern: several groups (regions) of points that are close together and have the "same statistical properties"
- clusters regions $\rightarrow$ approximate it using interacting small regions (random disks)
- local properties of the cluster pattern: small regions where locally there are a lot of farms with a high disease score value
- problem: pre-visualisation is difficult ...

(Stoica, Gay, Kretzschmar, 07)

T-tesselations: examples of natural spatial patterns

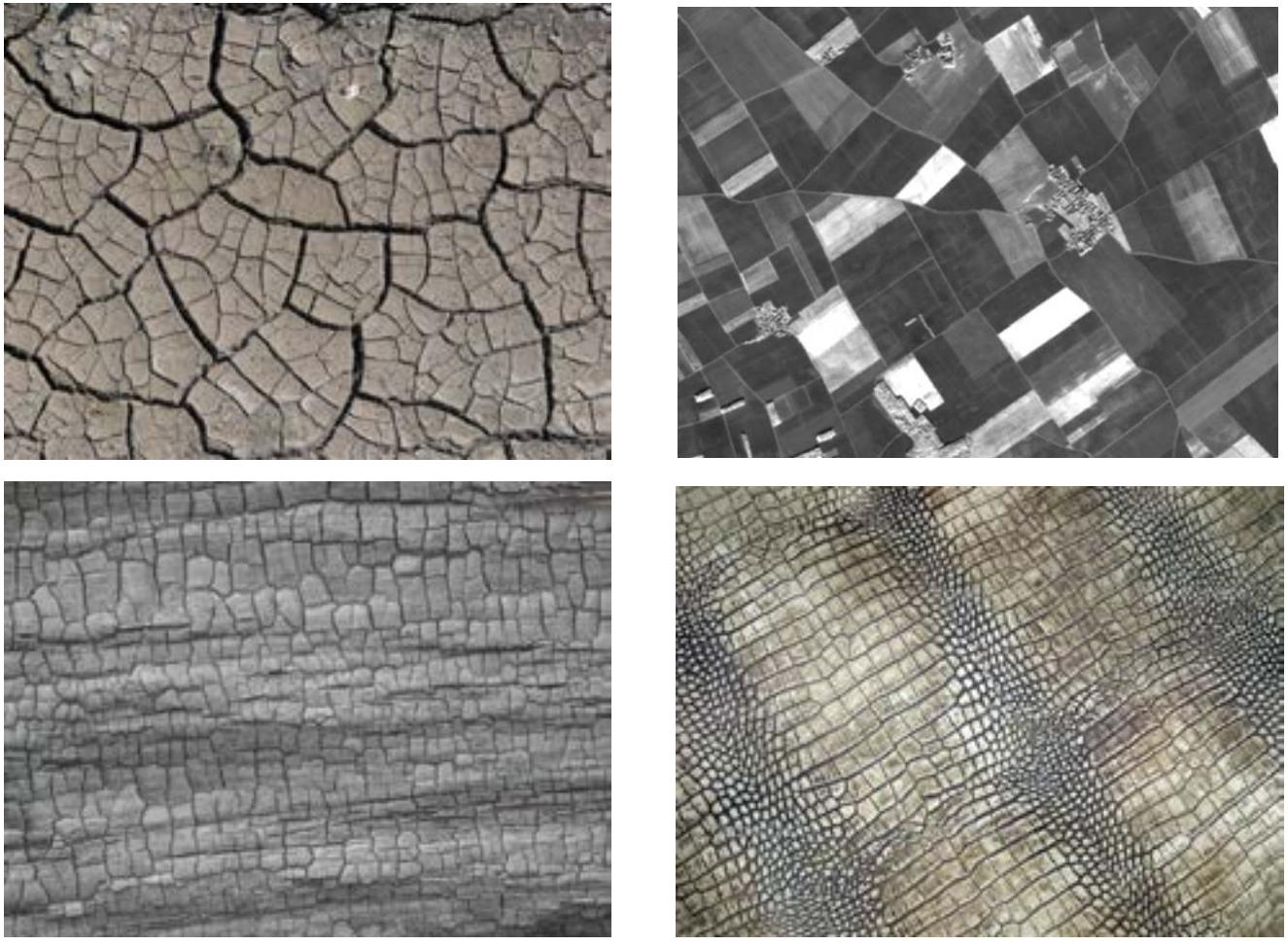


Figure: From upper left panel to lower right one: a cracked soil, a mosaic of agricultural fields, a texture of burnt wood, a fragment of alligator skin.

T-tesselation : agriculture landscape modeling

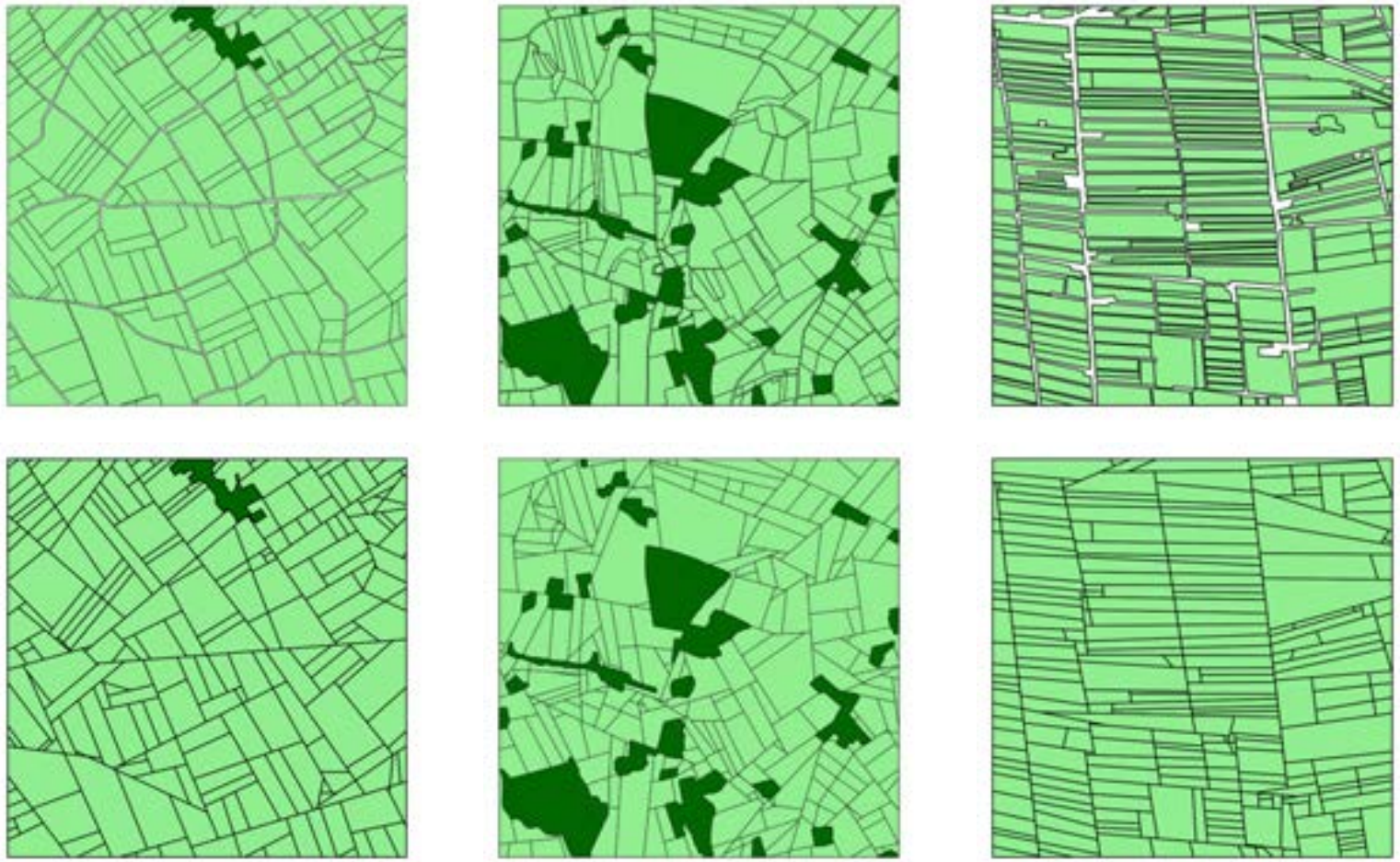


Figure: Upper row: landscapes of Selommes (a); Kervidy (b); and BVD (c). Light green polygons depict agricultural plots, dark green polygons stand for non-agricultural areas (mainly woods and villages). Lower row: landscapes approximated by T-tessellations.

Few words related to landscape modeling

- motivation: study polen dynamics
- methodology: probabilistic modeling, stochastic simulation and statistical inference
- key points: reference measure, convergence of the simulation dynamics, likelihood description
- posterior based inference: allow landscape classification

(Admaczyk et al., Stoica, 24)

Image analysis: road and hydrographic networks

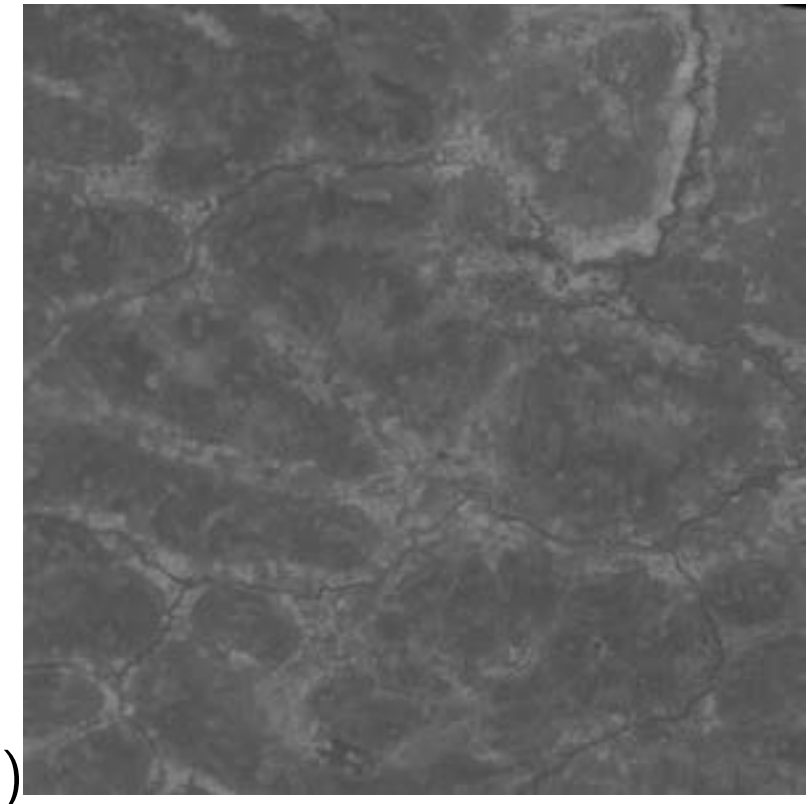
a)  b) 

Figure: a) Rural region in Malaysia (http://southport.jpl.nasa.gov), b) Forest galleries (BRGM).

Thin networks: some comments

- road and hydrographic networks $\rightarrow$ approximate it by connected random segments
- topologies: roads are "straight" while rivers are "curved"
- texture: locally homogeneous, different from its right and its left with respect a local orientation
- avoid false alarms: small fields, buildings,etc.
- local properties of the network: connected segments covered by a homogeneous texture

(Stoica et al., 02, 04)

Cosmology (1): spatial distribution of galactic filaments

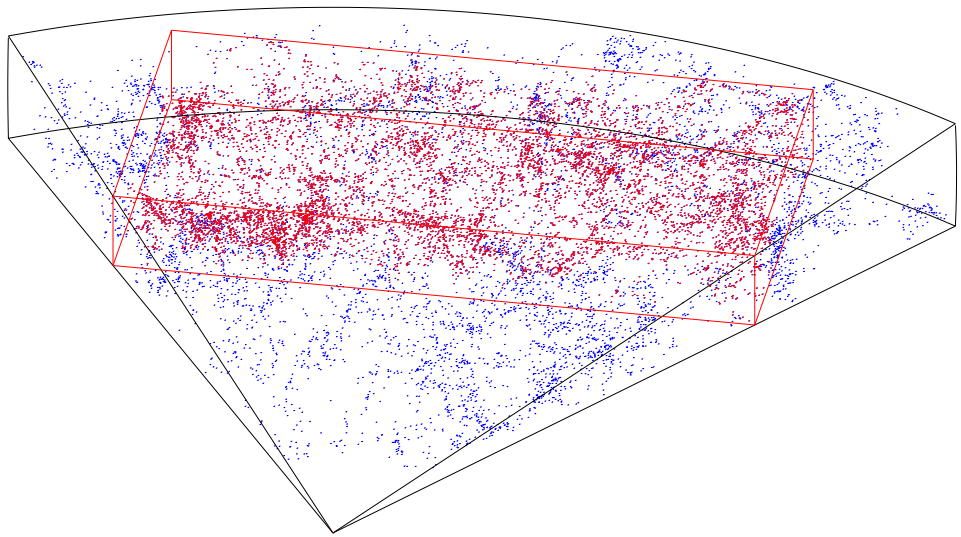
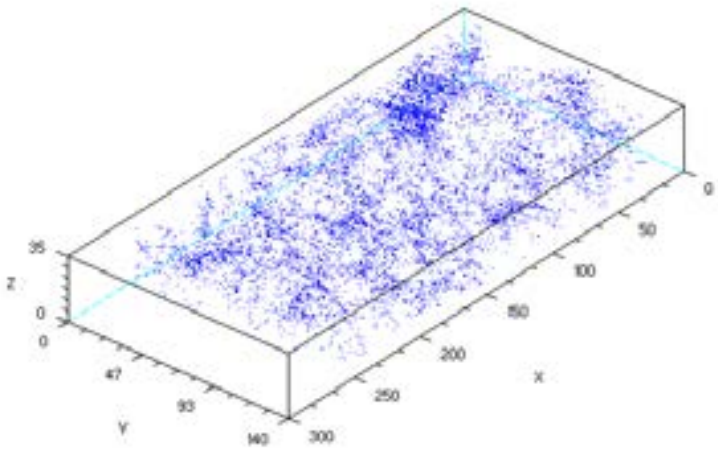
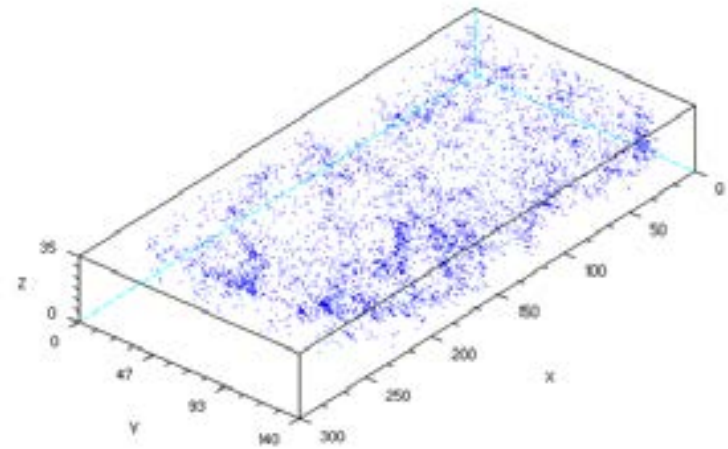


Figure: Cuboidal sample from the North Galactic Cap of the 2dF Galaxy Redshift Survey. Diameter of a galaxy $\sim 30 \times 3261.6$ light years.

Cosmology (2): study of mock catalogs



a)



b)

Figure: Galaxy distribution : a) Homogeneous region from the 2dfN catalog, b) A mock catalogue within the same volume

Cosmology (3): questions and observations

Few words about the 2dF GRS and SDSS catalogues

- filaments, walls and clusters : different size and contrast
- inhomogeneity effects (only the brightest galaxies are observed)
- filamentary network the most relevant feature
- local properties of the filamentary network : connecting random cylinders containing a "lot" of galaxies "along" its main axis

Mock catalogues

- how "filamentary" they are w.r.t the real observation ?
- how the theoretical models producing the synthetic data fits the reality ?

(Stoica, Martinez and Saar, 07,10)

Cosmology (4): cluster detection

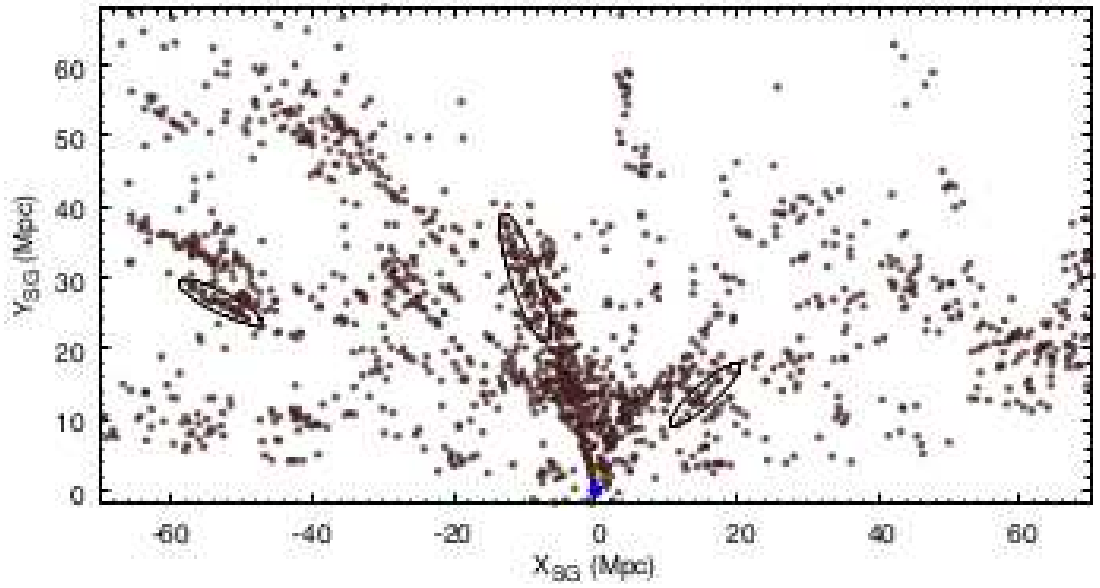


Figure: Distribution of galaxies in the 2MRS data set. Positions of galaxies are given in supergalactic coordinates, where observer is located at the origin of coordinates (marked as blue point on the figure). The thickness of the slice shown in the figure is 15 Mpc. Some galaxy clusters are marked with black ellipses to highlight the elongation of galaxy groups/clusters along the line of sight.

Cosmology (5): questions and observations

Few words about the 2MRS data set:

- more galaxies are observed
- price to pay : lack of precision for the third coordinates
- consequence : finger-of-God effect is much more important
- galaxy groups and clusters seem elongated along the line of sight
- inhomogeneity effects

(Tempel et al., Stoica, 18)

Cosmology (6): bias effects minimisation in cosmological data

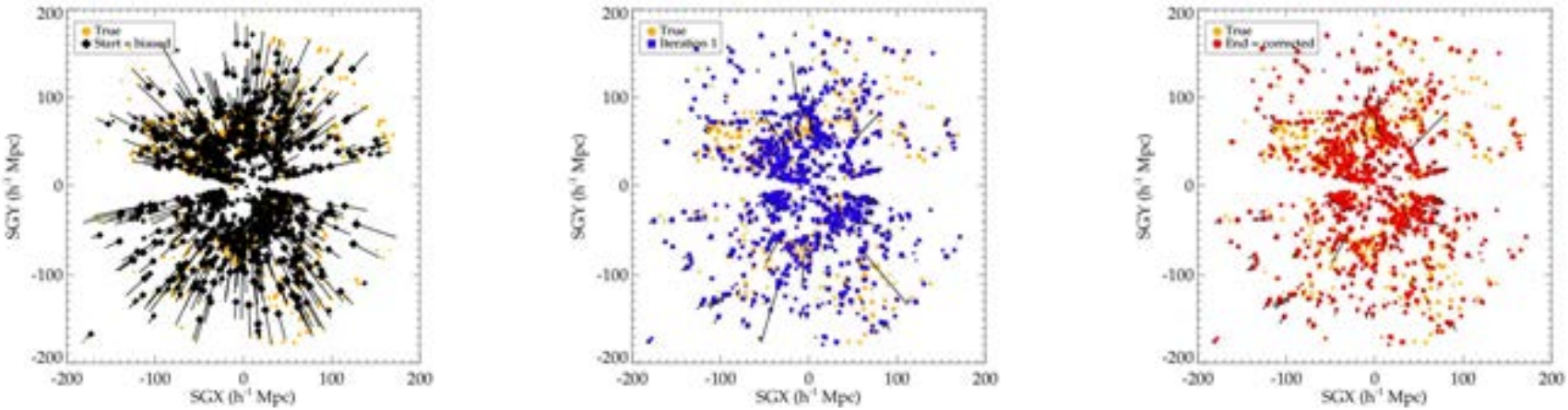


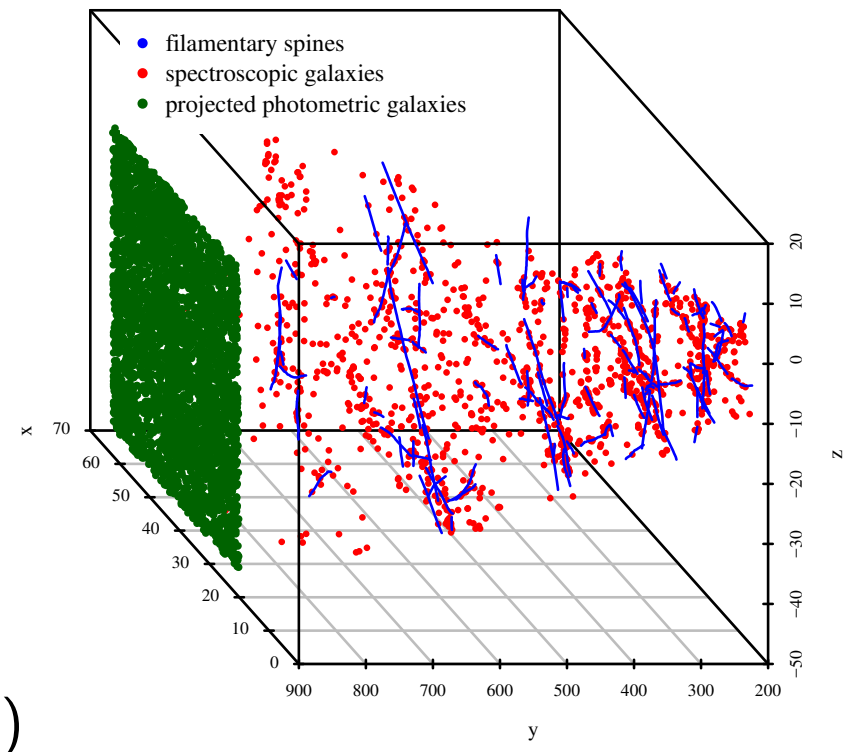
Figure: Results of the proposed method obtained on a mock catalogue (yellow). From top to bottom: initial perturbed configuration of galaxies (black), intermediate configuration after running a few steps of the MH algorithm (blue), final configuration after running the SA algorithm (red).

Some remarks

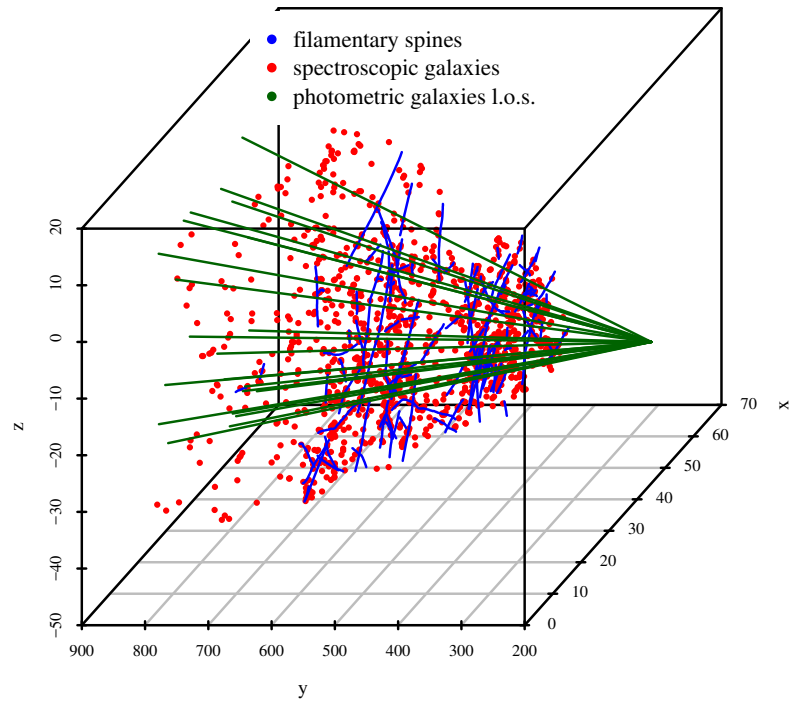
- in the previous section elongated structures appeared due to uncertainties that produce the "finger-of-God" effect
- peculiar velocity measurements are affected by several types of bias known together under the generic name of "Malmquist bias"
- propose model that "reverse" these effects and correct the data

(Sorce, Stoica, Tempel, 23)

Cosmology (7): influence of the new observations on the already detected structures



a)



b)

Figure: Photometric galaxies(green dots), spectroscopical galaxies (red dots) and filaments (blue): a) photometrical galaxies projected on a sphere, b) photometrical galaxies lines of sight

Few words about the SDSS Data Release 12 data set:

- much more galaxies are observed
- price to pay: bigger lack of precision for the third coordinates
- question: how this new data set is related to the already existing detected structures
- need to compare and to relate complex patterns in 2d and 3d

(Kruuse, Tempel, Kipper and Stoica, 19)

Cosmology (8): knowing the filamentary structure what is the galaxy distribution ?

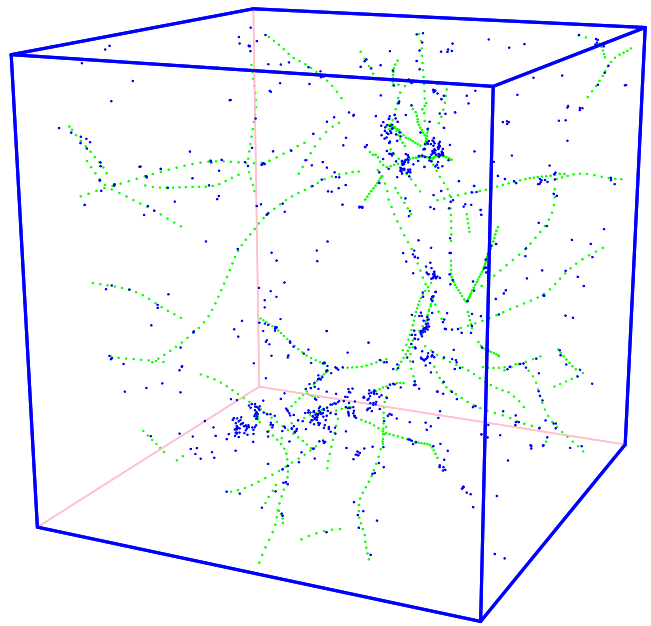


Figure: Data set: galaxy positions with spines (filaments main axes)

Some questions related to galaxy distribution

- galaxies are spread as pearls on a necklace: (Tempel et al., 2014)
- inhomogeneity: filaments presence or interactions ?
- which of these factors controls the galaxy distribution ?
- what type of interaction: gravitational, territorial, component oriented ?
- what is the interaction range ?
- what is the size of a cluster ?

(Hurtado, Stoica and al., 21)

Network science

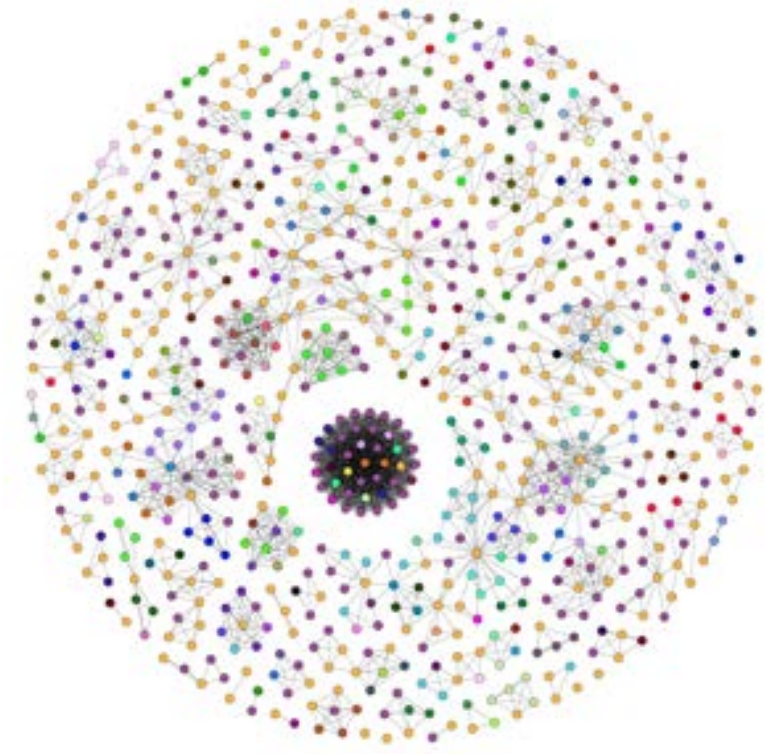


Figure: Collaborations among researchers within the Loria laboratory : HAL data set (2018).

Description of the network:

- each node represents a researcher,
- the edges are collaborative links
- nodes' color represent the affiliation to a laboratory.
- all Loria members are coloured in yellow, while the members of the other labs are differently coloured

Related questions:

- what determines the occurrence of a collaborative link ?
- how the presence of cooperating individuals can be characterised ?
- how to quantify the cooperation behaviour of a research team ?

(Laporte, Stoica et al., 22)

Geosciences: an interaction point process for Bayesian inference of multiple water sources chemical composition from hydrochemical data (1)

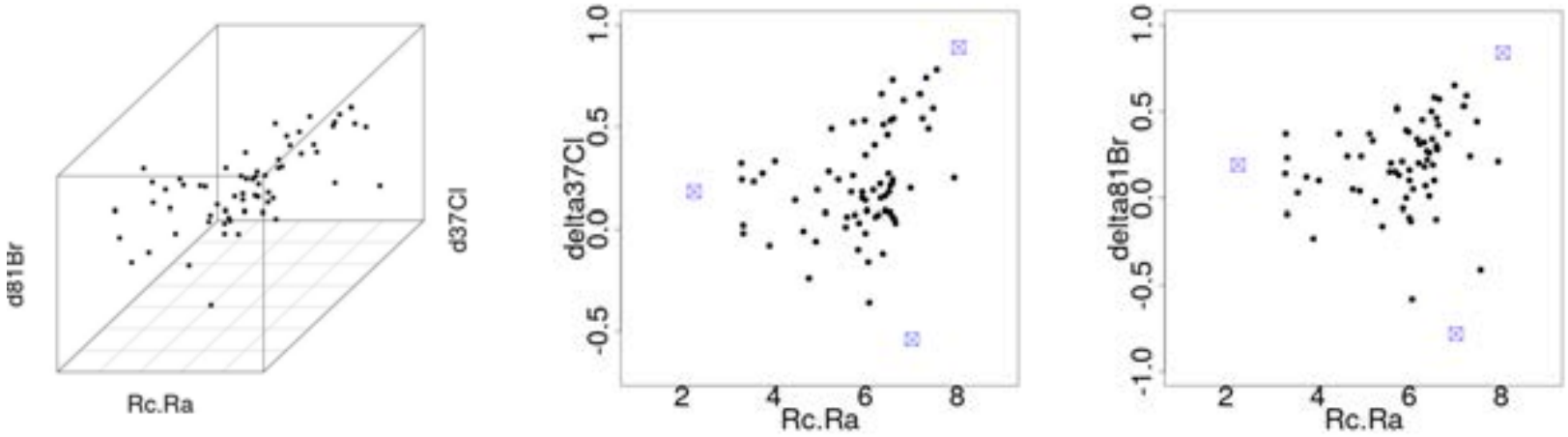


Figure: Data: deep geothermal fluids presented by Pinti et al. (2020). Detected sources via the proposed HUG model, a point process with interactions.

Some hypotheses and warnings:

- no chemical reaction
- the chemical composition of the sources is the same for all samples
- position in the data space $\Leftrightarrow$ composition
- distance in the data space $\Leftrightarrow$ difference in composition

Challlenges:

- multi-dimensionality
- recover the sources structures whenever working from projections
- data uncertainities
- take into account all the available data
- completeley unsupervised methods

(Reype, Stoica et al., 25)

Geosciences: stochastic seismic interpretation through marked point processes with interactions (2)

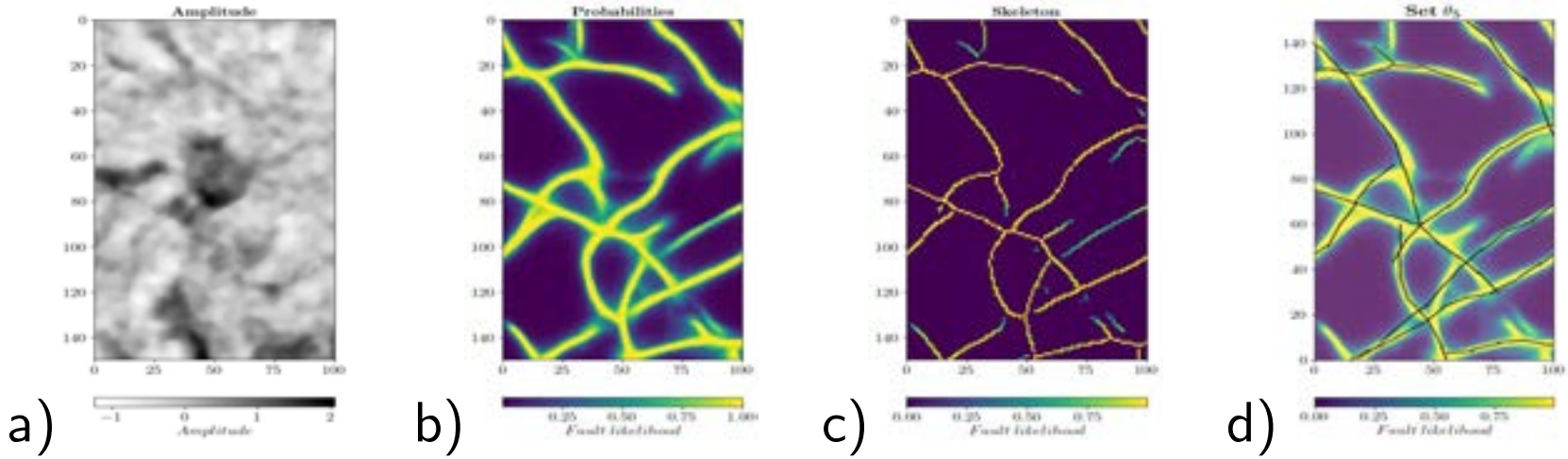


Figure: Seismic faults extraction: original data (a), likelihood computed using convolutional neural networks (b), faults skeleton or thinned likelihood of the faults (c), faults intepretation using the Candy model (random segments that connect and allign) (d)

(Taty, Stoica et al., 24)

Geosciences: fracture network characterization using stochastic simulations of marked point process and Bayesian inference (3)

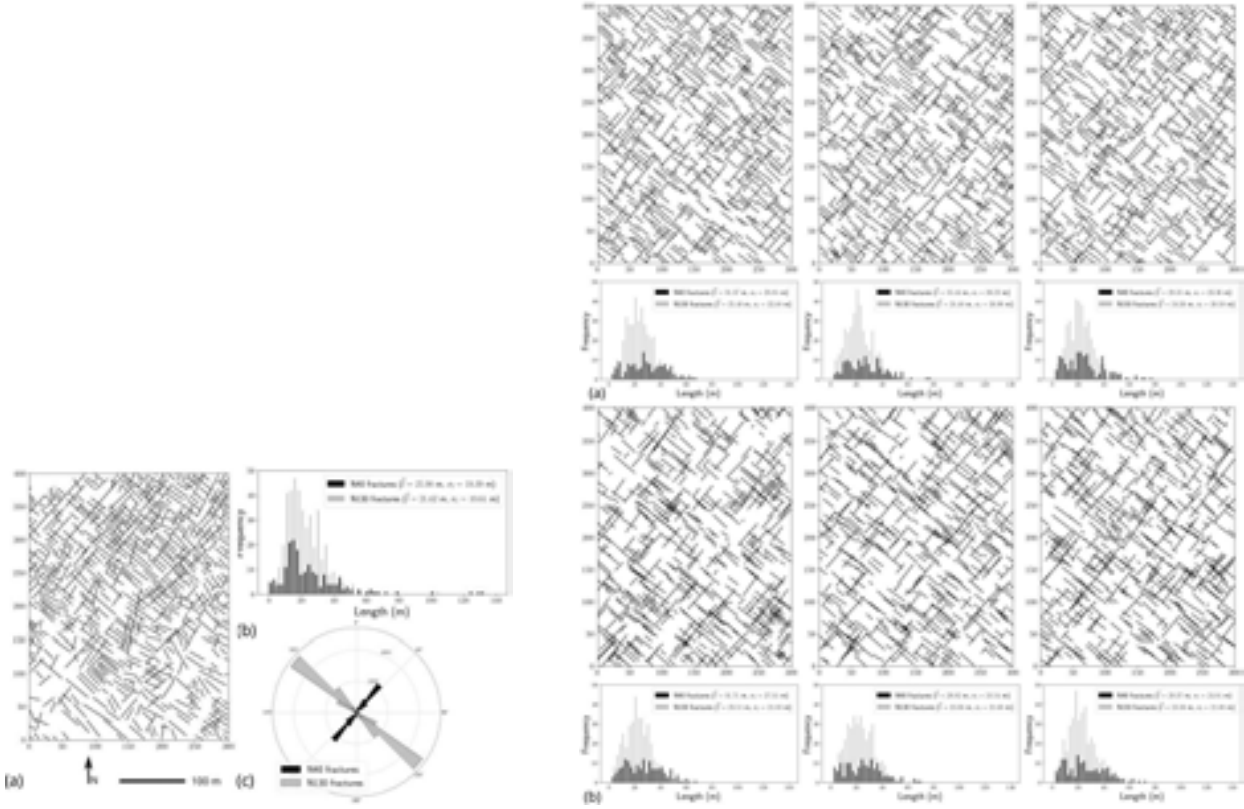


Figure: Left: fracture set with length and orientation distributions. Right: simulations using model parameters estimated from data.

(Bonneau, Stoica, Caumon, 25)

Spatio-temporal data

Time dimension available:

- the previous example may be considered snapshots
- more recent data sets have also a temporal coordinate
- question : what is the pattern hidden in the data and its spatio-temporal description ?

Roads dynamics in Central Africa region :

- in forest region with rare woods, road networks appear and disappear within the territory of an exploitation concession
- there is a difference between "classical" road networks and "exploitation" networks $\rightarrow$ mining galleries
- this roads dynamics may be relevant in many aspects : health of the forest, respect of rules for the enterprises, environmental behaviour and understanding
- characterize the distribution and the dynamics of the road network

(Kleinschroth, Stoica et al., 17)
$\rightarrow$ video roads dynamics

Study the spatio-temporal spread of failures in a water distribution network : failures (black points), detector's activation (red points)

- information available : position, activation date, alert type, etc.
- SEDIF data and questions : do the detectors work ? do the failures form a particular pattern ?
- how to integrate the temporal dynamics ?

(Dante, Stoica et al., 19)

Paleontology: Guérande salinas - fairy rings

- growing rings: territories occupied by cyano-bacterias
- the size of a ring is proportional with its age
- what determines the spatial distribution of the rings: edge effects, water arrival, interaction ?
- how to integrate the temporal dyanmics ?

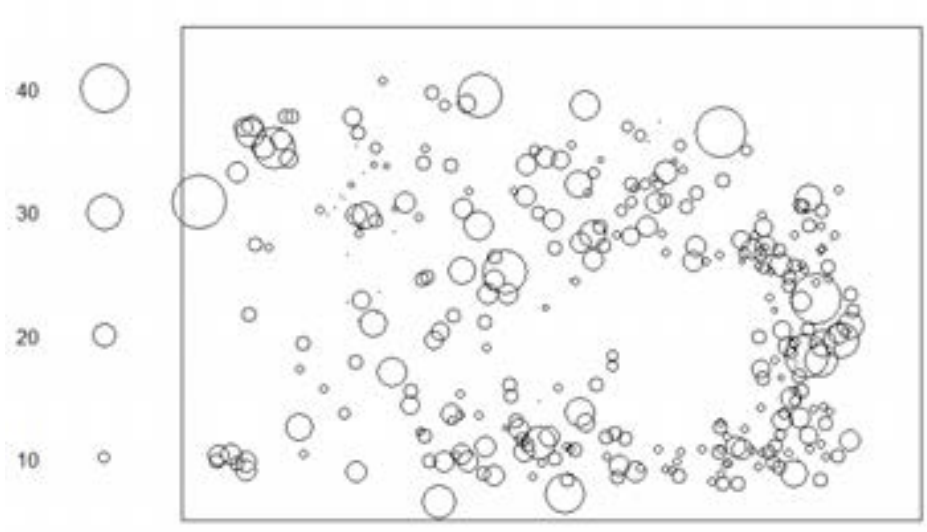


Figure: The distribution of the morphostructures and their sizes (diameter in cm)

(Astaburuaga, Stoica et al., 2024)

Sismology: characterisation of earthquakes occurences

- earthquakes: space-time events
- self-excitation phenomenon
- characterize and predict the sismic activity in different regions

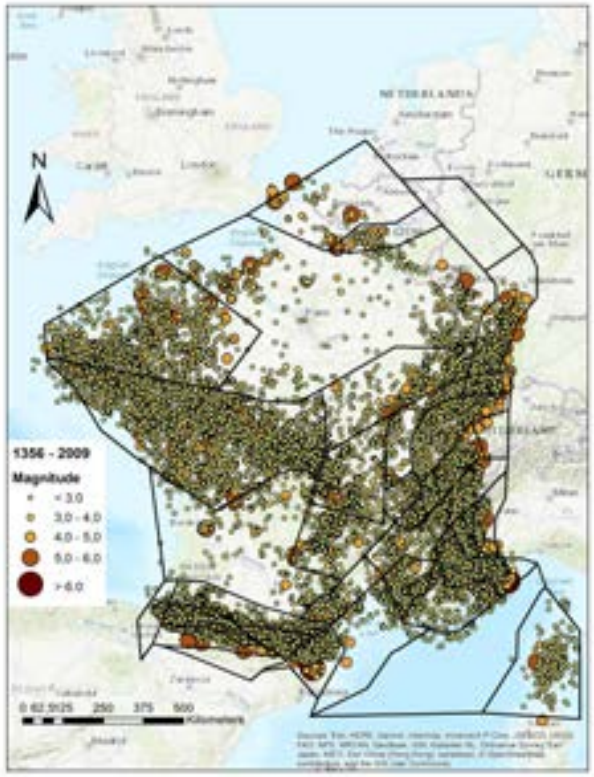


Figure: Representation of positions of earthquakes on the French territory

Synthesis

Hypothesis: the pattern to be detected or to be characterised is approximated by a configuration of random objects that interact

- marked points: repulsive or clustered objects (position + mark)
- clusters pattern: superposing random disks
- road network: connected and aligned segments
- cosmic filament network: connected and aligned cylinders

Important remark:

- the considered pattern exhibit a finite local number of objects $\rightarrow$ particular structure of the data and of the problems this approach is able to deal with

Marked point processes:

- probabilistic models for random points with random characteristics $\rightarrow$ interacting random objects
- origin $\rightarrow$ stochastic geometry
- the pattern is described by means of a probability density $\rightarrow$ stochastic modelling
- the probability density allows the computation of average quantities and descriptors (integrals) related to the pattern
- conversely, whenever a pattern is observed, the probabilistic framework allows the derivation of the law of parameters conditioned on the observation $\rightarrow$ Bayesian inference

Remarks:

- there exist also deterministic mathematical tools able to treat pattern recognition problems
- probability thinking framework offers simultaneously the analysis and the synthesis abilities of the proposed method
- probabilistic approach deeply linked with physics:
  - exploratory analysis
  - model formulation
  - simulation
  - statistical inference

- theoretical challenge: temporal dimension
  - spatio-temporal marked point processes
  - random sets theory
  - general idea: new data sets require new mathematics $\rightarrow$ stochastic processes and stochastic geometry
  - still, partial answers to these questions can be given using the tools presented in this course
- application challenge: big data
  - examples: cosmological data, geochemistry, particle tracking (quantum dots), etc.

For the purpose of this course: software and data sets are available

- R library : **spatstat** by A. Baddeley, R. Turner and contributors $\rightarrow$ www.spatstat.org
- C++ library : **DRLib** by D. Gemmerle, R. S. Stoica, C. Reype, N. Gillot and contributors $\rightarrow$ `https://gitlab.univ-lorraine.fr/labos/iecl/DRlib`

Mathematical background

Measure and integration theory $\rightarrow$ send notes

- $\sigma-$algebra
- measurable space, sets, functions
- measure
- measure space, integral with respect to a measure
- probability space, probability measure

Table of contents

Course 2. Point processes, Binomial and Poisson point processes

Construction of a point process : mathematical ingredients

- observation window : the measure space $(W, \mathcal{B}, \nu)$, with $W \subset \mathbb{R}^d$, $\mathcal{B}$ the Borel $\sigma-$algebra and $0 < \nu(W) < \infty$ the Lebesgue measure
- points configuration space : probability space $(\Omega, \mathcal{F}, \mathbb{P})$

Configuration space construction:

- state space Ω :

$$W_n \quad \text{is the set of all n-tuples} \quad \{w_1, \ldots, w_n\} \subset W$$
$$\Omega = \cup_{n=0}^{\infty} W_n, \quad n \in \mathbb{N}$$

- events space $\mathcal{F}$: the $\sigma-$ algebra given by

$$\mathcal{F} = \sigma(\{\mathrm{w} = \{w_1, \ldots, w_n\} \in \Omega : n(\mathrm{w}_B) = n(\mathrm{w} \cap B) = m\})$$

for any bounded $B \in \mathcal{B}$ and $m \in \mathbb{N}$

- probability measure $\mathbb{P}$: the model answering our questions

Definition

A point process in W is a measurable mapping from a probability space $(\mathcal{S}, \mathcal{A})$ in $(\Omega, \mathcal{F})$. Its distribution is given by

$$\mathbb{P}(X \in F) = \mathbb{P}\{\omega \in \mathcal{S} : X(\omega) \in F\},$$

with $F \in \mathcal{F}$. The realization of a point process is random set of points in W. We shall sometimes identify X and $\mathbb{P}(X \in F)$ and call them both a point process.

Remarks : point process $\Rightarrow$ random configuration of points w in a observation window W. In the following, it is considered that :

- a points configuration is $\mathrm{w} = \{w_1, w_2, \ldots, w_n\}$, with n the corresponding number of points
- the process is locally finite: $n(\mathrm{w} \cap B)$ is finite whenever $\nu(B)$ is finite
- the process is simple: $w_i \neq w_j$ for $i \neq j$

Marked point processes: attach characteristics to the points $\rightarrow$ extra-ingredient: marks probability space $(M, \mathcal{M}, \nu_M)$

Definition

A marked point process is a random sequence $\mathrm{x} = \{x_n = (w_n, m_n)\}$ *such that the points* w_n *are a point process in* W *and* m_n *are the marks corresponding for each* w_n*.*

Examples:

- random circles: $M = (0, \infty)$
- random segments: $M = (0, \infty) \times [0, \pi]$
- multi-type process: $M = \{1, 2, \ldots, k\}$

... and all the possible combinations ... $\rightarrow$ drawing

Stationarity and isotropy. A point process X on W is stationary if it has the same distribution as the translated proces X_w, that is

$$\{w_1, \ldots, w_n\} \stackrel{\mathcal{L}}{=} \{w_1 + w, \ldots, w_n + w\}$$

for any $w \in W$.
A point process X on W is isotropic if it has the same distribution as the rotated proces $\mathrm{r}X$, that is

$$\{w_1, \ldots, w_n\} \stackrel{\mathcal{L}}{=} \{\mathrm{r}w_1, \ldots, \mathrm{r}w_n\}$$

for any rotation matrix r.

- motion invariant: stationary and isotropic
- marked case: in principle easy to generalize, but take care ...
- counter example: a point process on a half plane is not stationary

Intuitive characterisation of a point process : being able to say how many points of the process X can be found in a neighbourhood in W.

Tools for point processes analysis: should be able to do the following

- count the points of a point process in a small neighbourhood of a point in W, and then extend the neighbourhood
- count the points of a point process in a small neighbourhood of a typical point of the process X, and then extend the neighbourhood
- "counting" means using a probability measure based counter

Let X be a point process on W, and let us consider the counting variable

$$N(B) = n(X_B), \quad B \in \mathcal{B},$$

representing the number of points "falling" in B.

Theorem

The distribution of a point process X on a complete, separable metric space (W, d) is determined by the finite dimensional distributions of its count function, i.e. the joint distribution of $N(B_1), \ldots, N(B_m)$ for any bounded $B_1, \ldots, B_m \in \mathcal{B}$ and $m \in \mathbb{N}$.

Theorem

The distribution of a simple point process on a complete, separable metric space (W, d) is uniquely determined by its void probabilities

$$v(B) = \mathbb{P}(N(B) = 0), \quad B \in \mathcal{B}.$$

Binomial point process

The trivial random pattern : a single random point x uniformly distributed in a compact W such that

$$\mathbb{P}(x \in B) = \frac{\nu(B)}{\nu(W)}$$

for all $B \in \mathcal{F}$.

More interesting point pattern : n independent points distributed uniformly such that

$$\begin{aligned} \mathbb{P}(x_1 \in B_1, \ldots, x_n \in B_n) = & \\ = & \ \mathbb{P}(x_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(x_n \in B_n) \\ = & \ \frac{\nu(B_1) \cdot \ldots \cdot \nu(B_n)}{\nu(W)^n} \end{aligned}$$

for Borel subsets $B_1, \ldots, B_n$ of the compact W.

$\rightarrow$ drawing

Properties

- this process earns its name from a distributional probability
- the r.v. $N(B)$ with $B \subseteq W$ follows a binomial distribution with parameters

$$n = N(W) = n(\mathrm{x}_W)$$

and

$$p = \frac{\nu(B)}{\nu(W)}$$

- the intensity of the binomial point process, or the mean number of points per unit volume

$$\rho = \frac{n}{\nu(W)}$$

- the mean number of points in the set B

$$\mathbb{E}(N(B)) = np = \rho\nu(B)$$

- the binomial point process is simple
- number of points in different subsets of W are not independent even if the subsets are disjoint

$$N(B) = m \Rightarrow N(W \setminus B) = n - m$$

- the distribution of the point process is characterized by the *finite dimensional distributions*

$$\mathbb{P}(N(B_1) = n_1, \ldots, N(B_k) = n_k) \quad \text{for} \quad k = 1, 2, \ldots$$

such that $n_1 + n_2 + \ldots + n_k \leq n$

- if the B_k are disjoint Borel sets with $B_1 \cup \ldots B_k = W$ and $n_1 + \ldots + n_k = n$, the finite-dimensional distributions are given by the multinomial probabilities

$$\begin{aligned} & \mathbb{P}(N(B_1) = n_1, \ldots, N(B_k) = n_k) \\ = & \quad \frac{n!}{n_1! \ldots n_k!} \frac{\nu(B_1)^{n_1} \ldots \nu(B_k)^{n_k}}{\nu(W)^n} \end{aligned}$$

- the void probabilities for the binomial point process are given by

$$v(B) = \mathbb{P}(N(B) = 0) = \frac{(\nu(W) - \nu(B))^n}{\nu(W)^n}$$

Stationary Poisson point process

Motivation : what happens if extend W towards $\mathbb{R}^d$?

- convergence binomial towards Poisson
- $\rightarrow$ drawing + blackboard

Definition : a stationary (homogeneous) Poisson point process X is characterized by the following fundamental properties

- Poisson distribution of points counts. The random number of points of X in a bounded Borel set B has a Poisson distribution with mean $\rho\nu(B)$ for some constant ρ, that is

$$\mathbb{P}(N(B) = m) = \frac{(\rho\nu(B))^m}{m!} \exp(-\rho\nu(B))$$

- Independent scattering. The number of points of X in k disjoint Borel sets form k independent random variables, for arbitrary k

Properties

- simplicity : no duplicate points
- the mean number of points in a Borel set B is

$$\mathbb{E}(N(B)) = \rho\nu(B)$$

- ρ : the intensity or density of the Poisson process, and it represents the mean number of points in a set of unit volume
- $0 < \rho < \infty$, since for $\rho = 0 \Rightarrow$ the process contains no points, while for $\rho = \infty$ we get a pathological case

- if $B_1, \ldots, B_k$ are disjoint Borel sets, then $N(B_1), \ldots, N(B_k)$ are independent Poisson variable with means $\rho\nu(B_1), \ldots, \rho\nu(B_k)$.Thus

$$\begin{aligned}
&\mathbb{P}(N(B_1) = n_1, \ldots, N(B_k) = n_k) \\
&= \frac{\rho^{n_1+\ldots+n_k}\nu(B_1)^{n_1}\ldots\nu(B_k)^{n_k}}{n_1!\cdot\ldots\cdot n_k!}\exp\left(-\sum_{i=1}^{k}\rho\nu(B_i)\right),
\end{aligned}$$

- this formula can be used to compute joint probabilities for overlapping sets
- the void probabilities for the Poisson point process are given by

$$v(B) = \mathbb{P}(N(B) = 0) = exp(-\rho(\nu(B)))$$

- the Poisson point process with $\rho = \text{ct.}$ is stationary and isotropic
- if the intensity is a function $\rho : W \to \mathbb{R}^+$ such that

$$\int_B \rho(w) d\nu(w) < \infty$$

for bounded subsets $B \subseteq W$, then we have a inhomogeneous Poisson process with mean

$$\mathbb{E}(N(B)) = \int_B \rho(w) d\nu(w) = \Upsilon(B)$$

- Υ is called the intensity measure
- we have already seen that for the stationary Poisson process : $\Upsilon(B) = \rho\nu(B)$

Theorem

Conditionning a Poisson point process. *Let X be a stationary Poisson point process on $\mathbb{R}^d$ with intensity $\rho > 0$ and W a bounded Borel set with $\nu(W) > 0$. Then, conditional on the event $\{N(W) = n\}$, X restricted to W is a binomial point process of n points.*

- proof of the theorem : $\rightarrow$ Exercice 1
- b: $\rightarrow$ Exercice 2, 3 and 4

Some properties of the Poisson point process

Theorem

Interval theorem. *Let X be a stationary point process on $(0, \infty)$ with intensity ρ and let the points of X be written in ascending order :*

$$0 < X_1 < X_2 < \ldots < X_n < \ldots \quad .$$

The the random variables :

$$Y_1 = X_1, Y_n = X_n - X_{n-1}$$

are independently, identically distributed according to $g(y) = \rho \exp(-\rho y)$ for $y > 0$.

- bus paradox: if to the initial process a symmetric independent copy on $(-\infty, 0)$ is added, then the interval between two consecutive points of the process containing 0 is longer than any other interval between two consecutive points
- no extension of this result for $d \geq 2$

Maybe most important marked Poisson point process: the unit intensity Poisson point process with i.i.d. marks on a compact W

- number of objects $\sim$ Poisson$(\nu(W))$
- locations and marks i.i.d. : $w_i \sim \frac{1}{\nu(W)}$ and $m_i \sim \nu_M$

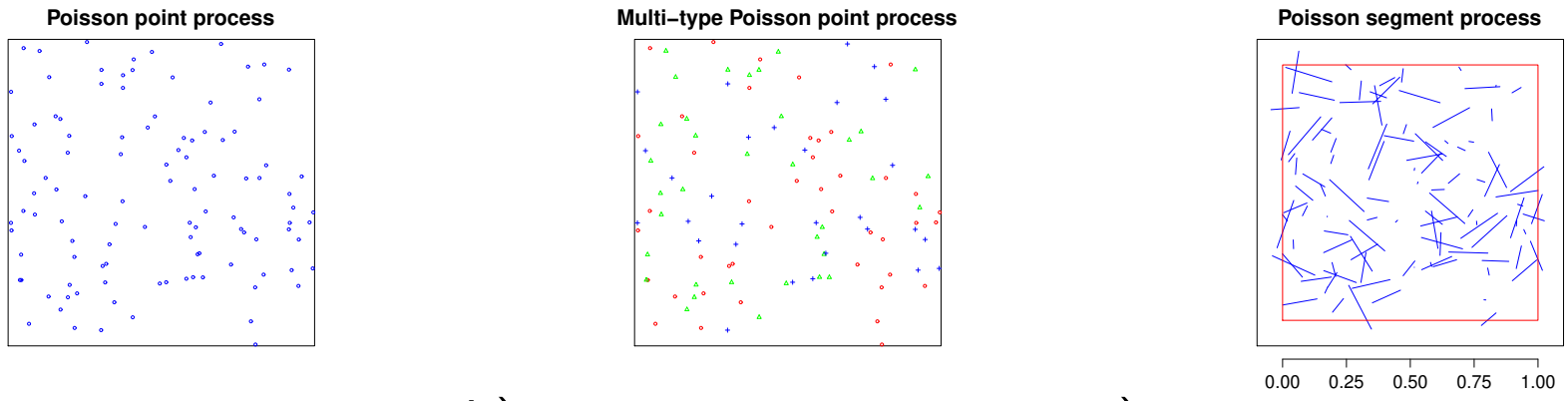
The corresponding probability measure : weighted 'counting" of objects

$$\begin{aligned}\mathbb{P}(X \in F) &= \sum_{n=0}^{\infty} \frac{e^{-\nu(W)}}{n!} \int_{W\times M} \cdots \int_{W\times M} 1_F\{(w_1,m_1),\ldots,(w_n,m_n)\} \\ &\quad \times d\nu(w_1)d\nu_M(m_1)\ldots d\nu(w_n)d\nu_M(m_)\end{aligned}$$

for all $F \in \mathcal{F}$.

Remark: the simulation of this process is straightforward, while the knowledge of the probability distribution allows analytical computations of the interest quantities

Simulations results of some Poissonian point processes: the domain is $W = [0,1] \times [0,1]$ and the intensity parameter is $\rho = 100$



a) b) c)

Figure: Poisson based models realizations : a) unmarked, b) multi-type and c) Poisson process of segments.

Definition

A disjoint union $\cup_{i=1}^{\infty} X_i$ of point processes $X_1, X_2, \ldots$ is called superposition.

Proposition

If $X_i \sim Poissson(W, \rho_i)$, $i = 1, 2, \ldots$ are mutually independent and if $\rho = \sum \rho_i$ is locally integrable, then with probability one, $X = \cup_{i=1}^{\infty} X_i$ is a disjoint union and est $X \sim Poisson(W, \rho)$.

$\rightarrow$ stable character of the Poisson process

Definition

Let be $q : W \to [0, 1]$ a function and X a point process on W. The point process $X_{thin} \subset X$ obtained by including the $\xi \in X$ in X_{thin} with probability $q(\xi)$, where points are included/excluded independently of each other, is said to be an independent thinning of X with retention probabilities $q(\xi)$.

Formally, we can set

$$X_{\text{thin}} = \{\xi \in X : R(\xi) \leq q(\xi)\},$$

with the random variables $R(\xi) \sim \mathcal{U}[0, 1]$, $\xi \in W$, mutually independent and independent of X.

Proposition

Suppose that $X \sim Poisson(W, \rho)$ *is subject to independent thinning with retention probabilities* $q(\xi), \quad \xi \in W$ *and let*

$$\rho_{thin} = q(\xi)\rho(\xi), \quad \xi \in W.$$

Then X_{thin} *and* $X \setminus X_{thin}$ *are independent Poisson processes with intensity functions* ρ_{thin} *and* $\rho - \rho_{thin}$*, respectively.*

Corollary

Suppose that $X \sim Poisson(W, \rho)$ *with* ρ *bounded by a positive constant* C*. Then* X *is distributed as independent thinning of a* $Poisson(W, C)$ *with retention probabilities* $q(\xi) = \rho(\xi)/C$*.*

Some general facts concerning the Poisson point processes

- the Poisson point process is as important for spatial statistics as the Gaussian process in classical probability theory
- the law is completely known $\rightarrow$ analytical formulas
- the Poisson process is invariant under independent thinning
- easy procedure for simulate non-stationary Poisson process
- completely random patterns : null or the default hypothesis that we want to reject
- independence $\rightarrow$ no interaction $\rightarrow$ no structure in the data

- two stationary Poisson point processes, they are not absolutely continuous with respect to each other, except if one process has unit intensity or if they have the same intensity
- two inhomogeneous Poisson point processes with strictly positive intensities, they are absolutely continuous with respect to each other
- more complicate models can be built $\rightarrow$ specifying a probability density $p(\mathrm{x})$ w.r.t. the reference measure given by the unit intensity Poisson point process. This probability measure is written as

$$\mathbb{P}(X \in F) = \int_F p(\mathrm{x})\mu(d\mathrm{x})$$

with μ the reference measure.

Remark : in this case the normalizing constant is not available from an analytical point of view. To check this replace in the expression of $\mu(\cdot)$ the indicator function $1_F\{\mathrm{y}\}$ with $p(\mathrm{y})$ ...

Few words about self-exciting point processes

Hawkes processes : a point process defined by its intensity of events conditional on the past $\rho^*(t) \rightarrow$ the intensity of the process evolves with the time depending on the points arrived in the configuration : no more independence

$$\rho^*(t) = \rho + \sum_{i=1}^{n} \mu(t - t_i)$$

such as

- $(t_i)_{1 \leq i \leq n}$ the sequence of arrival times of events that have occurred up to t.
- ρ background intensity.
- $\mu : [0, +\infty[\rightarrow [0, +\infty[$ excitation function.

Remarks:

- if $\mu = 0 \Rightarrow$ classical Poisson process
- modelling: several models available for the excitation functions
- simulation: thinning method
- inference: the conditional intensity allows the construction of a likelihood function
- application: sismology and epidemiology (extend the definition of the conditional intensity)

Exponential model: an example of excitation function for Hawkes processes

$$\mu(t) = \alpha \exp(-\beta t)$$

with $\alpha < \beta$. The parameter α gives the instantaneous influence of events and β the rate at which it decreases.

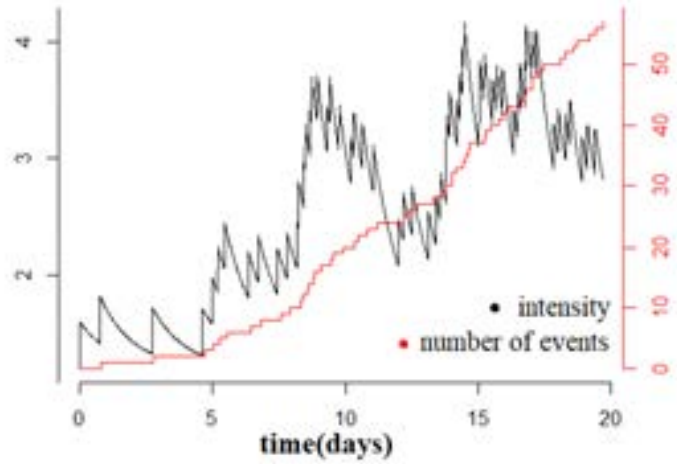


Figure: Number of events and conditional intensity of a Hawkes process, with exponential excitation function with parameters $\alpha = 0.6, \beta = 0.8$ et $\rho = 1.2$.

Table of contents

Cours 3. Tools for manipulating point processes.

Present context: + Wooclap

- mathematical background
- definition of a marked point process
- Binomial and Poisson point process
- important result: the point process law is determined by counts of points

Let X be a point process on W. The counts of points in bounded Borel regions of $B \subset W$, $N(B)$ characterize the point process and they are well defined random variables

- it is difficult to average the pattern X
- it is possible to compute moments of the $N(B)$'s

The appropriate mathematical tools are: blackboard explanations

- the moment measures
- the factorial moment measures
- the product densities
- the capacity functional ...

An extraordinary short story about their construction and very important results ...

Exterior conditioning:

- the probability of observing the point $\zeta \in W \times M$ knowing the configuration $\mathrm{y} \in F$ where $F \in \mathcal{F}$
- conditional intensity $\rightarrow$ interpretation

$$\lambda(\zeta; \mathrm{x}) d\zeta = \mathbb{P}(N(d\zeta) = 1 | \mathbb{X} \cap (d\zeta)^c = \mathrm{x} \cap (d\zeta)^c)$$

for $\zeta \in K \times M$ and $\zeta \notin \mathrm{x}$.

Interior conditioning:

- the probability of observing the configuration x $\in F$ knowing the point ζ belongs to the process
- Palm distributions $\rightarrow$ interpretation

$$\mathbb{P}_\zeta(F) = \mathbb{P}(\mathbb{X} \in F | N(\{\zeta\}) > 0)$$

- *reduced* Palm distributions $\rightarrow$ interpretation

$$\mathbb{P}^!_\zeta(F) = \mathbb{P}(\mathbb{X} \setminus \{\zeta\} \in F | N(\{\zeta\}) > 0)$$

Campbell - Mecke formula:

$$\mathbb{E}\left[\sum_{\zeta\in\mathbb{X}} g(\zeta, \mathbb{X}\setminus\{\zeta\})\right] = \int_W \int_\Omega g(\zeta, x) d\mathbb{P}^!_\zeta(x) d\mu(\zeta)$$

where $g : W \times \Omega \to \mathbb{R}$ is a measurable function and $\mu(A) = \mathbb{E}N(A)$ for any bounded Borel set $A \in W$.

Slivnyak Mecke formula : for stationary Poisson point processes the reduced Palm distribution and the process distribution coincides

Georgii - Nguyen - Zessin formula:

$$\mathbb{E}\left[\sum_{\zeta\in\mathbb{X}} g(\zeta, \mathbb{X}\setminus\{\zeta\})\right] = \mathbb{E}\left[\int_W g(\zeta,\mathbb{X})\lambda(\zeta;\mathbb{X})d\nu(\zeta)\right]$$

The ingredients for the proofs are: continuity of measures, standard proof argument and the detailed form the probability density of a point process.

Choquet theorem: the distribution of a random closed set X is entirely determined by knowledge of the capacity functional

$$T_{\mathbb{X}}(\mathcal{W}) = \mathbb{P}(\mathbb{X} \cap \mathcal{W} \neq \emptyset)$$

where $\mathcal{W}$ is any compact set in W.

- random closed sets: more general object than a point process $\rightarrow$ this object cannot be counted, but it can be observed through a compact window ...
- generalizes the characterization of a point process by the distributions of the void events
- a proof of the results can be found in (Molchanov, 2005)
- practical applications $\rightarrow$ Boolean model

Practical importance of the previous results

Summary statistics for point processes:

- construction: give particular values to functions and sets in the previous results
- compute their estimation does not require model simulation
- for Poissonian processes $\rightarrow$ analytical formulas
- statistical inference: hypothesis testing, parameter estimation and much more

Some known summary statistics for point processes

Assumption: $\mathbb{X}$ is a stationary point process

- the intensity or the first order product density function in

$$\mu(B) = \alpha(B) = \mathbb{E}[N(B)] = \int_B \rho(w) d\nu(w)$$

may be estimated by

$$\widehat{\rho(w)} = \sum_{x \in \mathbb{X}} \frac{1\{x \in B(w, \epsilon)\}}{\nu(B(w, \epsilon) \cap W)}$$

with $B(w, \epsilon)$ the closed ball centred in w of radius ϵ

- the second order product density function in

$$\alpha(B_1 \times B_2) = \int_{B_1} \int_{B_2} \rho(w_1, w_2) d\nu(w_1) d\nu(w_2) = \int \rho(0, r) dr$$

may be estimated by

$$\widehat{\rho(o, r)} = \frac{1}{\nu(B(r, \epsilon))} \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{X}}^{\neq} \frac{1\{y - x \in B(r, \epsilon)\}}{\nu(W \cap W_{y-x})}$$

where $\neq$ indicates that the sum is taken over all $(x, y) \in \mathbb{R}^2$ such that $x \neq y$ and $W_a = \{w + a : w \in W\}$.

- the pair correlation function given by

$$g(r) = \frac{\rho(x,y)}{\rho(z)} = \frac{\rho(0, r = |x-y|)}{\rho(z)}$$

is estimated by combining the previous two estimators

$$\widehat{g(r)} = \frac{1}{\nu(B(r,\epsilon))} \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{X}}^{\neq} \frac{1\{y - x \in B(r,\epsilon)\}}{\nu(W \cap W_{y-x}) \widehat{\rho(x)} \widehat{\rho(y)}},$$

one must remember that this function is sensitive to errors in $\widehat{\rho}$

- the empty space function

$$F(r) = \mathbb{P}(d(\zeta, \mathbb{X}) \leq r) = \mathbb{P}(\mathbb{X} \cap b(o, r) \neq \emptyset) = T_{\mathbb{X}}(b(o, r))$$

where $\mathbb{P}$ is the distribution of $\mathbb{X}$. An *uncorrected* estimator of F is given by

$$\widehat{F}_{un}(r) = \frac{1}{m} \sum_{j=1}^{m} 1\{d(w_j, \mathbb{X} \cap W) \leq r\}$$

with m the size ofthe grid $\{w_j\}_{1,\ldots,m}$ over W and $d(w, \mathbb{X} \cap W)$ the closest distance from the grid point w to the point pattern $\mathbb{X} \cap W$

- the nearest neighbour distance distribution function

$$G(r) = \mathbb{P}_o(d(o, \mathbb{X} \setminus \{o\}) \leq r) = \mathbb{P}_o^!(N(b(o, r) > 0)$$

where $\mathbb{P}_o$ and $\mathbb{P}_o^!$ the Palm and the reduced Palm distributions, respectively. An *uncorrected* estimator of G is given by

$$\widehat{G}_{un}(r) = \frac{1}{n(\mathbb{X})} \sum_{i=1}^{n(\mathbb{X})} 1\{d(x_i, \mathbb{X} \setminus x_i) \leq r\}$$

- the K function

$$\rho K(r) = \mathbb{E}_o^! \left[N(b(o, r)) \right]$$

where $\mathbb{E}_o^!$ is the expectation under the reduced Palm distribution and $N(b(o, r))$ is the number of points inside the ball centered at the origin o of radius r. An *uncorrected* estimator of K is given by

$$\widehat{K}_{un}(r) = \frac{\nu(W)}{n(\mathbb{X})(n(\mathbb{X}) - 1)} \sum_{i,j,i \neq j} 1\{d(x_i, x_j) \leq r\}$$

- the J function: compare nearest neighbour to empty distances

$$J(r) = \frac{1 - G(r)}{1 - F(r)}$$

defined for all $r > 0$ such that $F(r) < 1$, with the plug-in estimator given by

$$\widehat{J}(r) = \frac{1 - \widehat{G}(r)}{1 - \widehat{F}(r)}$$

Application 1: hypothesis testing $\rightarrow$ compare theoretical and estimated statistics to test complete randomness (Poisson) versus alternative

- formulas for the stationary Poisson process of intensity function ρ, on $W \subset \mathbb{R}^2$:

$$\begin{aligned} \rho(r) &= 1 \\ K(r) &= \pi r^2 \\ F(r) &= 1 - \exp[-\rho\pi r^2] \\ G(r) &= F(r) \\ J(r) &= 1 \end{aligned}$$

Application 2: study the morphological properties of the pattern

- J function:

$$J(r) \quad \text{is} \quad \begin{cases} = 1 & \text{Poisson: complete random} \\ \leq 1 & \text{clustering: attraction} \\ \geq 1 & \text{regular: repulsion} \end{cases}$$

- the pair correlation function:

$$\rho(r) \quad \text{is} \quad \begin{cases} = 1 & \text{Poisson: complete random} \\ \geq 1 & \text{clustering: attraction} \\ \leq 1 & \text{regular: repulsion} \end{cases}$$

Some comments remarks:

- the shown estimators are rather intuitive the principle of their construction but exhibit bias and edge effects
- see the cite references for existing correction methods: border correction, edge effects, Kaplan-Meyer, etc. (Baddeley et al. 2016, Illian et al. 2008, van Lieshout 2019, Møller and Waageptersen 2004)
- extensions: marks, non-stationary spaces, different observation spaces W is possible $\rightarrow$ Boolean model made of nice geometrical objects ...
- this list of statistics is not exhaustive : L, ρ, etc.
- parameter estimation $\Rightarrow$ these function are the "equivalent" of the moments in probability theory, they do not entirely determine the model to be estimated (Baddeley and Silverman, 1984)

Goodness-of-fit: envelope tests

- the null hypothesis H_0: the model fits the data
- the method is based on some functional summary characteristic $S(r)$: such as $F(r), G(r), J(r), K(r), \rho(r)$, etc.
- idea: compare empirical summary characteristic estimated from a point pattern in the observation window W to estimates of the summary characteristic for simulations from the model
- the model is simulated using the estimated parameters
- the simulations from the model are generated in the same window W

- the characteristic $S(r)$ is estimated by $\hat{S}(r)$ computed from the observed data
- the model is simulated k times
- estimates of $S(r)$, $\hat{S}_i(r)$ are computed for each sample $i = 1, \ldots, k$
- the extreme values are computed:

$$S_{\min}(r) = \min_{(i)} \hat{S}_i(r) \quad \text{and} \quad S_{\max}(r) = \max_{(i)} \hat{S}_i(r)$$

- if the inequality

$$S_{\min}(r) \leq \hat{S}(r) \leq S_{\max}(r)$$

is verified for all r, the H_0 is accepted, otherwise is rejected

- this test is regarded and intepreted as a significance test
- for fixed $r^\star$:
  - the error probability in one-sided testing is $\frac{1}{k+1}$
  - the error probability in two-sided testing is $\frac{2}{k+1}$
- thus for $k = 19 \rightarrow \alpha = 0.05$ or for $k = 99 \rightarrow \alpha = 0.01$
- the test can be modified and larger or smaller values can be used, rather than the minimum and maximum values
  - for $\alpha = 0.05$, $k = 999$ may be chosen and S_{min} is replaced by the 25th of the order S_i and S_{max} by the 975th

Summary statistics for an observed point pattern: the "cell" data set

- F, G, J and K functions
- regular pattern $\rightarrow$ repulsion
- R code

```
> data(cells)
> plot(allstats(cells))
```

or

```
> plot(Fest(cells)) ; plot(Gest(cells)) ;
plot(Jest(cells)) ; plot(Kest(cells))
```

Summary statistics – cells

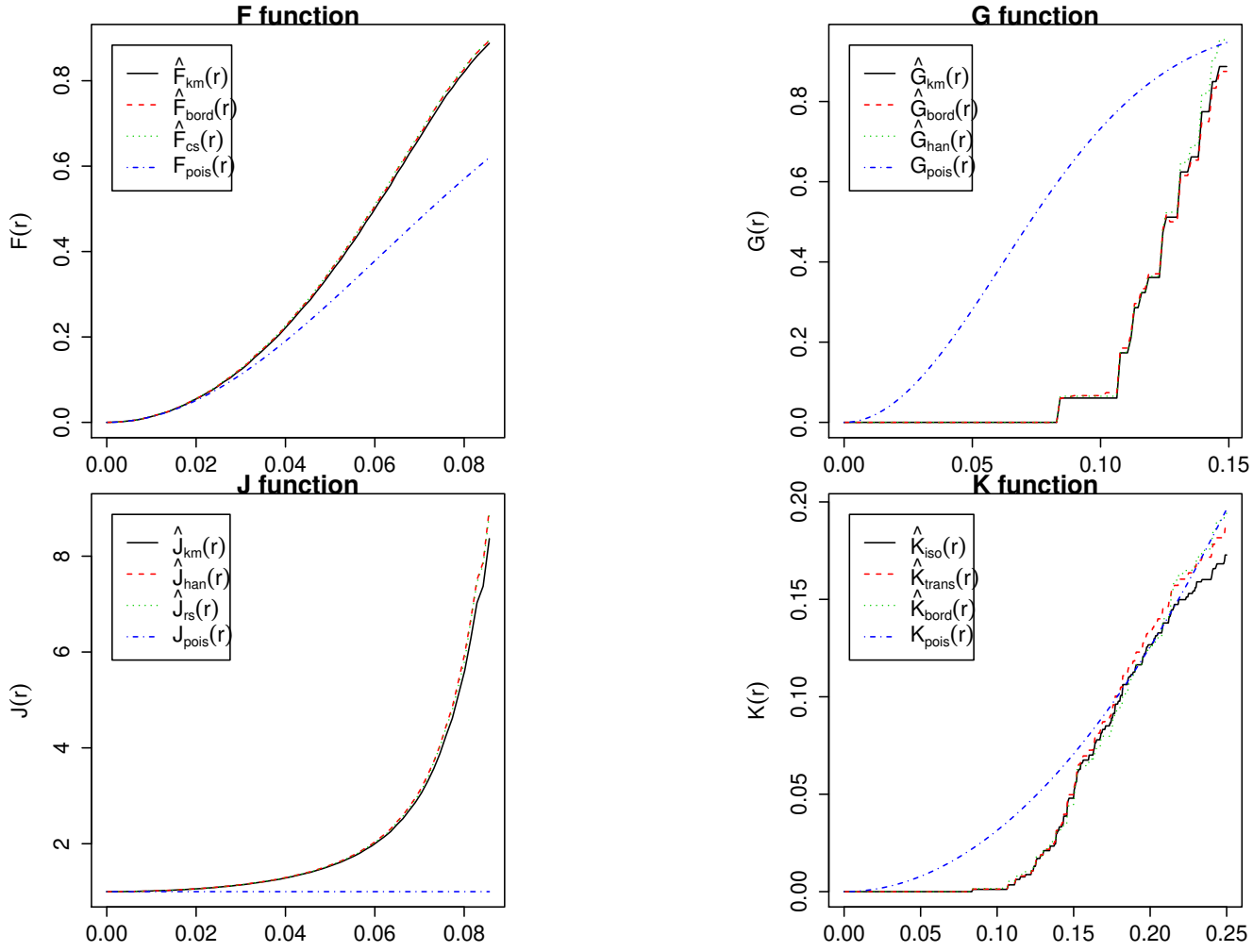


Figure: Estimated summary statistics: cell data set

Road network dynamics:

- data: spatio-temporal evolution of a road network due to logging activity → you have already seen the video ...
- questions:
  - do the exploitation companies respect the regulations for preserving the forest ?
  - does the economical activity affect the resilience capacity of the forest ?
- exploratory analysis: empty space function → road-less space measure
- challenge: build a stochastic model
- people: F. Kleinschroth, J. R. Healey, S. Gourlet-Fleury, F. Mortier, M. N. M. van Lieshout
- paper: (Kleinschroth et al., 2017)

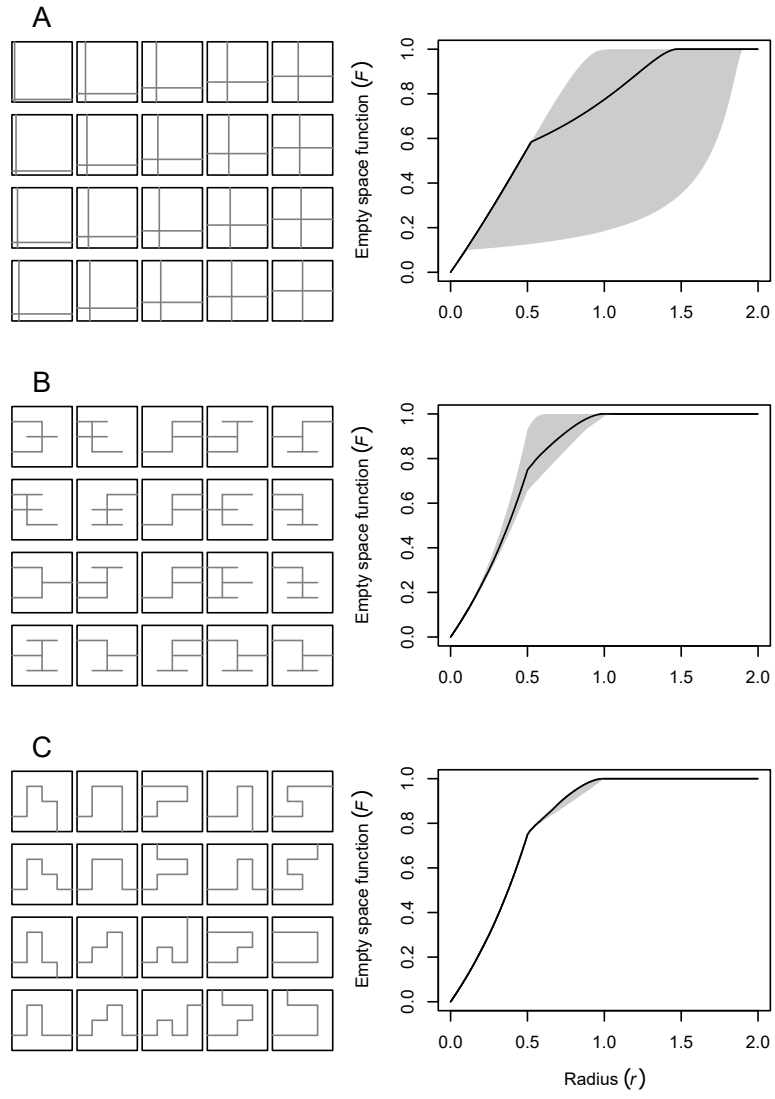


Figure: Toy model for explaining the behaviour of the empty space function: the simulated roads have the length, so the same density of roads per unit of surface.

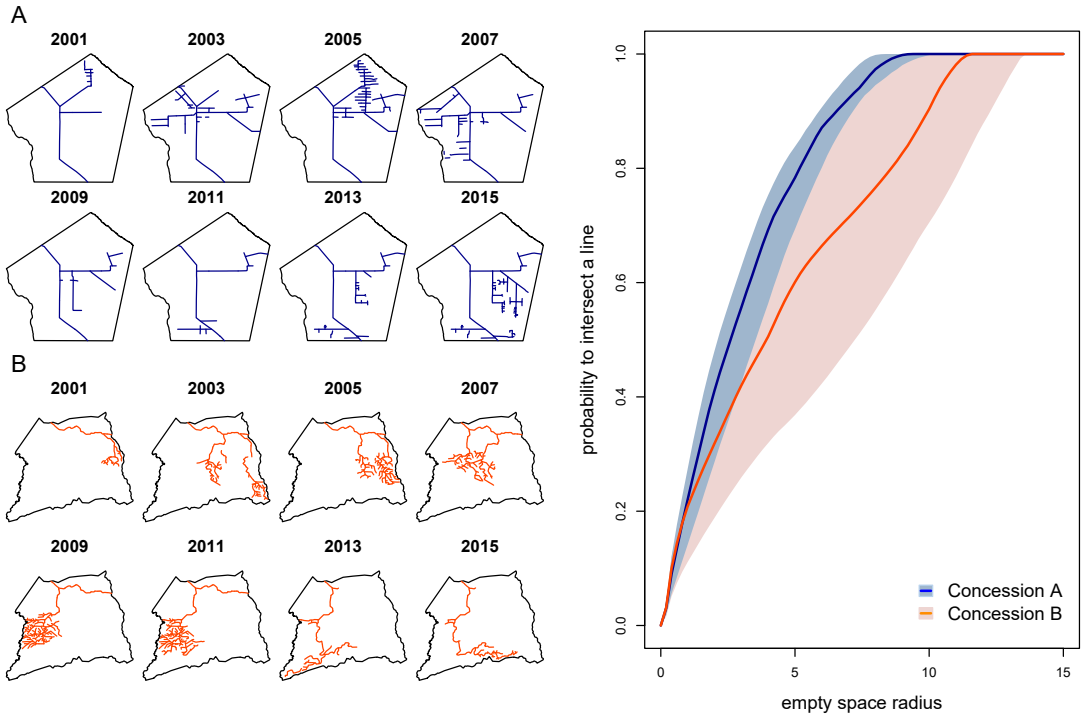


Figure: Empty space function for characterization the road network evolution in two logging companies.

Cosmology: influence of the new observations on the already detected structure

- data: SDSS Data Release 12 - photometrical galaxies
- question:
  - how this new data set is related to the already detected structure ?
- exploratory analysis: adapt the F, G and J function to establish possible dependence of different types of patterns
- challenge: the position of the photometrical galaxies is not entirely known
- people: M. Kruuse, E. Tempel, R. Kipper
- paper: (Kruuse et al., 2019)

Preliminary result :

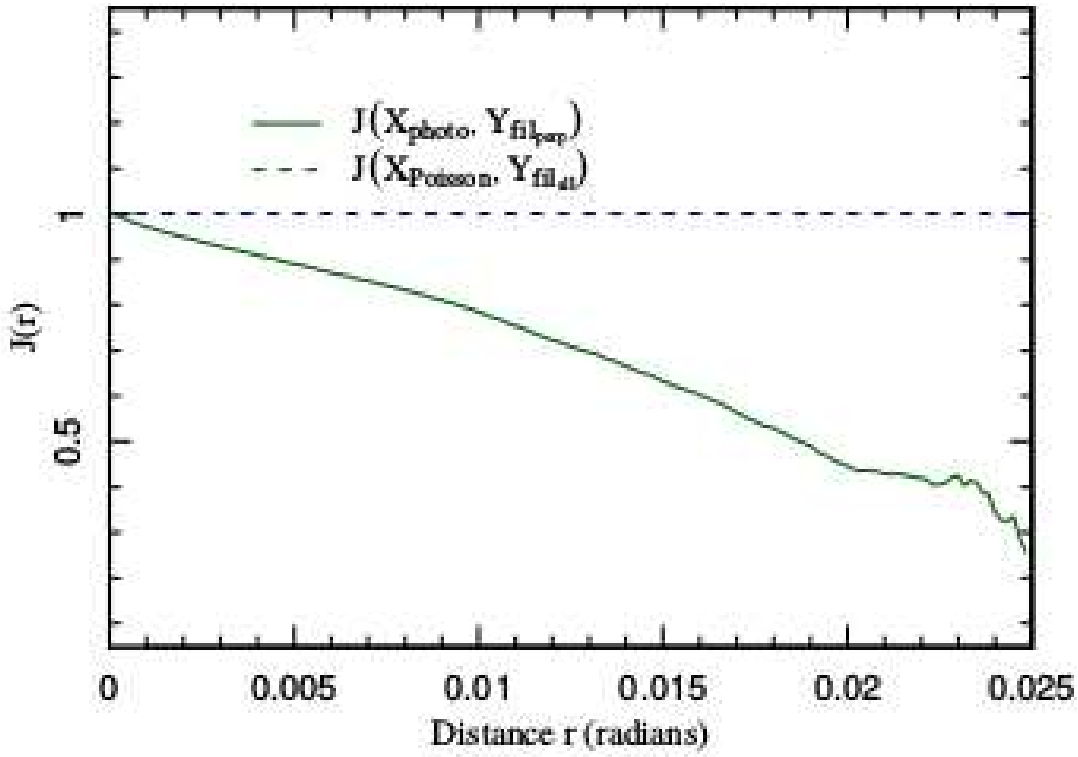


Figure: Estimation of the bivariate J function. The considered sets were the photometric galaxies and the projection on the sphere of the filamentary spines that are rather perpendicular on the line of sight. This result indicates positive association of these two patterns.

Synthesis :

- good exploring tool: **spatstat** provides also some 3d estimators
- different estimators
- numerical sensitivity, border effects
- parameter estimation: some limits
- statistical tests: envelopes, bootstrap, permutation, central limit theorem

$\Rightarrow$ the implementation of some statistical tests require simulation of the model under the null hypothesis

Table of contents

Previously in this course + WOOCLAP...

Direct application of counting measures and Palm distributions for point process analysis

- counting measures $\rightarrow$ summary statistics for point pattern characterization
- two categories: interpoint distances (F, G and J) and second order characteristics (ρ, K and L)
- possible extension of the summary statistics: marks, non-stationary processes, different observation spaces W case and spatio-temporal

- non-parametrical estimation of the summary statistics: kernel estimation and management of the border effects
- central limit available: statistical tests
- envelope simulation based tests: reject a particular model
- summary statistics for parameter estimation of a given model: minimum contrast method
  - these statistics are an "equivalent" of moments in probability theory
  - it is not always, that they entirely determine the model to be estimated
  - (Baddeley and Silverman, 1984), (Bedford and Berg, 1997)

- good exploring tool: **spatstat**
- outline important characteristics of a point pattern: clustering, repulsion, completely randomness
- it is difficult to differentiate between interaction and inhomogeneity if only one realisation is available
- need for models able to reproduce these characteristics

Cox processes

Definition

Let Υ be a random locally finite diffuse measure on $(W, \mathcal{B})$. If the conditional distribution of X given Υ is a Poisson process on W with intensity measure Υ, X is said to be a Cox point process with driving measure Υ. Sometimes X is also called doubly stochastic Poisson process.

Remarks :

- if there exists a random field $Z = \{Z(w), w \in W\}$ such that

$$\Upsilon(B) = \int_B Z(w) d\nu(w)$$

then X is a Cox process with driving function Z

- the conditional distribution of X given $Z = z$ is a distribution of the Poisson process with intensity function z $\Rightarrow$

$$\mathbb{E}[N(B)|Z = z] = \int_B z(w)d\nu(w)$$

- the first order factorial moment measure is obtained using the law of the total expectation

$$\begin{aligned}\mu^{(1)}(B) = \alpha^{(1)}(B) &= \mathbb{E}[N(B)] \\ &= \mathbb{E}\left[\mathbb{E}[N(B)|Z = z]\right] = \mathbb{E}\left[\int_B Z(w)d\nu(w)\right] \\ &= \mathbb{E}[\Upsilon(B)] = \int_B \mathbb{E}Z(w)d\nu(w)\end{aligned}$$

- if $\rho(w) = \mathbb{E}Z(w)$ exists then it is the intensity function

- smilarly, it can be shown that the second order factorial moment measure is

$$\begin{aligned}
\alpha^{(2)}(B_1 \times B_2) &= \mathbb{E}\left[\Upsilon(B_1)\Upsilon(B_2)\right] \\
&= \mathbb{E}\left[\int_{B_1} Z(u)d\nu(u)\int_{B_2} Z(v)d\nu(v)\right] \\
&= \mathbb{E}\left[\int_{B_1}\int_{B_2} Z(u)Z(v)d\nu(u)d\nu(v)\right] \\
&= \int_{B_1}\int_{B_2} \mathbb{E}\left[Z(u)Z(v)\right] d\nu(u)d\nu(v)
\end{aligned}$$

- if $\rho^{(2)}(u,v) = \mathbb{E}Z(u)Z(v)$ exists, then it is the product density

- the pair correlation function is

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} = \frac{\mathbb{E}\left[Z(u)Z(v)\right]}{\mathbb{E}\left[Z(u)\right]\mathbb{E}\left[Z(v)\right]}$$

- depending on Z it is possible to obtain analytic formulas for the second order characteristics (g, K and L) and the interpoint distance characteristic (F, G and J)

- the variance $VarN(B)$ is obtained using the total variance law, and it is

$$VarN(B) = \mathbb{E}N(B) + Var\left[\int_B Z(w)d\nu(w)\right] \geq \mathbb{E}N(B)$$

$\Rightarrow$ over - dispersion of the Cox process counting variables

- the void probabilities of Cox processes are

$$\begin{aligned}
\mathbb{P}(N(B) = 0)) &= \mathbb{E}1\{N(B) = 0\} \\
&= \mathbb{E}\left[\mathbb{E}1\{N(B) = 0\}|Z = z)\right] = \mathbb{E}\left[\mathbb{P}(N(B) = 0|Z = z)\right] \\
&= \mathbb{E}\left[\exp\left(-\int_B Z(w)d\nu(w)\right)\right] \\
&= \mathbb{E}\left[\exp\left(-\Upsilon(B)\right)\right]
\end{aligned}$$

Trivial Cox process : mixed Poisson processes

- $Z(w) = Z_0$ a common positive random variable for all locations $w \in W$
- $X|Z_0$ follows a homogeneous Poisson process with intensity Z_0
- the driving measure is $\Upsilon(B) = Z_0 \nu(B)$

Thinning of Cox processes :

- X is a Cox process driven by Z
- $\Pi = \{\Pi(w) : w \in W\} \subseteq [0, 1]$ is a random field which is independent of (X, Z)
- $X_{thin}|\Pi \rightarrow$ the point process obtained by independent thinning of the points in X with retention probabilities Π
- $\Rightarrow X_{thin}$ is a Cox process driven by $Z_{thin}(w) = \Pi(w)Z(w)$

Log Gaussian Cox processes

- introduced independently by astronomers (Coles and Jones, 1991) and statisticians (Møller et. al., 1998)
- consider $Y = \log Z$ is a Gaussian field
  - for any integer $n > 0$, locations $\xi_1, \ldots, x_n \in \mathbb{R}^d$ and numbers $a_1, \ldots a_n \in \mathbb{R}$, $\sum_{i=1}^{n} a_i Y(\xi_i)$ follows a normal distribution
- the Cox process X driven by $Z = exp(Y)$ is a *log Gaussian Cox Process* (LGCP)

- the distribution of (X, Y) is entirely determined by the mean and the covariance function

$$m(\xi) = \mathbb{E}Y(\xi) \quad \text{and} \quad c(\xi, \eta) = Cov(Y(\xi)Y(\eta))$$

- covariance function :
  - for simplicity it may be considered translation invariant

$$c(\xi, \eta) = c(\xi - \eta)$$

of the form

$$c(\xi) = \sigma^2 r(\xi/\alpha)$$

  - the function $r : \mathbb{R}^d \to [-1, 1]$ is a correlation function for a Gaussian field iif r is positive definite

$$\sum_{i=1}^{n} a_i a_j r(\xi_i, \xi_j) > 0 \quad \text{for all} \quad \xi_1, \ldots, x_n \in \mathbb{R}^d, a_1, \ldots a_n \in \mathbb{R}$$

- weak conditions are required on m and r in order to get

$$\Upsilon(B) = \int_B Z(\xi) d\nu(\xi)$$

for bounded $B \subset \mathbb{R}^d$. For instance, we may require $\xi \to Y(\xi)$ continuous almost surely

- as example, this is satisfied by continuous m and r such that

$$r(\xi) = \exp(- \parallel \xi \parallel^{\delta}), \quad 0 \leq \delta \leq 2$$

with δ controlling the smoothness of the realizations of the Gaussian field
  - $\delta = 1$: exponential correlation function
  - $\delta = 1/2$: stable correlation function
  - $\delta = 2$: Gaussian correlation function
- there is a one-to-one correspondence between (m, c) and $(g, \rho) \Rightarrow$ the distribution of (X, Y) is uniquely determined by (ρ, g)

$\to$ Exercise 11

Cluster processes

Definition

Let C be a point process (parent process), and for each $c \in C$ let X_c be a finite point process (daughter process). Then

$$X = \bigcup_{c \in C} X_c$$

is called a cluster point process.

Definition

Let X be a cluster point process such that C is a Poisson point process and conditional on C, the processes $X_c, c \in C$ are independent. Then X is called a Poisson cluster point process.

Neyman-Scott processes

Definition

Let X be a Poisson cluster point process such that centred daughter processes $X_c - c$ are independent of C. Given C, let the points of $X_c - c$ be i.i.d. with probability density function k on $\mathbb{R}^d$ and $N(X_c)$ be i.i.d. random variables. Then X is called a Neyman-Scott process. If moreover $N(X_c)$ given C has a Poisson distribution with intensity α, then X is a Neyman-Scott Poisson process.

$\rightarrow$ drawing + Exercice 12

Theorem

Let X be a Neyman-Scott Poisson process such that C is a stationary Poisson process with intensity κ. Then X is stationary process with intensity $\rho = \alpha\kappa$ and pair correlation function

$$g(u) = 1 + \frac{h(u)}{\kappa},$$

where

$$h(u) = \int k(v)k(u+v)d\nu(v)$$

is the density for the difference between two independent points distributed according to k.

Proof.

→ Exercise 13 □

Other very known cluster point processes

Matérn cluster process (Matérn 1960,1986)

$$k(u) = \frac{1\{\| u \| \leq r\}}{\omega_d r^d}$$

is the uniform density on the ball $b(o, r)$

Thomas process (Thomas 1949)

$$k(u) = \frac{\exp\left(-\frac{\|u\|^2}{2\omega^2}\right)}{(2\pi\omega^2)^{d/2}}$$

is the density for $\mathcal{N}_d(0, \omega^2 I_d)$, i.e. for d independent normally distributed variables with mean 0 and variance $\omega^2 > 0$

- both kernels are isotropic
- the Thomas process pair correlation function is

$$g(u) = 1 + \frac{1}{\kappa(4\pi\omega^2)^{d/2}} \exp\left[-\frac{\| u \|^2}{4\omega^2}\right]$$

and its $K-$function for $d = 2$ is

$$K(r) = \pi r^2 + \frac{1 - \exp[-r^2/(4\omega^2)]}{\kappa}$$

- other summary statistics can be also computed
- the expressions of the summary statistics are more complicated for the Matérn process

$\rightarrow$ drawing the processes ...

Remarks :

- usually in applications Z is unobserved
- one cannot distinguish a Cox process X from its corresponding Poisson process $X|Z$ whenever a single realisation of X is available
- open question: which of the two models might be most appropriate, i.e. whether Z should be random or "systematic"/deterministic

- prior knowledge of the observed phenomenon
- Bayesian setting of the intensity function $\Rightarrow$ Cox processes
- if we want to investigate the dependence of certain covariates associated to Z, these may be treated as systematic terms, while unobserved effects may be treated as random terms
- Cox process: more flexible models for clustered patterns than inhomogeneous Poisson point processes
- parameter estimation methods: minimum contrast, Palm distributions, composite likelihood
  - based on the K function and the Palm distributions
  - `spatstat:` the `kppm` function

$\rightarrow$ Exercise 14

Table of contents

Finite point processes

- the independence property of the Poisson based processes does not allow to introduce point interactions
- Gibbs models are flexible point processes that allow the specification of point interactions via a a probability density
- in the following: consider (W, d) a complete, separable metric space such that $W \subset \mathbb{R}^d$ and $0 < \nu(W) < \infty$
- the condition that W has to be finite is required in order to obtain a well defined probability density for the considered Gibbs model

- let μ be the unit intensity Poisson point process on W
- the probability density of a Gibbs model is a Radon-Nykodim derivative w.r.t the Poissonian reference measure μ
- within a statistical physics context, the probability density of such a point process has the form

$$p(\mathrm{x}) = \frac{\exp[-U(\mathrm{x}|\theta)]}{\alpha(\theta)}$$

with $U(\mathrm{x}|\theta)$ the energy function, θ the model parameters and

$$\alpha(\theta) = \int_{\Omega} \exp[-U(\mathrm{x}|\theta)] d\mu(\mathrm{x})$$

the normalizing constant or the partition function

- under these circumstances, the probability distribution of a Gibbs model writes as

$$\mathbb{P}(X \in F) = \int_F p(\mathrm{x}) d\mu(\mathrm{x})$$

and by introducing the expression of μ, it is further expressed as

$$\begin{aligned}\mathbb{P}(X \in F) = & \\ & = \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W 1(\{x_1, \ldots, x_n\} \in F) \times \\ & p(\{x_1, \ldots, x_n\}) d\nu(x_1) \ldots d\nu(x_n),\end{aligned}$$

whenever $n > 0$. If $n = 0$, we take $\exp[-\nu(W)]1(\emptyset \in F)p(\emptyset)$. If $\nu(W) = 0$, then $P(X = \emptyset) = 1$. For applications, we always assume that $\nu(W) > 0$.

- usually the probability density is known only up to a constant : $p \propto h = exp(-U)$
- the normalizing constant or the partition function is given by

$$\alpha = \int_{\Omega} h(x) d\mu(x)$$

that becomes

$$\alpha = \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)]}{n!} \int_W \cdots \int_W h(\{x_1, \ldots, x_n\}) d\nu(x_1) \ldots d\nu(x_n) \tag{1}$$

- the marked case writes in a similar way by introducing also the marks distribution ν_M

- the previous quantity is not always available under analytical closed form
- this is the main difficulty to be solved while ausing this approach ...

Normalizing constant for the Poisson process : Let ρ be the intensity function of a Poisson point process on W. Its probability density up to a normalizing constant is

$$p(\mathrm{x}) \propto \prod_{x_i \in \mathrm{x}} \rho(x_i).$$

Let $\Upsilon(B) = \int_B \rho(w) d\nu(w)$ be the associated intensity measure. By using (1), we get

$$\alpha = \exp[-\nu(W)] \sum_{n=0}^{\infty} \frac{\Upsilon(W)^n}{n!} = \exp[\Upsilon(W) - \nu(W)],$$

that gives for the complete probability density

$$p(\mathrm{x}) = \exp[\nu(W) - \Upsilon(W)] \prod_{x_i \in \mathrm{w}} \rho(x_i)$$

If the process is stationary $\rho(x) = \rho = \text{ct.}$, then the probability density is

$$p(\mathrm{x}) = \exp[(1-\rho)\nu(W)]\rho^n$$

Remarks :

- the probability density is specified only for finite point processes
- the extension to $\mathbb{R}^d$ of a finite point process specified by a probability density is possible under some conditions (see (Møller and Waagpetersen, 2004, section 6.4))
- phase transition - if such an extension is possible, does it surely leads to an unique probability measure ?

Construction of the probability density

- specify the interaction functions $\phi^{(k)} : \Omega \rightarrow \mathbb{R}^+$

$$\phi(x_{i_1}, \ldots, x_{i_k})^{(k)}$$

for any $k-$tuplet of objects

- the un-normalized probability density is the product of all these functions

$$h(\mathrm{x}) = \prod_{x_i \in \mathrm{x}} \phi(x_i)^{(1)} \cdots \prod_{\{x_{i_1}, \ldots, x_{i_k}\} \in \mathrm{x}} \phi(x_{i_1}, \ldots, x_{i_k})^{(k)} \quad (2)$$

- clearly, the energy function is obtained by taking $U(\mathrm{x}) = -\log h(\mathrm{x})$
- α the normalizing constant is difficult to be determined : untractable mathematical formula

- the un-normalized probability densities (2) are suitable for modelling provided they are integrable on Ω ; that is

$$\alpha = \int_{\Omega} h(x) d\mu(x) < \infty.$$

- the following results ensure the integrability of the probability density of a marked point process $\rightarrow$ the Ruelle stability conditions

Definition

Let X be a marked point process given by the un-normalized probability density h w.r.t the reference measure μ. The process X is stable in the sense of Ruelle, if it exists $\Lambda > 0$ such that

$$h(x) \leq \Lambda^{n(x)}, \quad \forall x \in \Omega. \tag{3}$$

Proposition

The un-normalized probability density of a stable point process is integrable.

Proof.

The integrability of $h(x)$ follows directly from the preceding condition :

$$\begin{aligned}\int_{\Omega} h(x)\mu(dx) &\leq \int_{\Omega} \Lambda^{n(x)}\mu(dx) \\ &= \sum_{n=0}^{\infty} \frac{\exp[-\nu(W)][\Lambda\nu(W)])^n}{n!} = \exp[\nu(W)(\Lambda - 1)].\end{aligned}$$

□

Definition

Under the same hypotheses as in Prop. 3, a marked point process is said to be locally stable if it exists $\Lambda > 0$ *such that*

$$h(\mathsf{x} \cup \{\eta\}) \leq \Lambda h(\mathsf{x}), \quad \forall \mathsf{x} \in \Omega, \eta \in W \times M \setminus \mathsf{x} \tag{4}$$

Proposition

A locally stable point process is stable in the sense of Ruelle.

Proof.

It is easy to show by induction that

$$h(x) \leq h(\emptyset)\Lambda^{n(x)}, \quad \forall x \in \Omega.$$

The local stability of a point process (4) implies its integrability (3). □

- the conditional intensity for a point process X with probability density p is

$$\lambda(\eta; \mathrm{x}) = \frac{p(\mathrm{x} \cup \{\eta\})}{p(\mathrm{x})} = \frac{h(\mathrm{x} \cup \{\eta\})}{h(\mathrm{x})}, \quad \mathrm{x} \in \Omega, \eta \in W \times M \setminus \mathrm{x},$$

  taking $a/0 = 0$ for $a \geq 0$
- the conditional intensity is also known in the literature as the Papangelou intensity condition (we have already meet it)
- we shall often consider functions $h : \Omega \to [0, \infty[$ which are hereditary

$$h(\mathrm{x}) > 0 \Rightarrow h(\mathrm{y}) > 0, \quad \text{for} \quad \mathrm{y} \subset \mathrm{x}.$$

- if p is hereditary, then there is a one-to-one correspondence between p and λ

Importance of the conditional intensity : key element in modelling

- plays a similar role as the conditional probabilities for Markov random fields
- integrability
- convergence properties of the MCMC algorithms used to sample from p
- the process X is attractive if $x \subseteq y$ implies

$$\lambda(\eta; x) \leq \lambda(\eta; y),$$

and repulsive otherwise

$$\lambda(\eta; x) \geq \lambda(\eta; y),$$

- attractive processes tend to cluster the points, while the repulsive ones tend to distance the points
- these conditions are important also for exact MCMC algorithms
- there exist processes that are neither attractive nor repulsive
- there are processes that are integrable but not locally stable : Lennard - Jones (statistical physics)

Markov point processes

The conditional intensity of an interacting point process is given by

$$\lambda(\eta; \mathrm{x}) = \phi(\eta)^{(1)} \prod_{x_i \in \mathrm{x}} \phi(x_i, \eta)^{(2)} \cdots \prod_{\{x_{i_1}, \ldots, x_{i_k}\} \in \mathrm{x}} \phi(x_{i_1} \ldots, x_{i_k}, \eta)^{(k+1)}$$

- difficult to manipulate
- possible simplifications : limit the order of interactions $\rightarrow$ only pairs of points for instance
- limit the range of the interaction : a point interact only with its closest neighbours

Let $\sim$ be a symmetrical and reflexive relation between points belonging to $W \times M$. This may be a typical neighbourhood relation based on a metric (Euclidean, Hausdorff) or on sets intersection.

Definition

A clique is a configuration $\mathrm{x} \in \Omega$ *such that* $\eta \sim \zeta$ *for all* $\eta, \zeta \in \mathrm{x}$. *The empty set is a clique.*

Definition

Let X be a marked point process on $W \times M$ with probability density p w.r.t the reference measure μ. The process X is Markov if for all $x \in \Omega$ such that $p(x) > 0$, the following conditions are simultaneously fulfilled :

(i) $p(y) > 0$ for all $y \subseteq x$ (hereditary)

(ii) $\frac{p(x \cup \{\zeta\})}{p(x)}$ depends only on ζ and $\partial(\zeta) \cap x = \{\eta \in x : \eta \sim \zeta\}$.

This process is known in the literature as the Ripley-Kelly Markov process.

Example : The probability density w.r.t to μ of a marked Poisson process on $W \times M$ with constant intensity function ($\rho(\eta) = \beta > 0$) is

$$p(\mathrm{x}) = \beta^{n(\mathrm{x})} \exp[(1-\beta)\nu(W)].$$

Clearly $p(\mathrm{x}) > 0$ for all configurations x. Its Papangelou conditional intensity is

$$\lambda(\eta; \mathrm{x}) = \beta 1\{\eta \notin \mathrm{x}\}.$$

Hence, the Poisson process is Markov, independently of the interaction functions $\phi^{(k)}$. This agrees with the choice of the Poisson process for modelling a completely random structure.

The following result is known as the spatial Markov property.
$\rightarrow$ drawing

Theorem

Let X be a Markov point process with density $p(\cdot)$ on W and consider a Borel set $A \subseteq W$. Then the conditional distribution of $X \cap A$ given $X \cap A^c$ depends only on X restricted to the neighbourhood

$$\partial(A) \cap A^c = \{u \in W \setminus A : u \sim a \quad \text{for some} \quad a \in A\}.$$

Proof.

The proof is given by (van Lieshout 2000, Thm. 2.1, pp. 47) and also by (Møller and Waagepetersen 2004, Prop. 6.1, pp. 93). $\square$

The following result is known as the Hammersley-Clifford theorem.

Theorem

A marked point process density $p : \Omega \to \mathbb{R}^+$ *is Markov with respect to the interaction relation* $\sim$ *if and only if there is a measurable function* $\phi_c : \Omega \to \mathbb{R}^+$ *such that*

$$p(x) = \prod_{cliques} \prod_{y \subseteq x} \phi_c(y), \quad \alpha = \phi(\emptyset) \tag{5}$$

for all $x \in \Omega$.

Proof.

This result is proved in (Ripley and Kelly 1977, Thm. 1, pp.189), (van Lieshout 2000, Thm. 2.2 pp. 49) and (Møller and Waagepetersen 2004, Thm. 6.1, pp. 89). □

Remarks :

- the previous result simplifies the writing of the probability density of an interacting point process
- taking $\phi_c(\mathrm{y}) = 1$ whenever y is not a clique leads us to the equivalence of (2) and (5)
- Markov point processes are known in physics community as Gibbs point processes

$$p(\mathrm{x}) = \frac{1}{Z} \exp\left[-U(\mathrm{x})\right] = \frac{1}{Z} \exp\left[-\sum_{\text{cliques}} \sum_{\mathrm{z} \subseteq \mathrm{x}} U_c(\mathrm{z})\right],$$

with Z the partition function, U the system energy and $U_c = \log \phi_c$ the clique potential

- all the Markov processes are Gibbs
- the reciprocal is not true

Poisson process as a Markov process : the probability density of a Poisson point process is

$$p(\mathrm{x}) = e^{(1-\beta)\nu(W)} \prod_{x \in \mathrm{x}} \beta.$$

Hence, the interactions functions applied to cliques are

$$\begin{aligned} \phi_c(\emptyset) &= e^{(1-\beta)\nu(W)} \\ \phi_c(\{u\}) &= \beta \end{aligned}$$

with $\phi_c \equiv 1$ for the cliques made of more than one object. The potential of the cliques made of a single object is

$$U_c(u) = -\log \beta,$$

while $U_c = 0$ otherwise. This confirms the lack of interaction in the Poisson process. It validates also, the choice of this process to model patterns exhibiting no particular morphological structure.

Distance interaction model - Strauss model : (Strauss, 1975), (Kelly and Ripley, 1976)

$$p(x) = \alpha\beta^{n(x)}\gamma^{s_r(x)}, \quad \alpha, \beta > 0, \gamma \in [0, 1]$$

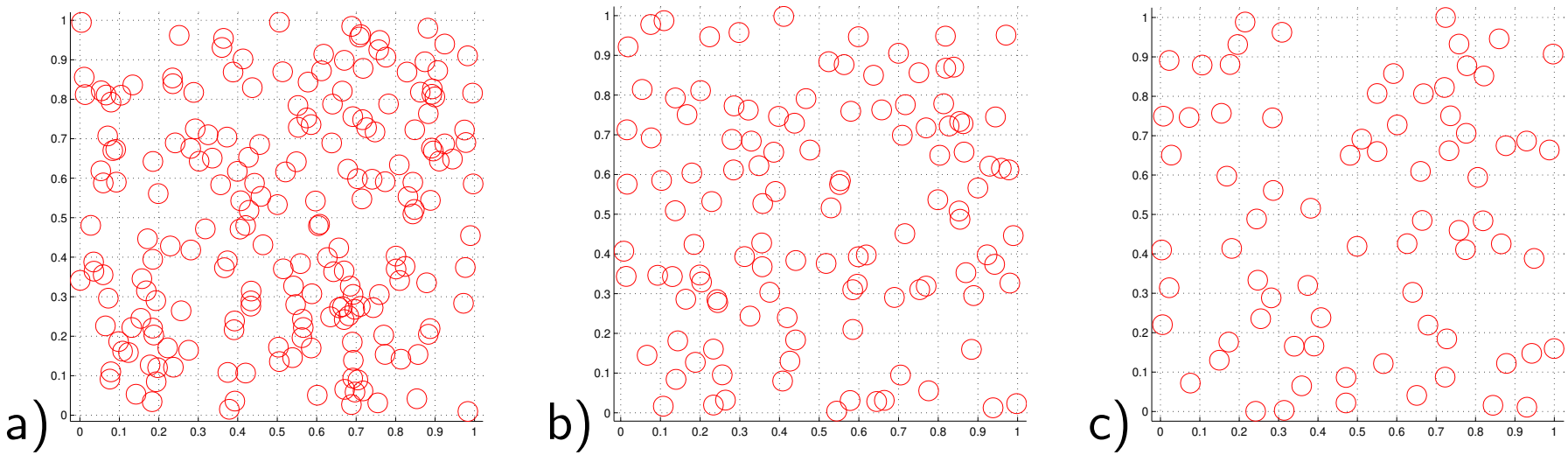


Figure: Strauss model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma = 0.5$ and c) $\gamma = 0.0$.

The interaction function $\gamma : W \times W \to [0, 1]$ is

$$\gamma(u, v) = \begin{cases} \gamma & \text{if } d(u, v) \leq r \\ 1 & \text{otherwise} \end{cases}$$

The conditional intensity of adding a point η to $\mathsf{x} \setminus \{\eta\}$ is

$$\lambda(u; \mathsf{x}) = \beta \gamma^{\mathrm{card}\partial(u)}$$

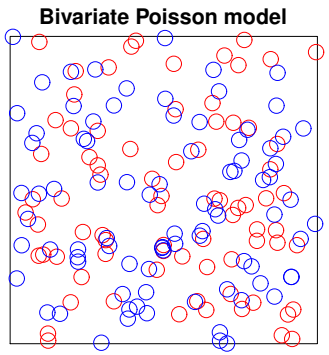
where $\partial(u) = \{v \in \mathsf{x} : d(u, v) \leq r\}$

The Strauss model is a locally stable model with $\Lambda = \beta$ and Markov with interaction range r.
The interaction functions applied to cliques are

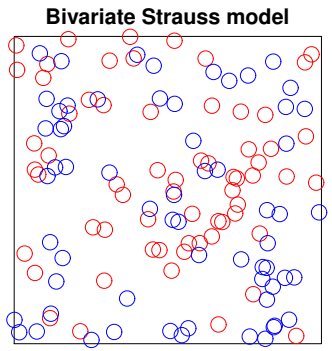
$$\begin{aligned} \phi_c(\emptyset) &= \alpha \\ \phi_c(\{u\}) &= \beta \\ \phi_c(\{u, v\}) &= \gamma(u, v) \end{aligned}$$

and $\phi_c \equiv 1$ if the cliques have three or more objects. The interaction potentials are obtained taking $U_c = -\log \phi_c$.

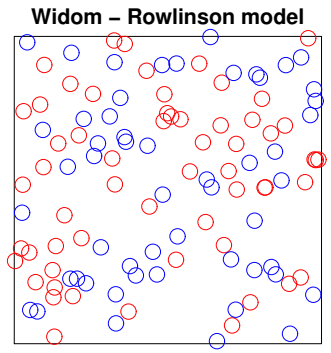
Multi-type pairwise interaction processes



a)



b)



c)

Figure: Bivariate pairwise interaction processes with $r = 0.05$ and : a) $\gamma_{1,2} = \gamma_{2,1} = 1.0$, b) $\gamma_{1,2} = \gamma_{2,1} = 0.75$ and c) $\gamma_{1,2} = \gamma_{2,1} = 0$. Circles around the points have a radius of 0.025.

Widom-Rowlinson or penetrable spheres model : this model is described by the mark space $M = \{1, 2\}$ and the density

$$p(\mathrm{x}) = \alpha \prod_{(w,m)\in \mathrm{x}} \beta_m \prod_{(u,1),(v,2)\in \mathrm{x}} 1\{\| u - v \| > r\} \qquad (6)$$

w.r.t the standard Poisson point process on $W \times M$ with $\nu_M(1) = \nu_M(2)$.
The parameters $\beta_1 > 0$ and $\beta_2 > 0$ control the number of particles of type 1 and 2, respectively.
The conditional intensity for adding $(w, 1) \notin \mathrm{x}$ to the configuration x is

$$\lambda((w, 1); \mathrm{x}) = \beta_1 1\{d(u, w) > r \text{ for all the } (u, 2) \in \mathrm{x}\}.$$

A similar expression is available for adding an object of type 2.

The Widom-Rowlinson is hereditary and locally stable with

$$\Lambda = \max\{\beta_1, \beta_2\}.$$

Furthermore, $\lambda((w,m);\mathrm{x}') \geq \lambda((w,m);\mathrm{x})$ for all $\mathrm{x}' \subseteq \mathrm{x}$ and $(w,m) \in W \times M$.
The interaction functions are

$$\begin{aligned} \phi_c(\emptyset) &= \alpha \\ \phi_c(\{(w,m)\}) &= \beta_m \\ \phi_c(\{(u,1),(v,2)\}) &= 1\{d(u,v) > r\} \end{aligned}$$

and $\phi_c \equiv 1$ if the cliques have two or more objects of the same type.

Multi-type pairwise interaction process : consider $M = \{1, \ldots, I\}$ with $I \in \mathbb{N}$ and ν_M the uniform distribution on M. The probability density w.r.t the standard multi-type process is

$$p(\mathrm{x}) = \alpha \prod_{(w,m)\in \mathrm{x}} \beta_m \prod_{(u,i)\neq(v,j)\in \mathrm{x}} \gamma_{ij}(d(u,v)). \tag{7}$$

- the parameters $\beta_m > 0$, $m \in M$ control the intensity of the points of type m.
- the measurable functions $\gamma_{ij} : [0, \infty) \to [0, 1]$ describe the interaction between each type pair of objects $i, j \in M$
- symmetric functions : $\gamma_{ij} \equiv \gamma_{ji}$ for all $i, j \in M$

For $(w, m) \notin \mathrm{x}$, the conditional intensity is

$$\lambda((w, m); \mathrm{x}) = \beta_m \prod_{(u,i) \in \mathrm{x}} \gamma_{im}(d(u, w)).$$

This process is locally stable with $\Lambda = \max_{m \in M} \beta_m$, anti-monotonic and Markov under smooth assumptions on the functions γ_{ij}.
The interaction functions are

$$\begin{aligned} \phi_c(\emptyset) &= \alpha \\ \phi_c(\{(w, m)\}) &= \beta_m \\ \phi_c(\{(u, i), (v, j)\}) &= \gamma_{ij}(d(u, v)) \end{aligned}$$

with $\phi_c \equiv 1$ for cliques of three objects and more.

Area interaction model :

(Baddeley and van Lieshout, 1995)

$$p(x) \propto \beta^{n(x)} \gamma^{-\nu[\Gamma(x)]}, \quad \beta, \gamma > 0 \tag{8}$$

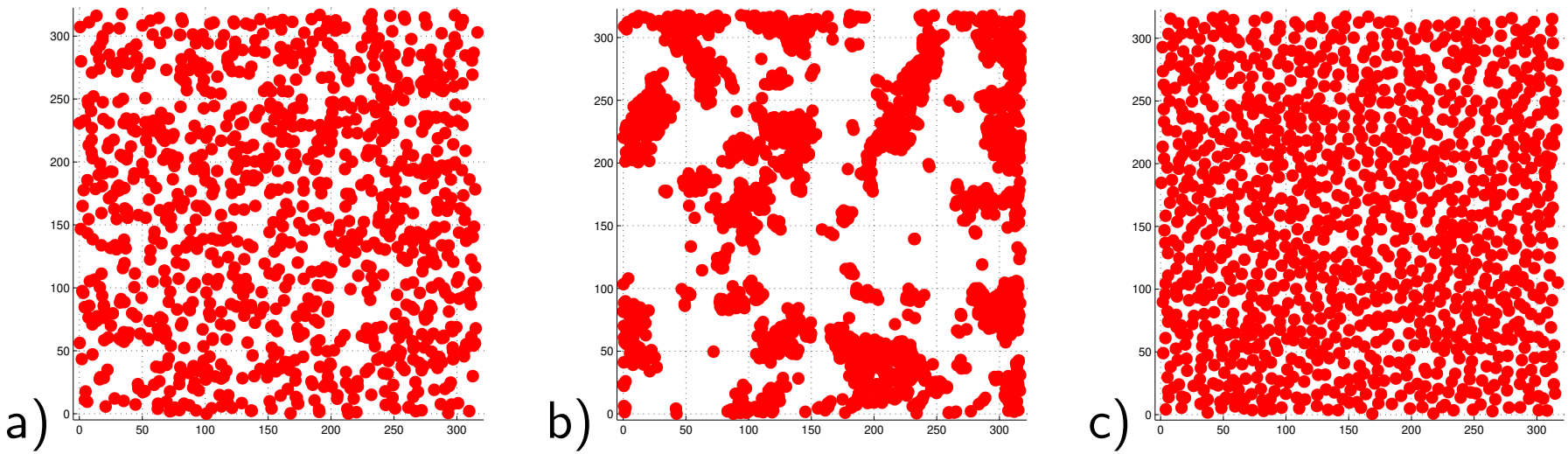


Figure: Area interaction model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma > 1.0$ and c) $\gamma < 1.0$.

Remarks :

- the first probability density based point process producing clusters $\rightarrow$ alternative to the Strauss process ...
- the model should be re-parametrized in order to be identifiable

Proposition

The area interaction process given by (8) *is a Markov point process.*

Proof.

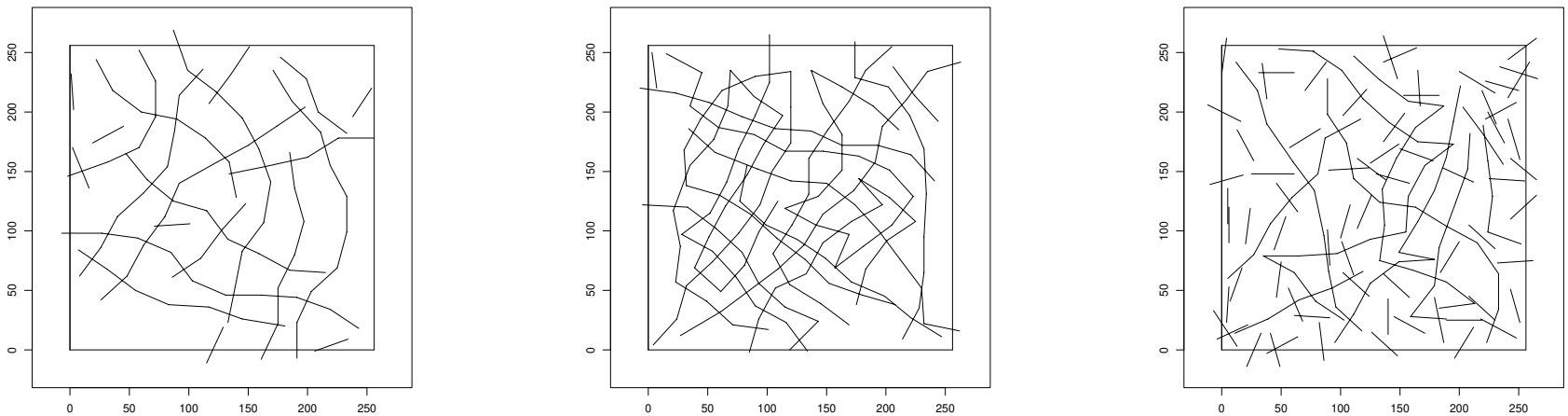
$\rightarrow$ left as an exercise $\square$

Candy model :

(van Lieshout and Stoica, 2003), (Stoica, Descombes and Zerubia, 2004)

$$p(x) \propto \gamma_f^{n_f(x)} \gamma_s^{n_s(x)} \gamma_d^{n_d(x)} \gamma_o^{n_o(x)} \gamma_r^{n_r(x)},$$

with $\gamma_f, \gamma_s, \gamma_d > 0$ and $\gamma_o, \gamma_r \in [0, 1]$



a) b) c)

Figure: Candy model realisations.

Bisous model :

(Stoica, Gregori and Mateu, 2005)

$$p(x) \propto \left[\prod_{s=0}^{q} \gamma_s^{n_s(x)}\right] \prod_{\kappa \in \Gamma \subset \mathcal{R}} \gamma_\kappa^{n_\kappa(x)} \quad \gamma_s > 0, \gamma_\kappa \in [0,1]$$

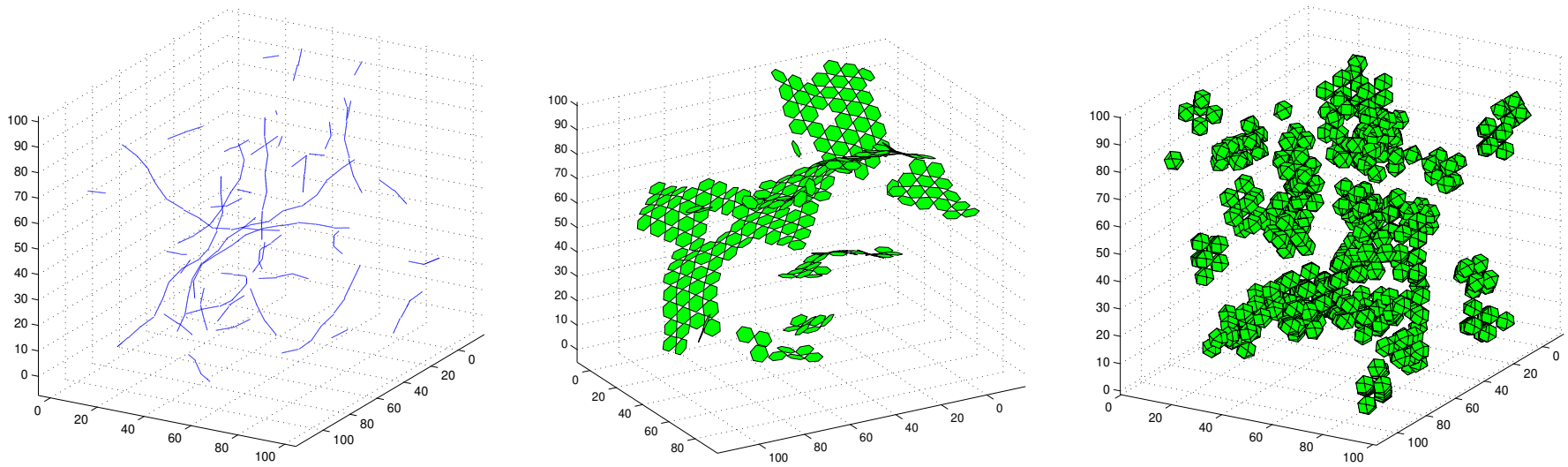


Figure: Random shapes generated with Bisous model.

Remarks :

- Candy and Bisous are based on compound interactions $\rightarrow$ drawing + explanations
- connections are produced by giving different weights for the repulsive interactions
- the conditional intensity is bounded

$$\lambda(\zeta; x) \leq \prod_{s=0}^{q} \max\{\gamma_s, \gamma_s^{-1}\}^{12} = \Lambda.$$

  this gives the name of the model $\rightarrow$ kissing number
- $\rightarrow$ blackboard - Candy
- Markov range : $4r_h + 2r_a$
- the models are locally stable but the exact simulation is sometimes difficult ...

Compare two random sets : idea inspired by work with M. N. M. van Lieshout and classical literature in mathematical morphology

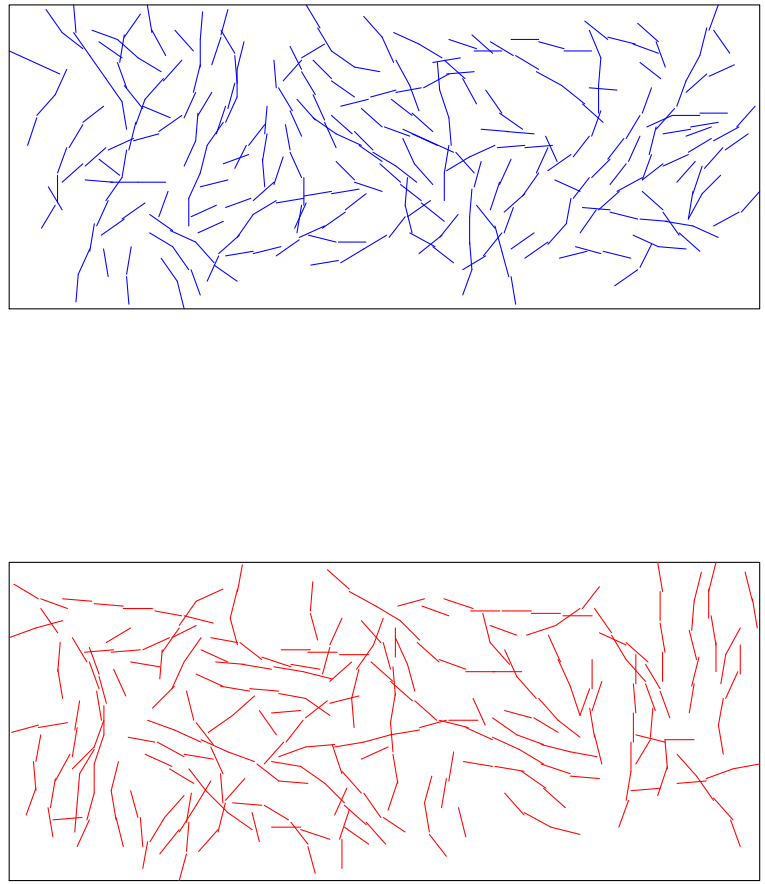


Figure: Realizations of the Candy model obtained with different samplers.

Empty space function : these probability distributions should be similar $\Rightarrow$ Kolmogorov-Smirnov $p-$ value is higher than 0.8

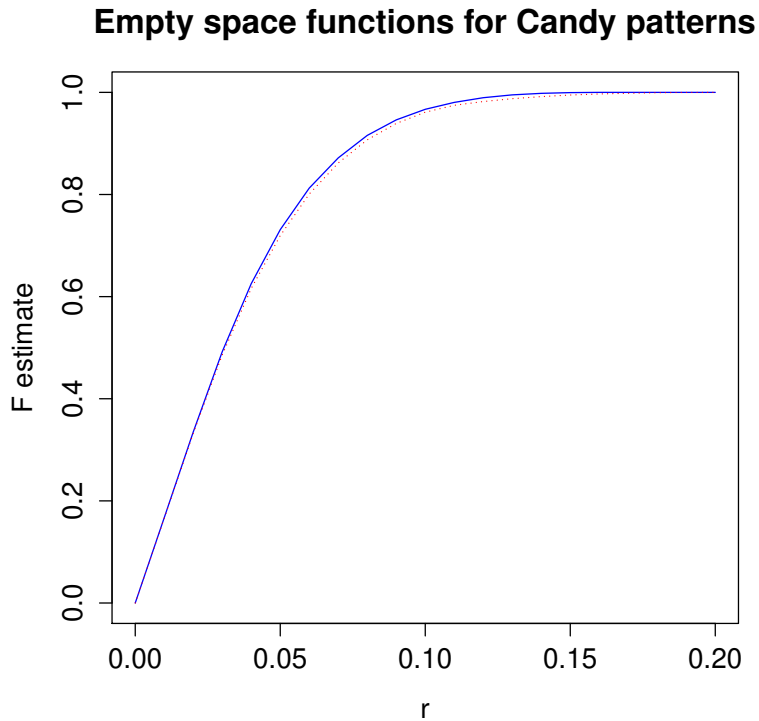


Figure: Estimation of the empty space function for the previous Candy realizations

Table of contents

Principles of Markov chain Monte Carlo simulation

+ WOOCLAP

Problem: sampling or simulation probability distributions

$$\pi(A) = \int_A p(\mathrm{x}) d\mu(\mathrm{x})$$

that are not available in closed form $\leftrightarrow$ normalising constant analytically intractable

Principle: simulate a Markov chain that has π as equilibrium distribution

Basic MCMC sampler

Algorithm

$x = \textit{My first MCMC sampler}\ (T)$

1. *choose an initial condition* x_0
2. *for* $i = 1$ *to* T, *do*
$$\{ \quad x_i = \textit{Update}(x_{i-1}) \quad \}$$
3. *return* x_T.

Principles of MCMC sampling algorithms :

- simulate a Markov chain
- the *Update* function reproduces the transition kernel of the considered Markov chain
- whenever $T \to \infty$, the output x_T is asymptotically distributed according to π, the distribution of interest
- if the Markov chain has good properties $\to$ statistical inference is possible from the obtained simulations
- several existing solutions : Gibbs sampler, Metropolis-Hastings, birth and death processes, stochastic adsorption, RJMCMC, exact simulation (CFTP, clan of ancestors, etc.)

Important properties of the Markov chain transition kernel: these properties guarantee the convergence of the Makov chain simulation algorithm

- irreducibility: all the states are accessible from any other state
- aperiodicity: no deterministic "loops"
- recurrence: all the states are visited often enough
- ergodicity: the convergence speed of the chain is good enough and the dependence of the initial conditions is controlled

Coarse interpretation of the role of these properties:

- irreducibility and aperiodicity: if the chain has an invariant distribution, then convergence towards the equilibrium distribution holds everywhere, exception of a set of initial conditions of null measure
- recurrence: gets rid of the previous null set - independence of the initial conditions
- ergodicity: Central Limit Theorem holds for the obtained samples

Remark:

- simulating or sampling marked point processes involves the construction of MCMC dynamics on a general state space $\rightarrow$ adapting the previous properties to this context

Metropolis-Hastings algorithm

Principle :

- consider the chain in the state $x_i = x$
- propose a new state $x_f = y$ using the proposal density $q(x_i \to x_f)$
- accept this new state with probability

$$\alpha(x, y) = \min\left\{1, \frac{p(y)q(y \to x)}{p(x)q(x \to y)}\right\}$$

  if not remain in the previous state
- iterate as many times as we need (... in theory till infinity ...)

Properties

- $\alpha(\cdot, \cdot)$ is a solution of the detailed balance equation $\rightarrow$ reversibility is preserved
- very few conditions are required for $q(\cdot \rightarrow \cdot)$ so that the chain has all the convergence properties
- $q(\cdot \rightarrow \cdot)$ should be simple to calculate and to simulate
- the knowledge of the normalizing constant of $p(\cdot)$ is not needed

$\rightarrow$ discuss the derivation of the acceptance probability and Exercise 17

MH algorithm for sampling marked point processes

Idea : taking into account the structure of the configuration space, the transition kernel propose to add an object to the configuration with probability p_b or propose to delete an object from the configuration with the probability p_d

Birth : add an object

- initial state : $x_i = \mathrm{x}$ an object configuration
- final state : $x_f = \mathrm{x} \cup \{\zeta\}$
- proposal density to add an object : choose uniformly its location in W and its mark independently according to ν_M

$$q(x_i \to x_f) = q(\mathrm{x} \to \mathrm{x} \cup \{\zeta\}) = p_b \frac{1\{\zeta_w \in W\}}{\nu(W)}$$

- proposal density to remove an object : choose uniformly an object from $\mathsf{x} \cup \{\zeta\}$

$$q(x_f \to x_i) = q(\mathsf{x} \cup \{\zeta\} \to \mathsf{x}) = p_d \frac{1\{\zeta \in \mathsf{x} \cup \{\zeta\}\}}{n(\mathsf{x}) + 1}$$

- acceptance probability

$$\alpha(\mathsf{x} \to \mathsf{x} \cup \{\zeta\}) = \min \left\{ 1, \frac{p_d p(\mathsf{x} \cup \{\zeta\})}{p_b p(\mathsf{x})} \times \frac{\nu(W)}{n(\mathsf{x}) + 1} \right\} \quad (9)$$

Death : remove an object

- the inverse movement of birth
- acceptance probability

$$\alpha(\mathsf{x} \to \mathsf{x} \setminus \{\zeta\}) = \min\left\{1, \frac{p_b p(\mathsf{x} \setminus \{\zeta\})}{p_d p(\mathsf{x})} \times \frac{n(\mathsf{x})}{\nu(K)}\right\} \qquad (10)$$

A transition kernel doing these transformations is

$$
\begin{aligned}
P(\mathrm{x}, A) = p_b \int_K b(\mathrm{x}, \eta)\alpha(\mathrm{x}, \mathrm{y} := \mathrm{x} \cup \{\eta\})1\{\mathrm{y} \in A\} d\sigma(\eta) \\
+ \quad p_d \sum_{\eta \in \mathrm{x}} d(\mathrm{x}, \eta)\alpha(\mathrm{x}, \mathrm{y} := \mathrm{x} \setminus \{\eta\})1\{\mathrm{y} \in A\} \\
+ \quad 1\{\mathrm{x} \in A\} \left[1 - p_b \int_K b(\mathrm{x}, \eta)\alpha(\mathrm{x}, \mathrm{x} \cup \{\eta\}) d\sigma(\eta) \right. \\
- \quad \left. p_d \sum_{\eta \in \mathrm{x}} d(\mathrm{x}, \eta)\alpha(\mathrm{x}, \mathrm{x} \setminus \{\eta\}) \right],
\end{aligned}
$$

where $K = W \times M$, $d\sigma(\eta) = d\sigma((w, m)) = d\nu(w) \times d\nu_M(m)$ et $0 < p_b + p_d \leq 1$. The birth rate is $b(\mathrm{x}, \eta) = \frac{1}{\nu(W)}$ and the death rate is $d(\mathrm{x}, \eta) = \frac{1}{n(\mathrm{x})}$

Remarks :

- the Papangelou intensity appears in the acceptance probability
- local stability property guarantees good convergence properties of the Markov chain
- $\rightarrow$ blackboard : discuss reversibility

Algorithm

$\mathrm{y} = \textit{Update}(\mathrm{x})$

1. *Choose "birth" or "death" with probabilities* p_b *and* p_d*, respectively.*
2. *If "birth" was chosen, then generate a new object following* $b(\mathrm{x}, \eta)$*. Accept the new configuration,* $\mathrm{y} = \mathrm{x} \cup \{\eta\}$ *with the probability* $\alpha(\mathrm{x}, \mathrm{y})$ *given by* (9)*.*
3. *If "death" was chosen, then select the object to be removed using* $d(\mathrm{x}, \eta)$*. Accept the new configuration,* $\mathrm{y} = \mathrm{x} \setminus \{\eta\}$ *with the probability* $\alpha(\mathrm{x}, \mathrm{y})$ *given by* (10)*.*
4. *Return the present configuration.*

Theorem. Let be b, d and q as described previously. Assume that $b(x, \eta)$ and $d(x, \eta)$ are strictly positive on their corresponding definition domain, respectively, and

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} \left[\sup_{\eta\in W\times M, x\in \Xi_n} \frac{d(x \cup \{\eta\}, \eta)}{b(x, \eta)} \right] \to 0.$$

Fix $p_b, p_d \in (0, 1)$ with $p_b + p_d \leq 1$ and let $p(x)$ be the probability density of a marked point process on $W \times M$. The point process is locally stable and $p(x)$ is built w.r.t the standard Poisson process μ. The MH sampler defined previously simulates a Markov chain with invariant measure $\pi = \int p d\mu$ who is $\phi-$irreducible, Harris recurrent and geometric ergodic.

$\to$ proof: Bonus C

Remark :

- the same result holds if change moves are introduced with care ... $\rightarrow$ explain ...

Optimality of the MH dynamics

- theoretical convergence properties
- local computation
- no need of the normalising constant
- highly correlated samples : only one element changed per accepted transition
- allows improvements : transition kernels that "help" the model

Tailored to the model proposal distribution

$$b(\mathrm{x}, \eta) = \frac{p_1}{\nu(K)} + p_2 b_a(\mathrm{x}, \eta),$$

with $p_1 + p_2 = 1$ and $b_a(\mathrm{x}, \eta)$ a probability density given by

$$b_a(\mathrm{x}, \eta) = \frac{1}{n(A(\mathrm{x}))} \sum_{x \in A(\mathrm{x})} \tilde{b}(x, \eta).$$

- the role of $b_a(\mathrm{x}, \eta)$: propose the birth of a new pointin those regions where the interactions between the new born and the other configuration members is favoured or not penalised by the model
- $A(\mathrm{x})$: the set of marked points in a configuration that are not exhibiting yet "good" interactions

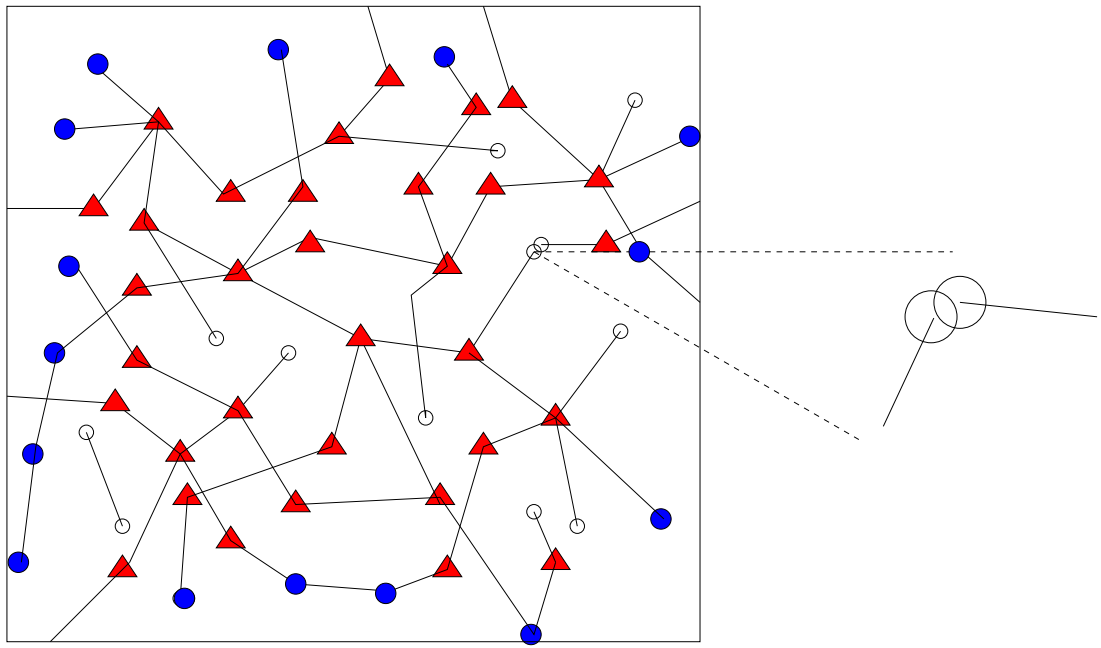
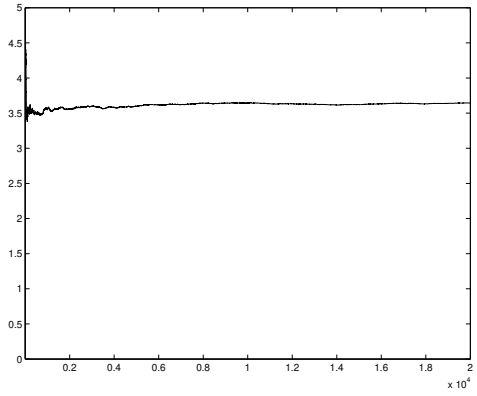
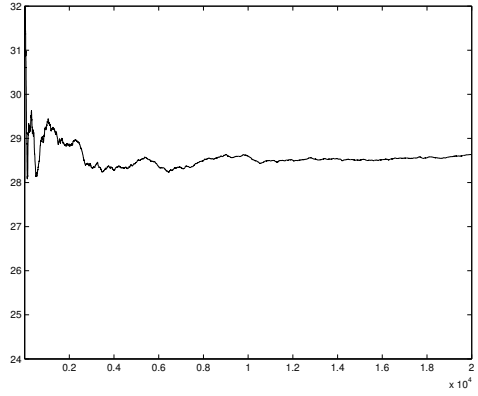


Figure: Extremities marked by triangles are connected and further than $\frac{1}{2}l_{\max} + r_c$ to the boundary, those labeled by a black disk are closer than $\frac{1}{2}l_{\max} + r_c$ to the boundary of K.

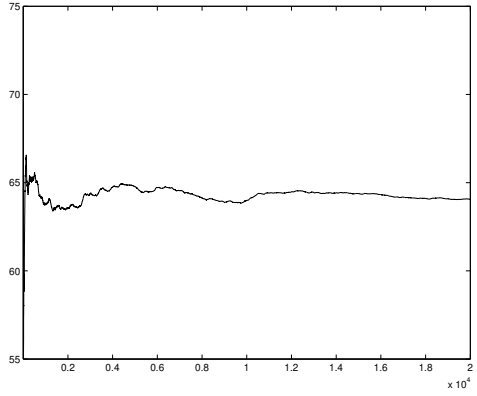
MH algorithm for sampling the Candy model : dynamics behaviour through the sufficient statistics analysis



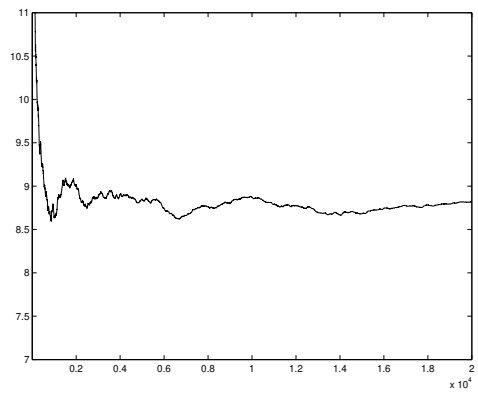
$\bar{n}_f = 3.64$



$\bar{n}_s = 28.63$



$\bar{n}_d = 64.07$



$\bar{n}_o = 8.82$

- great adaptability and theoretical convergence
- easy to use
- appropriate solutions need to be found for each new model
- the general framework, even if it has good theoretical properties, it is not always the most efficient from a numerical point of view
- still, if no good theoretical properties are available, the results will be always bad

→ Exercise 18

Open questions MCMC methods:

Classical algorithms - MH based dynamics

- good convergence properties but convergence at infinity
- burning-in methods + de-correlation techniques
- great adaptability: tailored to the model moves
- manipulate several objects during one move: work of X. Descombes
- link with RJMCMC: great adaptability, but difficult to state convergence proofs, hence difficult to use ...

Perfect simulation algorithms

- the simulated chain indicates by itself whenever convergence is reached: (van Lieshout and Stoica, 2006)
- parameter dependence: can be applied in practice only to a restricted range of parameters
- neither change moves, nor tailored moves
- study existing algorithm: Fill algorithm, forward simulation and simulated tempering
- challenging perspective: synthesis of both families of algorithms $\rightarrow$ exact algorithms able to be tailored to the model

Table of contents

Statistical inference problems

Problem I: pattern detection

- observe the data d and find x "hidden"
- the model parameters are: hidden, modelled, unknown
- open problem: the does the detected pattern really exist ?

Problem II: parameter estimation

- observe the pattern x and find the model parameters θ able to statistically reproduce it
- complete and incomplete data: pseudo-likelihood, Monte Carlo maximum likelihood, EM, ABC Shadow
- open problem : sampling $p(\theta|\mathrm{x})$

Pattern detection and characterisation

The pattern detection problem:

- a spatial data set is observed $\rightarrow$ pattern hidden in the data ?
- hypothesis : the pattern is the realization of a random process (MRFs, marked point processes, etc.)
- the Gibbsian modelling framework $\rightarrow$ write a probability density model

$$p(\mathrm{y}|\theta) = \frac{\exp[-U(\mathrm{x}|\theta)]}{c(\theta)}$$

with $U : \Omega \rightarrow \mathbb{R}^+$ the energy function, θ the model parameters and $c(\theta)$ the normalizing constant

- parameters knowledge $\rightarrow$ prior density $p(\theta)$
- the pattern estimator is

$$(\widehat{\mathrm{x}}, \widehat{\theta}) = \arg \max_{(\mathrm{x},\theta)\in\Omega\times\Theta} \{p(\mathrm{x}|\theta)p(\theta)\} = \arg \max_{(\mathrm{x},\theta)\in\Omega\times\Theta} \{p(\mathrm{x},\theta)\}.$$

Statistical pattern detection

Build the pattern model : probability density construction conditionally on the data observation

$$p(\mathrm{x}, \theta|\mathrm{d}) \propto \exp\left[-\frac{U_{\mathrm{d}}(\mathrm{x}|\theta) + U_i(\mathrm{x}|\theta)}{Z(\theta)} + \log p(\theta)\right]$$

- interaction energy $U_i(\mathrm{x}|\theta) \rightarrow$ objects interactions (geometrical shape of the structure)
- data energy $U_{\mathrm{d}}(\mathrm{x}|\theta)$ induced by the data field d $\rightarrow$ object locations
- if the interaction parameters are unknown $\rightarrow$ prior model $p(\theta)$

- role of the interaction and data energies :

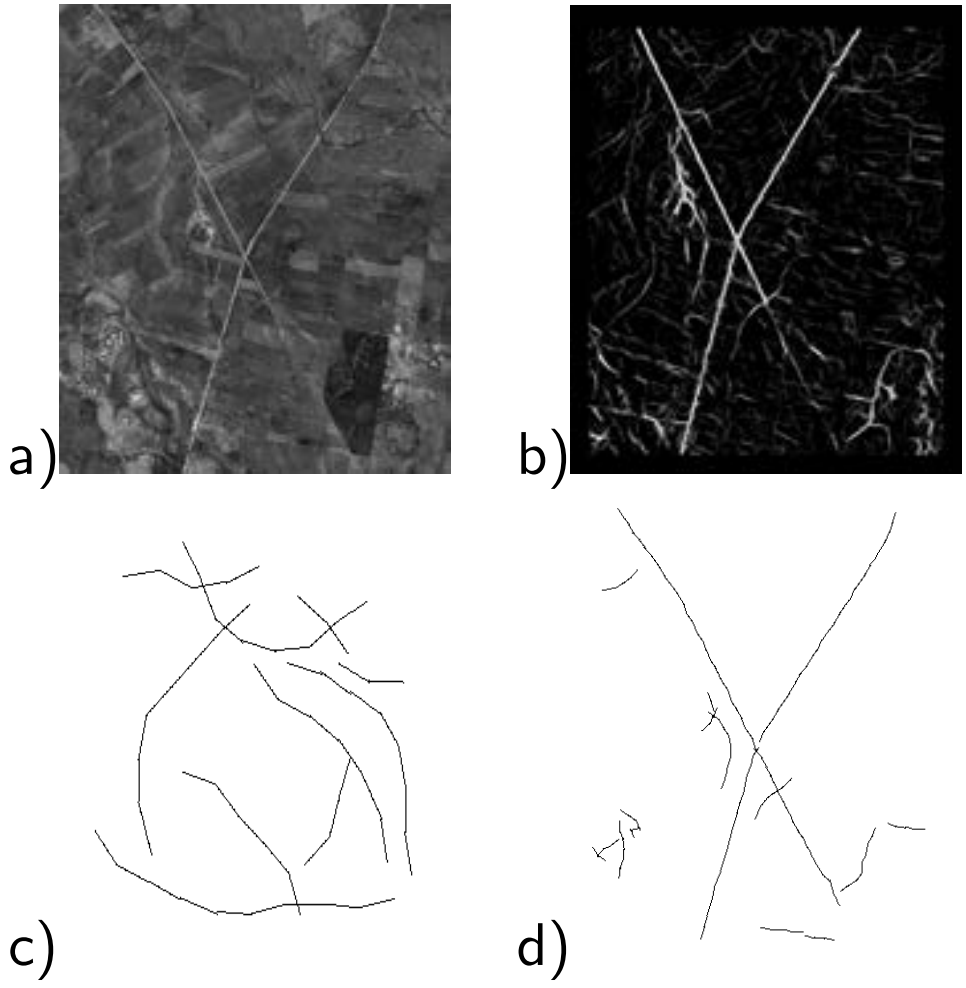


Figure: Influence of the energy components of the model : a) original image SPOT ; and results obtained using only : the data term (b), the interaction term (c), the complete model (d)

- setting the model parameters :

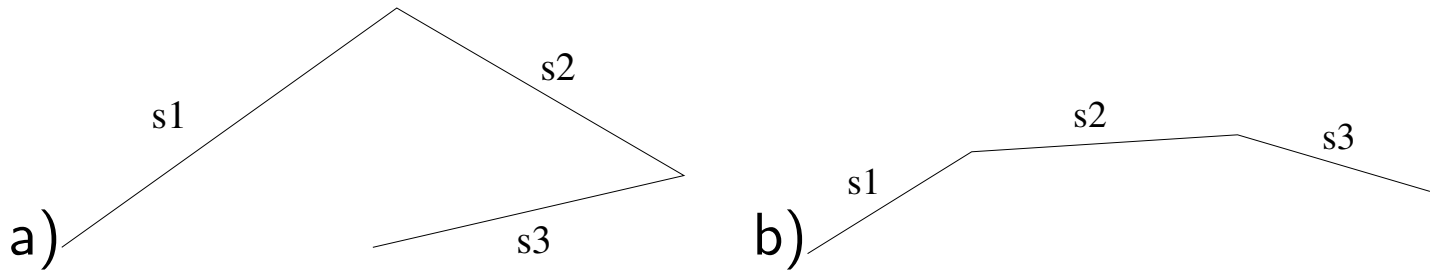


Figure: Two segments configurations : a) the connectivity is favored over alignment, b) connectivity and alignment have equivalent potentials.

Pattern estimator : the object configuration that maximises the probability density

$$(\widehat{\mathrm{x}}, \widehat{\theta}) = \arg \min_{\Omega \times \Psi} \left\{ \frac{U_{\mathrm{d}}(\mathrm{x}|\theta) + U_i(\mathrm{x}|\theta)}{Z(\theta)} - \log p(\theta) \right\}$$

with Ψ the model parameters space.

Simulated annealing : global optimisation technique

- sampling from $p(\mathrm{x}, \theta)^{1/T}$ while slowly $T \to 0$
- convergence towards the uniform distribution on the configuration subspace minimizing $U(\mathrm{x}, \theta)$ (Stoica, Gregori and Mateu, 2005)
- inhomogeneous Markov chain

Algorithm SA : $x = \text{Simulated Annealing}\ (T_0, \delta, T)$

1. choose an initial condition x_0
2. for $i = 1$ to T do
$$\{$$
$$x_i = \text{Update}\ (x_{i-1}, T_{i-1}, \delta)$$
$$T_i = T_0/[\log(i) + 1]$$
$$\}$$
3. return x_T.

- slow algorithm $\rightarrow$ an alternative cooling schedule :

$$T_{n+1} = cT_n \quad \text{with} \quad c \in [0.95, 1[$$

- simulated tempering : improving mixing properties

Level sets estimators :

- visit maps for compact regions in W :

$$\{T(w) > \alpha\} \Rightarrow \{T_n(w) > \alpha\}$$

  with $T(w) = \mathbb{P}(w \in X)$ the probability that the structure hits a point in W

- link with the capacity functional and volumic fraction

- two challenges : discretisation and Monte Carlo approximations
- Vorob'ev expectation : the level set with volume equal to the mean volume of the random set
  - demands the knowledge of the behaviour of an unknown random set, but still manageable in practice ...
- average behaviour of the pattern (fixed temperature)
- (Heinrich, Stoica and Tran, 2012) prove the convergence $\mathbb{L}^1$ of these estimators

Build the machine ...

Filaments detection in galaxies catalogues :

- interaction energy : Bisous model (random cylinders)
- data energy : local tests (density and spread of galaxies inside a cylinder)

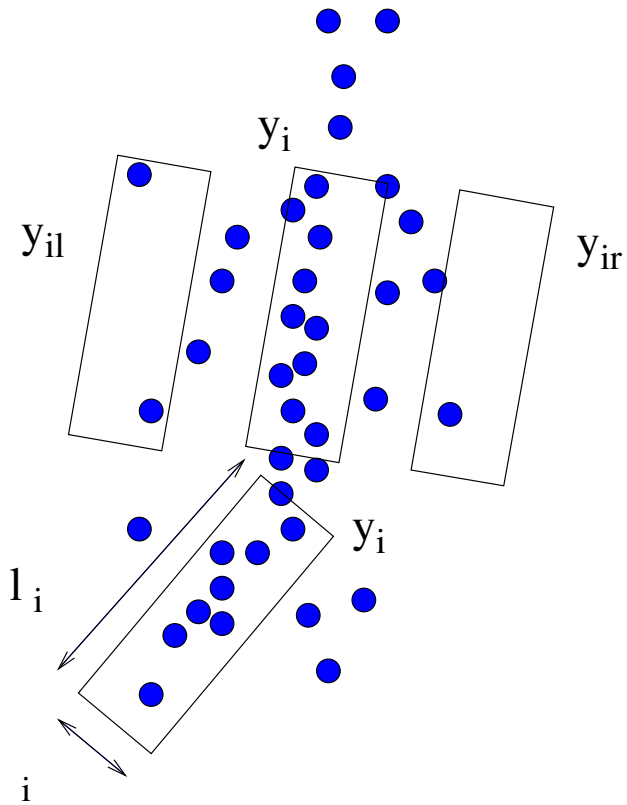


Figure: Locating interacting cylinders in a field of points.

Cluster detection in galaxies catalogues :

- interaction energy : Strauss and Area-interaction models (random object : trunck of a cone + two hal-spheres)
- data energy : local minimum number of galaxies insid the considered random object tests

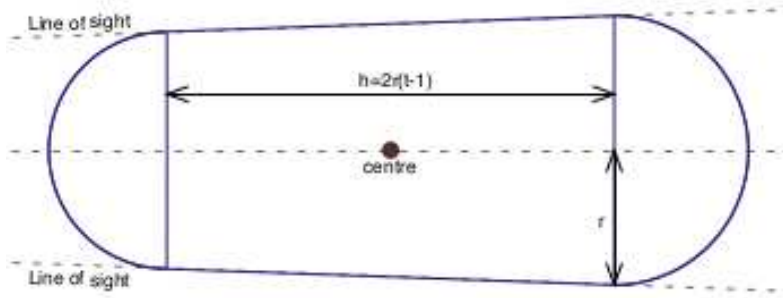


Figure: Cross-section of the considered random object : two half-spheres connected with a truncated cone. The object is fully determined by its centre position, radius r and shape parameter $t \geqslant 1$. Shape parameter t gives the aspect ratio of the object along and perpendicular to the line of sight; for $t = 1$ the object is a ball. For a given r and t the height of the truncated cone is defined as $h = 2r(t - 1)$. The shape of the truncated cone is defined by the lines of sights, which are indicated by dashed lines on the figure. The observer is located at far left from the object.

Cluster detection in epidemiological data :

- interaction energy : Strauss and Area-interaction models (random disks)
- data energy : local statistical test (the average score of the farms covered by a disk)

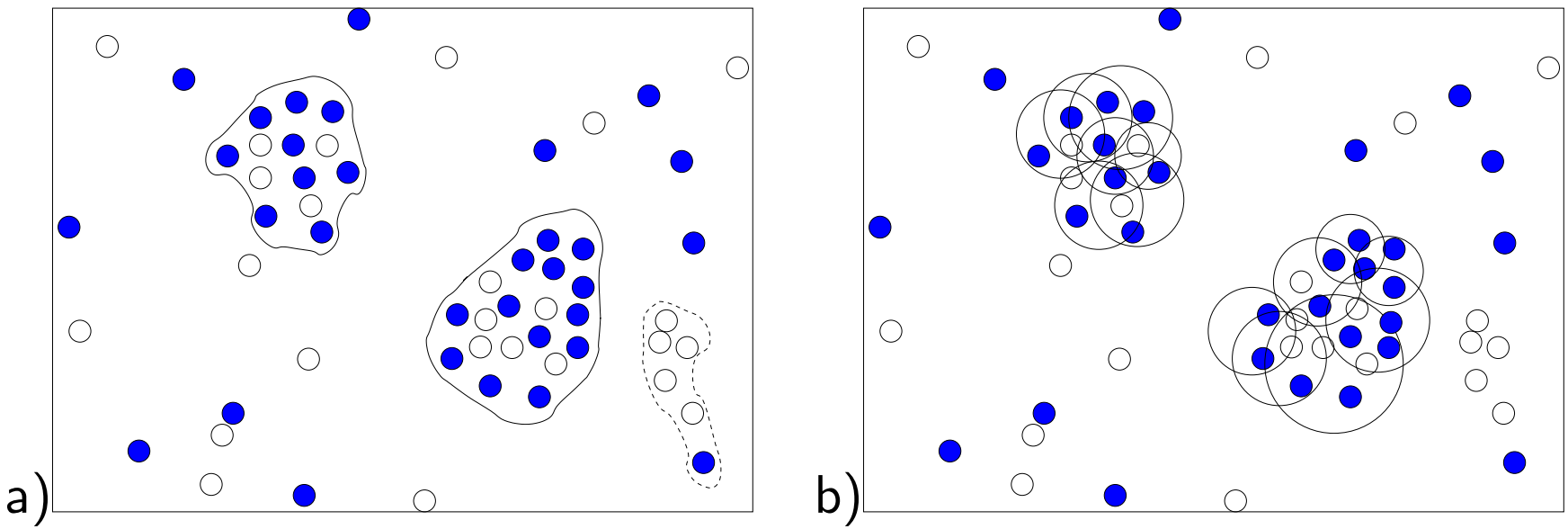


Figure: Data$\rightarrow$ field of marked points : a) observed clusters, b) clusters approximated by random disks.

Orbit determination for binary systems (1)

- interaction energy :

$$U_I(\theta) = \log p(\theta) = \sum_{i=1}^{7} \log p(\theta_i),$$

where $\theta = (a, e, i, \Omega, \omega, \tau, P)$ is the vector of orbital parameters

  - Jeffreys' principle - non-informative independent priors
  - our choice : uniform distributions over bounded intervals
  - the intervals were chosen taking into account the *a priori knowledge* of the objects to be detected
  - perspective : introduce dependence of the parameters

Orbit determination for binary systems (2)

- data energy : sum of the distances between the observed positions and the computed positions ; these last ones are computed using the given model parameters

$$U_{\mathrm{d}}(\varphi|\theta) = \sum_{i=1}^{n} \left[- \left(|x_i^o - x_i^c|^l + |y_i^o - y_i^c|^l \right)^{k/l} \right] =$$

where

- $\mathrm{d} = \{(x_i^o, y_i^o)\}$, $(i = 1, 2, \ldots, n)$: the n observed positions of the secondary asteroid with respect to the primary
- $\{(x_i^c, y_i^c)\}$: the computed positions at the same time i as the corresponding observations, given the current θ
- k, l : pre-fixed model parameters
  - $k = l = 2$: Gaussian character of the data model
  - $k = l = 1$: Laplacian character of the data model
- perspective : model choice

Road network extraction in satellite and aerial images

(Stoica, Descombes, van Lieshout and Zerubia, 2002)

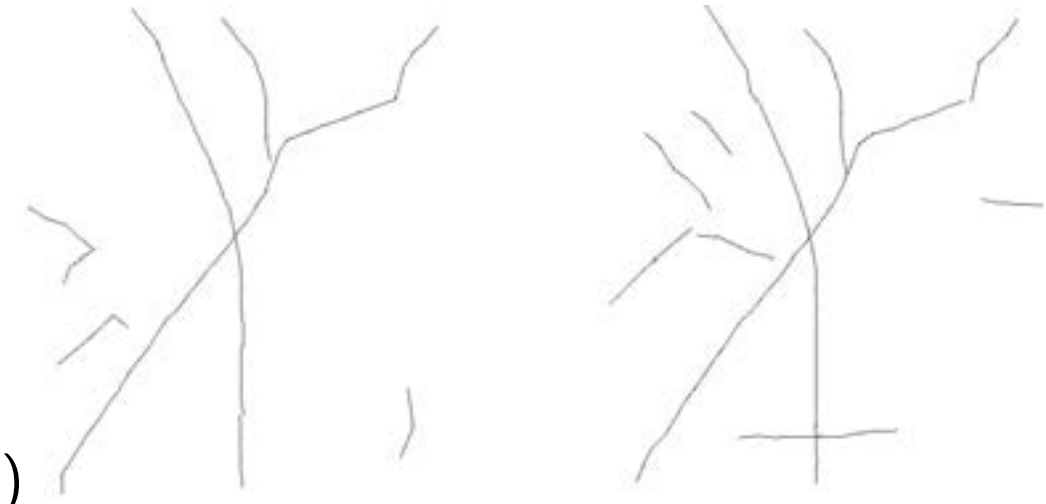
a)  b) 

Figure: Rural region in Malaysia : a) original image; b) obtained results.

Forest galleries : verifying the results

(Stoica, Descombes and Zerubia, 2004)

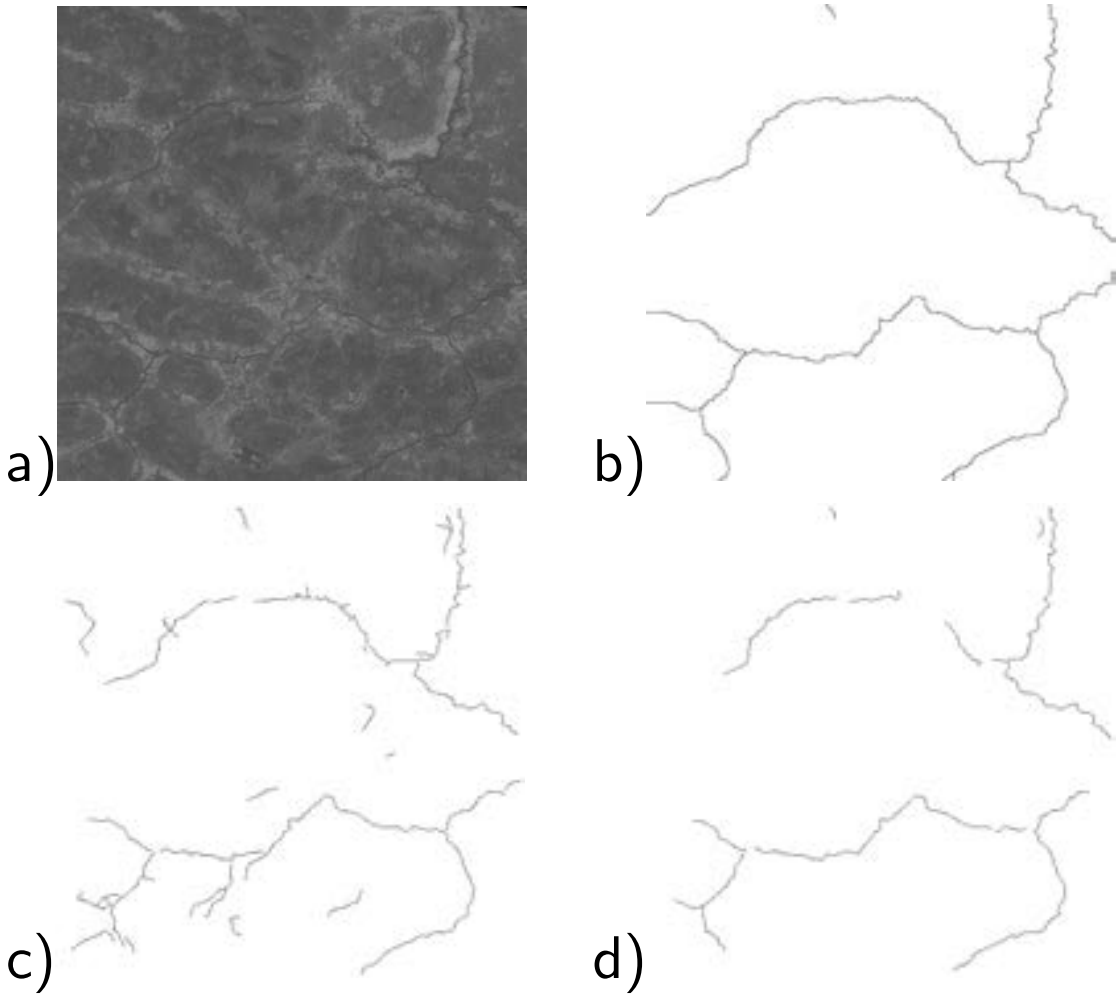


Figure: Forest galleries extraction : a) original image ; b) ground truth ; c)-d) obtained results. Data provided by BRGM.

Filaments detection (1) : (Tempel, Stoica et. al., 2014)

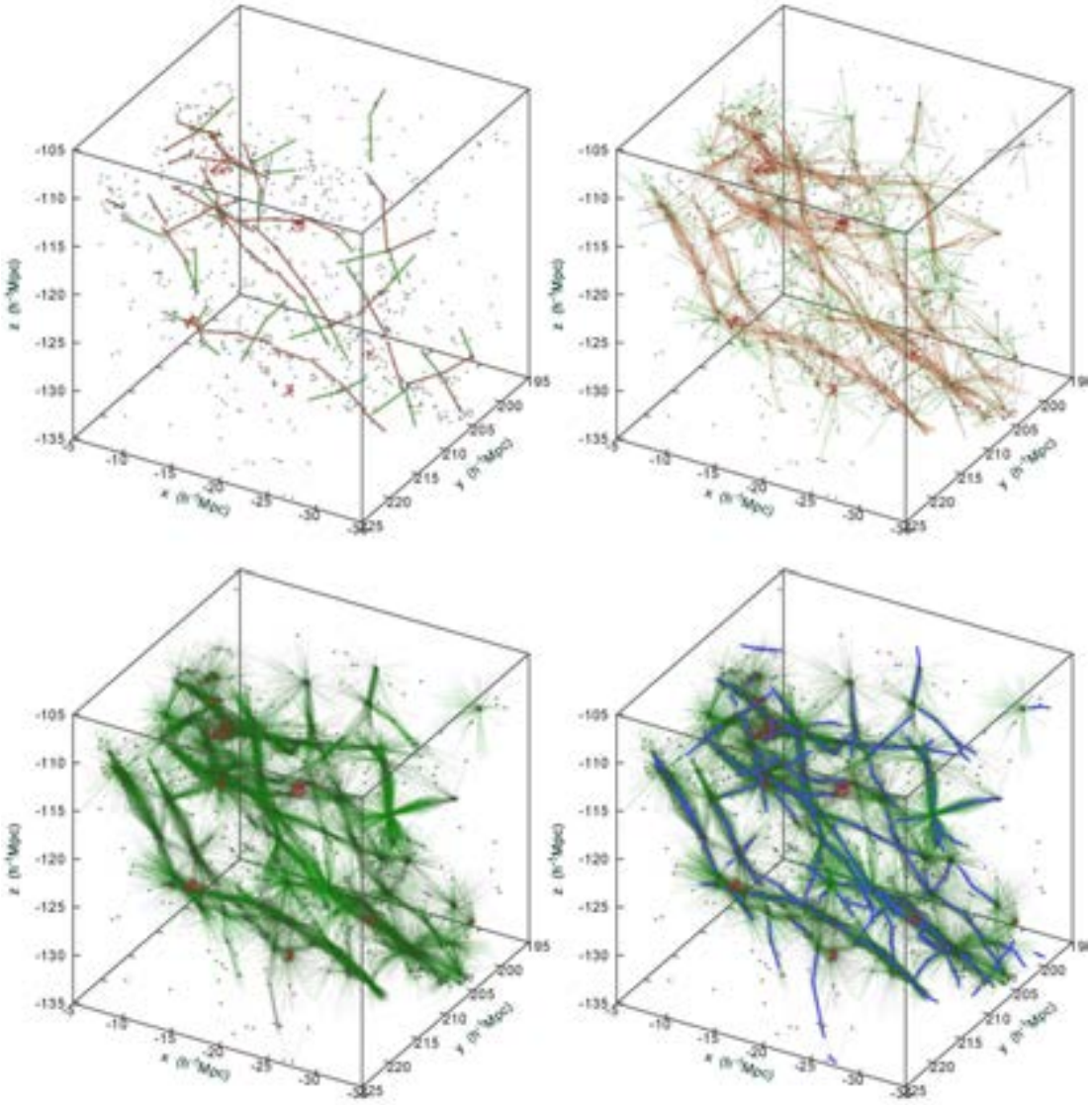


Figure: Detected filamentary pattern (cylinder axes) in a small sample volume within a pattern of galaxies (points).

Filaments detection (2)

(Tempel, Stoica et. al., 2014)
The movie, showing the MCMC in action is available at
:`http://www.aai.ee/ elmo/sdss-filaments/`

Cluster detection in cosmology : (Tempel, et al. and Stoica, 18)

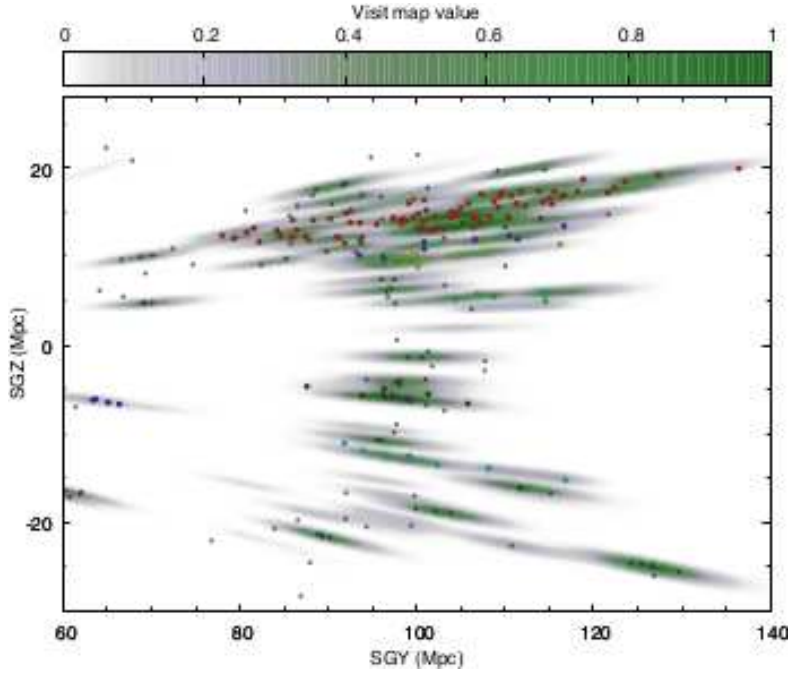
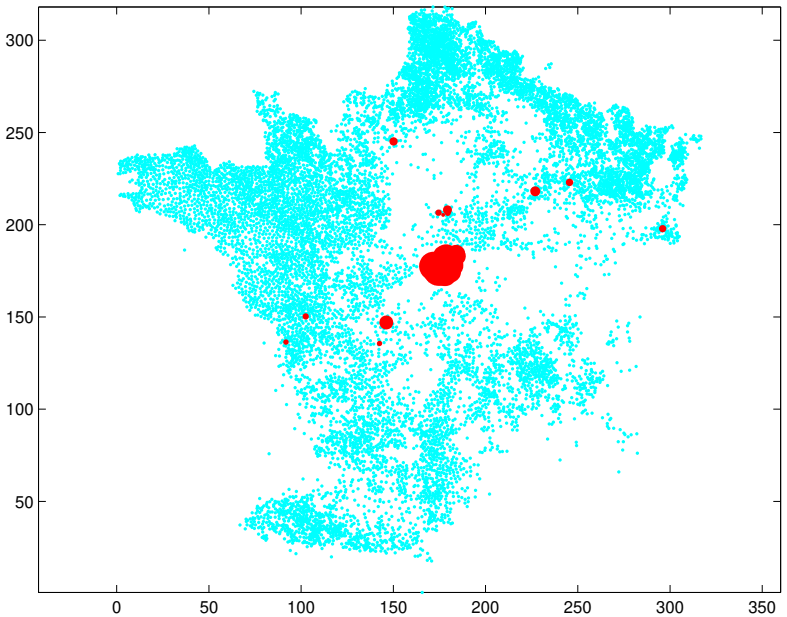


Figure: The distribution of galaxies in supergalactic coordinates (points) and the visit map obtained using the cluster detection process. The thickness of the slice is 4 Mpc around $\mathrm{SGX} = 0$. Red points show galaxies in the Coma cluster while the other coloured points show galaxies in some clusters with at least five members. The grey points show all remaining galaxies.

Cluster detection in epidemiology : sub-clinical mastitis data

(Stoica, Gay and Kretzschmar, 07)

a)



b)

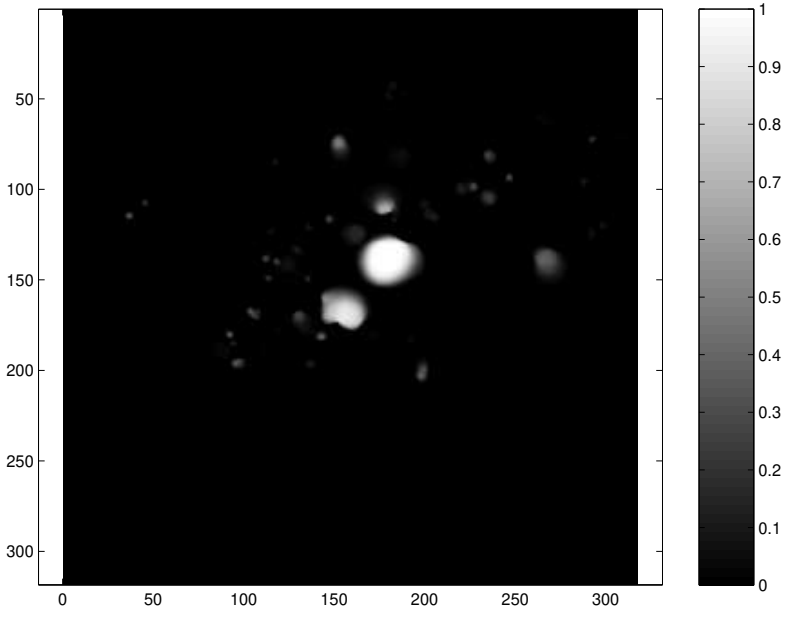


Figure: Disease data scores and coordinates for the year 1996 : a) disk configuration obtained using the simulated annealing algorithm ; b) cover probabilities.

Orbit determination (1) : (Kovalenko, Stoica and Emelyanov, 17)

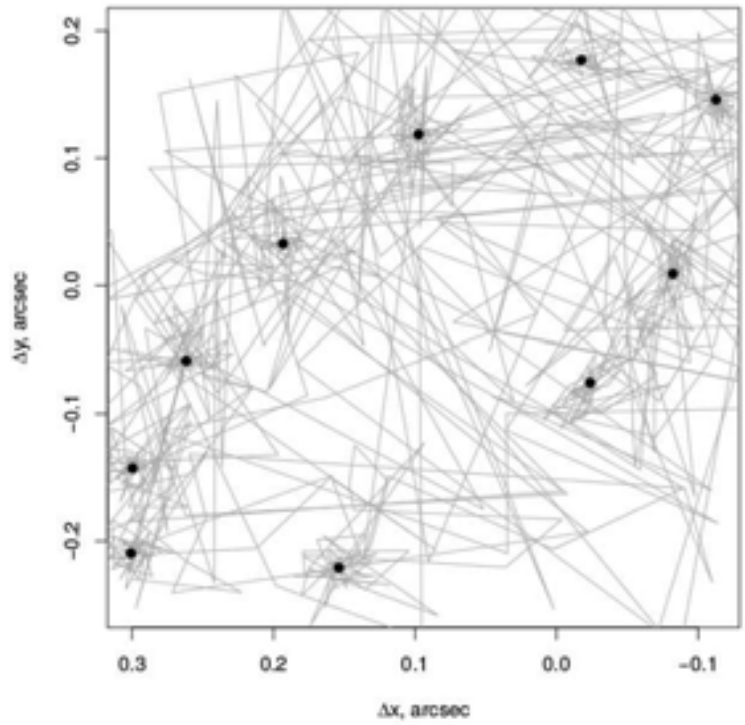


Figure: Simulated observations (black points): Δx and Δy correspond to relative positions of the secondary with respect to the primary. Lines show a search for the optimal solution during SA algorithm.

Orbit determination (2) :

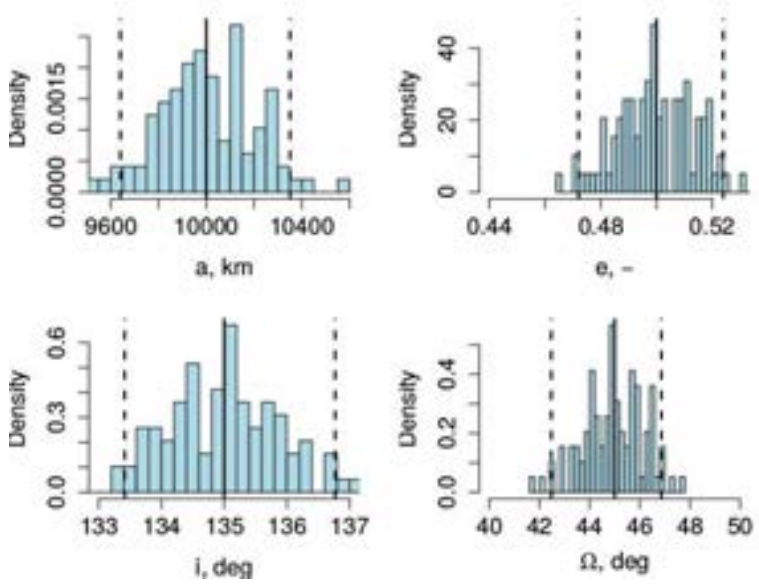


Figure: Resulting distributions of semi-major axis a, eccentricity e, inclination i and longitude of the ascending node Ω (referenced to J2000 equatorial frame) obtained for simulated observations with the likelihood model. Solid line represents the true parameter value. The doted lines represent the 2.5% and 97.5% quantiles of the resulting sample.

Does the detected pattern really exist ?

Idea : the sufficient statistics of the model $\rightarrow$ morphological descriptors of the shape hidden by the data

- turn the machine at constant temperature $T = 1$
- compute the average of the sufficient statistics
- compare with the maximum value obtained for the permuted data

Sufficient statistics :

- Bisous model (pattern of connected cylinders) : free cylinders, cylinders with one extremity connected, cylinders with both extremities connected

Test for the filaments existence in galaxy catalogs

Permuted data : keeping the same number of galaxies while spreading them uniformly (binomial point process)

Sufficient statistics	Data		
	NGP150	NGP200	NGP250
$\bar{n}_2$	4.13	5.83	9.88
$\bar{n}_0$	15.88	21.19	35.82
$\bar{n}_1$	21.35	35.58	46.49

Sufficient statistics	Simulated data (100 binomial catalogs)		
	NGP150	NGP200	NGP250
max $\bar{n}_2$	0.015	0.05	0.015
max $\bar{n}_0$	0.54	0.27	0.45
max $\bar{n}_1$	0.39	0.24	0.33

Test for the cluster existence epidemiological data

Permuted data : keeping the same farm locations while exchanging the score disease

Results :

- sufficient statistics for the data from the year 1996 :

$$\bar{n}(\mathrm{y}) = 74.10, \quad \bar{\nu}[Z(\mathrm{y})] = 312.46, \quad \bar{n}_o = 555.08$$

- maximum values of the sufficient statistics for 100 simulated data fields

$$\bar{n}(\mathrm{y}) = 2.36, \quad \bar{\nu}[Z(\mathrm{y})] = 13.83, \quad \bar{n}_o = 2.62$$

Interpretation : this test does not say if the pattern is well detected, but it says that there is something to be detected ...

Orbit determination validation : position prediction for the obtained parameter values

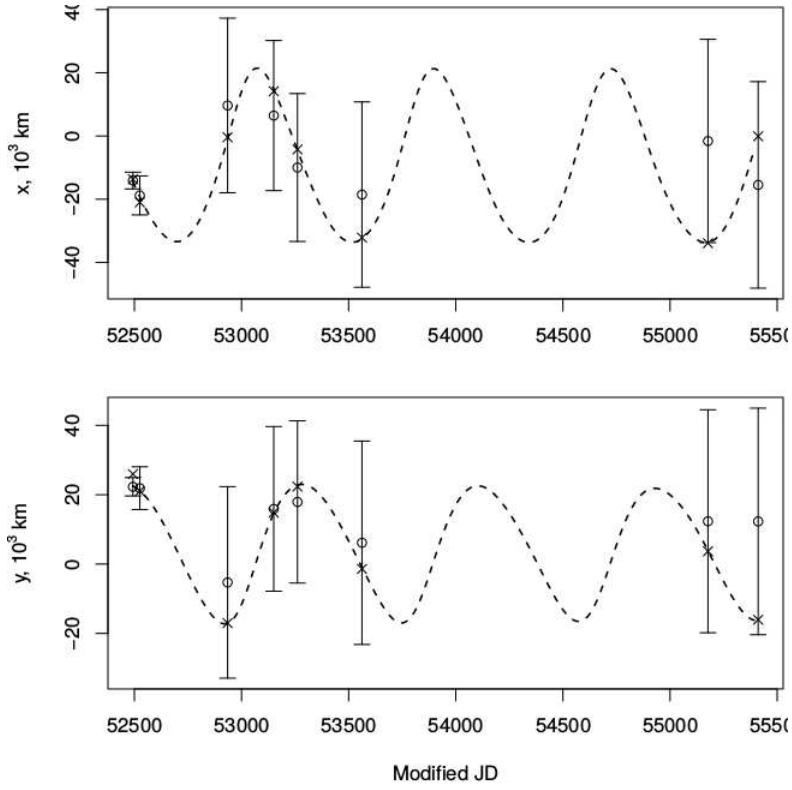


Figure: The calculated positions (black circles) are compared with given observed positions (crosses) by the x and y coordinates on sky-plane. Black bars denote the 2.5%-97.5% quantiles interval. Dotted line corresponds to the calculated positions for the orbit, obtained with the entire set of observations.

How similar are two data sets ?

Cosmology : compare the sufficient statistics for 22 mock catalogues with the ones for the observation (Stoica, Martinez and Saar, 10)

Discussion

- mock catalogues exhibit filaments
- mock filaments are generally shorter, more fragmented and more dense
- Bisous model : good for testing the filamentary structure

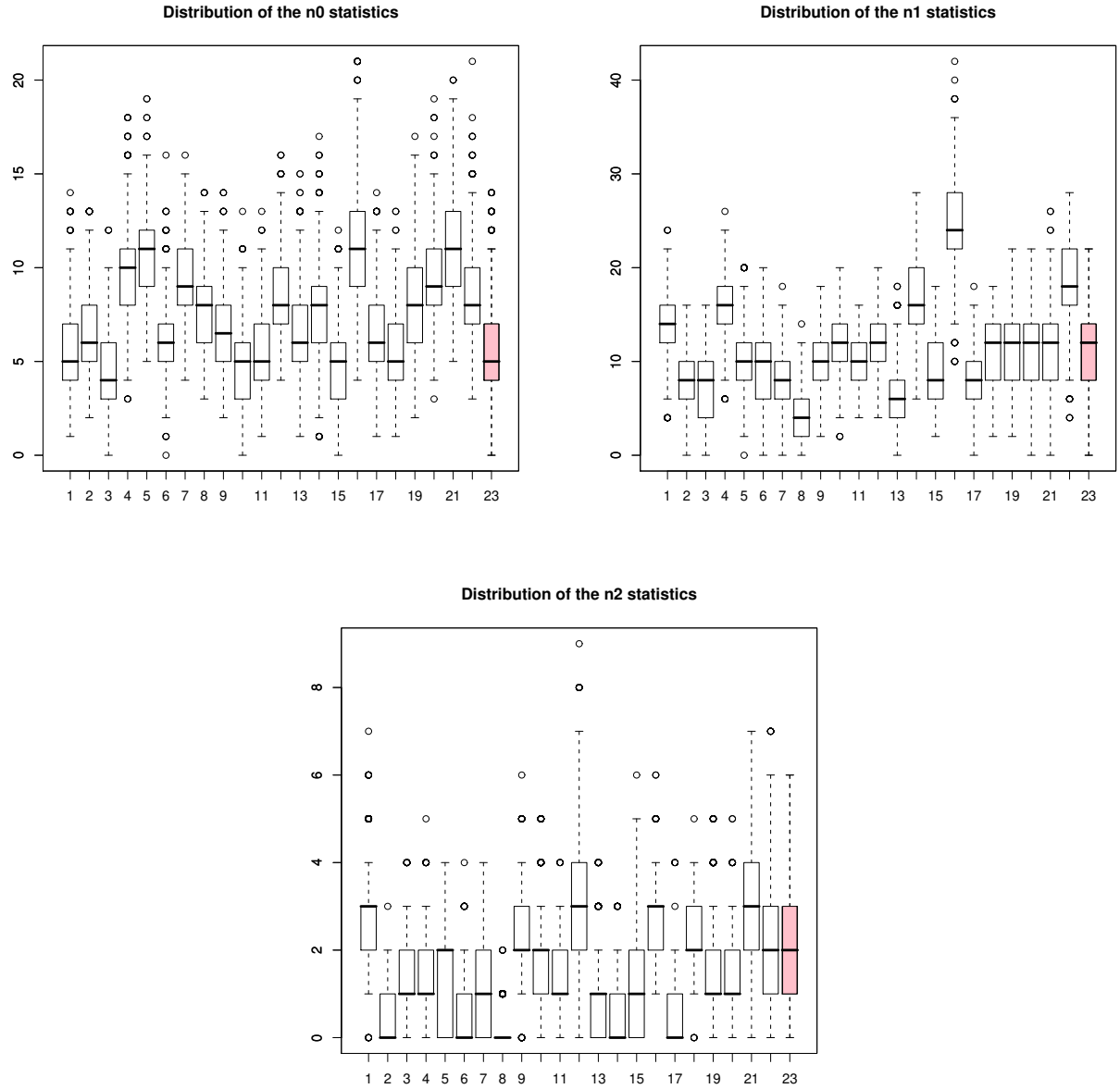


Figure: Comparison of the sufficient statistics distributions for the real data (dark box plot) and the mock catalogues.

Table of contents

Statistical inference problems - please do not forget the suspenders

Problem I: pattern detection

- observe the data d and find x "hidden"
- the model parameters are: hidden, modelled, unknown
- open problem: the does the detected pattern really exist ?

Problem II: parameter estimation

- observe the pattern x and find the model parameters θ able to statistically reproduce it
- complete and incomplete data: pseudo-likelihood, Monte Carlo maximum likelihood, EM, ABC Shadow
- open problem : sampling $p(\theta|\mathrm{x})$

Parameter estimation based on pseudo-likelihood

The pseudo-likelihood of a marked point process $\mathbb{X}$ with conditional intensity $\lambda_\theta(\zeta; \mathrm{x})$ observed on the bounded set W is expressed as

$$PL_W(\theta; \mathrm{x}) = \\ = \left[\prod_{x_i \in \mathrm{x}} \lambda_\theta(x_i; \mathrm{x})\right] \exp\left[-\int_{W \times M} \lambda_\theta((w, m); \mathrm{x}) \nu(dw) \nu_M(dm)\right].$$

The pseudo-likelihood estimator is given by the solution of the equation :

$$\frac{\partial PL_W(\theta; \mathrm{x})}{\partial \theta} = 0$$

Remarks :

- the PL is concave for exponential models
- no normalising constant needed
- it "amplifies" the interaction weights: check the formula - for a Strauss process the interactions are counted twice
- consistency and asymptotic normality of the estimator: under the assumption of a model, considering a realisation observed within a finite window that increases towards $\mathbb{R}^d$, the PL estimator converges towards the true parameters used to produce the observation on the "whole" window
- see (Winkler, 2003), (Jensen and Møller, 1991) and the cited references

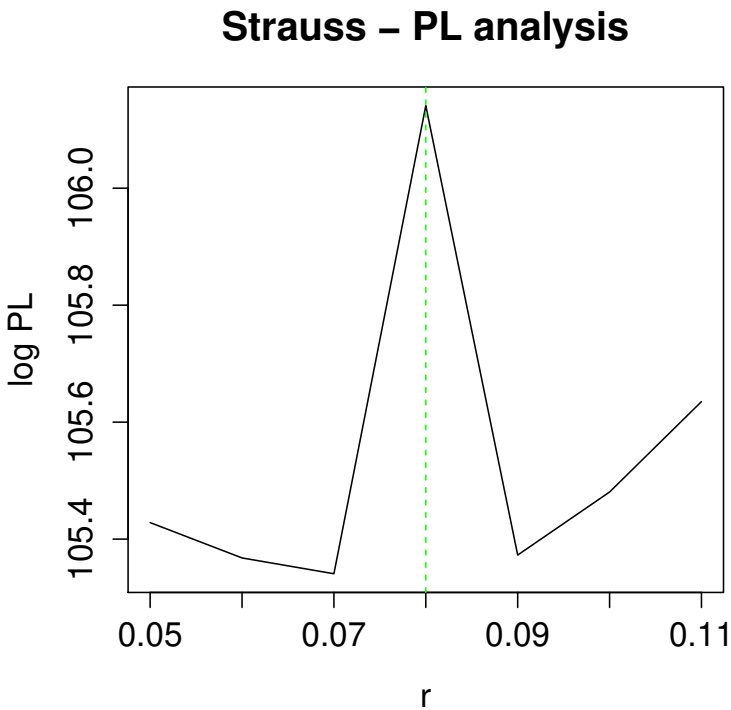
- it can be used as a rather “good” initial condition for other more elaborate methods
  - "good" results for mild interactions: (Mateu and Montes, 2001)
- no control of the estimation obtained from a finite window: one cannot say how far the obtained result is from the true parameters
  - exception: Poisson process - in this case the pseudo-likelihood is the true likelihood
- easy to be implemented :
  - this was the motivation to introduce it in the middle of 70s (Besag, 1975)
  - see (Baddeley, Rubak and Turner, 2016) for implementation details within the **spatstat** package

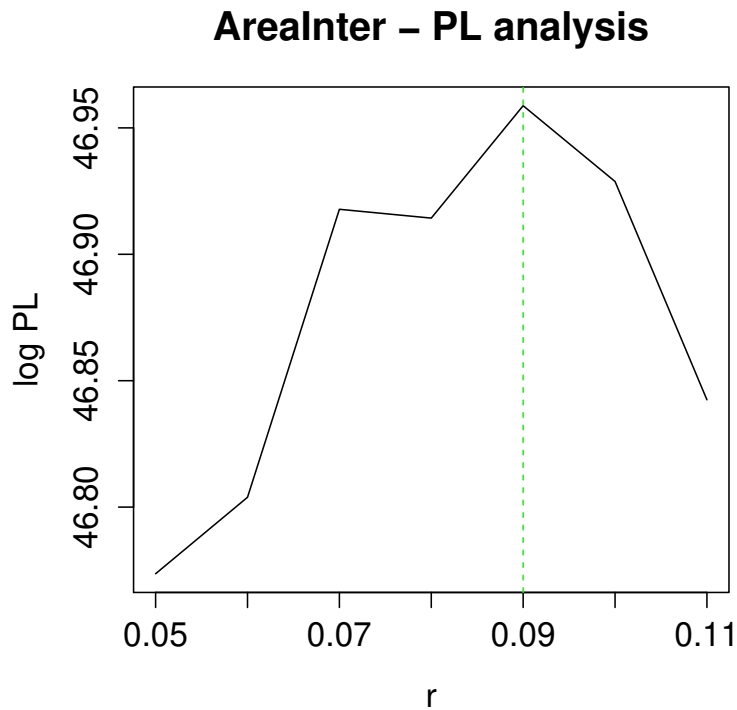
→ Exercicse 19 + Exercise 20

Example application: fitting a point process to real data

Pseudo-likelihood profile analysis : the range parameters

```
>radius = data.frame(r=seq(0.05,0.11, by=0.01))
>pradius = profilepl(radius, Strauss, japanesepines)
>plot(pradius,main="Strauss :  PL analysis")
```





Fitting the model to the pattern:

```
> ppm(japanesepines, 1,Strauss(r=0.08),rbord=0.08)
Stationary Strauss process
First order term:  beta 77.93567
Interaction:  Strauss process interaction distance:
0.08
Fitted interaction parameter gamma:  0.7953
```

Monte Carlo Maximum likelihood estimation

Exponential family models:

- very general framework
- the point processes models that were presented are given by

$$p(\mathrm{x}|\theta) = \frac{h(\mathrm{x}|\theta)}{Z(\theta)} = \frac{\exp\langle t(\mathrm{x}), \theta\rangle}{Z(\theta)}$$

where $h(\mathrm{x}|\theta)$, $t(\mathrm{x})$ and θ represent the un-normalized probability density w.r.t. the standard Poisson process, the sufficient statistics vector and the model parameters vector, respectively. The normalising constant $Z(\theta)$ is unknown.

The configuration x is entirely observed, hence the log-likelihood with respect a known parameter ψ can be written as follows :

$$l(\theta) = \langle t(\mathrm{x}), \theta - \psi \rangle - \log \frac{Z(\theta)}{Z(\psi)}$$

It is easy to check, that the normalizing constants ratio is

$$\frac{Z(\theta)}{Z(\psi)} = \mathbb{E}\left[\exp\langle t(\mathbb{X}), \theta - \psi \rangle\right],$$

since we have

$$\begin{aligned}
\frac{Z(\theta)}{Z(\psi)} &= \frac{1}{Z(\psi)} \int_{\Omega} h(\mathrm{x}|\theta) d\mu(\mathrm{x}) \\
&= \frac{1}{Z(\psi)} \int_{\Omega} h(\mathrm{x}|\theta) \frac{h(\mathrm{x}|\psi)}{h(\mathrm{x}|\psi)} d\mu(\mathrm{x}) \\
&= \int_{\Omega} \frac{h(\mathrm{x}|\theta)}{h(\mathrm{x}|\psi)} \frac{h(\mathrm{x}|\psi)}{Z(\psi)} d\mu(\mathrm{x}) \\
&= \mathbb{E}\left[\frac{h(\mathbb{X}|\theta)}{h(\mathbb{X}|\psi)}\right]
\end{aligned}$$

The Monte Carlo approximation of the normalizing constants ratio gives:

$$\frac{Z(\theta)}{Z(\psi)} \approx \frac{1}{n}\sum_{i=1}^{n} \exp\langle t(\mathbb{X}_i), \theta - \psi\rangle,$$

where $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n$ are samples obtained from $p(\mathrm{y}|\psi)$.
Hence, the Monte-Carlo counterpart of the log-likelihood is :

$$l_n(\theta) = \langle t(\mathrm{x}), \theta - \psi\rangle - \log\left(\frac{1}{n}\sum_{i=1}^{n} \exp\langle t(\mathbb{X}_i), \theta - \psi\rangle\right).$$

Theorem

The log-likelihood of an exponential family model is a convex function.

- proof : see (Monfort 1997, Thm.3, pp. 61)
- $l_n(\theta) \to l(\theta)$ almost sureley
- all these suggest that local optimisation procedures applied to $l_n(\theta)$ may give interesting results

$\to$ Exercise 21 + Exercise 22

MCMC local optimisation procedures

The gradient of the MCMC log-likelihood is

$$\nabla l_n(\theta) = t(\mathsf{x}) - \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})]$$

where

$$\mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})] = \frac{\sum_{i=1}^{n} t(\mathbb{X}_i) \exp\langle t(\mathbb{X}_i), \theta - \psi\rangle}{\sum_{i=1}^{n} \exp\langle t(\mathbb{X}_i), \theta - \psi\rangle}$$

that is the Monte Carlo importance sampling approximation of $\mathbb{E}_\theta t(\mathbb{X})$.

Similarly, the Hessian can be computed too:

$$-\nabla^2 l_n(\theta) = \mathbb{V}ar_{n,\theta,\psi}[t(\mathbb{X})]$$

where

$$\mathbb{V}ar_{n,\theta,\psi}[t(\mathbb{X})] = \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})t(\mathbb{X})^t] - \mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})]\mathbb{E}_{n,\theta,\psi}[t(\mathbb{X})^t].$$

Newton-Raphson method:

$$\theta_{k+1} = \theta_k - [\nabla^2 l_n(\theta_k)]^{-1} \nabla l_n(\theta_k) \tag{11}$$

for $k = 1, 2, \ldots,$

- $l_n(\cdot)$ is computed using n samples from $p(\mathrm{x}|\psi)$
- the computation of the gradient and Hessian inverse is numerically unstable
- useful only if the initial value is close enough from the solution

Iterative gradient method:

$$\begin{cases} l_n(\theta_k + \rho(\theta_k)\nabla l_n(\theta_k)) = \max_{\rho\in\mathbb{R}} l_n(\theta_k + \rho\nabla l_n(\theta_k)) \\ \theta_{k+1} = \theta_k + \rho(\theta_k)\nabla l_n(\theta_k) \end{cases}$$

where $\rho(\theta_k)$ is the optimal step (Descombes et al. '99, Stoica '01).

- re-sampling if $\| \theta_k - \psi \| >$ threshold
- obtain a reference value θ_0 close enough to the maximum likelihood estimator

Stochastic gradient:

$$\theta_{k+1} = \theta_k + \epsilon_k[t(\mathrm{x}) - t(\mathbb{X}_k)]$$

where $\epsilon_k > 0$ is a decreasing sequence while $\mathbb{X}_k$ is a sample of $p(\mathrm{x}|\theta_k)$

- very simple, but finding an optimal sequence $\{\epsilon_k\}$ is an open problem
- L. Younes, G. Winkler : Markov random fields
- R. Moyeed and A. Baddeley : point processes

Asymptotic results MCMCML estimation

The random variable $\sqrt{n}(\widehat{\theta}_n - \widehat{\theta})$ whenever $n \to \infty$, it converges in distribution towards a normal random variable of zero mean and variance $I(\widehat{\theta})^{-1}\Gamma I(\widehat{\theta})^{-1}$:

$$\sqrt{n}(\widehat{\theta}_n - \widehat{\theta}) \to \mathcal{N}(0, I(\widehat{\theta})^{-1}\Gamma I(\widehat{\theta})^{-1}).$$

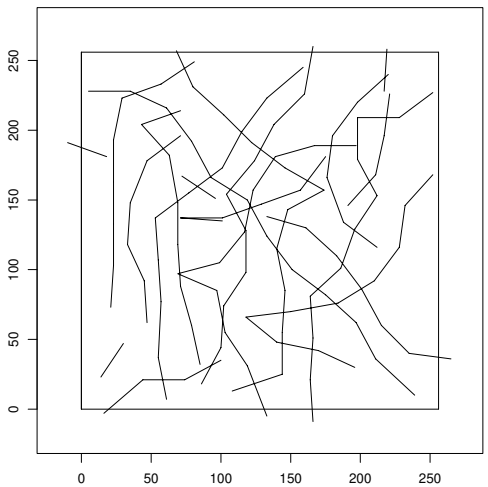
- the matrix
$$I(\widehat{\theta}) = \mathbb{V}ar_{\widehat{\theta}}[t(\mathbb{X})] = -\nabla^2 l(\widehat{\theta})$$
is the Fisher information of $\widehat{\theta}$
- the matrix Γ is the matrix of the asymptotic covariance of the normalised Monte Carlo gradient $\sqrt{n}\nabla l_n(\widehat{\theta})$

- the variance of the components of $\widehat{\theta} - \theta_0$ can be estimated by taking the diagonal elements of the inverse of $-\nabla^2 l_n(\widehat{\theta}_n)$
  - it represents the error between the maximum likelihood estimate and the true model parameters
- the variance of the components of $\sqrt{n}(\widehat{\theta}_n - \widehat{\theta})$ can be estimated by taking the diagonal elements of $I(\widehat{\theta})^{-1} \Gamma I(\widehat{\theta})^{-1}$
  - it represents the error between the maximum likelihood estimate and its Monte Carlo counterpart
- refer to (Monfort, 1997), (Geyer, 1999) and (van Lieshout and Stoica, 2003) for the computation of these matrices

$\rightarrow$ depending on the schedule $\rightarrow$ blackboard: incomplete data observation - EM algorithm

MCML example

Candy model: (van Lieshout and Stoica, 2003)



Model parameters
$\theta_f = -8.5$
$\theta_s = -3.0$
$\theta_d = 2.5$
$\theta_o = -2.5$
$\theta_r = -2.5$

Sufficient statistics
$n_f = 4$
$n_s = 34$
$n_d = 63$
$n_o = 12$
$n_r = 9$

Figure: Realization (left) of the reference model given by the parameters in the middle table. The observed values of the sufficient statistics are listed at right.

Results: estimation of the parameters from the reference configuration given by the Candy model

Initial parameters	Iterative method	Monte Carlo MLE
$\theta_f^i = -9.5$	$\hat{\theta}_f^0 = -8.37$	$\hat{\theta}_f^n = -8.32$
$\theta_s^i = -4.0$	$\hat{\theta}_s^0 = -2.74$	$\hat{\theta}_s^n = -2.73$
$\theta_d^i = 1.5$	$\hat{\theta}_d^0 = 2.46$	$\hat{\theta}_d^n = 2.47$
$\theta_o^i = -3.5$	$\hat{\theta}_o^0 = -2.13$	$\hat{\theta}_o^n = -2.17$
$\theta_r^i = -3.5$	$\hat{\theta}_r^0 = -2.42$	$\hat{\theta}_r^n = -2.42$

Asymptotics : estimation errors (central limit theorems available)

Asymptotic standard deviation of MLE	MCSE
0.51	0.002
0.23	0.003
0.17	0.001
0.30	0.002
0.31	0.005

Log-likelihood ratio approximation:

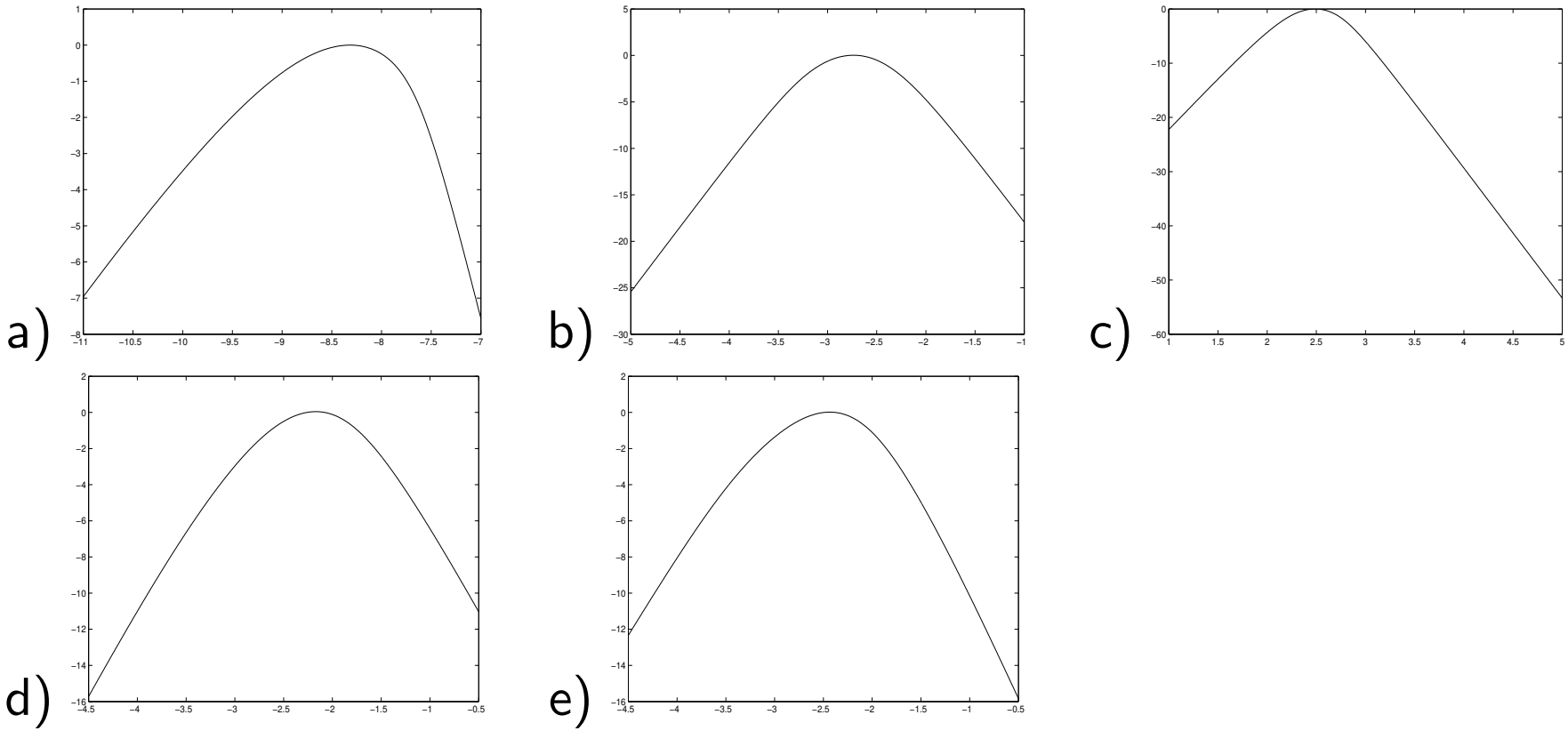


Figure: Monte Carlo approximation of the log likelihood function. The X axis represents the variation of a single component. The Y axis represents the values of the Monte Carlo log likelihood with all other components of $\hat{\theta}^0$ fixed : a - $\theta_f \in [-11, -7]$, b - $\theta_s \in [-5, -1]$, c - $\theta_d \in [1, 5]$, d - $\theta_o \in [-4.5, -0.5]$, e - $\theta_r \in [-4.5, -0.5]$.

Parameter estimation via posterior sampling

- let $p(\theta|\mathrm{y})$ be the conditional distribution of the model parameters given the patten

$$p(\theta|\mathrm{y}) = \frac{\exp[-U(\mathrm{y}|\theta)]p(\theta)}{Z(\mathrm{y})c(\theta)}$$

  with $p(\theta)$ the prior density for the model parameters and $Z(\mathrm{y})$ the normalization constant
- the posterior law is defined on Θ a compact region in $\mathbb{R}^m$; the parameter space Θ is endowed with the Borel algebra $\mathcal{T}_\Theta$

- the parameter estimator is

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \{p(\theta|\mathrm{y})\}$$

- optimisation procedure : requires sampling $p(\theta|\mathrm{y})$
- sampling the posterior law is not straightforward $\rightarrow$ requires the evaluation of the ratio $c(\theta)/c(\psi)$
- if $p(\theta)$ is the uniform distribution over Θ then $\widehat{\theta}$ is the maximum likelihood estimator

Remark: the model parameters taken into account by the posterior distribution are the "interaction" parameters. And not the "range" parameters ...

Sampling posterior laws: key element for parametric inference

- theoretical solution: auxiliary variable method given (Møller, Pettitt, Reeves and Berthelsen, 2006)
  - choice of the auxiliary variable density - behaviour of the simulated chain
  - exchange algorithms : (Murray et al.,2006), (Liang, 2010)
- practical solution: Approximate Bayesian Computation (Marin, Robert, Pudlo, Biau, Blum, etc.)
  - useful if enough samples are produced "close" to the observed pattern y
- ABC sampling strategies for point processes : Cox, determinantal and Gibbs
  - (Shirota and Gelfand, 2017), (Vihr, Møller and Gelfand,2021)
  - lack of practical and theoretical control and limited models (marked and multiple interactions)

Algorithm ABC: assume the observed pattern is y, fix a tolerance threshold ϵ and an integer value n.

1. For $i = 1$ to n do
   - Generate θ_i according to $p(\theta)$.
   - Generate x_i according to the probability density $p(\mathrm{x}|\theta_i) = \frac{\exp[-U(\mathrm{x}|\theta_i)]}{c(\theta_i)}$
2. Return all the θ_i's such that the distance between the statistics of the observation and those of the simulated pattern is small, that is

$$d(t(\mathrm{y}), t(\mathrm{x}_i)) \leq \epsilon$$

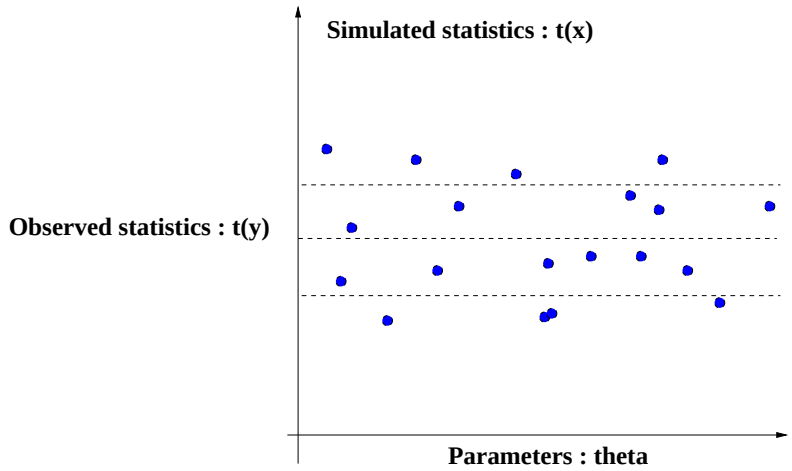


Figure: Graphical representation of the outputs of an ABC algorithm.

Theoretical result

- (Blum, 2009): gives the bias and the variance of the posterior distribution estimate
- (G. Biau, F. Cérous and A. Guyader , 2015): give asymptotic features of the outputs of a slightly different algorithm
- ideas: kernel and $k-$nearest neighbour estimation

Remarks:

- exact sampling from $p(\mathrm{x}|\theta)$ is needed
- choice of the statistics vector
  - exponential family models $\rightarrow$ the sufficient statistics
- appropriate setting :
  - distance d
  - precision parameter ϵ
  - number of neighbours k_n and bandwidth parameter h_n

Synthesis:

- ABC algorithms are useful if enough samples x_i are "close" to the observed pattern y

ABC Shadow algorithm

Key points

- need: an algorithm with outputs "close" enough to the posterior distribution
- tool: build a Markov chain evolving "close" to an equilibrium regime given by $p(\theta|\mathrm{y})$
- plan : use the auxiliary variable method ideas

Ideal MCMC sampling of the posterior: general MH algorithm

- assume the system is in the state θ
- choose a new value ψ according to a proposal density $q(\theta \to \psi)$
- the value ψ is accepted with probability

$$\alpha_i(\theta \to \psi) = \min\left\{1, \frac{p(\psi|\mathrm{y})p(\psi)}{p(\theta|\mathrm{y})p(\theta)} \frac{q(\psi \to \theta)}{q(\theta \to \psi)}\right\}$$

- consider the proposal density

$$q(\theta \to \psi) = q_\Delta(\theta \to \psi|\mathrm{x}) = \frac{f(\mathrm{x}|\psi)/c(\psi)}{I(\theta, \Delta, \mathrm{x})} 1_{b(\theta,\Delta/2)}\{\psi\}$$

  with
  - x : outcome of a marked point process driven by the probability density $p(\mathrm{x}|\upsilon)$ where υ is any value in Θ.
  - pattern detection context : $f(\mathrm{x}|\psi) = exp[-U(\mathrm{x}|\psi)]$
  - $\Delta > 0$: control parameter
  - $1_{b(\theta,\Delta/2)}\{\cdot\}$ is the indicator function over $b(\theta, \Delta/2)$, which is the ball of centre θ and radius $\Delta/2$
  - $I(\theta, \Delta, \mathrm{x}) = \int_{b(\theta,\Delta/2)} f(\mathrm{x}|\phi)/c(\phi)\, d\phi$.
- this choice guarantees the ideal chain to be uniformly ergodic and avoids the evaluation of the ratios $c(\theta)/c(\psi)$
- but, it requires the computation of integrals $I(\theta, \Delta, \mathrm{x})$ ...

$\to$ blackboard : drawing

Shadow chain: approximation of the ideal chain

Theorem : if $p(x|\theta)$ is a continuously differentiable function in θ

- For any fixed $\theta \in \Theta$ and $A \in \mathcal{T}_\Theta$, we have

$$\lim_{\Delta \to 0_+} \int_A |q_\Delta(\theta \to \psi) - U_\Delta(\theta \to \psi)| d\psi = 0$$

- For any fixed $\theta \in \Theta$, we have

$$\lim_{\Delta \to 0_+} \sup_{\psi \in \Theta} \left| \frac{q_\Delta(\theta \to \psi|x)}{q_\Delta(\psi \to \theta|x)} - \frac{\frac{f(x|\psi)}{c(\psi)} 1_{b(\theta,\Delta/2)}(\psi)}{\frac{f(x|\theta)}{c(\theta)} 1_{b(\psi,\Delta/2)}(\theta)} \right| = 0$$

uniformly in $\theta \in \Theta$, with

- V_Δ : the volume of the ball $b(\theta, \Delta/2)$
- $U_\Delta = \frac{1}{V_\Delta} 1_{b(\theta,\Delta/2)}\{\psi\}$: uniform probability density

Application: simulate the shadow chain that approximate the ideal chain

- first part : use $U_\Delta(\theta \to \psi)$ instead of $q_\Delta(\theta \to \psi)$ for proposing new values
- second part : approximates the computation of the proposal density ratio while simplifying the normalizing constant ratio
- the shadow Markov chain accepts new states with the probability :

$$\alpha_s(\theta \to \psi) = \min\left\{1, \frac{p(\psi|y)p(\psi)}{p(\theta|y)p(\theta)} \times \frac{f(x;\theta)c(\psi)1_{b(\psi,\Delta/2)}\{\theta\}}{f(x;\psi)c(\theta)1_{b(\theta,\Delta/2)}\{\psi\}}\right\}$$

Corollary: the acceptance probabilities of the ideal and shadow chains are uniformly as closed as desired whenever $\triangle \to 0_+$

Proposition : Let P_i and P_s be the transition kernels for the ideal and the shadow Markov chains using a general $\Delta > 0$ and a configuration $x \in \Omega$. Then, for every $\epsilon > 0$ and every $n \in \mathbb{N}$, there exists $\Delta_0 = \Delta_0(\epsilon, n) > 0$ such that for every $\Delta \leq \Delta_0$

$$|P_i^{(n)}(\theta, A) - P_s^{(n)}(\theta, A)| < \epsilon$$

uniformly in $\theta \in \Theta$ and $A \in \mathcal{T}_\Theta$.

Algorithm ABC Shadow: assume the observed pattern is y and fix values for Δ and n and the current state θ_0

1. Generate x according to $p(x|\theta_0)$
2. For $k = 1$ to n do
   - Generate a new candidate ψ following $U_\Delta(\theta_{k-1} \to \psi)$.
   - The new state $\theta_k = \psi$ is accepted with probability $\alpha_s(\theta_{k-1} \to \psi)$, otherwise $\theta_k = \theta_{k-1}$
3. Return θ_n

Remarks:

- if several samples are needed, re-start the procedure for the same Δ and n, with $\theta_0 = \theta_n$.
- depending on Δ, the algorithm approaches the equilibrium regime of the ideal chain :

$$\|P_s^{(n)}(\theta, A) - \pi(A)\| \leq M(\mathrm{x}, \Delta)\rho^n + \epsilon.$$

with $\pi(A) = \int_A p(\theta|\mathrm{y})d\theta$; M and ρ : ergodicity parameters of the ideal chain

- caution : this is not convergence

Application: sampling the posterior of a Gaussian model

The posterior of a Normal model with mean θ_1 and variance θ_2 is

$$p(\theta_1, \theta_2 | \mathrm{y} = \mathbb{Y}(\omega)) \propto \frac{\exp\left(\frac{\theta_1}{\theta_2}\mathbb{Y}(\omega) - \frac{\mathbb{Y}^2(\omega)}{2\theta_2}\right)}{c(\theta_1, \theta_2)} p(\theta_1, \theta_2)$$

with

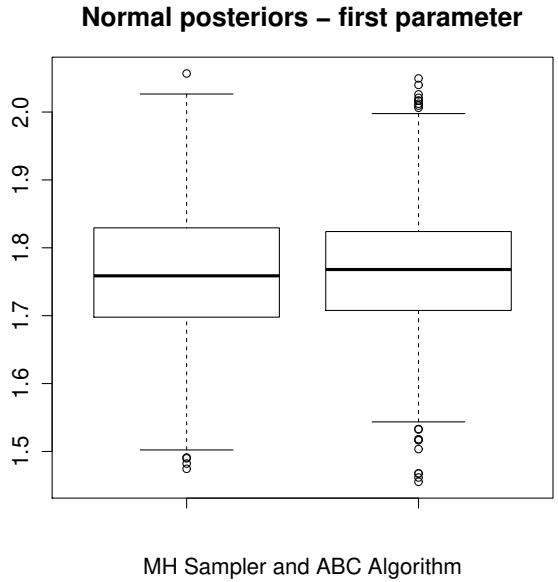
- $\mathrm{y} = \mathbb{Y}(\omega)$: observation issued from the supposed model
- $t(\mathrm{y}) = (\mathbb{Y}(\omega), \mathbb{Y}^2(\omega))$: the sufficient statistics vector
- if the sample size is m then:
  $t(\mathrm{y}) = \left(\sum_{i=1}^m \mathbb{Y}_i(\omega), \sum_{i=1}^m \mathbb{Y}_i^2(\omega)\right)$

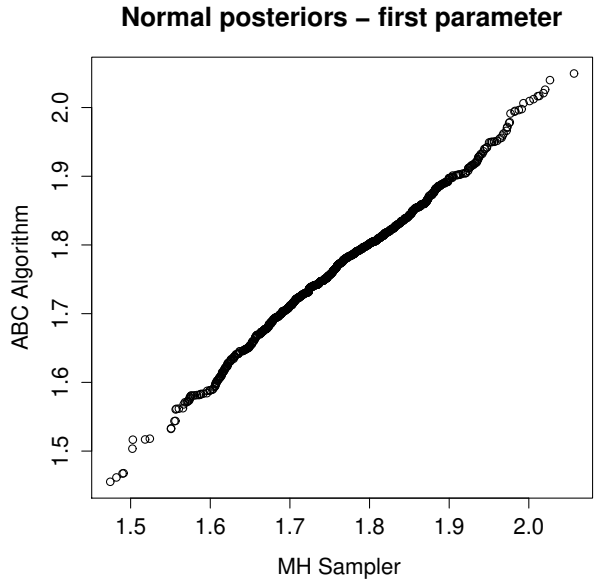
Experiment:

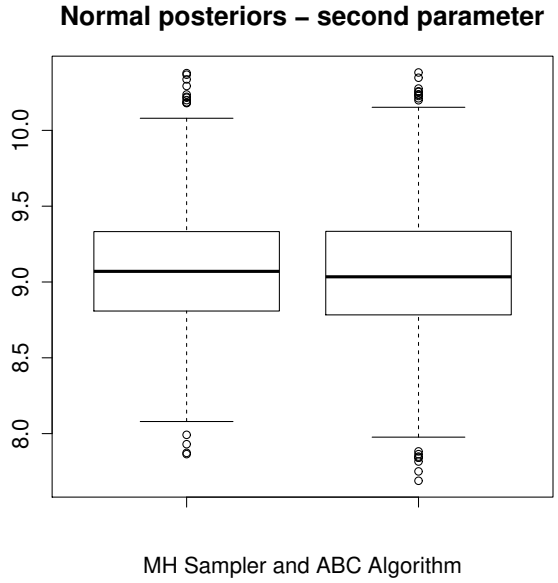
- simulate 1000 i.i.d. Normal r. v.'s with parameters $\theta = (\mu, \sigma^2) = (2, 9)$
- $t(\mathrm{y}) = (1765.45, 12145.83)$
- $p(\theta_1, \theta_2)$ the uniform distribution over $[-100, 100] \times [0, 200]$
- compare 1000 samples of the MH and ABC Shadow algorithms
- $\Delta = (0.005, 0.025)$ and $n = 500$

Summary statistics for Normal posterior sampling						
Algorithm	Q_5	Q_{25}	Q_{50}	$\bar{\theta}$	Q_{75}	Q_{95}
MH θ_1	1.60	1.69	1.75	1.76	1.82	1.92
ABC θ_1	1.60	1.70	1.76	1.76	1.82	1.91
MH θ_2	8.45	8.80	9.07	9.08	9.33	9.76
ABC θ_2	8.35	8.78	9.03	9.06	9.33	9.83

Table: Empirical quantiles and mean for the posterior of the Normal model.







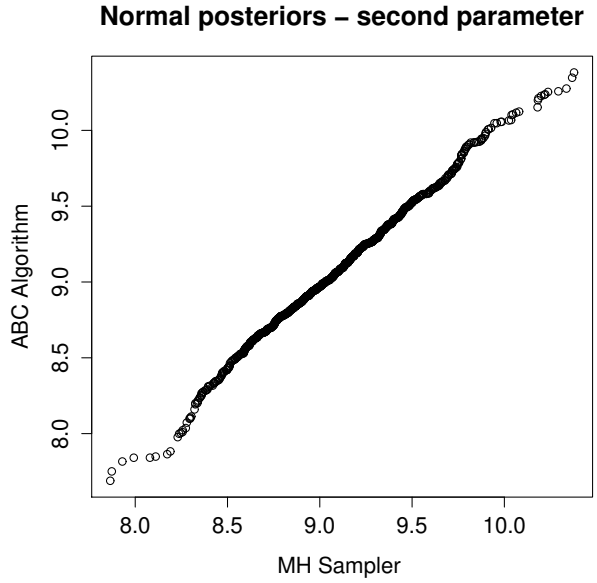


Figure: Boxplots and qqplots of the MH and ABC Shadow outpts.

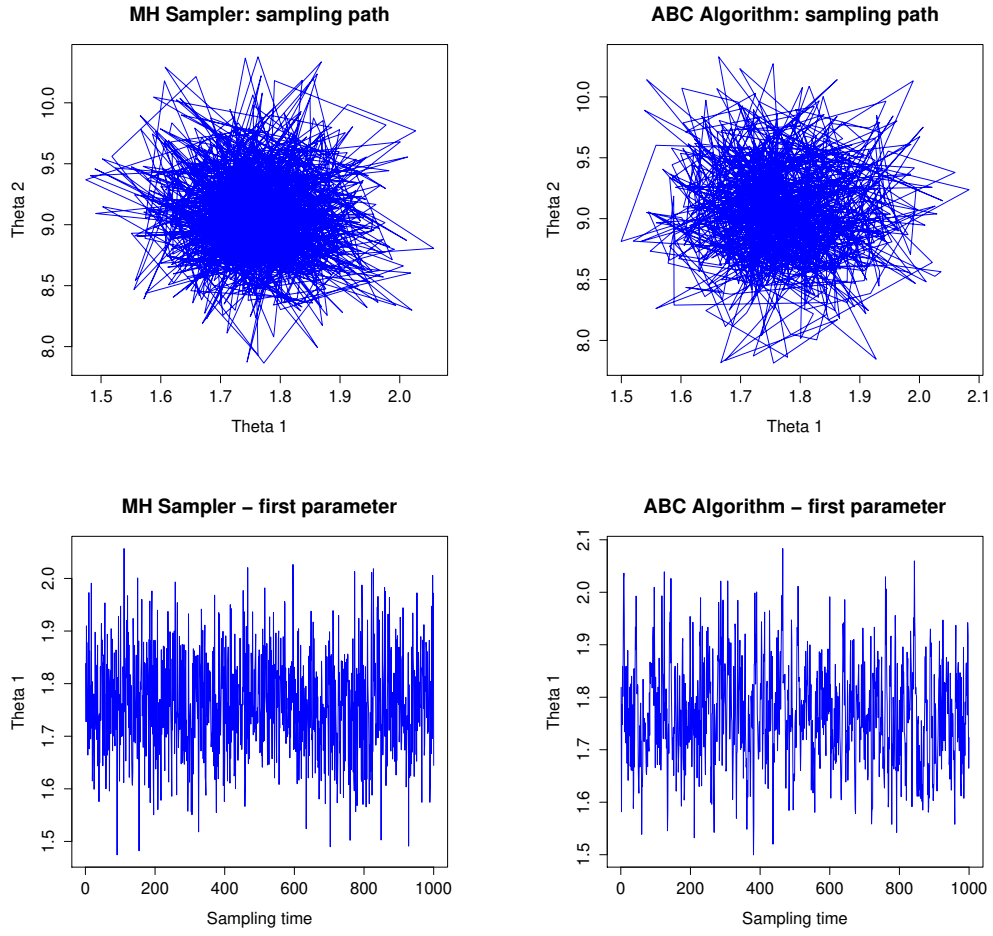


Figure: Sample path for the Normal posterior. Left colum: the MH algorithm results - from top to bottom the joint parameter path and the θ_1 time series. Right column: the ABC Shadow procedure - from top to bottom the joint parameter path and the θ_1 time series.

Application - Strauss model: (Strauss, 1975), (Kelly and Ripley, 1976)

$$
\begin{aligned}
p(\mathrm{x}|\theta) \propto \beta^{n(\mathrm{x})}\gamma^{s_r(\mathrm{x})} = \quad & \beta, r > 0, \gamma \in [0,1] \\
&= \exp\left[n(\mathrm{x})\log\beta + s_r(\mathrm{x})\log\gamma\right] \\
&= \exp\langle t(\mathrm{x}), \theta\rangle.
\end{aligned}
$$

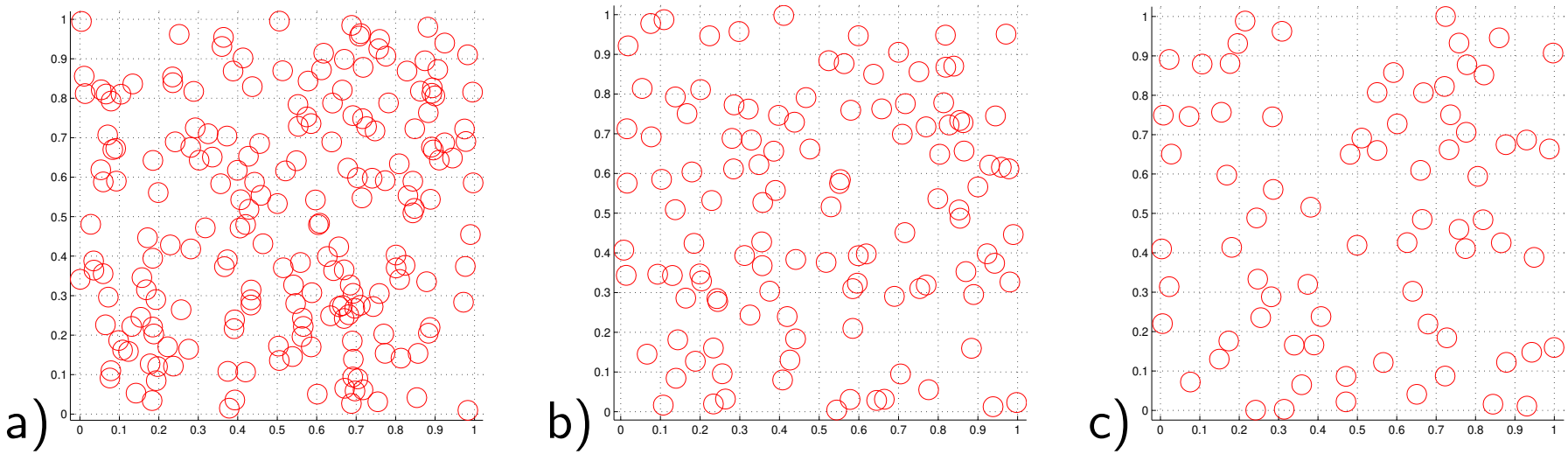


Figure: Strauss model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma = 0.5$ and c) $\gamma \to 0.0$.

- a sample made of m independent point patterns is observed.
- it is further assumed that the observed patterns are realisations of a Strauss model with parameters r and θ.
- the range parameter is considered known
- it is easy to observe that the maximum likelihood estimate $\widehat{\theta}$ satisfies the equation

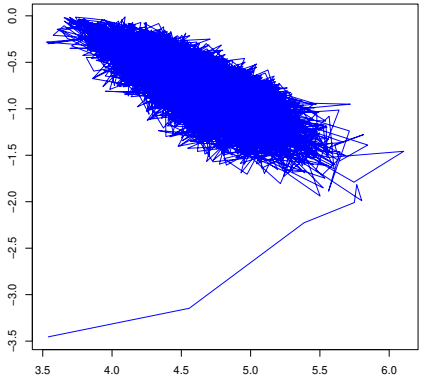
$$\frac{1}{m}\sum_{i=1}^{m} t(\mathrm{y}_i) - \mathbb{E}_{\widehat{\theta}} t(\mathbb{X}) = 0$$

with $\mathrm{y}_1, \ldots, \mathrm{y}_m$ the observed point patterns forming the sample.

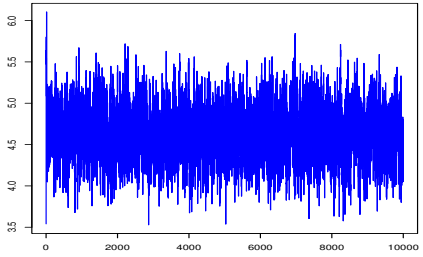
Numerical experiments: verification of the ABC Shadow algorithm

- domain $W = [0, 1] \times [0, 1]$, range $r = 0.1$
- simulate 1000 realisations of the model $\theta = (4.6, -0.69)$ using the CFTP algorithm
- the observed empirical means of the sufficient statistics $\bar{t}(\mathrm{y}) = (45.30, 17.99)$
- $\Delta = (0.01, 0.01)$, $m = (200, 200)$, $n = 10^6$
- $p(\theta)$: uniform distribution over the parameter domain $\Theta = [0, 7] \times [-7, 0]$

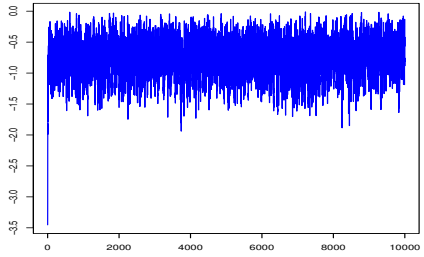
Posterior approximation for the Strauss process using the ABC Shadow algorithm:



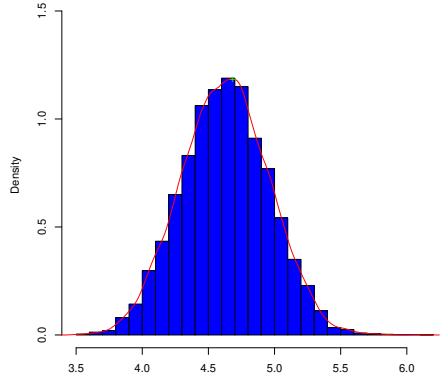
a)



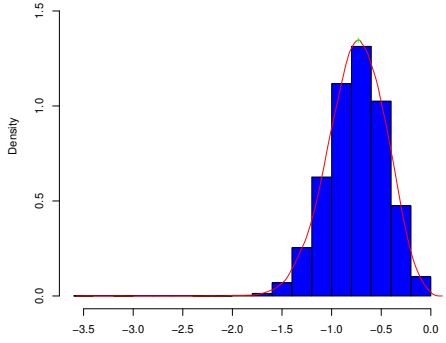
b)



c)



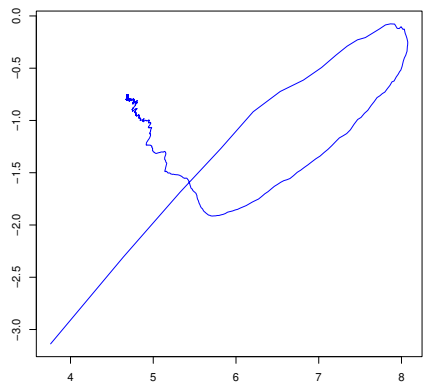
d)



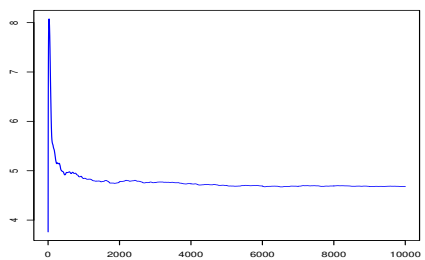
e)

Figure: Parametric inference for the Strauss process throught posterior approximation: joint distribution (a), marginal distributions (b,c) and histograms (d,e). The approximated MAP is $(4.68, -0.73)$.

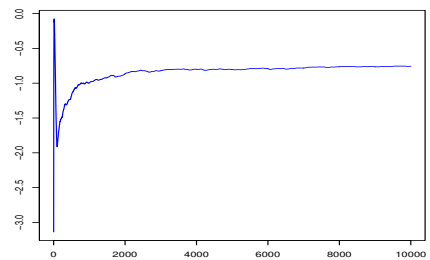
Comparison of the approximated MAP obtained via the ABC Shadpw with the approximated MLE computed via the stochastic gradient:



a)



b)



c)

Figure: Estimation of parameters using the stochastic gradient algorithm: joint dynamics of the parameters (a), marginals (b,c). The approximated MLE is $(4.68, -0.75)$.

Shadow Simulated Annealing: parameter estimation

(Stoica, Deaconu, Hurtado, Philippe, 2021)

Fix $\Delta = \Delta_0$, $T = T_0$, m and $k_\Delta, k_T : \mathbb{R}^+ \to \mathbb{R}^+$ two positive functions. Assume the observed pattern is y and the current state is θ_0.

1. Generate x according to $p(\mathrm{x}|\theta_0) = \frac{f(\mathrm{x}|\theta_0)}{c(\theta_0)}$.
2. For $k = 1$ to m do
   - Generate a new candidate ψ following the density $U_\Delta(\theta_{k-1} \to \psi)$ defined by

$$U_\Delta(\theta \to \psi) = \frac{1}{V_\Delta} 1_{b(\theta, \Delta/2)}\{\psi\},$$

representing the uniform probability density over the ball $b(\theta, \Delta/2)$ of volume V_Δ.

- The new state $\theta_k = \psi$ is accepted with probability given by

$$\alpha_s(\theta_{k-1} \to \theta_k) =$$

$$= \min\left\{1, \left[\frac{p(\theta_k|\mathrm{y})}{p(\theta_{k-1}|\mathrm{y})} \times \frac{f(\mathrm{x}|\theta_{k-1})}{f(\mathrm{x}|\theta_k}\right]^{1/T} \times \frac{1_{b(\theta_k,\delta/2)}\{\theta_{k-1}\}}{1_{b(\theta_{k-1},\delta/2)}\{\theta_k\}}\right\}$$

$$= \min\left\{1, \left[\frac{f(\mathrm{y}|\theta_k)p(\theta_k)}{f(\mathrm{y}|\theta_{k-1})p(\theta_{k-1})} \times \frac{f(\mathrm{x}|\theta_{k-1})}{f(\mathrm{x}|\theta_k)}\right]^{1/T}\right\}$$

  otherwise $\theta_k = \theta_{k-1}$.

3. Return θ_m
4. Stop the algorithm or go to step 1 with $\theta_0 = \theta_m$, $\delta_0 = k_\delta(\delta)$ and $T_0 = k_T(T)$.

Discussion of the convergence properties

- key idea: build a Shadow Simulated Annealing process while varying the control parameter of the ABC Shadow dynamics in order to obtain a convergent algorithm
- key tool: Dobrushin's coefficient (Iosifescu and Theodorescu, 1969)
- principle of the classical SA algorithm result: the iteration of the transition kernel of an inhomogeneous Markov chain with a slow varying Dobrushin coefficient converges weakly towards the uniform distribution over the subspace of configurations containing the global optima of the function of interest (Haario and Saksman, 1991, 1992)

Convergence result

Let $\{\theta_j, j \geq 1\}$ be a realisation of the SSA process. Then there exists a sequence $\{\Delta_j = \Delta_j(\varepsilon, m), j \geq 1\}$ corresponding to each θ_j such that for $j \geq j_{\max}(\varepsilon, m)$

$$\{\theta_j, j \geq j_{\max}\} \subset b(\theta_{opt}, \varepsilon)$$

where θ_{opt} is the global optimum.

We have

$$\Delta_j = \frac{K(x_j, T_j, m)}{j} \tag{12}$$

with $K(x_j, T_j, m)$ a constant depending on the auxiliary variable x_j, the temperature T_j and the number of iterations m.

Interpretation: to a decreasing sequence of balls around the problem solution, it is possible to associate a sequence of Δ parameters in order to get as close as desired to the desired global optimum.

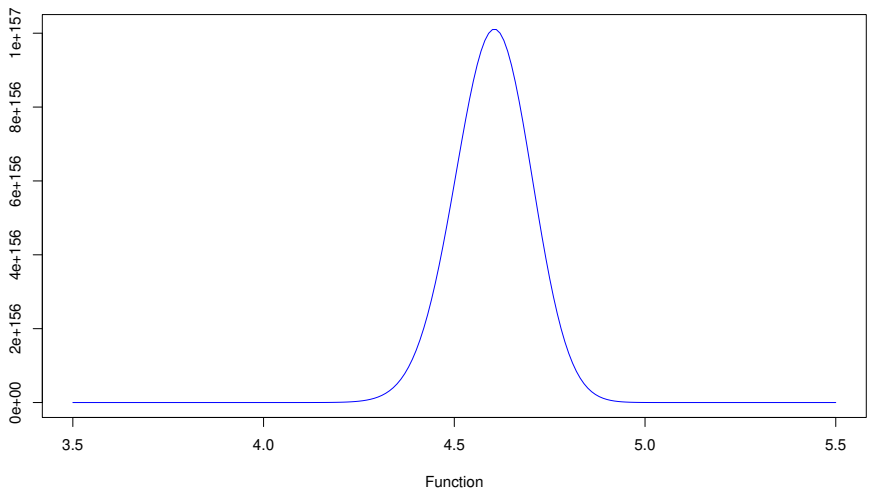
Simulated data: marked Gibbs point processes

Homogeneous Poisson process:

$$p(\mathrm{y}|\theta) \propto \exp\langle \theta n(\mathrm{y}) \rangle \tag{13}$$

with $n(\mathrm{y})$ the observed number of points and θ the intensity parameter

- $W = [0,1]^2$, the observed statistics $n(\mathrm{y}) = 100$
- the maximum likelihood estimate of the intensity is $\widehat{\theta} = \log(100) \approx 4.6$
- the posterior with a uniform prior is equivalent with a restriction of the likelihood function



- consider the following prior for θ:

$$p(\theta) = \frac{1}{\sqrt{0.045\pi}} \left\{ 0.6 \exp\left[-\frac{(\theta - 4.3)^2}{0.0225} \right] + 0.4 \exp\left[-\frac{(\theta - 4.9)^2}{0.0225} \right] \right\}$$

- the maximum likelihood estimate of the intensity is $\widehat{\theta} = 4.46$
- the likelihood function in this case is not concave

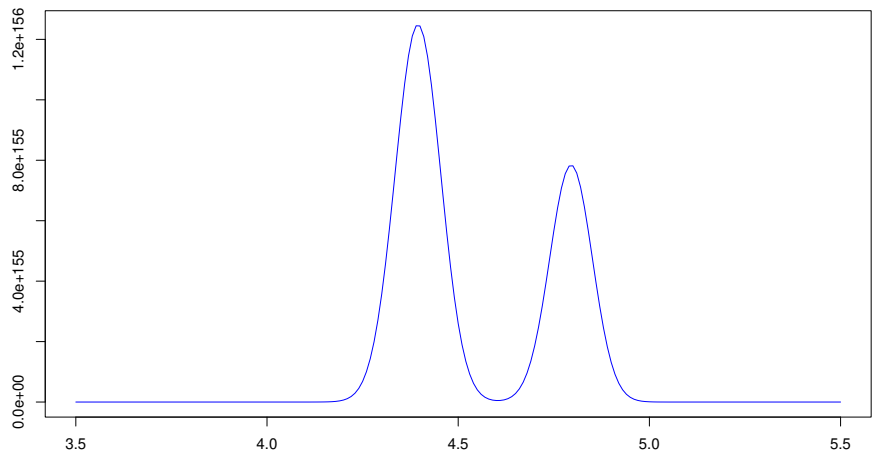
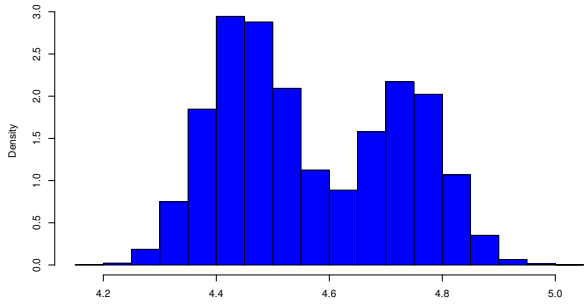
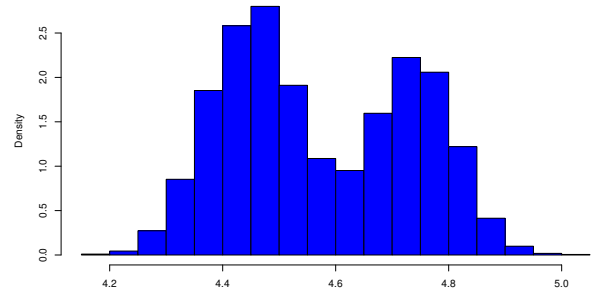


Figure: Log of the posterior of an homogeneous Poisson point process with non-uniform prior.

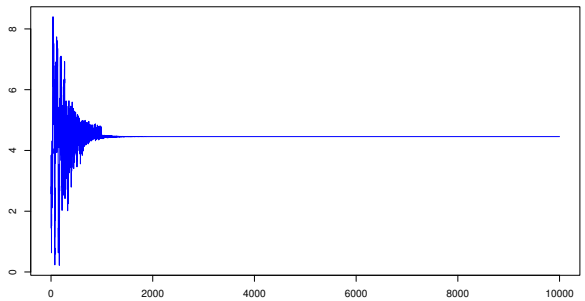


a)

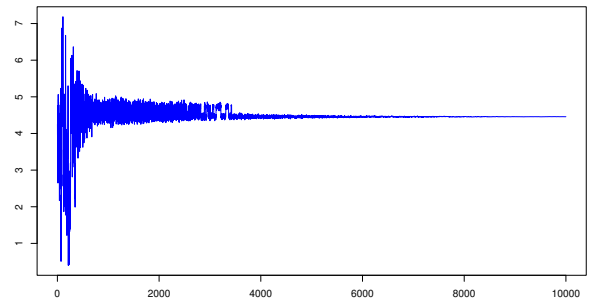


b)

Figure: Posterior sampling: a) MH dynamics, b) ABC Shadow dynamics



a)



b)

Figure: MAP estimation: a) SA based on MH dynamics $\widehat{\theta} = 4.456$, b) SSA dynamics $\widehat{\theta} = 4.457$

Candy model:

$$p(\mathrm{y}|\theta) \propto \exp\langle \theta_d n_d(\mathrm{y}) + \theta_s n_s(\mathrm{y}) + \theta_f n_f(\mathrm{y}) + \theta_r n_r(\mathrm{y}) \rangle \qquad (14)$$

with

- $t(\mathrm{y}) = (n_d(\mathrm{y}), n_s(\mathrm{y}), n_f(\mathrm{y}), n_r(\mathrm{y}))$: sufficient statistics
- $\theta = (\theta_d, \theta_s, \theta_f, \theta_r)$: model parameters

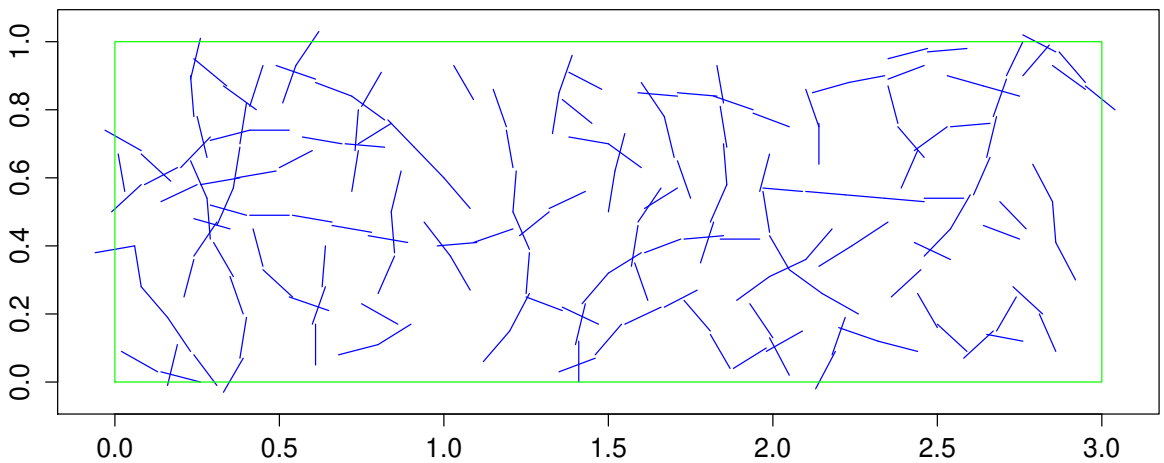


Figure: Realisation of the Candy model.

Experiment :

- domain $W = [0, 3] \times [0, 1]$, segment length $l = 0.12$, connection range $r_c = 0.01$, curvature parameters $\tau_c = \tau_r = 0.5$ radians
- simulate 1000 realisations of the model $\theta = (10, 7, 3, -1)$ using an Adapted MH algorithm
- the observed empirical means of the sufficient statistics $\bar{t}(\mathrm{y}) = (51.10, 101.06, 19.97, 72.89)$
- $\Delta_0 = (0.01, 0.01, 0.01, 0.01)$, $T_0 = 10^4$, $m = 500$, $n = 10^6$
- cooling schedules: $T_n = k_T T_{n-1}$, $\Delta_n = k_\Delta \Delta_{n-1}$ with $k_T = 0.9999$ and $k_\Delta = 0.999995$
- $p(\theta)$: uniform distribution over the parameter domain $\Theta = [2, 12]^3 \times [-7, 0]$

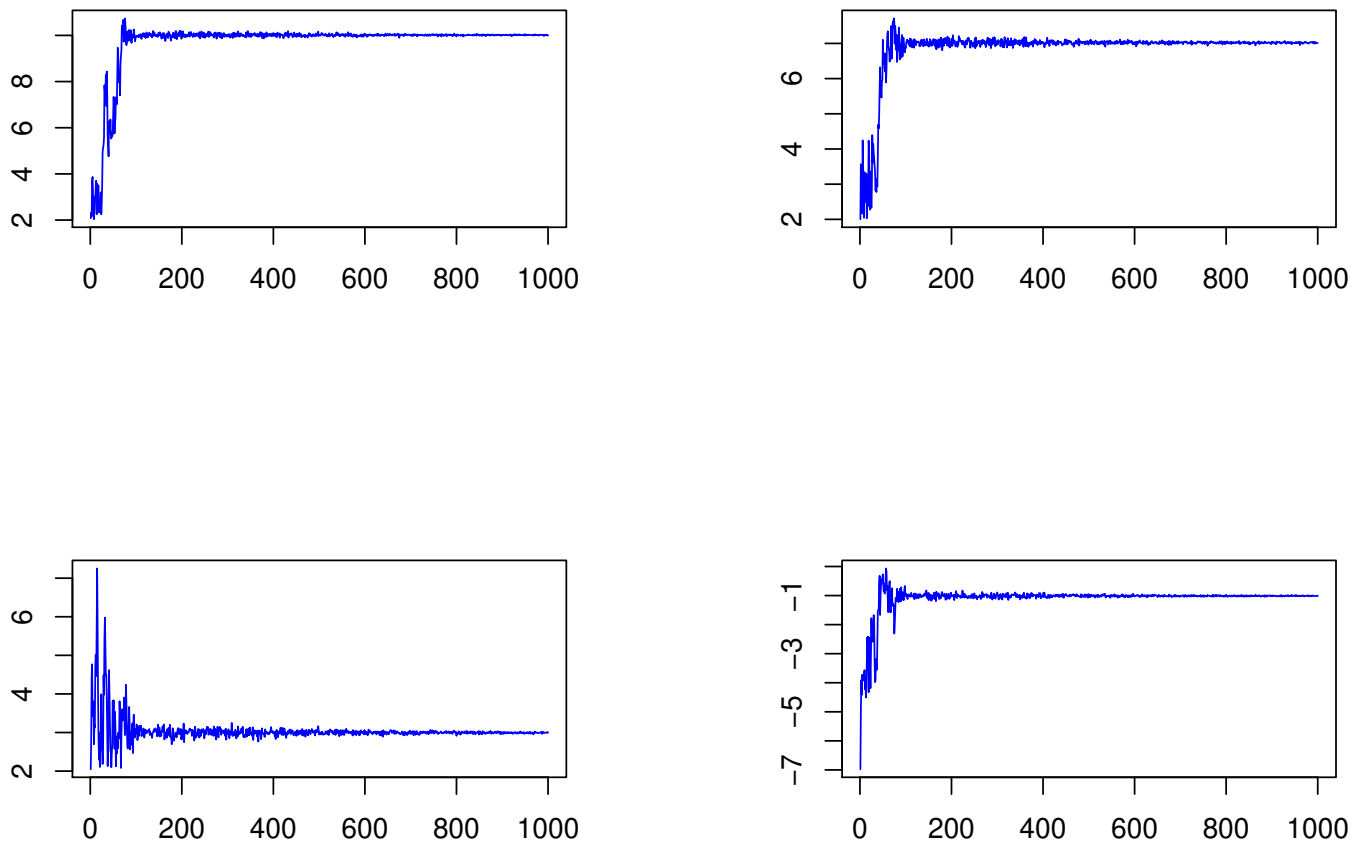


Figure: SSA Candy parameter estimation: algorithm dynamics. Maximum of the kernel estimated density is $\widehat{\theta} = (10.012, 7.029, 2.994, -1.011)$.

Summary statistics for SSA Candy estimation			
Algorithm	Q_{25}	Q_{50}	Q_{75}
SSA $\log \gamma_d$	9.975	10.01	10.04
SSA $\log \gamma_s$	6.985	7.020	7.045
SSA $\log \gamma_f$	2.968	2.999	3.032
SSA $\log \gamma_r$	-1.034	-1.012	-0.991

Table: Empirical quartiles for the SSA Candy model estimation.

Real data: galaxies distribution knowing the filaments

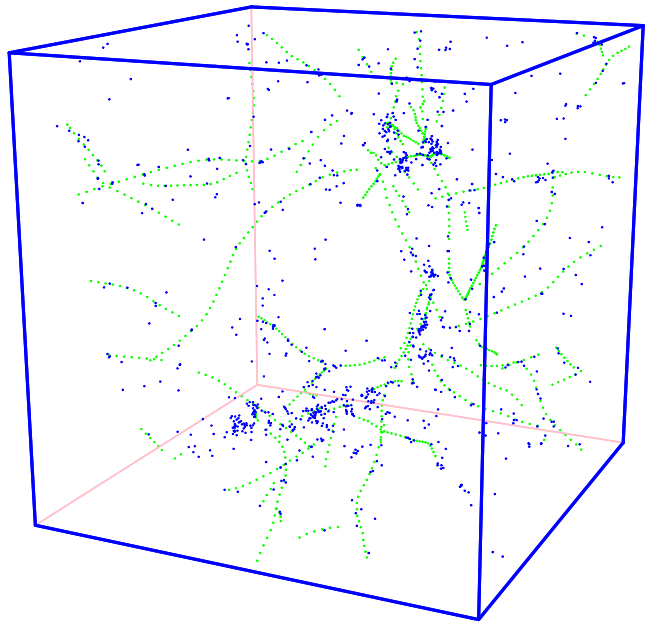


Figure: Galaxies positions (blue) and the induced filaments pattern or spines (green).

(Hurtado, Stoica, et al., 2021)

Proposed model: inhomogeneous interacting point process

Such a model can be represented by the point process given by the following probability density

$$p(\mathrm{y}|\theta, F) \propto \beta_1^{n(\mathrm{y})} \beta_2^{d_F(\mathrm{y})} \gamma^{-a_r(\mathrm{y})} \tag{15}$$

with the model parameters vector given by

$$\theta = (\log \beta_1, \log \beta_2, \log \gamma),$$

and the sufficient statistics vector

$$t(\mathrm{y}) = (n(\mathrm{y}), d_F(\mathrm{y}), a_r(\mathrm{y})).$$

The model works as it follows:

- the parameter β_1 controls the number of galaxies $n(\mathrm{y})$
- the parameter β_2 controls the distance to the filamentary pattern, that is

$$d_F(\mathrm{y}) = -\sum_{i=1}^{n(\mathrm{x})} d(y_i, F)$$

  with $d(y_i, F)$ the minimum distance from the galaxy position y_i to the spines pattern F
- the parameter γ controls the volume occupied by the galaxy pattern, that is

$$a_r(\mathrm{y}) = \frac{3A(\mathrm{y})}{4\pi r^3}, A(\mathrm{y}) = \nu[\bigcup_{i=1}^{n(\mathrm{y})} b(y_i, r)]$$

  where $b(y, r)$ is the ball centred in y with radius r and $A(\mathrm{y})$ is the volume of the object resulting from the set union of the spheres of radius r and that are centred in the points given by the configuration y

Computing the data for the SSA algorithm:

Data for the galaxy pattern							
r	0.5	1	1.5	2	2.5	3	3.5
$n(\mathrm{y}) = 1024,\ d_F(\mathrm{y}) = -\sum_{i=1}^{n(\mathrm{y})} d(y_i, F) = -1180.05$							
$a_r(\mathrm{y})$	724.29	484.01	357.16	263.10	195.08	142.86	105.30

Table: The sufficient statistics computed for the observed galaxy pattern, depending on the range parameter r.

SSA algorithm's outputs:

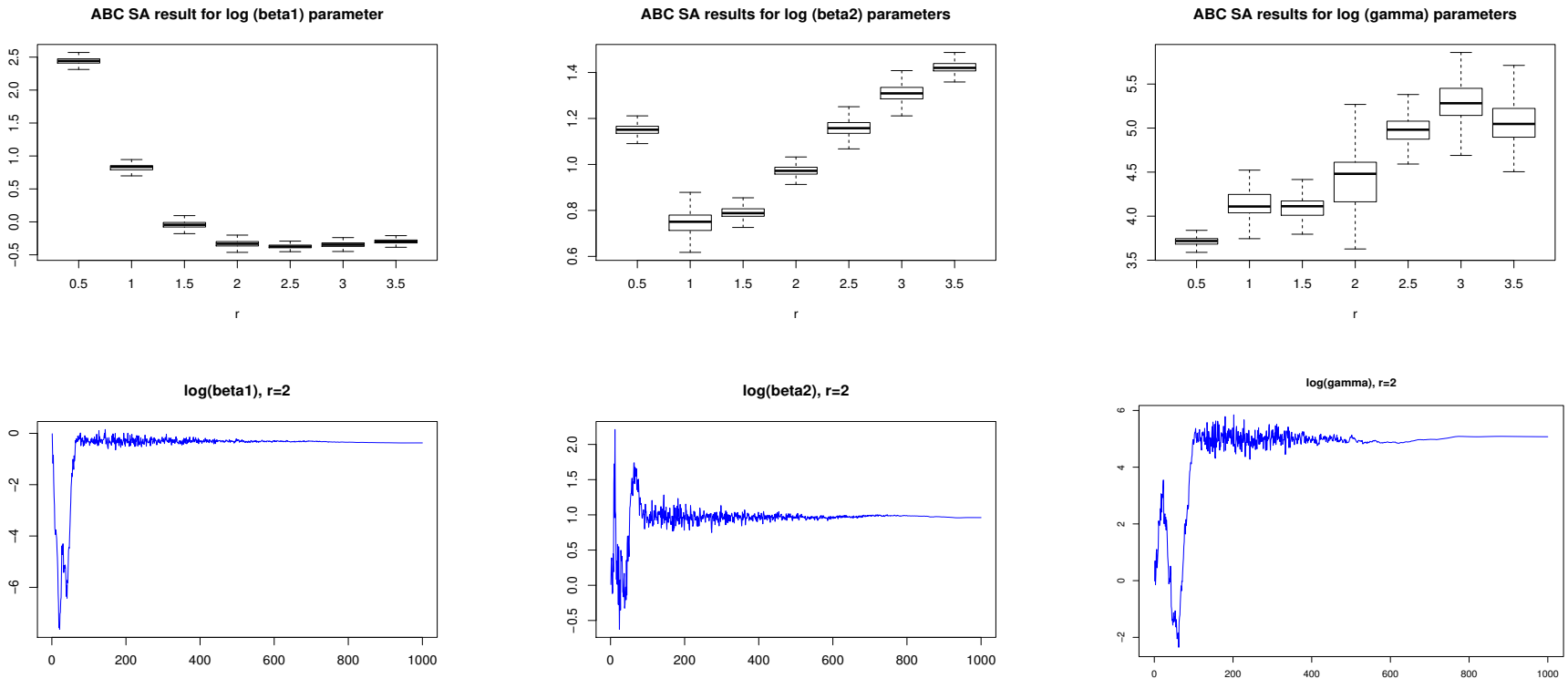


Figure: SSA outputs for the MAP estimates computation of the inhomogeneous area-interaction model parameters fitted to the considered SDSS sample. Right column: box plots of SSA algorithm outputs for each parameter depending on the interaction radius. Left column: the time series of the outputs of the SSA algorithm for $r = 2$.

Results verification: aymptotics

Asymptotic errors for the SSA estimates						
r	$\sigma_{\log \beta_1}$	$\sigma^{MC}_{\log \beta_1}$	$\sigma_{\log \beta_2}$	$\sigma^{MC}_{\log \beta_2}$	$\sigma_{\log \gamma}$	$\sigma^{MC}_{\log \gamma}$
0.5	0.04	1e-4	0.03	7e-5	0.07	2e-4
1	0.03	1e-4	0.03	9e-5	0.09	5e-4
1.5	0.05	1e-4	0.04	1e-4	0.12	1e-3
2	0.05	1e-4	0.04	2e-4	0.17	2e-3
2.5	0.05	2e-4	0.04	2e-4	0.24	5e-3
3	0.05	2e-4	0.04	5e-4	0.43	0.016
3.5	0.05	2e-4	0.04	9e-4	0.66	0.039

Table: Asymptotics errors for the SSA MAP estimates of the model (15) fitted to the considered cosmological sample. For each corresponding radius a model was fitted, and the asymptotic errors were computed for each model. For the computation of the MCSE 15×10^3 samples from the fitted model were used.

Results verification: statistical tests

- the SSA estimate for $r = 2$ is

$$\widehat{\theta} = (\widehat{\log \beta_1}, \widehat{\log \beta_2}, \widehat{\log \gamma}) = (-0.33, 0.98, 4.57)$$

- firstly, a Student like test was carried on to check whether the mean of the posterior distribution is different from the SSA algorithm output
- secondly, another Student like test was conducted in order to verify whether the obtained parameter values are significantly different from 0.

Tests results:

- sampling using the posterior distribution was carried on using the ABC Shadow algorithm (Stoica et al., 2017)
- first test: the smallest p-value was 0.77
- second test: the p-value corresponding to $\log \beta_1$ was 0.16 $\rightarrow$ this parameter is not significatif ; all the other p-values were smaller than 10^{-5}

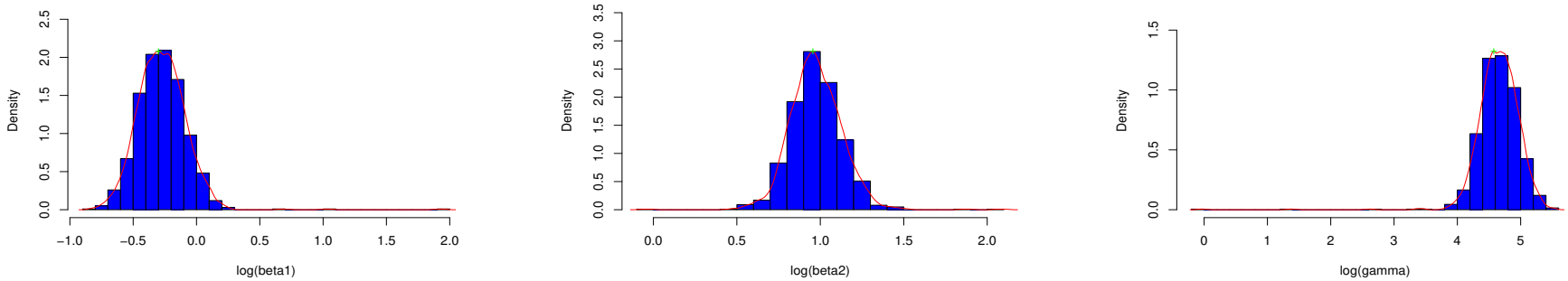


Figure: ABC Shadow outputs for the approximate posterior sampling of the inhomogeneous area interaction model fitted to the SDSS sample (the range parameter is $r = 2$). The obtained MAP estimate is $\widehat{\theta} = (-0.29, 0.95, 4.57)$.

Synthesis parameter estimation

Pseudo-likelihood estimation:

- easy to compute
- good alternative whenever nothing else can be done
- consistency and central limit theorems: difficult to interpret
- no properties concerning the sufficient statistics of the model using the PL estimates of the parameters
- work of J. Mateu and P. Montes: comparison with maximum likelihood

Monte Carlo maximum likelihood:

- general statistical framework
- numerically unstable $\rightarrow$ but re-sampling is guaranteed to convergence, since the log-likelihood is convex
- the asymptotics are related to the true model
- property: the expectation of the sufficient statistics computed by the model with the ML parameters equals the observed sufficient statistics

ABC Shadow parameter estimation :

- approximate algorithm that samples "close" to the posterior: reliable numerical results
- theoretical control: convergence of the Shadow Simulated Annealing algorithm (Stoica et al., 21)
- relatively low computational cost: MCMCML needs re-sampling
- no theoretical need for exact simulation
- direct statistical tests using the estimated posterior
- perspectives: incomplete data + link with existing methods
- open problems: range parameters
- application proved: astronomy, network sciences, geology - (Hurtado, Stoica et al., 21), (Laporte, Stoica et al.,22), (Bonneau, Stoica and Caumon, 25)

Model validation: residual analysis for point processes

Let $\mathbb{X}$ be a locally stable marked point process on $W \times M$.

h−Innovations: for nonnegative functions h and $A \subseteq W \times M$

$$I(A, h, \lambda) = \sum_{x_i \in \mathbb{Y}_A} h(x_i, \mathbb{X} \setminus x_i) - \int_A \lambda(\eta; \mathbb{X}) h(\eta, \mathbb{X})(\nu \times \nu_M)(d\eta)$$

- assuming the sum and the integral in the definition have finite expectations, the Georgii-Nguyen-Zessin formula gives

$$\mathbb{E} I(A, h, \lambda) = 0$$

- I is a signed measure
- $\triangle I(x_i) = h(x_i, \mathbb{X} \setminus \eta)$: the innovation increment ('error') attached to a point $\eta \in \mathbb{X}$
- $dI(\eta) = -\lambda(\eta; \mathbb{X}) h(\eta, \mathbb{X})$: the innovation increment attached to a background location $\eta \in W \times M$

$h-$Residuals: for $h \geq 0$ functions and $A \subseteq W \times M$

$$R(A, \widehat{h}, \widehat{\theta}) = I(A, \widehat{h}, \widehat{\lambda})$$
$$= \sum_{x_i \in \mathrm{x}_A} \widehat{h}(x_i, \mathrm{x} \setminus x_i) - \int_A \widehat{\lambda}(\eta; \mathrm{x}) \widehat{h}(\eta, \mathrm{x}) (\nu \times \nu_M)(d\eta)$$

since the function h may depend on the model, $\widehat{h}$ denotes an estimate.

Application idea :

- consider a parametric model for a marked point process $\mathbb{X}$ observed within A
- estimate the model parameters (maximum likelihood, pseudo-likelihood)
- expect the residuals $R(A)$ to be close to 0 if the model is appropriate

Building residuals: several possible choices for h

- raw residuals $h(\eta, \mathsf{x}) = 1$

$$R(A, 1, \widehat{\theta}) = n(\mathsf{x} \cap A) - \int_A \widehat{\lambda}(\eta; \mathsf{x})(\nu \times \nu_M)(d\eta)$$

- inverse residuals $h(\eta, \mathsf{x}) = 1/\lambda(\eta; \mathsf{x})$ (equivalent with the Stoyan-Grabarnik diagnostic)

$$R(A, \frac{1}{\widehat{\lambda}}, \widehat{\theta}) = \sum_{x_i \in \mathsf{x}_A} \frac{1}{\widehat{\lambda}(x_i; \mathsf{x}_A)} - \int_A 1\{\widehat{\lambda}(\eta; \mathsf{x}) > 0\}(\nu \times \nu_M)(d\eta)$$

- Pearson residuals $h(\eta, \mathrm{x}) = 1/\sqrt{\lambda(\eta; \mathrm{x})}$ (analogy with Poisson log-linear regression)

$$R(A, \frac{1}{\sqrt{\widehat{\lambda}}}, \widehat{\theta}) = \sum_{x_i \in \mathrm{x}_A} \frac{1}{\sqrt{\widehat{\lambda}(x_i; \mathrm{x}_A)}} - \int_A \sqrt{\widehat{\lambda}(\eta; \mathrm{x})}(\nu \times \nu_M)(d\eta)$$

Remark : the inverse and Pearson residuals we need $\lambda_{\theta(\mathrm{x})}(x_i; \mathrm{x}) > 0$ for all $x_i \in \mathrm{x}$ for any pattern x, while $\lambda_{\theta(\mathrm{x})}(\eta; \mathrm{x}) = 0$ is allowed for $\eta \notin \mathrm{x}$

Properties:

- expectation

$$\mathbb{E}\left[R(A,\widehat{h},\widehat{\theta})\right] = \int_A \mathbb{E}\left[h_{\widehat{\theta}(\mathbb{X}\cup\{\eta\})}(\eta,\mathbb{X})\lambda(\eta,\mathbb{X}) - h_{\widehat{\theta}(\mathbb{X})}(\eta,\mathbb{X})\lambda_{\widehat{\theta}(\mathbb{X})}(\eta,\mathbb{X})\right]$$

- variance: more complicate structures but very nice formulas for Poisson processes (Baddeley, Moller and Pakes 2008)
- these residuals do not have independent increments $\rightarrow$ the raw innovations for Markov point processes are conditionnaly independent and uncorrelated (Baddeley, 2005)
- consistency and asymptotic normality for the residuals of stationary Gibbs point processes (Coeurjolly and Lavancier, 2013)

Application: smoothed residuals to test several models for japanesepines datasets

- Strauss process : only repulsion
- area-interaction process : repulsion or attraction (competition for ressources)

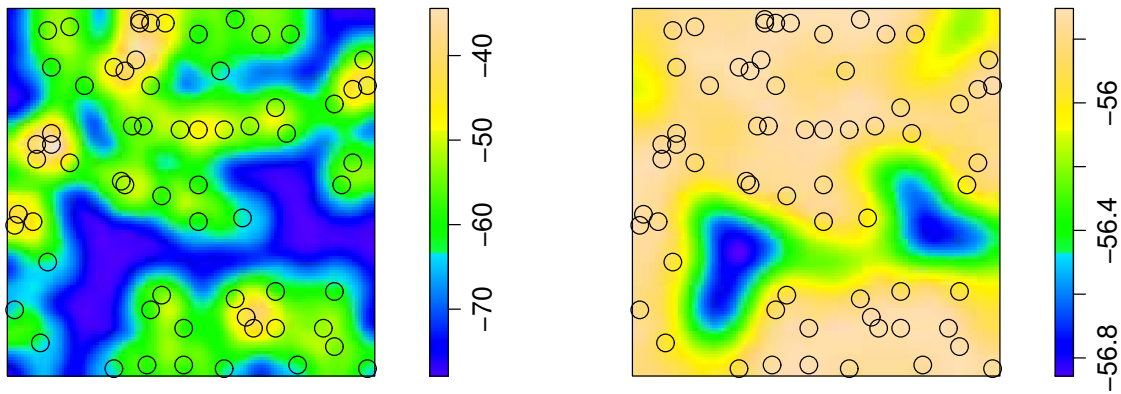


Figure: Raw residual analysis, from left to right : Strauss(r=0.08) and AreaInt(r=0.09)

- R code:
  ```
  >mjp=ppm(japanesepines, 1,Strauss(r=0.08),rbord=0.08)
  >rjp=residuals(mjp,type="raw") >plot(rjp)
  ```

QQ plots: comparison of empirical quantiles of the smoothed residuals with the expected quantiles under the estimated model

- interpretation in the spirit of K and F functions
- if the data pattern is more clustered than the model: heavier tails especially in the left-hand tail
- if the data pattern is more inhibited than the model: lighter tails especially in the right-hand tail
- R code : `qqplot.ppm(rjp, type="raw")`

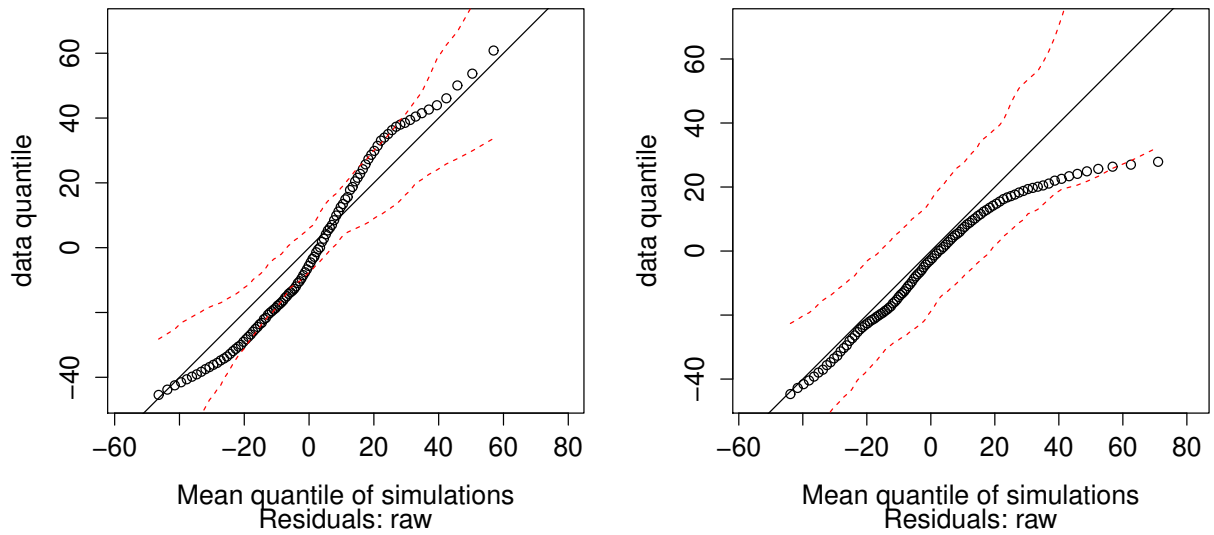


Figure: Q-Q plot analysis, from left to right : Strauss(r=0.08) and AreaInt(r=0.09)

- Strauss(r=0.08): over-estimates repulsion, but under-estimates close attraction
- AreaInter(r=0.09): very well for the close attraction, underestimate the repulsion
- the best model for the entire data set : polynomial inhomogeneity and soft-core interaction

Remarks :

- the theory is wonderful
- but the numerical results are obtained using the PL estimators ...
- see the remark of J. Besag
- visualisation of residuals difficult for higher dimensions
- the qq plots very informative $\rightarrow$ link with the central limit theorems for computing confidence intervals
- open question: validating pattern detection result ... ?

$\rightarrow$ Exercise 23

Table of contents

Conclusions and perspectives (1)

Pattern detection: inference related to the distribution of the pattern knowing the parameters $p(x|\theta)$

- Markov point processes allow statistical and morphological description of the detected patterns
- good synthesis properties: compute moments and integrals
- limitations: models remain just models ...

Figure: Creation by Laurent Ballesta. Winner of the Wildlife Photographer of the Year 2021. Royal Ontario Museum.

Conclusions and perspectives (2)

Pattern characterisation: inference related to the distribution $p(\theta|\mathrm{x})$ of the parameters knowing the pattern

- ABC and SSA Shadow algorithms: alternative option whenever exact simulation is not available
- theoretical control: equivalent with classical stochastic approximation methods (stochastic gradient)
- pros: complementary tools, not traped in local minima, more freedom regarding the choice of priors
- and cons: range parameters, models remain just models ...

Simultaneous pattern detection and characterisation: since sampling from $p(\mathrm{y}|\theta)$ and $p(\theta|\mathrm{y})$ is possible, what about sampling $p(\mathrm{y}, \theta)$?

Current projects:

- three on-going phd thesis: 1 math (N. Gillot) + 1 maths and geology (A. Fratani) + 1 maths and cosmology (F. Nørby)
- Horizon European project with Estonian, Dutch and German cosmologists: autumn 2024 - autumn 2027
- DRlib: C++ library for marked point processes based inference `https://gitlab.univ-lorraine.fr/labos/iecl/drlib`

Some perspectives:

- open problems: missing data, exact simulation, number of objects in a configuration, dimensionality, dynamical behaviour
- develop the research directions: modelling, simulation, statistical inference within each new context
- applications domains but not only: astronomy, geology, computer science, entreprises, etc.
- new research projects together ? : parameter estimation, model validation, detection and characterisation of latent structures

Thank you very much for your invitation. Questions, comments, remarks ?