Marked point processes : a tool for morphological and statistical characterisation of patterns in digital images and spatial data

Radu Stoica
Université Lille 1 - Laboratoire Paul Painlevé
radu.stoica@math.univ-lille1.fr

Workshop on Stochastic Geometry and its Applications
University of Rouen, March 28-30 2012

How to tackle pattern characterisation ?

Problem formulation :

- study physical phenomenas $\rightarrow$ register data
- spatial data $\rightarrow$ data element $=$ location $+$ characteristic ("where the phenomenon take place" $+$ "what it is measured")
- characterise the shape outlined by the observed data

Methodology : probability, statistics and a good ... feeling ...

- exploratory statistics $\rightarrow$ do we see any pattern in the data ?
- formulate hypothesis and modeling $\rightarrow$ define the shape we would like to find ...
- Monte Carlo simulation $\rightarrow$ build the shape in the middle of the data field ...
- inference $\rightarrow$ describe the pattern characterising the shape,
- validate the results $\rightarrow$ does the detected shape really exist ?

Remote sensing

Road and hydrographic networks

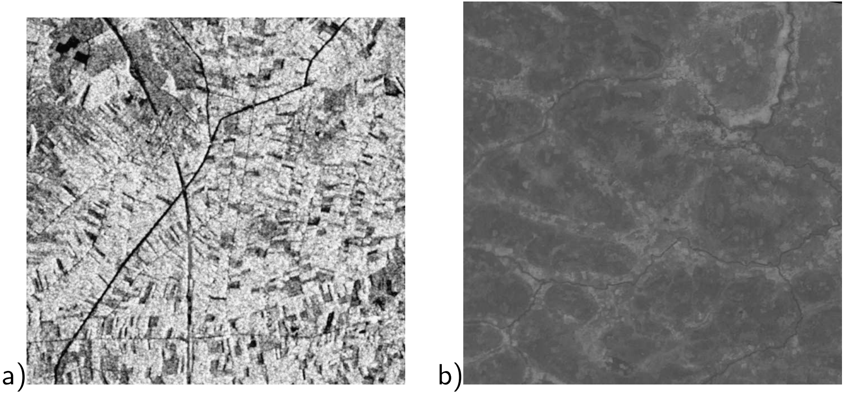


Figure: a) Rural region in Malaysia (http://southport.jpl.nasa.gov), b) Forest galleries (BRGM).

Cosmology : spatial distribution of galactic filaments (1)

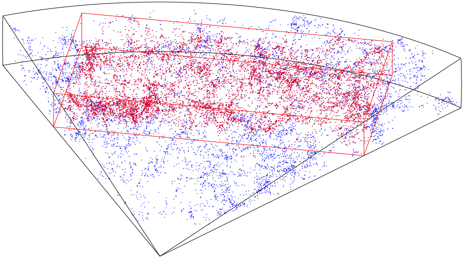
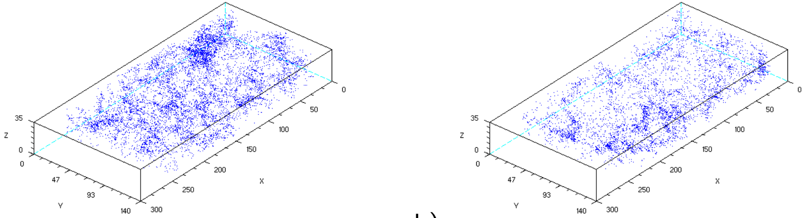


Figure: Cuboidal sample from the North Galactic Cap of the 2dF Galaxy Redshift Survey. Diameter of a galaxy $\sim 30 \times 3261.6$ light years.

Cosmology : study of mock catalogues (2)



a) b)

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

Epidemiology (veterinary context)

Disease : 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

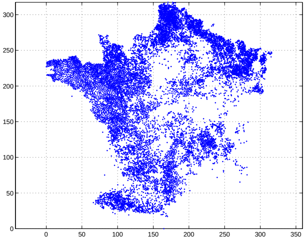


Figure: The spatial distribution of the farms outlines almost the entire French territory (INRA Avignon).

Markovian modeling of the pattern

Main hypothesis : the hidden pattern is a complicate entity made of simple interacting objects

Examples :

- road and hydrographic networks, cosmological filamentary structure
- "invisible" regions : clusters

Two "classical" families of models

- Markov random fields (Li, Winkler, Geman, Younes, etc.)
- Markov marked point processes (Baddeley, Lieshout, Møller, Stoyan,etc.)
- key point : local specification of the model $\rightarrow$ break the complexity of the pattern

Marked point processes : definition, properties, examples

Key hypothesis reformulated :

the shape we are looking for is the realization of a marked point process

Ingredients :

- measure space : $(K, \mathcal{B}, \nu)$, $K \subset \mathbb{R}^d$, $\mathcal{B}$ the Borel $\sigma-$algebra and $0 < \nu(K) < \infty$ the Lebesgue measure
- probability space: $(\Omega, \mathcal{F}, \mu)$

Configuration space :

$$\Omega = \cup_{n=0}^{\infty} K_n, \quad n \in \mathbb{N}$$
$$K_n = \{k_1, \ldots, k_n\} \subset K$$

Definition :

a point process in K is measurable mapping from a probability space in $(\Omega, \mathcal{F})$.

Marked point processes :

a random sequence $\mathbf{y} = \{y_n = (k_n, m_n)\}$ such that the points k_n are a point process in K and m_n are the marks corresponding for each k_n. Let $(M, \mathcal{M}, \nu_M)$ be the marks space where $\nu_M(M) = 1$.

In pattern recognition :

- k_n : objects locations
- m_n : objects characteristics (geometrical shape, texture, but also : species,age, disease,etc.)

The simplest marked point process is the Poisson object process :

- number of objects $\sim$ Poisson$(\nu(K))$
- locations and marks independent : $k_i \sim \frac{1}{\nu(K)}$ and $m_i \sim \nu_M$

Poisson object process

- reference probability measure

$$\mu(F) = \sum_{n=0}^{\infty} \frac{e^{-\nu(K)}}{n!} \int_{K\times M} \cdots \int_{K\times M} \mathbf{1}_F\{(k_1, m_1), \ldots, (k_n, m_n)\}$$
$$\times d\nu(k_1) d\nu_M(m_1) \ldots d\nu(k_n) d\nu_M(m_))$$

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

- Boolean model : analytical formulas, null hypothesis $\leftrightarrow$ the law is completely known
- no interaction $\rightarrow$ no shape ...
- more complicate models $\rightarrow$ specifying a probability density $p(\mathbf{y})$ such that :

$$P(Y \in F) = \int_F p(\mathbf{y}) \mu(d\mathbf{y})$$

- in this case the normalising constant is not known

A word on random sets (1)

Intuitive definition : general mathematical tool dealing with sets of objects having random characteristics

- marked point processes $\leftrightarrow$ random sets

Capacity functional : local characterisation of the distribution of a random set Ξ

$$T_{\Xi}(\mathcal{K}) = \mathbb{P}(\Xi \cap \mathcal{K} \neq \emptyset)$$

where $\mathcal{K}$ is any compact set in $\mathbb{R}^d$

A word on random sets (2)

Choquet theorem : the distribution of a random set is completely determined by knowledge of the capacity functional

• important tools for exploratory statistics : spherical contact distribution, volumic fraction, mean quantities (perimeter, area, volume), covariance, moments, etc.

• Boolean models only $\rightarrow$ analytical closed form of these quantities

• ... not always very informative ...

• this may justify the need of the probability densities for more complex models

Interacting marked point processes (1)

Construction of the probability density :

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

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

  for any $k-$tuple of objects

- the density is the product of all these functions

$$p(\mathbf{y}) = \alpha \prod_{y_i \in \mathbf{y}} \phi(y_i)^{(1)} \cdots \prod_{\{y_{i_1}, \ldots, y_{i_k}\} \in \mathbf{y}} \phi(y_{i_1}, \ldots, y_{i_k})^{(k)}$$

- α the normalising constant is now known

Interacting marked point processes (2)

Papangelou conditional intensity : local specification of the model

$$\lambda(\zeta; \mathbf{y}) = \frac{p(\mathbf{y} \cup \{\zeta\})}{p(\mathbf{y})}$$

for $\zeta \in K \times M$ and $\zeta \notin \mathbf{y}$

• interpretation : probability (energy contribution) of adding a new object to the configuration

• plays a similar role as the conditional probabilities for Markov random fields

Interacting marked point processes (3)

Properties of the probability density : a lot of freedom for specifying models

• integrability (Ruelle stability)

$$p(\mathbf{y}) \leq \Lambda^{n(\mathbf{y})}$$

• local stability $\rightarrow$ implies Ruelle stability ; important for MCMC dynamics convergence

$$\lambda(\zeta; \mathbf{y}) \leq \Lambda$$

• monotonic or anti-monotonic : order relation on the configuration space (inclusion)

Markov point processes (1)

Neighbourhood system :

- $\sim$: symmetric, reflexive neighbourhood relation $K \times M$

- clique : a configuration $\mathbf{y} \in \Omega$ such that $u \sim v$ for any $u, v \in \mathbf{y}$; the empty set is a clique

Examples : distance (Euclidean, Hausdorff), set intersection, etc.

Definition (Ripley and Kelly '77) : A point process Y is Markov w.r.t. the symmetric, reflexive relation $\sim$ on $K \times M$, if for all $\mathbf{y}$ such that $p(\mathbf{y}) > 0$:

- $p(\mathbf{z}) > 0$ for all $\mathbf{z} \subset \mathbf{y}$

- $\frac{p(\mathbf{y} \cup \{\zeta\})}{p(\mathbf{y})}$ depends only on ζ and $\partial(\zeta) \cap \mathbf{y} = \{\eta \in \mathbf{y} : \eta \sim \zeta\}$

Markov point processes (2)

Example

the density of Poisson object process with intensity β is :

$$p(\mathbf{y}) = \beta^{n(\mathbf{y})} \exp[(1 - \beta)\nu(K)]$$

with respect to $\mu(\cdot)$.

- $p(\mathbf{y}) > 0$ for any configuration $\mathbf{y}$
- $\lambda(\zeta; y) = \beta \mathbf{1}\{\zeta \notin \mathbf{y}\}$

$\Rightarrow$ $p(\cdot)$ is Markov for any choice of the neighbourhood system

In accordance with the fact that a Poisson object process represents complete spatial randomness

Spatial Markov property

Theorem

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

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

Remarks :

- a drawing to better understand ...
- compare with the result obtained for Markov random fields.

Hammersley-Clifford factorisation

Theorem

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

$$p(\mathbf{y}) = \prod_{\textit{cliques} \quad \mathbf{z} \subseteq \mathbf{y}} \phi_c(\mathbf{z}), \quad \alpha = \phi(\emptyset)$$

for all $\mathbf{y} \in \Omega$.

Gibbs point processes :

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

where Z is the partition function, U is the system energy and U_c is the clique potential ($U_c(\emptyset) = 0$).

Example

Poisson object process with density

$$p(\mathbf{y}) = e^{(1-\beta)\nu(K)} \prod_{y \in \mathbf{y}} \beta$$

- the clique interaction functions are given by :

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

and $\phi_c \equiv 1$ for cliques of two or more objects

- the clique potentials

$$U_c(\cdot) = -\log \phi_c(\cdot)$$

- confirmation of the lack of interaction

Distance interaction model - Strauss model :

(Strauss, 1975), (Kelly and Ripley, 1976)

$$p(\mathbf{y}) \propto \beta^{n(\mathbf{y})}\gamma^{s_r(\mathbf{y})}, \quad \beta > 0, \gamma \in [0, 1]$$

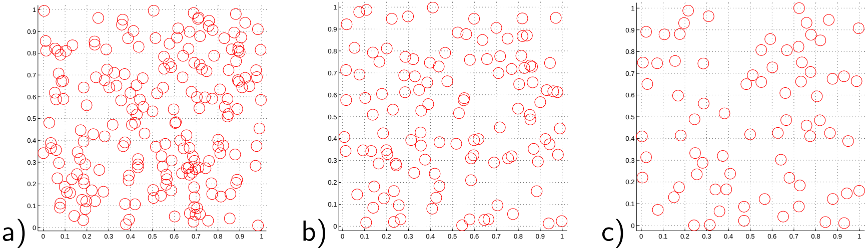


Figure: Strauss model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma = 0.5$ and c) $\gamma = 0.0$.

Area interaction model :

(Baddeley and van Lieshout, 1995)

$$p(\mathbf{y}) \propto \beta^{n(\mathbf{y})} \gamma^{-\nu[Z(\mathbf{y})]}, \quad \beta, \gamma > 0$$

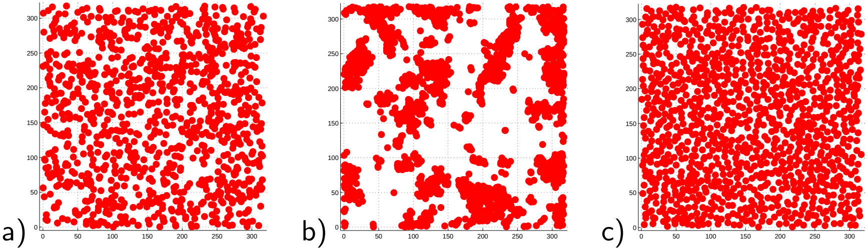


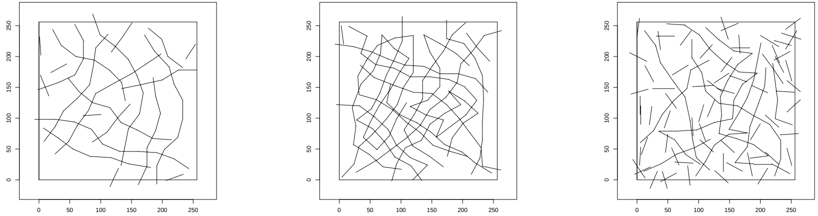
Figure: Area interaction model realisations for different parameter values : a) $\gamma = 1.0$, b) $\gamma > 1.0$ and c) $\gamma < 1.0$.

Candy model :

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

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

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)

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

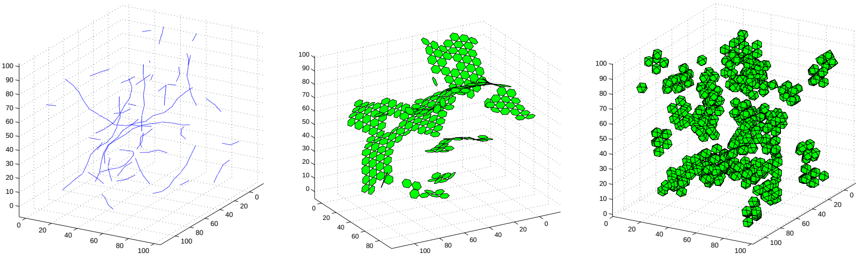


Figure: Random shapes generated with Bisous model.

Markov chain Monte Carlo algorithms

Problem : sampling probability distribution that are not available in closed form (*e.g.* normalising constant not available)

Solution : Monte Carlo method

- simulating a Markov chain $\rightarrow$ building a reversible transition kernel
- the equilibrium distribution of the chain is the distribution we want to sample
- statistical inference is possible
- several solutions : Gibbs sampler, Metropolis-Hastings, birth an death processes, stochastic adsorption, RJMCMC, exact simulation (CFTP, clan of ancestors, etc.)
- open problems : convergence at infinity (almost all the methods) and parameter dependence ("perfect" methods)

Adapted MH dynamics

- theoretical convergence properties $\rightarrow$ in practice *burning-in time*
- local computations
- allows improvements : transition kernels that "help" the model

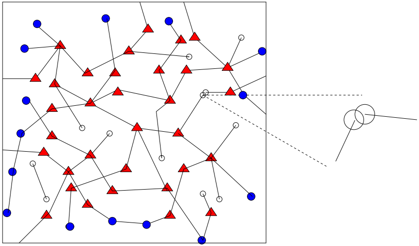


Figure: Extremities marked by triangles are connected and further than $\frac{1}{2} l_{\max} + r_c$ to the boundary, those labeled by a black disk are closer than $\frac{1}{2} l_{\max} + r_c$ to the boundary of K.

Perfect sampling algorithms

Exact simulation : CFTP, clan of ancestors, exact Metropolis-Hastings, Gibbs

- the simulated chain indicates by itself whenever convergence is reached
- model parameters should have "very, very nice" values
- can be applied in practice only to a restricted range of parameters

Strauss model : convergence speed for exact sampling methods

(van Lieshout and Stoica, 2006)

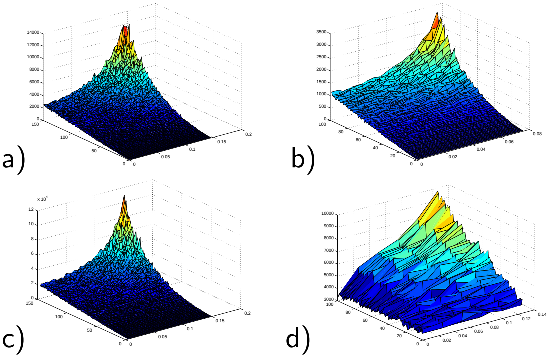


Figure: Exact simulation algorithms applied to Strauss model : a) CFTP, b) clan of ancestors, c) Metropolis-Hastings and d) Gibbs.

Statistical inference problems

Problem I : parameter estimation

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

Problem II : pattern detection

- observe the data $\mathbf{d}$ and find $\mathbf{y}$ "hidden"
- the model parameters are : hidden, modeled, unknown
- open problem : the detected pattern does it really exist ...?

Problem III : shape modeling

- observe a phenomenon and propose a model doing the "same" ...
- needs the time dimension
- open problem : time, what it is ? "crystal ball" ?

Statistical pattern detection (1)

Build the model : probability density construction

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

- interaction energy $U_i(\mathbf{y}|\theta) \rightarrow$ objects interactions
- data energy $U_{\mathbf{d}}(\mathbf{y}|\theta)$ induced by the data field $\mathbf{d}$ $\rightarrow$ object locations
- if the interaction parameters are unknown $\rightarrow$ prior model $p(\theta)$

Pattern estimator :

the object configuration that maximises the probability density

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

with Ψ the model parameters space

Statistical pattern detection (2)

Simulated annealing : global optimisation technique

- sampling from $p(\mathbf{y}, \theta)^{1/T_{sa}}$ while slowly $T_{sa} \to 0$
- convergence towards the uniform distribution on the configuration subspace minimising $U(\mathbf{y}, \theta)$ (Stoica, Gregori and Mateu, 2005)

Level sets estimators :

- visit maps for compact regions in K (Heinrich, Stoica and Tran, 2012) :

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

- two challenges : discretisation and Monte Carlo approximations
- average behaviour of the pattern (fixed temperature)

Build the machine ...

Remotely sensed images :

- interaction energy : Candy model (random segments)
- data energy : local hypothesis tests (checking the pixels covered by a segment)

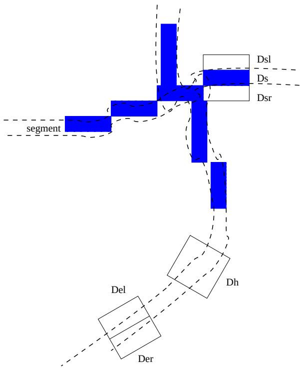


Figure: Connected segments approximating a thin network.

Galaxies catalogues :

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

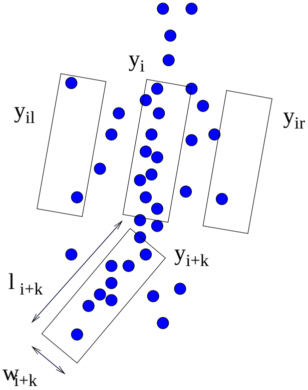


Figure: Locating interacting cylinders in a field of points.

Epidemiological data :

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

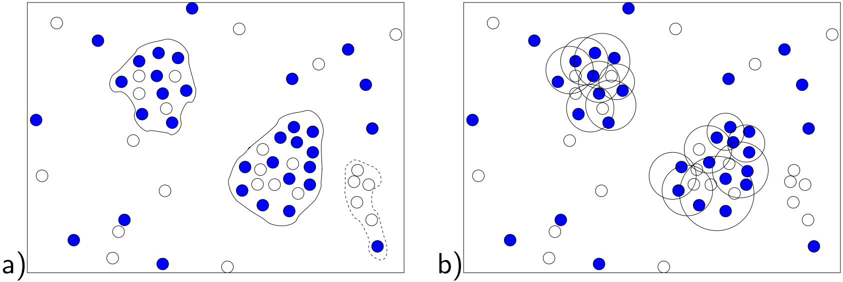


Figure: Data→ field of marked points : a) observed clusters, b) clusters approximated by random disks.

Road network extraction

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

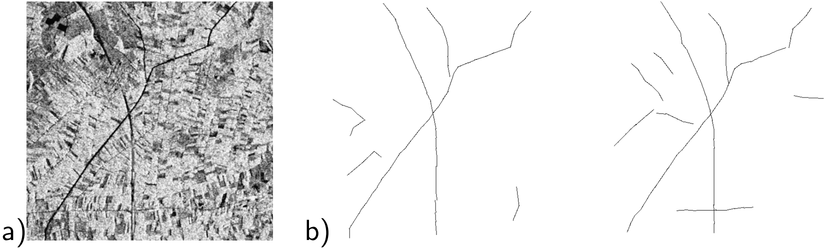


Figure: Rural region in Malaysia : a) original image; b) obtained results.

Forest galleries : verifying the results

(Stoica, Descombes and Zerubia, 2004)

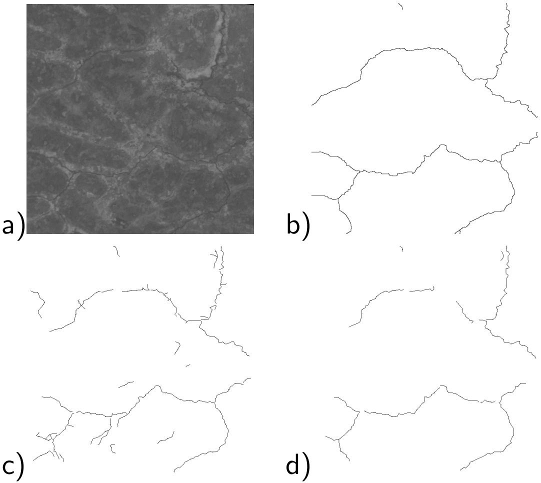


Figure: Forest galleries extraction : a) original image ; b) ground truth ; c)-d) obtained results. Data provided by BRGM.

Catalogue NGP150 (1)

(Stoica, Martinez and Saar, 2007)

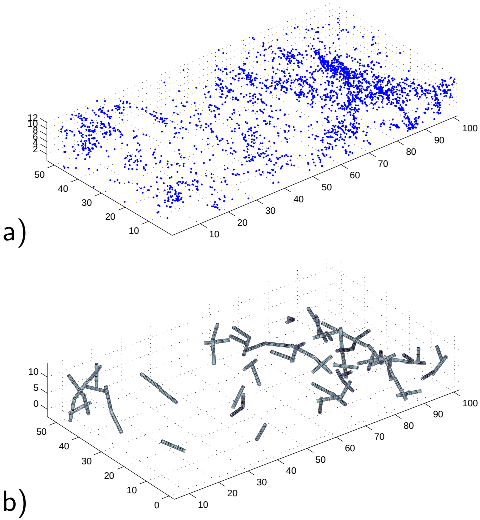


Figure: a) Original data. b) Cylinder configuration obtained after running the simulated annealing algorithm.

Catalogue NGP150 (2)

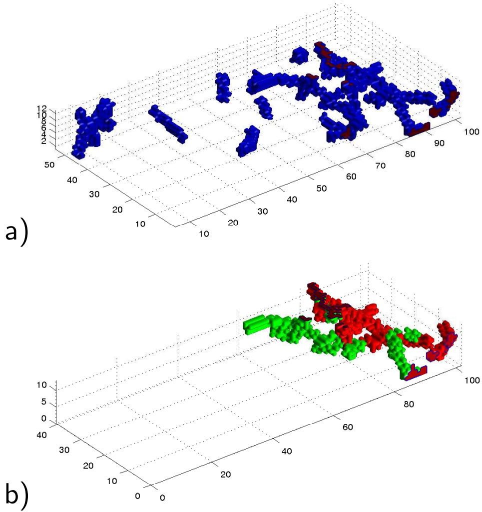


Figure: a) Cover probability thresholded at 95%. b) Structure of a filament : green shading shows the filament obtained with $\mathbb{P}_W = 0.5$, red shading $\mathbb{P}_W = 0.95$.

Epidemiology : sub-clinical mastitis data

(Stoica, Gay and Kretzschmar, 2007)

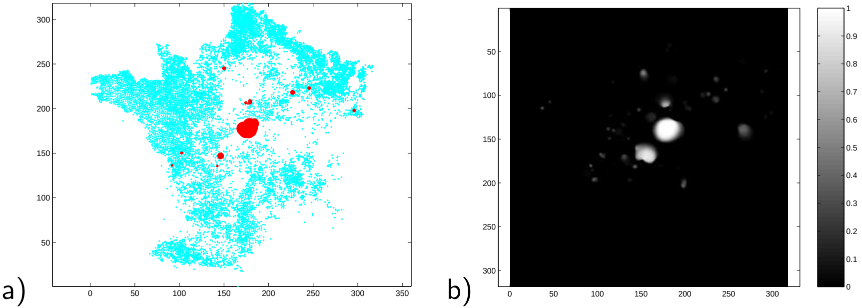


Figure: Disease data scores and coordinates for the year 1996 : a) disk configuration obtained using the simulated annealing algorithm ; b) cover probabilities.

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 :

- cosmology : free cylinders, cylinders with one extremity connected, cylinders with both extremities connected
- epidemiology : number of pairs of overlapping disks, surface occupied by the disks

Test for the galaxy catalogues

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 catalogues)		
	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 epidemiological data

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

Results :

- sufficient statistics for the data from the year 1996 :

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

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

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

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

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, 2010)

Discussion

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

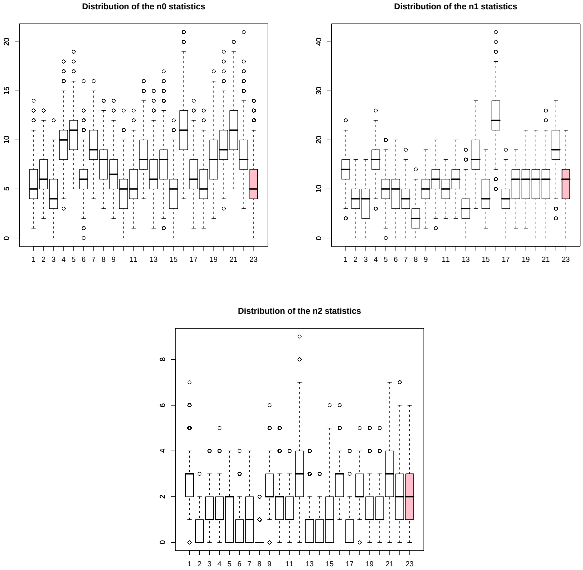


Figure: Comparison of the sufficient statistics distributions for the real data (dark box plot) and the mock catalogues.

Spatial Markov models :

- marked point processes allow statistical and morphological description of the pattern
- good synthesis properties
- limitations : models remain just models ...

Perspectives :

- random geometry (marked point processes, random fields) $\rightarrow$ modeling, simulation, statistics and also temporal dimension ...
- applications : astronomy, cosmology, geology, environmental sciences

Acknowledgements :

this work was done together with wonderful co-authors and also with help of some very generous people ... Some of them are today with us :) ...