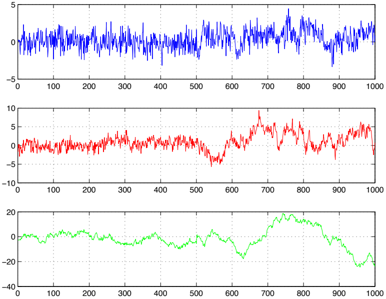
Statistical study of spatial dependences in long memory random fields on a lattice, point processes and random geometry.

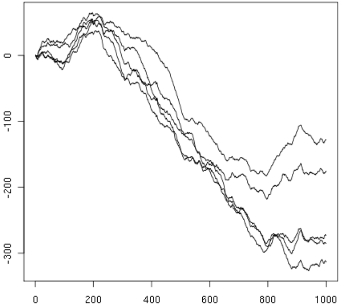
Frédéric Lavancier,
Laboratoire de Mathématiques Jean Leray, Nantes

9 décembre 2011.

Introduction

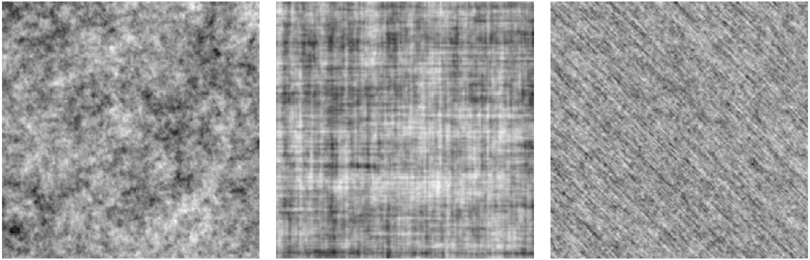
Long memory, self-similarity in (multivariate) time series :





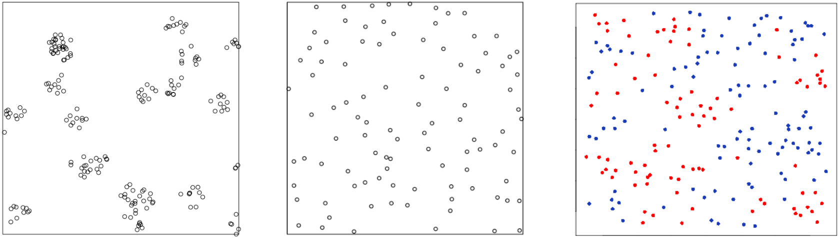
Introduction

Long memory, self-similarity in images :



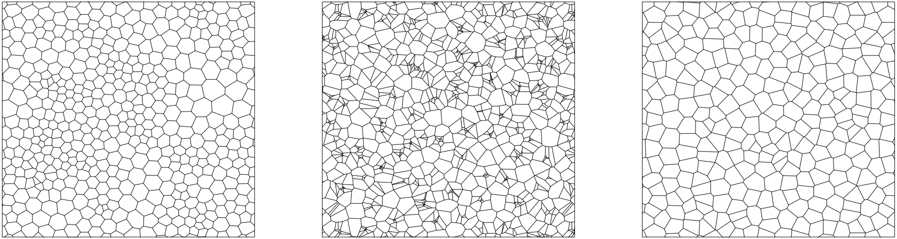
Introduction

Attraction, repulsion between points :



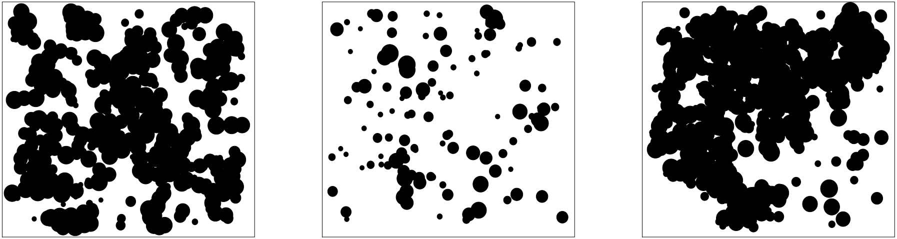
Introduction

Dependence between cells of a tesselation :



Introduction

Random sets as a union of interacting balls :



Introduction

General motivations

Developping mathematical models that

- respect some prescribed features (e.g. self-similiarity, long range dependence, repulsion or attraction between points,...) ;
- are flexible enough (through few parameters) ;
- we can simulate.

Introduction

General motivations

Developping mathematical models that

- respect some prescribed features (e.g. self-similiarity, long range dependence, repulsion or attraction between points,...) ;
- are flexible enough (through few parameters) ;
- we can simulate.

From a stastical point of view :

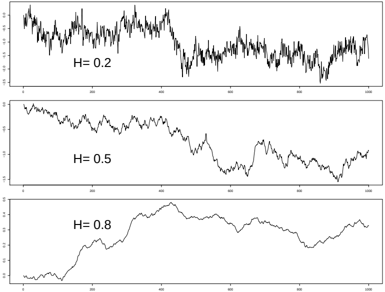
- fitting these models to data (inference problem) ;
- assessing the theoretical quality of inference (consistency, limiting law, optimality,...) ;
- providing some diagnostic tools :
  - adequation of the model to data,
  - change-point problems.

1. Self-similarity, Long memory
    - Vector Fractional Brownian Motion
    - Long memory time series
    - Long memory random fields (images)
2. Point processes, random geometry
    - Estimation of Gibbs point processes
    - Model validation for Gibbs point processes

1. Self-similarity, Long memory
   - Vector Fractional Brownian Motion
   - Long memory time series
   - Long memory random fields (images)

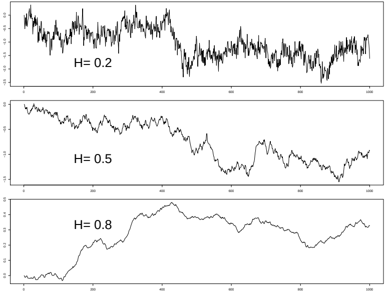
Fractional Brownian Motion

Some univariate examples :

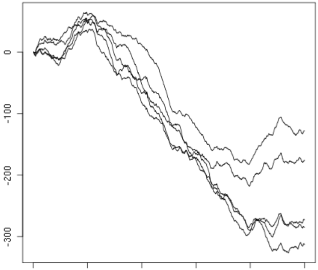


Fractional Brownian Motion

Some univariate examples :



Aim : correlating p fractional Brownian motions
$\rightarrow$ Vector (or Multivariate) FBM



Vector FBM with P.-O. Amblard, J.-F. Coeurjolly, A. Philippe, D. Surgailis

A multivariate process $\mathbf{B}(t) = (B_1(t), \ldots, B_p(t))$ is a Vector FBM with parameter $\mathbf{H} = (H_1, \ldots, H_p)$ if $\mathbf{B}(0) = 0$ and

- it is Gaussian ;
- it is **H**-self-similar, i.e.

$$\forall c > 0, \; (B_1(ct), \ldots, B_p(ct))_{t\in\mathbb{R}} \stackrel{\mathcal{L}}{=} (c^{H_1} B_1(t), \ldots, c^{H_p} B_p(t))_{t\in\mathbb{R}};$$

- it has stationary increments.

Vector FBM with P.-O. Amblard, J.-F. Coeurjolly, A. Philippe, D. Surgailis

A multivariate process $\mathbf{B}(t) = (B_1(t), \dots, B_p(t))$ is a Vector FBM with parameter $\mathbf{H} = (H_1, \dots, H_p)$ if $\mathbf{B}(0) = 0$ and

- it is Gaussian ;
- it is $\mathbf{H}$-self-similar, i.e.

$$\forall c > 0, \; (B_1(ct), \dots, B_p(ct))_{t \in \mathbb{R}} \stackrel{\mathcal{L}}{=} (c^{H_1} B_1(t), \dots, c^{H_p} B_p(t))_{t \in \mathbb{R}};$$

- it has stationary increments.

Lamperti type result :
If there exist a vector process $(Y_1(t), \dots, Y_p(t))_{t \in \mathbb{R}}$ and real functions $a_1, \dots, a_p$ such that

$$(a_1(n) Y_1(nt), \dots, a_p(n) Y_p(nt)) \xrightarrow[fidi]{n \to \infty} \mathbf{Z}(t),$$

then the vector process $(\mathbf{Z}(t))$ is self-similar.

Convergence of partial sums :
Any Vector FBM can be obtained as the limit of partial sums of some superlinear processes.

Comprehensive characterisation : $(p = 2)$

1. Let $(B_1(t), B_2(t))$ be a (H_1, H_2)-VFBM, then (when $H_1 + H_2 \neq 1$) :

- B_1 is a H_1-FBM and B_2 is a H_2-FBM
- for $0 \leq s \leq t$, the cross-covariance is

  $$\mathbb{E}B_1(s)B_2(t) \;\propto\; (\rho + \eta)s^{H_1+H_2} + (\rho - \eta)t^{H_1+H_2} - (\rho - \eta)(t - s)^{H_1+H_2}$$

  with $\rho = \mathrm{corr}(B_1(1), B_2(1))$, $\eta = \frac{\mathrm{corr}(B_1(1),B_2(-1)) - \mathrm{corr}(B_1(-1),B_2(1))}{2 - 2^{H_1+H_2}}$.

Comprehensive characterisation : $(p = 2)$

1. Let $(B_1(t), B_2(t))$ be a (H_1, H_2)-VFBM, then (when $H_1 + H_2 \neq 1$) :

- B_1 is a H_1-FBM and B_2 is a H_2-FBM
- for $0 \leq s \leq t$, the cross-covariance is

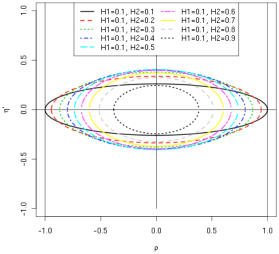
$$\mathbb{E}B_1(s)B_2(t) \propto (\rho+\eta)s^{H_1+H_2} + (\rho-\eta)t^{H_1+H_2} - (\rho-\eta)(t-s)^{H_1+H_2}$$

with $\rho = \mathrm{corr}(B_1(1), B_2(1))$, $\eta = \frac{\mathrm{corr}(B_1(1),B_2(-1)) - \mathrm{corr}(B_1(-1),B_2(1))}{2-2^{H_1+H_2}}$.

2. Conversely any Gaussian process with the above covariance function is a VFBM iff for some known $R > 0$

$$\rho^2 \sin\left(\frac{\pi}{2}(H_1+H_2)\right)^2 + \eta^2 \cos\left(\frac{\pi}{2}(H_1+H_2)\right)^2 \leq R$$

$\rightarrow$ not possible to set up arbitrary correlated FBM



Comprehensive characterisation : $(p = 2)$

1. Let $(B_1(t), B_2(t))$ be a (H_1, H_2)-VFBM, then (when $H_1 + H_2 \neq 1$) :

- B_1 is a H_1-FBM and B_2 is a H_2-FBM
- for $0 \leq s \leq t$, the cross-covariance is

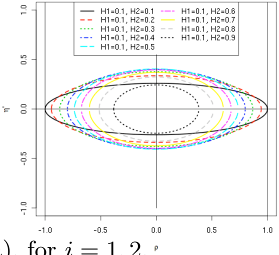
$$\mathbb{E}B_1(s)B_2(t) \;\propto\; (\rho+\eta)s^{H_1+H_2} + (\rho-\eta)t^{H_1+H_2} - (\rho-\eta)(t-s)^{H_1+H_2}$$

with $\rho = \mathrm{corr}(B_1(1), B_2(1))$, $\eta = \frac{\mathrm{corr}(B_1(1),B_2(-1)) - \mathrm{corr}(B_1(-1),B_2(1))}{2-2^{H_1+H_2}}$.

2. Conversely any Gaussian process with the above covariance function is a VFBM iff for some known $R > 0$

$$\rho^2 \sin\left(\frac{\pi}{2}(H_1+H_2)\right)^2 + \eta^2 \cos\left(\frac{\pi}{2}(H_1+H_2)\right)^2 \leq R$$

$\rightarrow$ not possible to set up arbitrary correlated FBM



3. Consider the increments $\Delta B_i(n) = B_i(n+1) - B_i(n)$, for $i = 1, 2$.

Comprehensive characterisation : $(p=2)$

1. Let $(B_1(t), B_2(t))$ be a (H_1, H_2)-VFBM, then (when $H_1 + H_2 \neq 1$) :

- B_1 is a H_1-FBM and B_2 is a H_2-FBM
- for $0 \leq s \leq t$, the cross-covariance is

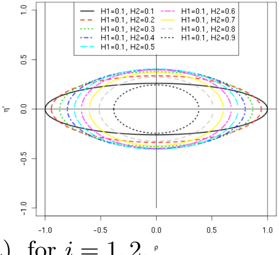
$$\mathbb{E}B_1(s)B_2(t) \;\propto\; (\rho+\eta)s^{H_1+H_2} + (\rho-\eta)t^{H_1+H_2} - (\rho-\eta)(t-s)^{H_1+H_2}$$

with $\rho = \mathrm{corr}(B_1(1), B_2(1))$, $\eta = \frac{\mathrm{corr}(B_1(1),B_2(-1)) - \mathrm{corr}(B_1(-1),B_2(1))}{2-2^{H_1+H_2}}$.

2. Conversely any Gaussian process with the above covariance function is a VFBM iff for some known $R > 0$

$$\rho^2 \sin\left(\frac{\pi}{2}(H_1+H_2)\right)^2 + \eta^2 \cos\left(\frac{\pi}{2}(H_1+H_2)\right)^2 \leq R$$

$\rightarrow$ not possible to set up arbitrary correlated FBM



3. Consider the increments $\Delta B_i(n) = B_i(n+1) - B_i(n)$, for $i = 1, 2$. The two components are either non-correlated components or

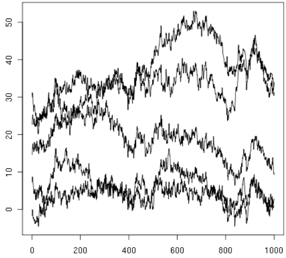
$$\mathbb{E}\Delta B_1(n)\Delta B_2(n+h) \sim \kappa |h|^{H_1+H_2-2}$$

$\rightarrow$ Very constrained cross-correlation
(For instance $H_1 + H_2 \geq 1 \Rightarrow$ long-range cross-dependence)

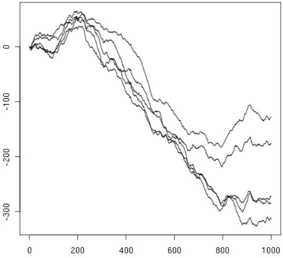
Some examples :

Simulations are achieved thanks to a Wood and Chan algorithm.

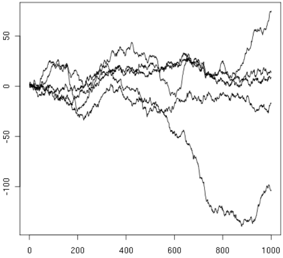
Below, the correlation between any couple of FBM is $\rho = 0.6$



$H \in [0.3, 0.4]$ (decentered)



$H \in [0.8, 0.9]$



$H \in [0.4, 0.8]$

1. Self-similarity, Long memory
   - Vector Fractional Brownian Motion
   - Long memory time series
   - Long memory random fields (images)

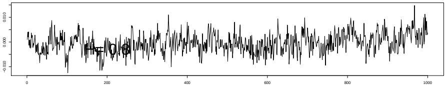
Definition :
A stationary, L^2, time series $(X(n))_{n\in\mathbb{Z}}$ exhibits long memory if

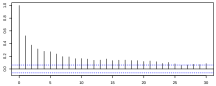
$$\sum_{n\in\mathbb{Z}} |r(n)| = +\infty$$

where r denotes the covariance function of X.

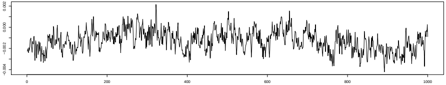
Typically : for some $0 < d < 0.5$, $r(n) \sim_\infty \kappa |n|^{2d-1}$.
The parameter d is called the *long memory parameter.*

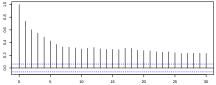
$d = 0.3$:





$d = 0.45$:



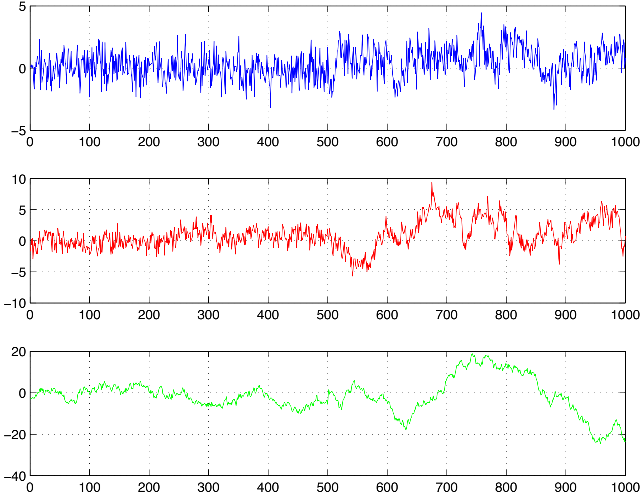


Examples :

- If B is a FBM, $X(n) = B(n+1) - B(n)$ with $d = H - 0.5$.
- An $I(d)$ time series : $X(n) = (1-L)^{-d}\epsilon(n)$
  (where ϵ is a white noise and L the lag operator : $L\epsilon_n = \epsilon_{n-1}$)

Change point problem with R. Leipus, A. Philippe, D. Surgailis

Constant vs non-constant long memory parameter ? ex : $I(d_1) \to I(d_2)$



Change point problem with R. Leipus, A. Philippe, D. Surgailis

Basically, under $\mathbf{H}_0$: $\underline{d \text{ is constant}}$:

$$n^{-d-0.5}\sum_{k=1}^{[nt]}X(k) \xrightarrow{\mathcal{D}([0,1])} \kappa B_{d+0.5}(t) \quad \text{and} \quad Var\left(\sum_{k=1}^{[nt]}X(k)\right) \approx n^{2d+1}$$

Under $\mathbf{H}_1$: $\underline{d \text{ increases}}$, for some t_0

$$Var\left(\sum_{k=1}^{[nt_0]}X(k)\right) \ll Var\left(\sum_{k=[nt_0]}^{n}X(k)\right)$$

$\Rightarrow$ To test for $\mathbf{H}_0$ against $\mathbf{H}_1$, we estimate and compare these variances.

Change point problem with R. Leipus, A. Philippe, D. Surgailis

Basically, under $\mathbf{H}_0$: <u>d is constant</u> :

$$n^{-d-0.5}\sum_{k=1}^{[nt]} X(k) \xrightarrow{\mathcal{D}([0,1])} \kappa B_{d+0.5}(t) \quad \text{and} \quad Var\left(\sum_{k=1}^{[nt]} X(k)\right) \approx n^{2d+1}$$

Under $\mathbf{H}_1$: <u>d increases</u>, for some t_0

$$Var\left(\sum_{k=1}^{[nt_0]} X(k)\right) \ll Var\left(\sum_{k=[nt_0]}^{n} X(k)\right)$$

$\Rightarrow$ To test for $\mathbf{H}_0$ against $\mathbf{H}_1$, we estimate and compare these variances.

Let $S_j = \sum_{k=1}^{j} X(k)$ and $S^*_{n-j} = \sum_{k=j+1}^{n} X(k)$. Define

- the *forward* variance : $V_k = \widehat{Var}(S_1, \dots, S_k)$
- the *backward* variance : $V^*_{n-k} = \widehat{Var}(S^*_{n-k+1}, \dots, S^*_1)$

Test statistic, Consistency

$$I_n = \int_0^1 \frac{V^*_{n-[nt]}}{V_{[nt]}} \mathrm{d}t$$

Under $\mathbf{H}_0$: $I_n \to I(B_{d+0.5})$ (simulable) ; Under $\mathbf{H}_1$: $I_n \to +\infty$

1. Self-similarity, Long memory
   - Vector Fractional Brownian Motion
   - Long memory time series
   - Long memory random fields (images)

Definition :
A stationary, L^2, random field $(X(n))_{n\in\mathbb{Z}^d}$ exhibits long memory if

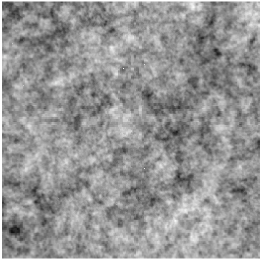
$$\sum_{n\in\mathbb{Z}^d} |r(n)| = +\infty$$

where r denotes the covariance function of X.

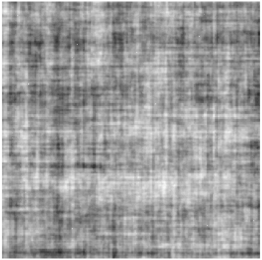
Main difference with time series : possible occurence of anisotropy.

Examples $(d = 2)$: for $0 < \alpha < 1$

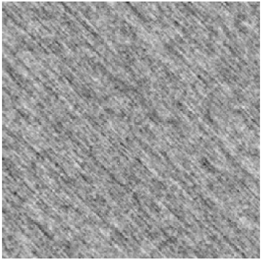
$r(n_1, n_2) \sim_\infty (n_1^2 + n_2^2)^{-\alpha}$



$r(n_1, n_2) \sim_\infty |n_1|^{-\alpha}|n_2|^{-\alpha}$



$r(n_1, n_2) \sim_\infty |n_1 + n_2|^{-\alpha}$



Investigations on long memory random fields

Modelling : long memory random fields appear

- in similar models as for time series (increment of fractional Brownian sheet ; aggregation of short memory random fields ; fractional filtering of a white noise)
- in some Gibbs processes on $\mathbb{Z}^d$ in phase transition
  ex : Ising model at the critical temperature

Investigations on long memory random fields

Modelling : long memory random fields appear

- in similar models as for time series (increment of fractional Brownian sheet ; aggregation of short memory random fields ; fractional filtering of a white noise)
- in some Gibbs processes on $\mathbb{Z}^d$ in phase transition
  ex : Ising model at the critical temperature

Limit theorems, statistical applications :

- for partial sums $\longrightarrow$ non Central Limit Theorems
  $\Rightarrow$ Testing for the presence of long memory.
- for the empirical process $\longrightarrow$ asymptotic degeneracy
  $\Rightarrow$ Asymptotics for U-Statistics.
- for some quadratic forms (with A. Philippe)
  $\longrightarrow$ non-CLTs for $\sum g(i-j)X(i)X(j)$ where $(X(i))_{i\in\mathbb{Z}^2}$ is Gaussian.
  $\Rightarrow$ Asymptotics of empirical covariance functions.

Partial sums of long memory random fields

Let $X(n) = \sum_{k \in \mathbb{Z}^d} a_k \, \epsilon(n-k)$, where $(a_k) \in \ell^2$ and ϵ is a Gaussian white noise.

Theorem

Denote $a(x) = \sum_{k \in \mathbb{Z}^d} a_k \mathrm{e}^{i\,k.x}$. If $a \in L^2$ and $\forall \lambda$, $a(\lambda x) = |\lambda|^{-\alpha} a(x)$, $0 < \alpha < d$, then denoting $A_n = \{1, \dots, n\}^d$

$$\frac{1}{n^{d/2+\alpha}} \sum_{k \in A_{[nt]}} X_k \xrightarrow{\mathcal{D}([0,1]^d)} \int_{\mathbb{R}^d} a(x) \prod_{j=1}^{d} \frac{\mathrm{e}^{it_j x_j} - 1}{i x_j} \mathrm{d}Z(x)$$

The limit is the Fractional Brownian Sheet only when $a(x) = \prod_{i=1}^{d} |x_i|^{-H_i}$.

Partial sums of long memory random fields

Let $X(n) = \sum_{k\in\mathbb{Z}^d} a_k\,\epsilon(n-k)$, where $(a_k) \in \ell^2$ and ϵ is a Gaussian white noise.

Theorem

Denote $a(x) = \sum_{k\in\mathbb{Z}^d} a_k \mathrm{e}^{i\,k.x}$. If $a \in L^2$ and $\forall\lambda$, $a(\lambda x) = |\lambda|^{-\alpha} a(x)$, $0<\alpha<d$, then denoting $A_n = \{1,\dots,n\}^d$

$$\frac{1}{n^{d/2+\alpha}} \sum_{k\in A_{[nt]}} X_k \xrightarrow{\mathcal{D}([0,1]^d)} \int_{\mathbb{R}^d} a(x) \prod_{j=1}^d \frac{\mathrm{e}^{it_j x_j}-1}{ix_j} \mathrm{d}Z(x)$$

The limit is the Fractional Brownian Sheet only when $a(x) = \prod_{i=1}^d |x_i|^{-H_i}$.

$$\frac{1}{n^{d/2+\alpha}} \sum_{k\in A_n} X_k = \frac{1}{n^{d/2+\alpha}} \sum_{k\in A_n} \int_{[-\pi,\pi]^d} a(x)\mathrm{e}^{i\,k.x}\mathrm{d}Z(x) \quad Z : \text{spectral measure of } \epsilon$$

$$= \int_{[-\pi,\pi]^d} n^{-\alpha} a(x) \sum_{k\in A_n} \mathrm{e}^{i\,k.x} \frac{\mathrm{d}Z(x)}{n^{d/2}}$$

$$= \int_{[-n\pi,n\pi]^d} a(x) \prod_{j=1}^d \frac{\mathrm{e}^{ix_j}-1}{n(\mathrm{e}^{ix_j/n}-1)} \mathrm{d}Z(x) \xrightarrow{L^2} \int_{\mathbb{R}^d} a(x) \prod_{j=1}^d \frac{\mathrm{e}^{ix_j}-1}{ix_j} \mathrm{d}Z(x)$$

2 Point processes, random geometry

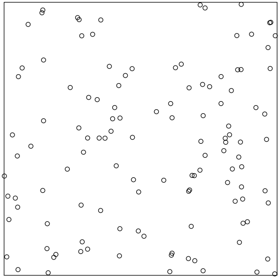
- Estimation of Gibbs point processes
- Model validation for Gibbs point processes

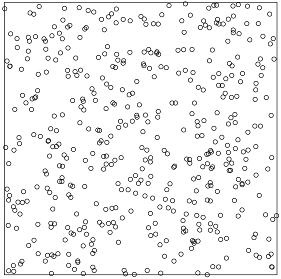
Point processes

Notation :

- Denote by φ a point pattern on $\mathbb{R}^d$, i.e. $\varphi = \bigcup_{i \in \mathcal{I}} x_i$, for $\mathcal{I} \subset \mathbb{N}^*$
- A point process Φ is a random variable on the space $\Omega = \{\varphi\}$.

Example : the Poisson point process $\rightarrow$ independance in locations





How to introduce dependencies between the location of points ?

Gibbs point processes (basic definition)

They are absolutely continuous w.r.t the Poisson point process with density

$$f_\theta(\varphi) = \frac{1}{c_\theta} \mathrm{e}^{-V_\theta(\varphi)}, \qquad \text{for some parameter } \theta \in \mathbb{R}^p.$$

$V_\theta(\varphi)$: energy of φ also called Hamiltonian (belongs to $\mathbb{R} \cup \{+\infty\}$)

- φ is more likely to occur if $V_\theta(\varphi)$ is small.
- if $V_\theta(\varphi) = +\infty$, then φ is forbidden.

Gibbs point processes (basic definition)

They are absolutely continuous w.r.t the Poisson point process with density

$$f_\theta(\varphi) = \frac{1}{c_\theta} e^{-V_\theta(\varphi)}, \qquad \text{for some parameter } \theta \in \mathbb{R}^p.$$

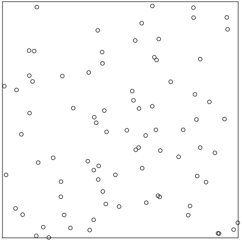
$V_\theta(\varphi)$: energy of φ also called Hamiltonian (belongs to $\mathbb{R} \cup \{+\infty\}$)

- φ is more likely to occur if $V_\theta(\varphi)$ is small.
- if $V_\theta(\varphi) = +\infty$, then φ is forbidden.

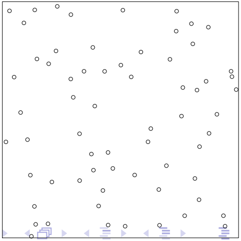
Example : Strauss process with range of interaction $R = 0.05$ on $[0,1]^2$

$$V_\theta(\varphi) = \theta \sum_{(x,y)\in\varphi} \mathbb{1}_{|y-x|<R}, \quad \theta > 0$$

$\theta = 0.35$:



$\theta = 2.3$:



Geometric interactions

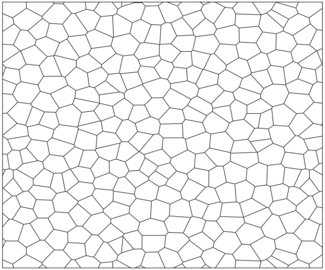
Geometric interactions

For a point pattern φ, denote $\mathrm{Vor}(\varphi)$ the associated Voronoï tessellation.

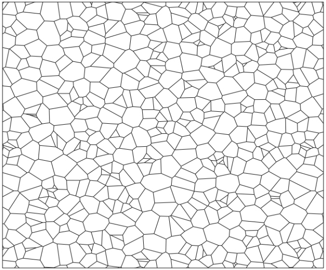
Gibbs Voronoï tessellation : one example

$$V_\theta(\varphi) = \sum_{C \in \mathrm{Vor}(\varphi)} V_1(C) + \theta \sum_{\substack{C, C' \in \mathrm{Vor}(\varphi) \\ C \text{ and } C' \text{are neighbors}}} \left| vol(C) - vol(C') \right|$$

$$V_1(C) = \begin{cases} +\infty & \text{if the cell is too "irregular",} \\ 0 & \text{otherwise} \end{cases}$$



$\theta > 0$



$\theta < 0$

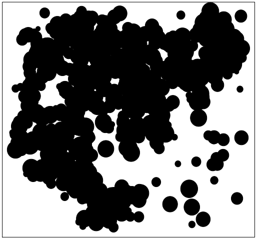
Geometric interactions

Quermass model

Each point $x \in \varphi$ is associated with a random radius r.

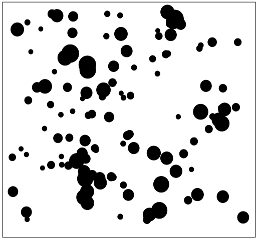
$$V_\theta(\varphi) = \theta_1 \ \mathcal{P}(\Gamma) + \theta_2 \ \mathcal{A}(\Gamma) + \theta_3 \ \mathcal{E}(\Gamma) \quad \text{where} \quad \Gamma = \bigcup_{x \in \varphi} B(x, r)$$

$\mathcal{P}$: perimeter $\quad \mathcal{A}$: area $\quad \mathcal{E}$: EP characteristic (nb connected sets - nb holes).



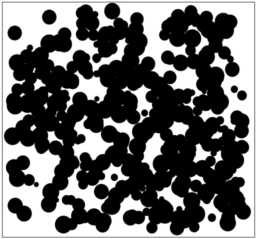
$\theta_1 > 0$

$(\theta_2 = \theta_3 = 0)$



$\theta_2 > 0$

$(\theta_1 = \theta_3 = 0)$



$\theta_3 > 0$

$(\theta_1 = \theta_2 = 0)$

Statistical issues

Let φ be a point pattern observed (or Vor(φ), or Γ,...) on a domain Λ.

Assumption : φ is the realisation of a Gibbs point process associated to V_θ for some (unknown) $\theta \in \mathbb{R}^p$.

Statistical issues

Let φ be a point pattern observed (or Vor(φ), or Γ,...) on a domain Λ.

Assumption : φ is the realisation of a Gibbs point process associated to V_θ for some (unknown) $\theta \in \mathbb{R}^p$.

Caution :

The family $(V_\theta)_\theta$ must lead to a well defined Gibbs process for all θ

(For the geometric interactions above, see D. Dereudre, R. Drouilhet, H.-O. Georgii)

Statistical issues

Let φ be a point pattern observed (or $\mathrm{Vor}(\varphi)$, or Γ,...) on a domain Λ.

Assumption : φ is the realisation of a Gibbs point process associated to V_θ for some (unknown) $\theta \in \mathbb{R}^p$.

Caution :

The family $(V_\theta)_\theta$ must lead to a well defined Gibbs process for all θ

(For the geometric interactions above, see D. Dereudre, R. Drouilhet, H.-O. Georgii)

1. How to estimate θ ?

   From $f_\theta(\varphi) = \frac{1}{c_\theta} e^{-V_\theta(\varphi)}$: maximum likelihood procedure.

   $\rightarrow$ Problem : c_θ is intractable and (prohibitive) time consuming simulations are required to approach it.

Statistical issues

Let φ be a point pattern observed (or $\mathrm{Vor}(\varphi)$, or Γ,...) on a domain Λ.

Assumption : φ is the realisation of a Gibbs point process associated to V_θ for some (unknown) $\theta \in \mathbb{R}^p$.

Caution :

The family $(V_\theta)_\theta$ must lead to a well defined Gibbs process for all θ

(For the geometric interactions above, see D. Dereudre, R. Drouilhet, H.-O. Georgii)

1. How to estimate θ ?
   From $f_\theta(\varphi) = \frac{1}{c_\theta} \mathrm{e}^{-V_\theta(\varphi)}$: maximum likelihood procedure.
   $\rightarrow$ Problem : c_θ is intractable and (prohibitive) time consuming simulations are required to approach it.

2. Is the above assumption reasonable ?

Takacs-Fiksel method

Campbell equilibrium equation (Georgii, Nguyen, Zessin)

Let $V_\theta(x|\varphi) = V_\theta(\varphi \cup x) - V_\theta(\varphi)$ (energy needed to insert x in φ)
Φ is a Gibbs point process with energy V_θ if and only if, for all function h,

$$\mathbb{E}_\Phi \left(\int_{\mathbb{R}^d} h\,(x, \Phi)\, \mathrm{e}^{-V_\theta(x|\Phi)} \mathrm{d}x \right) = \mathbb{E}_\Phi \left(\sum_{x \in \Phi} h\,(x, \Phi \setminus x) \right).$$

Takacs-Fiksel method

Campbell equilibrium equation (Georgii, Nguyen, Zessin)

Let $V_\theta(x|\varphi) = V_\theta(\varphi \cup x) - V_\theta(\varphi)$ (energy needed to insert x in φ)
Φ is a Gibbs point process with energy V_θ if and only if, for all function h,

$$\mathbb{E}_\Phi \left(\int_{\mathbb{R}^d} h\left(x, \Phi\right) \mathrm{e}^{-V_\theta(x|\Phi)} \mathrm{d}x \right) = \mathbb{E}_\Phi \left(\sum_{x \in \Phi} h\left(x, \Phi \setminus x\right) \right).$$

Empirical counterpart : **Takacs Fiksel estimation**
Let $h_1, \ldots, h_k$ be K test functions (to be chosen),

$$\hat{\theta} = \arg\min_\theta \sum_{k=1}^{K} \left[\int_\Lambda h_k(x, \varphi) \mathrm{e}^{-V_\theta(x|\varphi)} \mathrm{d}x - \sum_{x \in \varphi} h_k(x, \varphi \setminus x) \right]^2 .$$

TF estimation includes pseudo-likelihood estimation as a particular case

Takacs-Fiksel method

Campbell equilibrium equation (Georgii, Nguyen, Zessin)

Let $V_\theta(x|\varphi) = V_\theta(\varphi \cup x) - V_\theta(\varphi)$ (energy needed to insert x in φ)
Φ is a Gibbs point process with energy V_θ if and only if, for all function h,

$$\mathbb{E}_\Phi \left(\int_{\mathbb{R}^d} h(x, \Phi) \, \mathrm{e}^{-V_\theta(x|\Phi)} \mathrm{d}x \right) = \mathbb{E}_\Phi \left(\sum_{x \in \Phi} h(x, \Phi \setminus x) \right).$$

Empirical counterpart : **Takacs Fiksel estimation**
Let $h_1, \ldots, h_k$ be K test functions (to be chosen),

$$\hat{\theta} = \arg\min_\theta \sum_{k=1}^{K} \left[\int_\Lambda h_k(x, \varphi) \mathrm{e}^{-V_\theta(x|\varphi)} \mathrm{d}x - \sum_{x \in \varphi} h_k(x, \varphi \setminus x) \right]^2.$$

TF estimation includes pseudo-likelihood estimation as a particular case

Contributions : with J.-F. Coeurjolly, D. Dereudre, R. Drouihet, K. Stankova-Helisova

- Identifiability ($K > dim(\theta)$), consistency, asymptotic normality ;
- Extension to a two-step procedure in presence of (possible non-hereditary) hardcore interactions ;
- Application to Gibbs Voronoï tessellations ;
- Application to Quermass model, where points are not observed.

Model Validation with J.-F. Coeurjolly

Let φ a point pattern supposed to be the observation on Λ of a Gibbs process with parametric potential V_θ. Let $\hat{\theta}$ be an estimate of θ from φ.

Model Validation with J.-F. Coeurjolly

Let φ a point pattern supposed to be the observation on Λ of a Gibbs process with parametric potential V_θ. Let $\hat{\theta}$ be an estimate of θ from φ.

The residuals assess the Campbell equilibrium : for any h,

$$R_\Lambda(h) = |\Lambda|^{-1} \left(\int_\Lambda h(x,\varphi) \mathrm{e}^{-V_{\hat{\theta}}(x|\varphi)} \mathrm{d}x - \sum_{x \in \varphi} h(x, \varphi \setminus x) \right).$$

If the model is well specified, we expect, for any h, $R_\Lambda(h) \approx 0$.

Model Validation with J.-F. Coeurjolly

Let φ a point pattern supposed to be the observation on Λ of a Gibbs process with parametric potential V_θ. Let $\hat{\theta}$ be an estimate of θ from φ.

The residuals assess the Campbell equilibrium : for any h,

$$R_\Lambda(h) = |\Lambda|^{-1}\left(\int_\Lambda h(x,\varphi)\mathrm{e}^{-V_{\hat{\theta}}(x|\varphi)}\mathrm{d}x - \sum_{x\in\varphi} h(x,\varphi\setminus x)\right).$$

If the model is well specified, we expect, for any h, $R_\Lambda(h) \approx 0$.

We have proved : as $\Lambda \to \mathbb{R}^d$, $R_\Lambda(h) \to 0$ and $R_\Lambda(h) \sim \mathcal{N}(0,\Sigma)$.

Towards χ^2 goodness of fit tests : 2 frameworks

$$\begin{cases} R_\Lambda(h_1) \\ \vdots \\ R_\Lambda(h_j) \\ \vdots \\ R_\Lambda(h_s) \end{cases}$$

$\sum_{j=1}^s R_\Lambda(\varphi, h_j)^2 \sim \chi^2$

$R_{\Lambda_1}(h)$	$R_{\Lambda_2}(h)$	$\dots$
$\dots$	$\dots$	$\dots$
$\dots$	$\dots$	$R_{\Lambda_q}(h)$

$\sum_{j=1}^q R_{\Lambda_j}(\varphi, h)^2 \sim \chi^2$

Theoretical Ingredients

For the estimation and the validation procedures, our asymptotic results rely on

- the Campbell equilibrium equation,
- an ergodic theorem (Nguyen and Zessin),
- the central limit theorem below.

Theoretical Ingredients

For the estimation and the validation procedures, our asymptotic results rely on

- the Campbell equilibrium equation,
- an ergodic theorem (Nguyen and Zessin),
- the central limit theorem below.

CLT for a linear functional $U(\Phi_\Lambda)$ where $\Phi_\Lambda = \Phi \cap \Lambda$

Assume that if $\Lambda = \cup_{k=1}^{n} \Delta_k$ for disjoint Δ_k's then $U(\Phi_\Lambda) = \sum_{k=1}^{n} U(\Phi_{\Delta_k})$

Basically, if

(i) Conditioned centering : $\quad \mathbb{E}\left[U(\Phi_{\Delta_k}) | \Phi_{\Delta_j}, \, j \neq k\right] = 0$

(ii) Convergence of empirical covariances : $\quad \frac{1}{n} \sum_{k=1}^{n} \sum_{k'=1}^{n} U(\Phi_{\Delta_k}) U(\Phi_{\Delta_{k'}}) \xrightarrow{L^1} \Sigma$

then $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} U(\Phi_{\Delta_k}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$

The key assumption (i) allows us to go **without mixing assumptions** (which typically do not hold for all values of θ for a Gibbs process).

Alternatives to Gibbs point processes ?

Gibbs processes introduce interactions in a very natural way, but

- they can be tedious to simulate
- there are many remaining challenges for inference

Alternatives to Gibbs point processes ?

Gibbs processes introduce interactions in a very natural way, but

- they can be tedious to simulate
- there are many remaining challenges for inference

Alternatives to Gibbs processes :

- **Cox processes** : Poisson processes with random intensity function.
  $\Rightarrow$ They induce clustered point patterns.

Alternatives to Gibbs point processes ?

Gibbs processes introduce interactions in a very natural way, but

- they can be tedious to simulate
- there are many remaining challenges for inference

Alternatives to Gibbs processes :

- **Cox processes** : Poisson processes with random intensity function.
  $\Rightarrow$ They induce clustered point patterns.
- **Determinantal point processes** (with J. Møller and E. Rubak)
  Their joint intensities depend on a covariance function C.

  $\Rightarrow$ They are repulsive point processes, for $g(h) = 1 - \frac{C(h)^2}{C(0)^2}$.

  Appeals :

  - ▷ Perfect and fast simulation is available
  - ▷ Flexible models : just consider a parametric family of covariance functions
  - ▷ Inference is feasible by standard methods (maximum likelihood, contrast functions,...)

Thank you.