MCMC SIMULATION OF A RANDOM SET

Kateřina Helisová *) and Jesper Møller **)

*) Czech Technical University in Prague

**) Aalborg University (Denmark)

10th December 2009

Outline

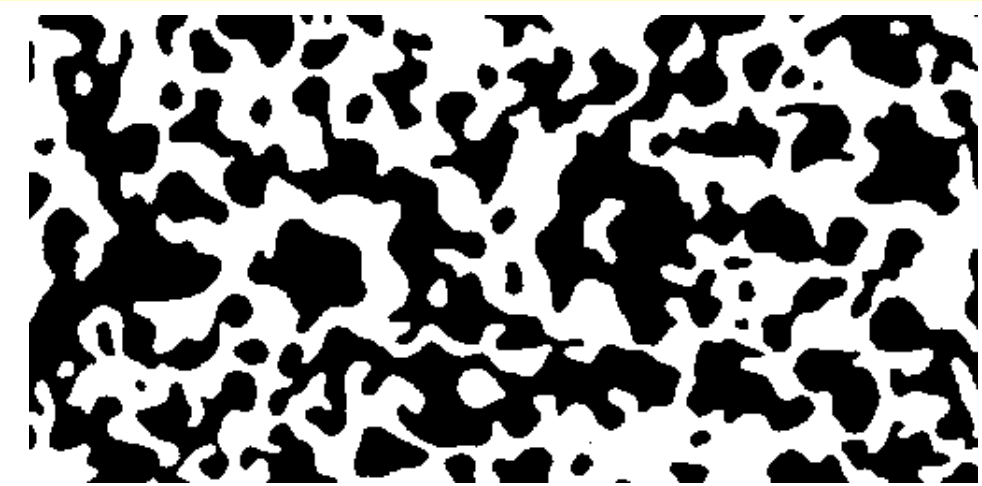
1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

Outline

1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

Back
Close

Motivation



Heather dataset first presented by Peter Diggle in 1981. The image shows the presence of heather (indicated by black) in a 10×20 m region at Jädraås, Sweeden.

Outline

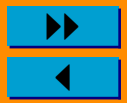
1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

Back

Close

Point processes

Definition Consider N the system of locally finite subsets of $\mathbb{R}^d$ with the σ-algebra $\mathcal{N} = \sigma(\{\mathbf{x} \in N : \sharp(\mathbf{x} \cap A) = m\} : A \in \mathcal{B}, m \in \mathbf{N}_0)$. *A point process* X defined on $\mathbb{R}^d$ is a measurable mapping from some probability space $(\Omega, \mathcal{F}, P)$ to $(N, \mathcal{N})$.



Back

Close

Point processes

Definition Consider N the system of locally finite subsets of $\mathbb{R}^d$ with the σ-algebra $\mathcal{N} = \sigma(\{\mathbf{x} \in N : \sharp(\mathbf{x} \cap A) = m\} : A \in \mathcal{B}, m \in \mathbf{N}_0)$. *A point process* X defined on $\mathbb{R}^d$ is a measurable mapping from some probability space $(\Omega, \mathcal{F}, P)$ to $(N, \mathcal{N})$.

Definition A locally finite, diffusion measure μ on $\mathcal{B}$ satisfying $\mu(A) = \mathrm{E}X(A)$ for all $A \in \mathcal{B}$ is called *the intensity measure*.

Definition If there exists a function $\rho(x)$ for $x \in \mathbb{R}^d$ such that $\mu(A) = \int_A \rho(x)dx$, then $\rho(x)$ is called *the intensity function*.

Definition A point process is called *finite point process* if $\mu(\mathbb{R}^d) < \infty$.

Poisson point process

Definition *The Poisson process* Y is the process which satisfies:

- for any finite collection $\{A_n\}$ of disjoint sets in $\mathbb{R}^d$, the numbers of points in these sets, $Y(A_n)$, are independent random variables,
- for each $A \subset \mathbb{R}^d$ such that $\mu(A) < \infty$, $Y(A)$ has Poisson distribution with parameter $\mu(A)$, i.e. $P[Y(A) = k] = \frac{\mu(A)^k}{k!} e^{-k}$, where μ is the intensity measure.

Back

Close

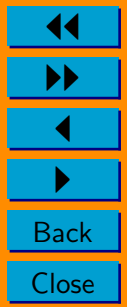
Point process given by the density with respect to Poisson process

Let Y be the Poisson process with an intensity measure μ.

For $F \in \mathcal{N}$, denote $\Pi(F) = P(Y \in F)$.

Definition A point process X is given by density f with respect to the Poisson process Y if

$$P(X \in F) = \int_F f(\mathbf{x}) \Pi(d\mathbf{x}).$$



Model

Denoting $b = b(u, r)$ a disc with centre in $u \in \mathbb{R}^2$ and radius $r \in (0, \infty)$, we have a process of discs $\cup b_i = \cup b(u_i, r_i)$. Then, we identify b with the point $x = (u, r)$ in $\mathbb{R}^2 \times (0, \infty)$ and the process of discs $\cup b_i = \cup b(u_i, r_i)$ with a point process in $\mathbb{R}^2 \times (0, \infty)$.

Back

Close

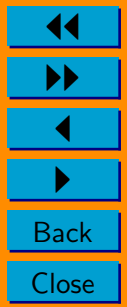
Model

Denoting $b = b(u, r)$ a disc with center in $u \in \mathbb{R}^2$ and radius $r \in (0, \infty)$, we have a process of discs $\cup b_i = \cup b(u_i, r_i)$. Then, we identify b with the point $x = (u, r)$ in $\mathbb{R}^2 \times (0, \infty)$ and the process of discs $\cup b_i = \cup b(u_i, r_i)$ with a point process in $\mathbb{R}^2 \times (0, \infty)$.

The reference process: A Poisson point process Y (so that the reference Boolean model is the random set given by the union of discs in Y) with intensity measure $\rho(u)\,\mathrm{d}u\,Q(\mathrm{d}r)$ on $\mathbb{R}^2 \times (0, \infty)$.

Model: The process of discs $\mathbf{X}$ such that the corresponding point process is absolutely continuous with respect to the reference Poisson process Y, and given by density $f(\mathbf{x})$ for a finite configurations $\mathbf{x} = \{x_1, \ldots, x_n\}$.

Assumption: $\mathbf{X}$ is a finite point process defined on $S \times (0, R)$, where S denotes a given bounded planar region such that $\int_S \rho(u)\,\mathrm{d}u > 0$ and $R < \infty$.



Exponential family density

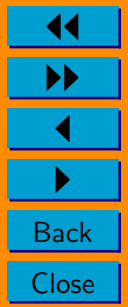
General form of the density:

$$f_\theta(\mathbf{x}) = \exp\left(\theta \cdot T(\mathcal{U}_\mathbf{x})\right)/c_\theta$$

Set $T = (A, L, N_{\mathrm{cc}}, N_{\mathrm{h}})$, where
$A = A(\mathcal{U}_\mathbf{x})$...the area
$L = L(\mathcal{U}_\mathbf{x})$...the perimeter
$N_{\mathrm{cc}} = N_{\mathrm{cc}}(\mathcal{U}_\mathbf{x})$...the number of connected components
$N_{\mathrm{h}} = N_{\mathrm{h}}(\mathcal{U}_\mathbf{x})$...the number of holes,

i.e. the density is of the form

$$f_\theta(\mathbf{x}) = \frac{1}{c_\theta} \exp\left(\theta_1 A(\mathcal{U}_\mathbf{x}) + \theta_2 L(\mathcal{U}_\mathbf{x}) + \theta_3 N_{\mathrm{cc}}(\mathcal{U}_\mathbf{x}) + \theta_4 N_{\mathrm{h}}(\mathcal{U}_\mathbf{x})\right).$$

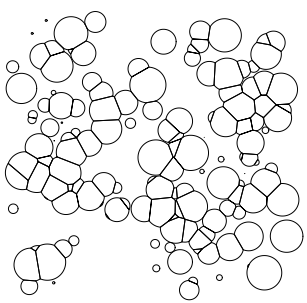
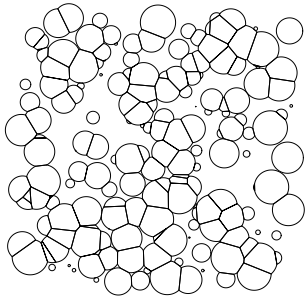
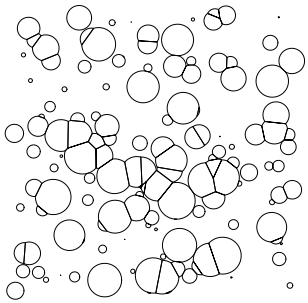


Outline

1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

Back
Close

Example of simulations

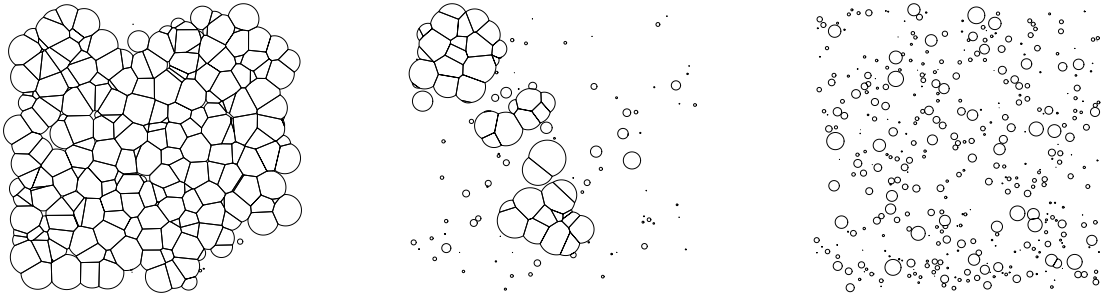




A realization of the reference Poisson process with Q the uniform distribution on the interval $[0, 2]$, $\rho(u) = 0.2$ on a rectangular region $S = [0, 30] \times [0, 30]$, and $\rho(u) = 0$ outside S (left) and A-interaction model with parameters $\theta_1 = 0.1$ (middle), resp. $\theta_1 = -0.1$ (right).

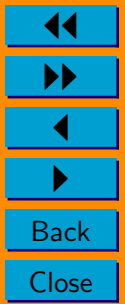
Back

Close

Example of simulations



(A, L, N_{cc})-interaction process, where $N_{\mathrm{cc}}(\mathcal{U}_{\mathbf{x}})$ is the number of connected components, with parameters $(0.6, -1, 1)$ (left), $(0.6, -1, 2)$ (middle) and $(0.6, -1, 5)$ (right).



Papangelou conditional intensity

Definition For finite $\mathbf{x} \subset S \times (0, \infty)$ and $v \in S \times (0, \infty) \setminus \mathbf{x}$, *Papangelou conditional intensity* is defined as

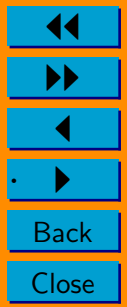
$$\lambda_\theta(\mathbf{x}, v) = f_\theta(\mathbf{x} \cup \{v\}) / f_\theta(\mathbf{x}).$$

Denoting

$$\begin{aligned} A(\mathbf{x}, v) =& A(\mathbf{x} \cup v) - A(\mathbf{x}), \\ L(\mathbf{x}, v) =& L(\mathbf{x} \cup v) - L(\mathbf{x}), \\ & \vdots \end{aligned}$$

we get

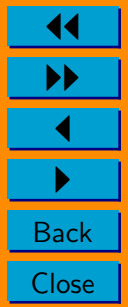
$$\lambda_\theta(\mathbf{x}, v) = \exp\left(\theta_1 \cdot A(\mathbf{x}, v) + \theta_2 \cdot L(\mathbf{x}, v) + \theta_3 \cdot N_{cc}(\mathbf{x}, v) + \theta_4 \cdot N_h(\mathbf{x}, v)\right).$$



MCMC algorithm

1. Suppose that in time t, we have a configuration $\mathbf{x_t} = \{x_1, \ldots, x_n\}$
2. Proposal in time $t+1$:
   (a) with probability $1/2$, the proposal is $\mathbf{x_t} \cup \{x_{n+1}\}$
      i. we accept the proposal with probability $min\{1; H(\mathbf{x_t}, x_{n+1})\}$ and set $\mathbf{x_{t+1}} = \mathbf{x_t} \cup \{x_{n+1}\}$
      ii. else we set $\mathbf{x_{t+1}} = \mathbf{x_t}$
   (b) else, the proposal is $\mathbf{x_t} \backslash \{x_i\}$
      i. we accept the proposal with probability $min\{1; 1/H(\mathbf{x_t} \backslash \{x_i\}, x_i)\}$ and set $\mathbf{x_{t+1}} = \mathbf{x_t} \backslash \{x_i\}$
      ii. else $\mathbf{x_{t+1}} = \mathbf{x_t}$

where $H(\mathbf{x_t}, x_{n+1}) = \lambda_\theta(\mathbf{x_t}, x_{n+1}) \frac{|S|}{\rho(x_{n+1}) \cdot (n+1)}$
and $H(\mathbf{x_t} \backslash \{x_i\}, x_i) = \lambda_\theta(\mathbf{x_t} \backslash \{x_i\}, x_i) \frac{|S|}{\rho(x_i) \cdot n}$.



Outline

1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

Back

Close

Power tessellation of a union of discs

Assume a union of discs $\mathcal{U} = \cup_I b_i$ in the general position.

For each disc b_i $(i \in I)$ with ghost sphere s_i, let $s_i^+ = \{(y_1, y_2, y_3) \in s_i : y_3 \geq 0\}$ denote the corresponding upper hypersphere.

For $u \in b_i$, let $y_i(u)$ denote the unique point on s_i^+ those orthogonal projection on $\mathbb{R}^2$ is u.

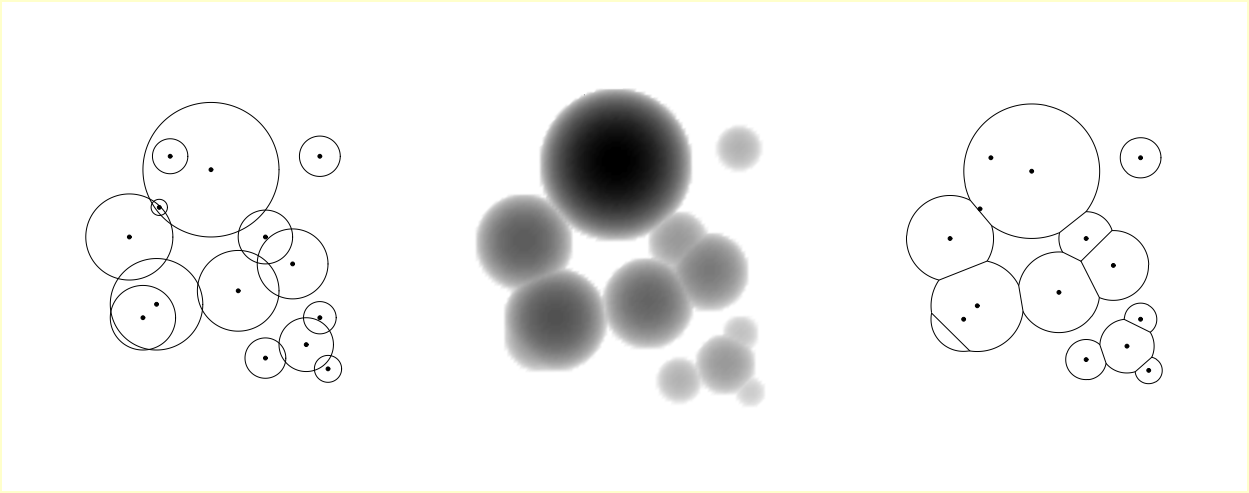
Define

$$C_i = \{y_i(u) : u \in b_i,\ \|u - y_i(u)\| \geq \|u - y_j(u)\| \text{ for } u \in b_j,\ j \in I\}.$$

Denote B_i the orthogonal projection of C_i on $\mathbb{R}^2$.

Definition The system $\mathcal{B}$ of all sets B_i is called a *power tessellation of a union of discs*.

Power tessellation of a union of discs



Left: A configuration of discs in general position. Middle: The upper hemispheres as seen from above. Right: The power tessellation of the union of discs.

Usefulness of power tessellation in MCMC algorithm

1. Calculation of $A(\mathcal{U}_{\mathbf{x}})$: instead of

$$A(\mathcal{U}_{\mathbf{x}}) = \sum_{i} A(b_i) - \sum_{\{i_1,i_2\}} A(b_{i_1} \cap b_{i_2}) + \dots$$

$$+(-1)^{n+1} \sum_{\{i_1,\dots,i_n\}} A(b_{i_1} \cap \dots \cap b_{i_n})$$

we use

$$A(\mathcal{U}_{\mathbf{x}}) = \sum_{i} A(B_i).$$

2. Analogously we calculate $L(\mathcal{U}_{\mathbf{x}})$.

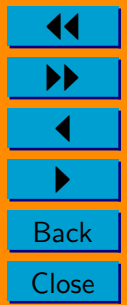
Usefulness of power tessellation in MCMC algorithm

3. For calculation of $N_{\mathrm{h}}(\mathcal{U}_{\mathbf{x}})$, we use Euler-Poincaré characteristic $\chi(\mathcal{U}_{\mathbf{x}})$ satisfying $\chi(\mathcal{U}_{\mathbf{x}}) = N_{\mathrm{cc}}(\mathcal{U}_{\mathbf{x}}) - N_{\mathrm{h}}(\mathcal{U}_{\mathbf{x}})$:
   from its definition $\chi(K_i) = 1$ for K_i compact convex and $\chi(K) = \sum_{k=1}^{N}(-1)^{k+1}\sum_{\{i_1,\ldots,i_k\}}\chi(K_{i_1}\cap\cdots\cap K_{i_k})$ for $K = \cup_{i=1}^{N}K_i$, we have that

$$\chi(\mathcal{U}_{\mathbf{x}}) = N_{\mathrm{c}}(\mathcal{U}_{\mathbf{x}}) - N_{\mathrm{ie}}(\mathcal{U}_{\mathbf{x}}) + N_{\mathrm{iv}}(\mathcal{U}_{\mathbf{x}}),$$

   where N_{c} is the number of cells, N_{ie} the number of interior edges and N_{iv} the number of interior vertices.

4. All the calculations are local.



Outline

1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

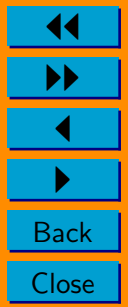
Estimating parameters

Denote $f_\theta(\mathbf{x}) = h_\theta(\mathbf{x})/c_\theta$ (i.e. $h_\theta(\mathbf{x}) = \exp(\theta \cdot T(\mathcal{U}_\mathbf{x}))$ is the unnormalized density).

For an observation $\mathbf{x}$, the log likelihood function is given by

$$l(\theta) = \log h_\theta(\mathbf{x}) - \log c_\theta = \theta \cdot T(\mathcal{U}_\mathbf{x}) - \log c_\theta.$$

Problem: c_θ has no explicit expression.



Estimating parameters

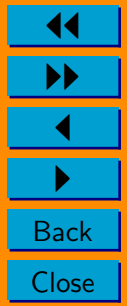
For fixed θ_0, the log likelihood ratio

$$l(\theta) - l(\theta_0) = \log(h_\theta(\mathbf{x})/h_{\theta_0}(\mathbf{x})) - \log(c_\theta/c_{\theta_0})$$

can be approximated by

$$l(\theta) - l(\theta_0) = \log(h_\theta(\mathbf{x})/h_{\theta_0}(\mathbf{x})) - \log \frac{1}{n} \sum_{m=0}^{n-1} h_\theta(Y_m)/h_{\theta_0}(Y_m),$$

where Y_m are realizations from $f_{\theta_0}(\mathbf{x})$ obtained from MCMC simulations.



Outline

1. Motivation
2. Describing the model
3. Simulations
4. Power tessellation of a union of discs
5. Estimating the parameters by MCMC MLE
6. Results

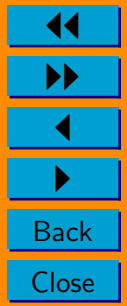
Reference processes

The three reference Boolean models are specified by

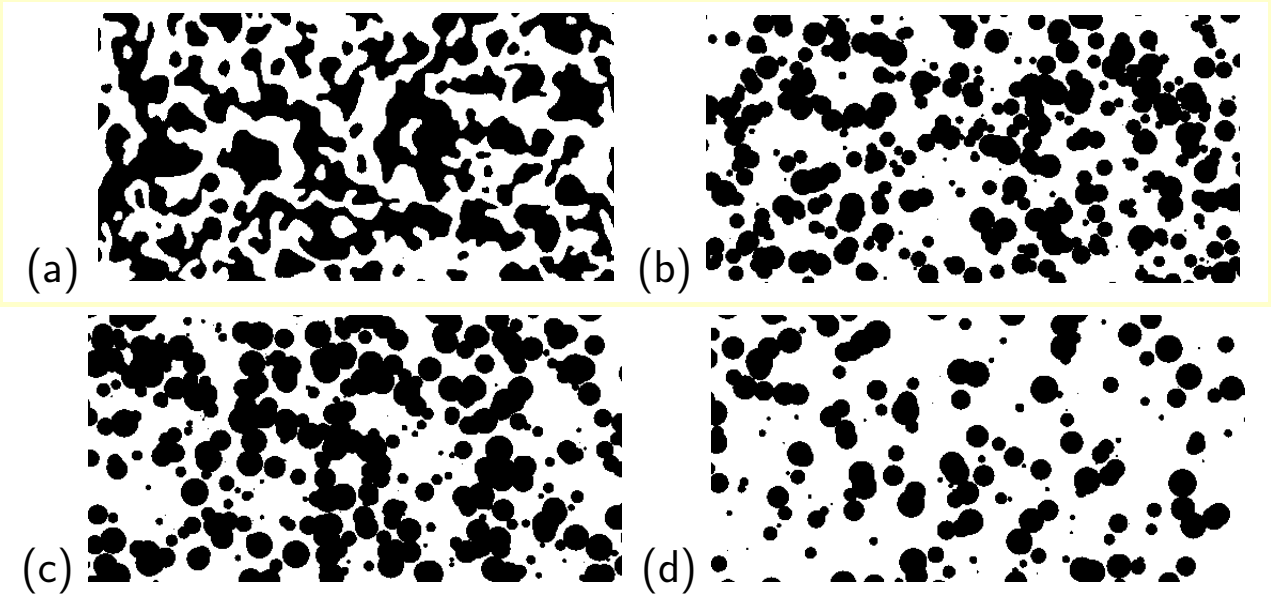
(R1) $\rho = 2.45$ and Q is the restriction of $N(0.26, 0.16^2)$ to the interval $[0, 0.50]$,

(R2) $\rho = 2.45$ and Q is the uniform distribution on $[0, 0.53]$,

(R3) $\rho = 1.16$ and Q is the uniform distribution on $[0, 0.53]$.

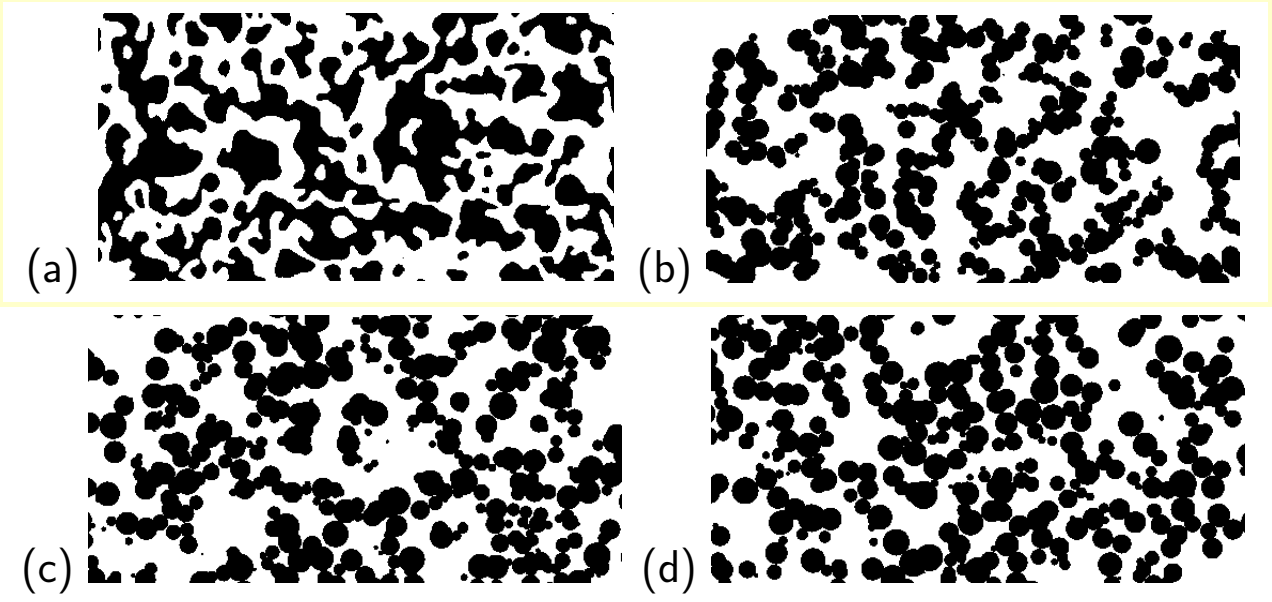


Reference processes



Compare the data (a) with a realizations of the reference Boolean models (R1) (b), (R2) (c) and (R3) (d).

Simulations



Compare the data (a) with realizations of the fitted $(A, L, N_{\mathrm{cc}})-$interaction models with respect to the reference process (R1) (b), (R2) (c) and (R3) (d).

Back
Close

Thank you for your attention!